VICERRECTORADO DE INVESTIGACIÓN INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSTGRADOS

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN SISTEMA DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

TEMA: IDENTIFICAR UN MODELO DE DATA MINING PARA DESARROLLAR UN ANÁLISIS PREDICTIVO EN LA ADMINISTRACIÓN INTEGRAL DEL TRABAJO Y EMPLEO DE LAS EMPRESAS ECUATORIANAS

AUTORES: AYALA ROSERO, EDISON JAVIER
LOGACHO FERNÁNDEZ, ANA ISABEL

DIRECTORA: MSC. DUQUE CRUZ, LORENA GUESELLE

SANGOLQUÍ

2018

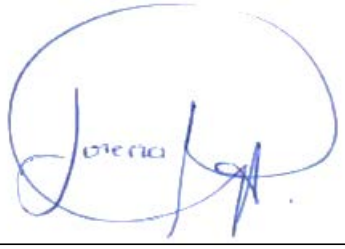
VICERRECTORADO DE INVESTIGACIÓN INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, ***"IDENTIFICAR UN MODELO DE DATA MINING PARA DESARROLLAR UN ANÁLISIS PREDICTIVO EN LA ADMINISTRACIÓN INTEGRAL DEL TRABAJO Y EMPLEO DE LAS EMPRESAS ECUATORIANAS"*** fue realizado por los señores ***Ayala Rosero, Edison Javier*** *y* ***Logacho Fernández, Ana Isabel*** el mismo ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 23 de mayo del 2018



Msc. Duque Cruz, Lorena Gueselle

C.C.: 171192525

VICERRECTORADO DE INVESTIGACIÓN INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

AUTORÍA DE RESPONSABILIDAD

Nosotros, ***Ayala Rosero, Edison Javier***, con cédula de ciudadanía n° 1722551049 y ***Logacho Fernández, Ana Isabel***, con cédula de ciudadanía n° 1714367495 declaramos que el contenido, ideas y criterios del trabajo de titulación: ***"Identificar un modelo de data mining para desarrollar un análisis predictivo en la administración integral del trabajo y empleo de las empresas ecuatorianas"*** es de nuestra autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí, 23 de mayo del 2018

Ayala Rosero, Edison Javier

C.C. 1722551049

Logacho Fernández, Ana Isabel

C.C. 1714367495

VICERRECTORADO DE INVESTIGACIÓN INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

AUTORÍZACIÓN

Nosotros, **Ayala Rosero, Edison Javier** *con C.C. n° 1722551049* y **Logacho Fernández, Ana Isabel** *con C.C. n° 1714367495* autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación **"Identificar un modelo de data mining para desarrollar un análisis predictivo en la administración integral del trabajo y empleo de las empresas ecuatorianas"** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, 23 de mayo del 2018

Ayala Rosero, Edison Javier
C.C. 1722551049

Logacho Fernández, Ana Isabel
C.C. 1714367495

DEDICATORIA

Dedico este trabajo a mi esposo e hijos, por su continuo apoyo y ánimo que me brindan día con día para alcanzar nuevas metas, tanto profesionales como personales.

Logacho Fernández, Ana Isabel

AGRADECIMIENTOS

Gracias a Dios por permitirme alcanzar una meta más en mi vida, gracias a mi familia por apoyarme en cada decisión y proyecto.

Agradezco a la Universidad del Fuerzas Armadas – ESPE por haberme dado la oportunidad de formar parte de esta Maestría.

Logacho Fernández, Ana Isabel

DEDICATORIA

Este proyecto está dedicado a mi familia, quienes me han dado la fuerza para seguir adelante en mi vida profesional y todo el apoyo para alcanzar este nuevo objetivo profesional.

Ayala Rosero, Edison Javier

AGRADECIMIENTOS

Agradezco a Dios por permitirme alcanzar este nuevo objetivo profesional, gracias a mi esposa por apoyarme en cada decisión y proyecto.

Agradezco a nuestra Directora de Tesis, por sus conocimientos y experiencia que fueron una guía durante todo el proceso.

.

Ayala Rosero, Edison Ayala

ÍNDICE

CARATULA

CERTIFICADO DEL DIRECTOR i

AUTORÍA DE RESPONSABILIDAD ii

AUTORÍZACIÓN iii

DEDICATORIA iv

AGRADECIMIENTOS v

DEDICATORIA vi

AGRADECIMIENTOS vii

ÍNDICE DE CONTENIDOS viii

ÍNDICE DE FIGURAS xvi

ÍNDICE DE TABLAS xiv

RESUMEN xvi

ABSTRACT xxii

CAPÍTULO I

ASPECTOS GENERALES 1

1.1. Introducción 1

1.1. Motivación 2

1.2. Planteamiento del Problema 3

1.2.1. Descripción del Problema 3

1.2.2. Formulación del Problema 4

1.2.3. Preguntas de Investigación 4

1.3. Justificación e Importancia 5

1.4. Alcance 6

1.5. Objetivos 6

1.5.1. Objetivo General 6
1.5.2. Objetivos Específicos 7
1.6. Metodología 7
1.6.1. Metodología para la Construcción de un Data Warehouse 7
1.6.2. Metodología de Minería de Datos 9

CAPÍTULO II
MARCO TEORICO 11
2.1. Estado del Arte 11
2.2. Metodología Data Warehouse 13
2.2.1. Metodología Inmon 13
2.2.2. Metodología Kimball 14
2.2.3. Metodología Hefesto 16
2.3. Minería de Datos 18
2.4. Métodos de Minería de Datos 18
2.4.1. Métodos Descriptivos 19
2.4.2. Métodos Predictivos 19
2.5. Técnicas de Minería de Datos 20
2.5.1. Redes Neuronales 20
2.5.2. Regresión Lineal 21
2.5.2.1. Regresión Lineal Simple 21
2.5.2.2. Regresión Lineal Múltiple 22
2.5.3. Naive Bayes 22
2.5.4. Árboles de Decisión 23
2.5.5. Máquina de Soporte Virtual (SVM) 24
2.5.6. Regresión Logística 24

2.5.7. Algoritmo A priori 25
2.6. Herramientas para Minería de Datos 26
2.6.1. Knime 26
2.6.2. RapidMiner 26
2.6.3. Weka 27
2.6.4. Lenguaje R 28
2.7. Metodología CRISP-DM 28
2.7.1. Comprensión del Negocio 30
2.7.2. Comprensión de los Datos 31
2.7.3. Preparación de los Datos 33
2.7.4. Modelado 34
2.7.5. Evaluación 36
2.7.6. Implementación o Despliegue 37

CAPÍTULO III
DESARROLLO DE LA SOLUCIÓN 39
3.1. Construcción Data Warehouse con la Metodología Kimball 39
3.1.1. Planificación del Proyecto 39
3.1.2. Definición de Requerimientos del Negocio 40
3.1.3. Diseño de la arquitectura técnica 41
3.1.4. Selección e Instalación de Herramientas 42
3.1.5. Modelado Dimensional 43
3.1.5.1. Estructura de Datos 43
3.1.5.2. Tablas de Hechos y Dimensiones 46
3.1.5.2.1. Tablas de Dimensiones 46
3.1.5.2.2. Tablas de Hechos 56

3.1.5.2.3. Contexto y Universo 60
3.1.6. Especificación de aplicaciones BI 64
3.1.7. Diseño Físico 65
3.1.8. Diseño e Implementación del ETL 66
3.1.8.1. Diseño del ETL 66
3.1.8.2. Implementación del ETL 68
3.1.9. Exploración de Herramientas BI 96
3.1.10. Mantenimiento y Crecimiento del Data Warehouse 96
3.2. Construcción Modelo de Minería de Datos aplicando la Metodología CRISP-DM 97
3.2.1. Comprensión del negocio 97
3.2.1.1. Determinar los Objetivos del Negocio 97
3.2.1.2. Evaluación de la Situación 99
3.2.1.3. Determinar los Objetivos de la Minería de Datos 101
3.2.1.4. Generación plan de proyecto 102
3.2.2. Comprensión de los datos 107
3.2.2.1. Recolectar los Datos Iniciales 107
3.2.2.2. Descripción de los Datos 107
3.2.2.3. Exploración de los Datos 108
3.2.2.4. Verificar la Calidad de los Datos 116
3.2.3. Preparación de los datos 117
3.2.3.1. Selección de los datos 117
3.2.3.2. Limpieza de los Datos 119
3.2.3.3. Construcción de los Datos 120

3.2.3.4. Integración de los Datos 123
3.2.3.5. Formateo de los Datos 123
3.2.4. Modelado 123
3.2.4.1. Selección Técnica de Modelado 123
3.2.4.2. Generación del diseño de Pruebas 124
3.2.4.3. Construcción del Modelo 126
3.2.5. Evaluación del Modelo 141
3.2.5.1. Evaluación de Resultados 142
3.2.6. Despliegue 143
3.2.6.1. Plan de Despliegue 143
3.2.6.2. Plan de Monitoreo y Mantenimiento 144
3.2.6.3. Informe Final 145

CAPÍTULO IV
ANÁLISIS E INTERPRETACIÓN DE RESULTADOS 146
4.1. Empresas Inspeccionadas o No Inspeccionadas 146
4.1.1. Primer Modelo 146
4.1.2. Segundo Modelo 148
4.1.3. Tercer Modelo 151
4.1.3.1. Arboles de Decisión 154
4.1.3.2. Regresión Logística 157
4.1.3.3. Redes Neuronales 160
4.2. Reglas de Asociación 164
4.2.1. Primer Modelo 164
4.2.2. Segundo Modelo 166
4.3.3. Tercer Modelo 167

4.3.4. Cuarto Modelo 169

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES 171

5.1. Conclusiones 171

5.2. Recomendaciones 172

BIBLIOGRAFÍA 173

GLOSARIO 178

ÍNDICE DE TABLAS

Tabla 1. *Comparativa metodologías construcción data warehouse*.....8
Tabla 2. *Comparativas metodologías de minería de datos* 9
Tabla 3. *Artículos Revisados* 11
Tabla 4. *Planificación del Proyecto* 40
Tabla 5. *Requerimientos del Negocio* 40
Tabla 6. *Dimensión Discapacidad*.....46
Tabla 7. *Dimensión Género* 46
Tabla 8. *Dimensión Etnia*.....47
Tabla 9. Dimensión Actividad Económica 47
Tabla 10. *Dimensión Ubicación* 48
Tabla 11. *Dimensión Contrato* 49
Tabla 12. *Dimensión Tipo de Contrato*.....50
Tabla 13. *Dimensión Institución* 50
Tabla 14. *Dimensión Empleado* 51
Tabla 15. *Dimensión Fecha* 52
Tabla 16. *Dimensión Tipo Empresa* 52
Tabla 17. *Dimensión Motivo de Salida*.....53
Tabla 18. *Dimensión Grupo Ocupacional*.....54
Tabla 19. *Dimensión Acta de Finiquito*.....55
Tabla 20. *Dimensión Boletas* 55
Tabla 21. *Dimensión Tipo Trámite* 56
Tabla 22. *Dimensión Trámite* 56
Tabla 23. *Tabla de Hechos Contratos*.....57
Tabla 24. *Tabla de Hechos Acta de Finiquito*.....58
Tabla 25. *Tabla de Hechos Boletas* 59
Tabla 26. *Tabla de Hechos Trámite* 59
Tabla 27. *Tabla de Hechos Cumplimiento* 60
Tabla 28. *Características Hardware*.....100
Tabla 29. *Detalle Recurso Humano* 101

Tabla 30. *Plan del Proyecto* 103
Tabla 31. *Selección de Técnica de Minería de Datos* 106
Tabla 32. *Transformación de Campos* 120
Tabla 33. *Pivoteo de Campos* 121
Tabla 34. *Cumplimiento de los Objetivos de Negocio* 142
Tabla 35. *Cumplimiento de los Objetivos de Minería de Datos* 142
Tabla 36. *Matriz de Ponderación de Variables* 147
Tabla 37. *Porcentaje Incumplimiento* 147
Tabla 38. *Error Técnicas Seleccionadas Primer Modelo* 148
Tabla 39. *Tiempo de Ejecución Técnicas Segundo Modelo* 150
Tabla 40. *Error Técnicas Seleccionadas Primer Modelo* 150
Tabla 41. *Tiempo de Ejecución Técnicas Tercer Modelo* 151
Tabla 42. *Resumen de Indicadores más representativos* 163

ÍNDICE DE FIGURAS

Figura 1. Encuesta uso de metodologías de minería de Datos.....10
Figura 2. Porcentaje de Artículos por Áreas de Aplicación 12
Figura 3. Metodología Inmon.....14
Figura 4. Tareas de la Metodología de Kimball.....16
Figura 5. Metodología HEFESTO.....17
Figura 6. Proceso Minería de Datos 18
Figura 7. Red Neuronal Multicapa 21
Figura 8. Modelo Regresión lineal Simple 22
Figura 9. Curva de Regresión Logística.....25
Figura 10: Ciclo de Vida – Metodología CRISP-DM 29
Figura 11. Fase 1. Comprensión del Negocio.....31
Figura 12. Fase 2: Comprensión de los Datos.....32
Figura 13. Fase 3. Preparación de los Datos.....34
Figura 14. Fase4. Modelado.....36
Figura 15. Fase 5. Evaluación 37
Figura 16. Fase 6. Implementación o Despliegue.....38
Figura 17. Diseño de la Arquitectura Técnica 42
Figura 18. Herramientas Seleccionadas.....42
Figura 19. Modelo Entidad Relación Sistema SAITE.....44
Figura 20. Modelo Entidad Relación Sistema SINACOI.....45
Figura 21. Modelo Entidad Relación Sistema SGI 45
Figura 22. Contexto Actas de Finiquito.....61
Figura 23. Contexto Contratos.....62
Figura 24. Contexto Trámites 62
Figura 25. Contexto Boletas 63
Figura 26. Contexto Factores de Cumplimiento.....63
Figura 27. Universo 64
Figura 28. Esquema de la Base de Datos en PostgreSQL 66
Figura 29. Carga de Datos en las Tablas de Dimensiones 67

Figura 30. Carga de Datos en las Tablas de Hechos....68
Figura 31. Proceso Carga de Datos Dimensión Empleado....69
Figura 32. Datos en la Tabla Dimensión Empleado....70
Figura 33. Proceso Carga de Datos Dimensión Institución....71
Figura 34. Datos en la Tabla Dimensión Institución....71
Figura 35. Proceso Carga de Datos Dimensión Ubicación....72
Figura 36. Datos en la Tabla Dimensión Ubicación....73
Figura 37. Proceso Carga de Datos Dimensión Género....73
Figura 38. Datos en la Tabla Dimensión Género....74
Figura 39. Proceso Carga de Datos Dimensión Actividad Económica....74
Figura 40. Datos en la Tabla Dimensión Actividad Económica....75
Figura 41. Proceso Carga de Datos Dimensión Tipo Contrato....75
Figura 42. Datos en la Tabla Dimensión Tipo Contrato....76
Figura 43. Proceso Carga de Datos Dimensión Fecha....76
Figura 44. Datos en la Tabla Dimensión Fecha....77
Figura 45. Proceso Carga de Datos Dimensión Etnia....77
Figura 46. Datos en la Tabla Dimensión Etnia....78
Figura 47. Proceso Carga de Datos Dimensión Contrato....79
Figura 48. Datos en la Tabla Dimensión Contrato....79
Figura 49. Proceso Carga de Datos Dimensión Discapacidad....80
Figura 50. Datos en la Tabla Dimensión Discapacidad....81
Figura 51. Proceso Carga de Datos Dimensión Tipo Contrato....81
Figura 52. Datos en la Tabla Dimensión Tipo Contrato....82
Figura 53. Proceso Carga de Datos Dimensión Acta de Finiquito....83
Figura 54. Datos en la Tabla Dimensión Tipo Contrato....83
Figura 55. Proceso Carga de Datos Dimensión Motivo de Salida....84
Figura 56. Datos en la Tabla Dimensión Motivo de Salida....84
Figura 57. Proceso Carga de Datos Dimensión Grupo Ocupacional....85
Figura 58. Datos en la Tabla Dimensión Grupo Ocupacional....85
Figura 59. Proceso Carga de Datos Dimensión Trámites....86

Figura 60. Datos en la Tabla Dimensión Grupo Ocupacional 87
Figura 61. Proceso Carga de Datos Dimensión Boletas 88
Figura 62. Datos en la Tabla Dimensión Boletas 89
Figura 63. Proceso Carga de Datos Tabla de Hechos Contratos 90
Figura 64. Datos en la Tabla de Hechos Contratos 90
Figura 65. Proceso Carga de Datos Tabla de Hechos Actas de Finiquito 91
Figura 66. Datos en la Tabla de Hechos Actas de Finiquito 91
Figura 67. Proceso Carga de Datos Tabla de Hechos Trámites 92
Figura 68. Datos en la Tabla de Hechos Trámites 93
Figura 69. Proceso Carga de Datos Tabla de Hechos Boletas 94
Figura 70. Datos en la Tabla de Hechos Boletas 94
Figura 71. Proceso Carga de Datos Tabla de Hechos Factores Cumplimiento 95
Figura 72. Datos en la Tabla de Hechos Factores Cumplimiento 96
Figura 73. Cuadrante Mágico de Gartner para Plataformas de Ciencia de Datos y Aprendizaje Automático 104
Figura 74. Empresas Registradas por Provincia 108
Figura 75. Empresas No Inspeccionadas por Provincia 109
Figura 76. Empresas No Inspeccionadas por Provincia 109
Figura 77: Distribución de Empleados de Acuerdo al Género en Relación a la Actividad Económica de la Empresa 110
Figura 78. Trabajo Juvenil por Provincia 111
Figura 79. Rango de Edades por Actividad Económica 112
Figura 80. Tipo de Contratos por Provincia 113
Figura 81. Distribución de Contratos por Provincia 113
Figura 82. Actas Registradas por Motivo de Salida y Rango de Edad 114
Figura 83. Actas de Finiquito Registrada por Provincia 115
Figura 84. Tipo de Trámites 115
Figura 85. Boletas Generadas por Región 116
Figura 86. Componente X-Partitioner 125
Figura 87. Configuración X-Partitioner 125

Figura 88. Componente X-Agregator....126
Figura 89. Configuración X-Partitioner....126
Figura 90. Árbol de Decisión....127
Figura 91. Configuración Nodo de Aprendizaje....129
Figura 92. Árbol de Decisión Parte Superior....130
Figura 93. Parte Izquierda Árbol de Decisión – Primera Parte....131
Figura 94. Parte Izquierda Árbol de Decisión – Segunda Parte....132
Figura 95. Parte Derecha Árbol de Decisión....133
Figura 96. Regresión Logística....134
Figura 97. Configuración de Ajustes Iniciales....135
Figura 98. Configuración de Ajustes Avanzada....136
Figura 99. Coeficientes de Variables....137
Figura 100. Redes Neuronales....138
Figura 101. Configuración Red Neuronal....139
Figura 102. Regla de Asociación 1....140
Figura 103. Regla de Asociación 2....141
Figura 104. Regla de Asociación 3....141
Figura 105. Regla de Asociación 4....141
Figura 106. Matriz de Confusión Árboles de Decisión....155
Figura 107. Tabla de Estadísticas de Precisión Árboles de Decisión....155
Figura 108. Curva ROC Árboles de Decisión....157
Figura 109. Matriz de Confusión Árboles de Decisión....158
Figura 110. Tabla de Estadísticas de Precisión Árboles de Decisión....159
Figura 111. Curva ROC Árboles de Decisión....160
Figura 112. Matriz de Confusión Redes Neuronales....161
Figura 113. Tabla de Estadísticas de Precisión Redes Neuronales....162
Figura 114. Curva ROC Redes Neuronales....163
Figura 115. Reglas de Asociación del Primer Modelo....165
Figura 116. Reglas de Asociación del Segundo Modelo....166
Figura 117. Reglas de Asociación del Tercer Modelo....168

Figura 118. Reglas de Asociación del Tercer Modelo.....169

RESUMEN

El Ministerio del Trabajo al ser la institución rectora de las políticas de trabajo y empleo a través de las Inspectorías de Trabajo a nivel nacional realiza inspecciones a las empresas ecuatorianas del sector privado con el fin de promover y garantizar el cumplimiento de la normativa legal vigente en el ámbito laboral por parte de las empresas hacia los empleados. El Ministerio pone a disposición de las empresas sistemas transaccionales como SAITE, SINACOI y SGI, los mismos que generan grandes cantidades de información que se almacenan en un único repositorio para luego someterlos a un análisis y el resultado que se obtiene sea un apoyo para la toma de decisiones a las autoridades del Ministerio. El proyecto de tesis presenta el resultado de la investigación realizada para identificar un modelo de data mining con la finalidad de desarrollar un análisis predictivo en la administración integral del trabajo y empleo de las empresas ecuatorianas, así como reconocer los patrones de comportamiento que se tienen las empresas en la contratación de talento humano. La metodología Kimball nos guía en la construcción del data warehouse, los datos que se almacenan en el data warehouse nos sirve como insumo para determinar el modelo predictivo de minería de datos el cual es determinado con la aplicación de la metodología CRISP-DM en todas sus seis fases. Knime es la herramienta elegida para construir los modelos, las técnicas aplicadas son: árboles de decisión, regresión logística y redes neuronales para predecir si una empresa debe ser inspeccionada o no inspeccionada, de igual manera para determinar los patrones de comportamiento en la contratación del talento humano se utiliza reglas de asociación.

Palabras Clave

- **REPOSITORIO**
- **MINERÍA DE DATOS**
- **MODELO PREDICTIVO**
- **ÁRBOLES DE DECISIÓN**
- **REGRESIÓN LOGÍSTICA**

ABSTRACT

The Labor Ministry is an institution that governs labor laws, primarily through the work of inspectors who report to the national level. The inspectors' purpose is to inspect Ecuadorian enterprises in the private sector in order to promote and guarantee compliance with current legal regulations in the work environment. The Ministry dispenses transactional systems to companies, such as SAITE, SINACOI and SGI. These transactional systems generate large amounts of information that are stored in a single repository. After generating the information, the systems conduct their analysis and offer a solution to The Labor Ministry. These authorities can use the results to make a more informed decision. This thesis project presents the findings of an investigation that sought to identify a data mining model in which developed a predictive analysis of administration in Ecuadorian enterprises. This project also aimed to recognize performance patterns in the companies' hiring practices. Kimball Methodology guides in the construction of the data warehouse. The data storied in the data warehouse serves as input to determine a predictable model of data mining, which is determined by the application of CRISP-DM Methodology in all its six stages. Knime is the tool chosen to build the models; its applied techniques are decision trees, logistical regression, and neuronal networks in order to predict whether a company should be inspected. It also determines the performance patterns in the companies' hiring practices.

KEYWORDS

- **ARCHIVE**
- **DATA MINING**
- **PREDICTABLE MODEL**
- **DECISION TREES**
- **LOGISTICAL REGRESSION**

CAPÍTULO I: ASPECTOS GENERALES

1.1. Introducción

El Ministerio del Trabajo MDT[1] al ser el ente rector de políticas de trabajo y empleo debe garantizar en el sector privado y público, la estabilidad y la armonía en las relaciones laborales entre el trabajador y empleador, enfocándose en la protección de los derechos fundamentales del trabajador ha implementado Sistemas de Información que permiten llevar un control en el cumplimiento de la Normativa Legal Vigente para las empresas privadas del Ecuador.

El Sistema de Administración Integral de Trabajo y Empleo (SAITE[2]) tiene como objetivo “Gestionar información integral de Trabajo y Empleo por parte del empleador y empleado, este sistema se encuentra dividido en: Registros de Contratos de Trabajo, Registro de Actas de Finiquito, Seguridad y Salud en el Trabajo, Reglamentos Internos”, para nuestro caso de estudio se trabaja los módulos de Registros Contratos de Trabajo y Registro de Actas de Finiquito.

El Sistema Nacional de Control de Inspectores (SINACOI[3]) tiene como objetivo registrar los trámites de denuncias, reclamos laborables del trabajador y empresa, inspecciones integrales, pago de multas, consignaciones, el sistema se encuentra dividido en módulos: Inspectoría, Erradicación Trabajo Infantil y Financiero para el caso de estudio se trabaja con el módulo de Inspectoría específicamente con Trámites y Reclamos Laborales.

[1] Ministerio del Trabajo
[2] Sistema de Administración Integral del Talento Humano
[3] Sistema Nacional de Control de Inspectores

El Sistema Inspector Integral 2.0 (SGI[4]) tiene como objetivo registrar si una empresa fue inspeccionada o no inspeccionada de acuerdo a ciertos parámetros y al cruce de información con otras instituciones como SRI, IESS, Superintendencia de Compañías y DINARDAP.

Estos sistemas generan grandes cantidades de datos que nos permitirá analizar la información haciendo uso de técnicas de minería de datos, la finalidad es construir y evaluar un modelo de minería de datos (data mining) confiable, el cual ayude al Ministerio del Trabajo a crear información y generar conocimiento a partir de los datos obtenidos.

1.1. Motivación

En los últimos años el MDT ha realizado grandes esfuerzos por realizar seguimiento al cumplimiento de la normativa legal por parte de las empresas ecuatorianas, es así que ha automatizado varios servicios, como es el caso de Registro de Contratos y Actas de Finiquito en línea, es decir, las empresas realizan el registro sin necesidad de acercarse a las instalaciones del Ministerio.

Tomando en cuenta que no se ha encontrado artículos similares o sobre el estudio propuesto, se propone la construcción de un modelo predictivo para determinar si una empresa debe o no ser inspeccionada, haciendo uso de diferentes técnicas de minería de datos.

El presente trabajo ayudara a las autoridades del Área de Inspectoría del Trabajo a tomar decisiones sobre las inspecciones que se deben realizar a las empresas,

[4] Sistema Inspector Integrar 2.0

optimizando el tiempo de los inspectores del trabajo y enfocándose en las ciudades donde se concentran el mayor incumplimiento de la Normativa Legal Vigente.

1.2. Planteamiento del Problema

1.2.1. Descripción del Problema

El MDT registra a diario un gran volumen de datos a través de sus sistemas transaccionales SAITE, SINACOI y SGI, los cuales recopilan datos de las empresas públicas de los funcionarios que están bajo el Código del Trabajo y empresas privadas, pero el Ministerio en la actualidad no cuenta con una herramienta que permita recolectar, analizar y mostrar la información a las autoridades del Ministerio, motivo por el cual se propone un proyecto de construcción de data warehouse y minería de datos, mediante la generación de un modelo predictivo y técnicas de minería de datos, que permiten extraer datos desde distintas fuentes, realizar la respectiva transformación de los parámetros requeridos, encontrar el modelo más apropiado para el análisis de la administración integral del talento humano en las empresas ecuatorianas. Todo esto aplicado a los sistemas SAITE, SINACOI y SGI, de tal manera que se pueda disponer de información clara, veraz, oportuna y precisa, facilitando la toma de decisiones de una manera rápida, ya que los datos son confiables y estables, estos se ponen a disposición de los usuarios que necesiten siempre que cuenten con las respectivas autorizaciones.

Las áreas que demandan información realizan enormes esfuerzos para recopilar, transcribir y cambiar el formato de los datos con el fin que sean entendibles al usuario, y las áreas pueden generar los reportes con un margen de error elevado, esto ha ocasionado que las áreas que toman decisiones en el Ministerio reciban la información

muy tarde, con un peso muy grande, con la misma información las áreas tienen diferentes resultados, no es de relevancia ni tampoco confiable, con esto se puede concluir que no existe un adecuado criterio de selección, ni uso de estándares.

Para analizar la información relevante, se solicita al área tecnología la generación de reportes, dependiendo del reporte solicitado se procede a vincular tablas las cuales contienen dichos datos, estos poseen muchas propiedades o atributos que causan que sea difícil de visualizar y en otras circunstancias se obtiene una cantidad de datos muy elevada; en este proceso los resultados se ven propensos a pérdida e inconsistencia, lo cual podría generar un reporte con información incoherente y no fiable para la toma de decisiones. Adicionalmente no se dispone de herramientas que permitan verificar en los datos almacenados, patrones que ayuden a visualizar las tendencias y por ende no permite realizar proyecciones de los procesos relacionados con el trabajo y empleo.

1.2.2. Formulación del Problema

El proyecto de tesis que se presenta tiene como finalidad ofrecer una respuesta al siguiente cuestionamiento:

¿Es posible predecir haciendo uso de las técnicas de minería de datos el comportamiento de las empresas privadas en el Ecuador en el cumplimiento de la Normativa Legal Vigente?

1.2.3. Preguntas de Investigación

El presente proyecto de tesis debe responder a las siguientes preguntas:

¿Cómo mejorar el tratamiento de los datos para convertirlos en información y proporcionar un recurso para la toma de decisiones en el ámbito laboral?

¿Cuáles son las tendencias y patrones que se presentan en los datos analizados?

¿Cuáles son las técnicas de minería de datos más apropiadas para la administración integral del trabajo y empleo de las empresas ecuatorianas?

1.3. Justificación e Importancia

Las técnicas de minería de datos que se aplican en los procesos de registros de contratos de Trabajo, registro de actas de finiquito, registro de trámites de denuncias, reclamos laborables del trabajador y empresa, colaboran en mejorar los servicios que presta el MDT a las empresas y trabajadores del sector privado a través de los sistemas SAITE, SINACOI y SGI, cabe indicar que a partir de agosto del 2016 se implementó el sistema SGI donde se almacena el registro de empresas inspeccionadas tomando datos almacenados en las bases de datos de la institución así como datos externos que son proporcionados por las siguientes entidades: SRI[5], IESS[6], Superintendencia de Compañías y DINARDAP[7].

Para nuestro caso de estudio se tomarán datos almacenados en las bases de la institución que son registrados a través de los sistemas transaccionales SAITE y SINACOI, para luego contrastar la validez del modelo con los resultados de inspecciones registradas en el sistema transaccional SGI, con la finalidad de tener una mejor planificación de las inspecciones enfocándose en las empresas que no han cumplido con la Normativa Legal Vigente, minimizando el tiempo que se invierte en la verificación en

[5] Servicio de Rentas Internas
[6] Instituto Ecuatoriano de Seguridad Social
[7] Dirección Nacional de Registro de Datos Públicos

forma manual, con esto se garantiza que el empleador cumpla con las obligaciones que indica la ley para con los trabajadores, velando por su bienestar laboral.

Con la aplicación del proyecto de minería de datos se detecta e identifica los patrones de comportamiento que no se visualizan a simple vista de los datos almacenados en las bases de datos SAITE, SINACOI y SGI, extrayendo información relevante que sea de utilidad y soporte para la toma de decisiones por parte de las autoridades del MDT centrándose en el área de inspecciones de trabajo.

1.4. Alcance

El presente proyecto tiene como finalidad extraer información de los datos históricos de las bases de datos de los Sistemas SAITE, SINACOI y SGI, al realizar un análisis profundo de la información nos permitirá observar el comportamiento de los empleadores y trabajadores, esto nos ayudará a identificar hechos relevantes, relaciones entre los datos, tendencias, patrones y anomalías. Las bases de datos que serán utilizadas para el presente proyecto guardan información desde el año 2012 con fecha de corte enero 2018.

1.5. Objetivos

1.5.1. Objetivo General

Construir un modelo de minería de datos que permita predecir y caracterizar los datos de los sistemas transaccionales (Sistema de Administración Integral de Trabajo y Empleo (SAITE), Sistema Nacional de Control de Inspectores (SINACOI) y Sistema Integral

Inspector 2.0 (SGI), para su posterior implementación y obtención de conocimiento que permita dar respuesta a indicadores y comportamiento de la información.

1.5.2. Objetivos Específicos

1. Diseñar un marco teórico que fundamente las teorías utilizadas para el desarrollo de la propuesta.
2. Definir las necesidades institucionales, identificando la información necesaria para la ejecución de los procesos relacionados con el área de trabajo y empleo.
3. Desarrollar un proceso de depuración de la información de una manera adecuada para la obtención de resultados confiables.
4. Evaluar y determinar las técnicas de minería de datos que mejor se ajusten a la administración integral del trabajo y empleo.

1.6. Metodología

1.6.1. Metodología para la Construcción de un Data Warehouse

Existen varias guías de diseño y construcción de Data Warehouse, sin embargo, pocas se han consolidado como metodologías. El término metodología se entiende como un conjunto de procedimientos ordenados para el logro de objetivos planteados que dirigen una investigación. Se ha decidido evaluar tres metodologías de las más conocidas que han sido documentadas por Ralph Kimball[8], Bill Inmon[9] y Bernabeu Ricardo Darío (Hefesto)[10].

[8] Propone el modelado dimensional para data warehouse
[9] Reconocido como el padre del concepto de data warehouse
[10] Metodología para la construcción de una data warehouse

HEFESTO es una metodología propia, está fundamentada en la investigación, comparación de otras metodologías, experiencias propias en la construcción de Data Warehouse (almacenes de datos). La metodología se encuentra en constante evolución y ha tomado en cuenta la retroalimentación de quienes han utilizado la metodología.

Bill Inmon define el Data Warehouse como: “Es una colección de datos orientados al tema, integrados, no volátiles e historiados que facilitan la toma de decisiones”. (Inmon, 2005)

Ralph Kimball define un Data Warehouse como: “Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”. (Vargas, 2016)

Tabla 1.

Comparativa metodologías construcción data warehouse

METODOLOGIA	HEFESTO	INMON	KIMBALL
Autor	Bernabeu Ricardo Darío	Bill Inmon	Ralph Kimball
Año	2010	1990	1997
Tipo de Empresa	Pequeño y mediano	Pequeño, mediano y grande	Pequeño, mediano y grande
Arquitectura	Ascendente (Bottom - up)	Descendente (Top –down)	Ascendente (Bottom - up)
Enfoque empresarial	Análisis de objetivos y establecimiento de indicadores.	Análisis corporativo	Análisis por departamento
Énfasis	Data Warehouse y Data Mart	Data Warehouse	Data Mart
Integración de Datos		Todos los sistemas transaccionales de la organización	Áreas del negocio en forma individual
Perspectiva	Estrella / Copo de nieve	Relacional	Estrella
Flexibilidad	Si	No	Si
Costo de Implementación	Bajo	Alto	Bajo

Por las características del proyecto se ha seleccionada la Metodología de Kimball por su versatilidad y haciendo uso de pocos recursos, permite implementar pequeños data marts en áreas específicas para ir integrándolos de acuerdo al avance del proyecto cuya finalidad es obtener un data warehouse.

1.6.2. Metodología de Minería de Datos

Para la selección de la metodología de minería de datos se realizó una comparación entre dos metodologías CRISP-DM[11] y SEMMA[12], esta comparación

Tabla 2.

Comparativas metodologías de minería de datos

METODOLOGIA	CRISP-DM	SEMMA
Comprensión del negocio	Si	No
Selección y preparación de datos	Si	Si
Modelado	Si	Si
Evaluación	Si	Si
Implementación	Si	No
Número de fases	6	5
Elección libre de herramientas	Si	No
Fases relacionadas	Si	No
Detalle en pasos a seguir para cada fase	Si	No
Metodología Estructurada	Si	Si
Estabilidad de la Metodología	Si	Si
Uso Amplio	Si	No

Fuente: (Grández, 2017)

[11] Cross Industry Standard Process for Data Mining
[12] Sample Explore Modify Model and Asses

De acuerdo a los resultados que se obtienen en la Tabla 2, la metodología que se adapta a nuestro proyecto de tesis es la metodología CRISP-DM. Otro punto a favor para seleccionar esta metodología es la encuesta del 2014 publicada por el sitio web KDNuggets[13], la misma que es comparada con la encuesta del 2007 sobre el uso de metodologías y Minería de Datos, donde se observa que la metodología CRISP-DM sigue ocupando el primer lugar, seguida por metodologías propias y en tercer lugar SEMMA como se muestra en la Figura 1.

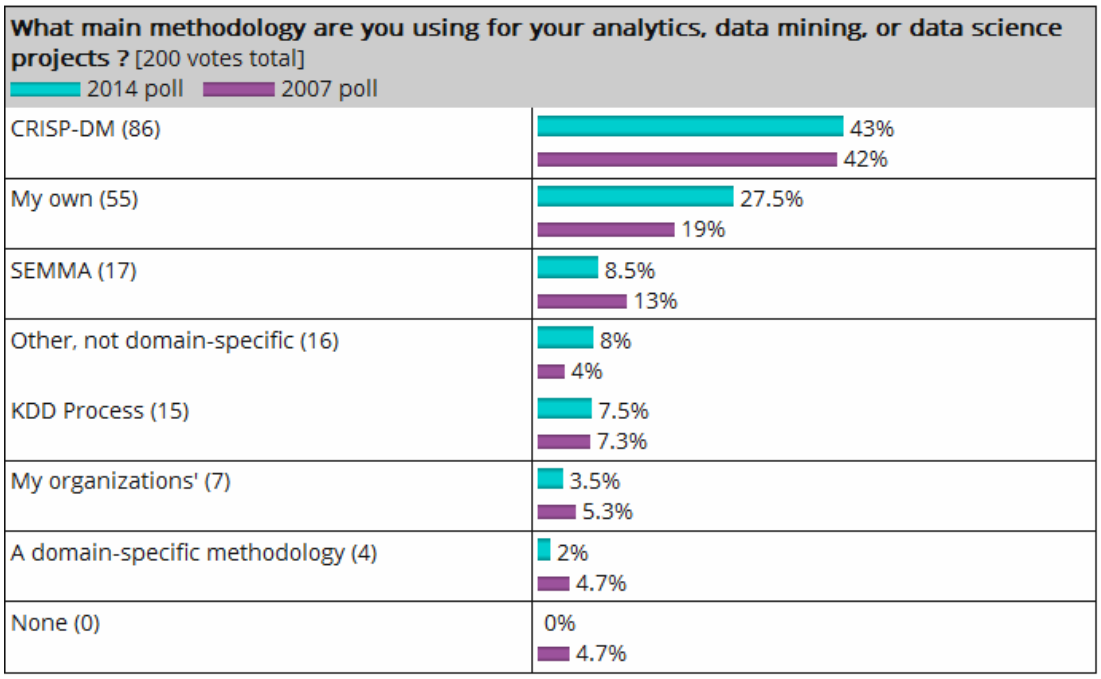


Figura 1. Encuesta uso de metodologías de minería de Datos

Fuente: (Piatetsky, 2014)

[13] Sitio web creado por algunos fundadores de Minería de Datos en los años 90 como Gregory Piatetsky Shapiro y Evangelos Simoudis.

CAPÍTULO II: MARCO TEORICO

2.1. Estado del Arte

Se realizó la búsqueda en diferentes repositorios virtuales de trabajos relacionados con el ámbito laboral; se realizaron búsquedas con varias cadenas, con la que mejores resultados se obtuvo es la siguiente cadena de búsqueda: ((("Document Title":data mining) AND classification techniques) AND labor), donde se obtuvo la siguiente información:

- Se encontraron 104 artículos que hacen referencia a la cadena de búsqueda planteada.
- Se realizó una clasificación de los artículos y se puede observar en la Tabla 3, que no existió ningún artículo que haga referencia al tema planteado en el presente documento.

Tabla 3.

Artículos Revisados

Áreas	Número de Artículos
Educación	14
Salud	15
Financiero	5
Tecnología	10
Otras categorías	53
Agricultura	7
Total de Artículos	104

En la Figura 2, se observa que, del total de artículos encontrados con la cadena de búsqueda aplicada, un 14% pertenece al área de salud, seguida muy de cerca por el área de educación, luego por tecnología con un 10%, agricultura 7% y financiero en un 5%, existen otras categorías con un 51% en esta se encuentran inmersas áreas que no tiene relación entre sí, ni con las áreas indicadas anteriormente.

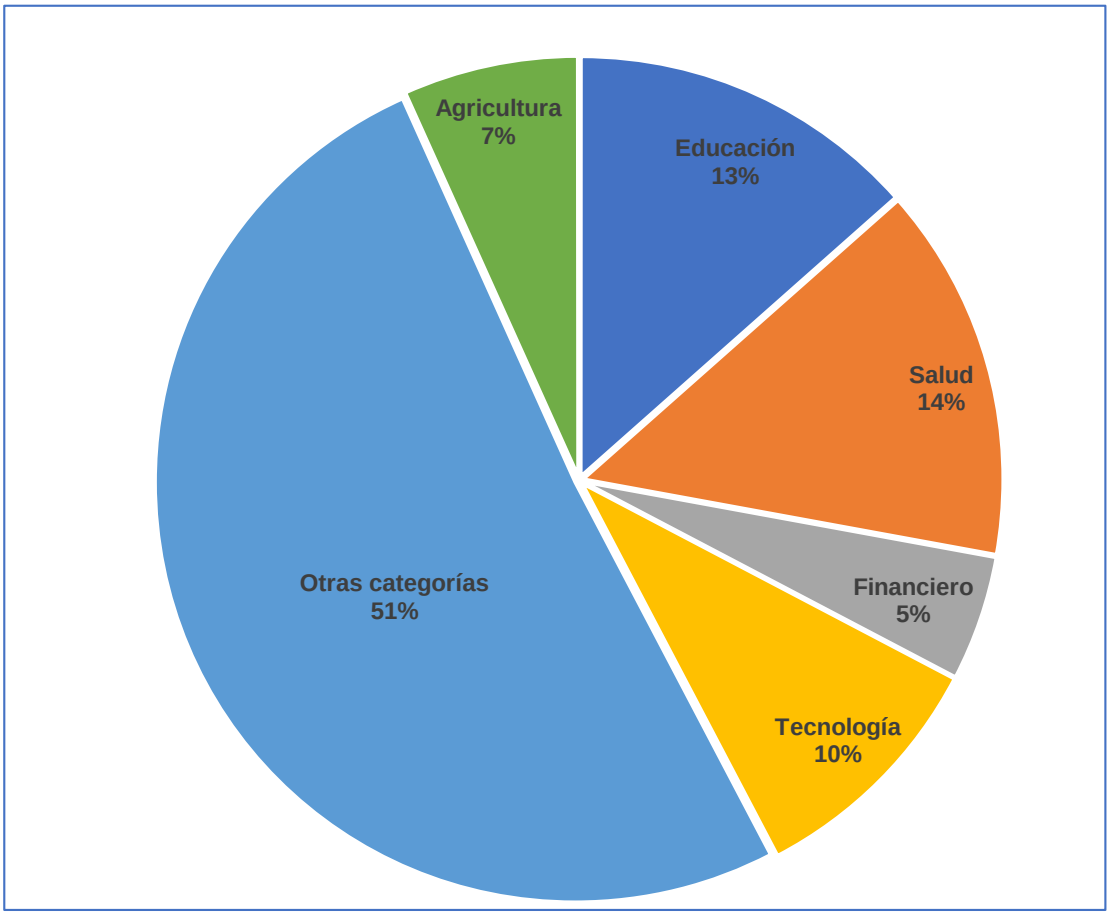


Figura 2. Porcentaje de Artículos por Áreas de Aplicación

De la revisión realizada se pudo observar que en todos los artículos evaluados tienen como objetivo obtener modelos predictivos haciendo de técnicas de clasificación, no se encontró estudios dentro del Ecuador ni fuera del país, que

hagan referencia o que nos permitan construir un modelo predictivo para determinar si una empresa debe ser inspeccionada o no, tomando como base la normativa legal vigente en el Ecuador. El proyecto tiene como finalidad realizar una propuesta que nos permita utilizar los datos que se encuentran almacenados en las diferentes bases de datos con las que cuenta el Ministerio del Trabajo. Ante la ausencia de estudios referentes al tema, es necesario realizar la propuesta de un modelo predictivo de minería de datos que sirva como apoyo a la Inspectoría de Trabajo para realizar una planificación de las inspecciones a efectuarse de una manera rápida y eficiente.

2.2. Metodología Data Warehouse

El término metodología es un conjunto de procedimientos ordenados para el logro de objetivos planteados que dirigen una investigación. Existen varias metodologías de diseño y construcción Data Warehouse.

2.2.1. Metodología Inmon

Bill Inmon define el Data Warehouse (DWH) como: “Es una colección de datos orientados al tema, integrados, no volátiles e historiados que facilitan la toma de decisiones”. (Inmon, 2005).

El enfoque es global, es decir, desarrolla todo y después se realiza el detalle, durante el desarrollo usa el esquema Entidad/Relación. El enfoque Inmon defiende la metodología descendente de trabajo “Top-Down[14]”, donde se consideran todos los

[14] Enfoque de análisis conocida como de arriba-abajo, es decir va de mayor complejidad a menor complejidad.

requerimientos y datos para el diseño del Data Warehouse y posteriormente crea los Data Marts, esto evita que existan incoherencias en la comparación de datos de los diferentes departamentos.

Con esta metodología el modelo de datos debe estar en tercera forma normal, con lo cual se evita la redundancia de datos, se mantiene la integridad, facilidad en el mantenimiento de las tablas, así como la disminución del tamaño de la base de datos. El inconveniente se refleja en las consultas que requieren queries más complejos dificultando el análisis de la información de manera directa, este problema se resuelve con la construcción de Data Marts basados en los modelos dimensionales de estrella o copo de nieve.

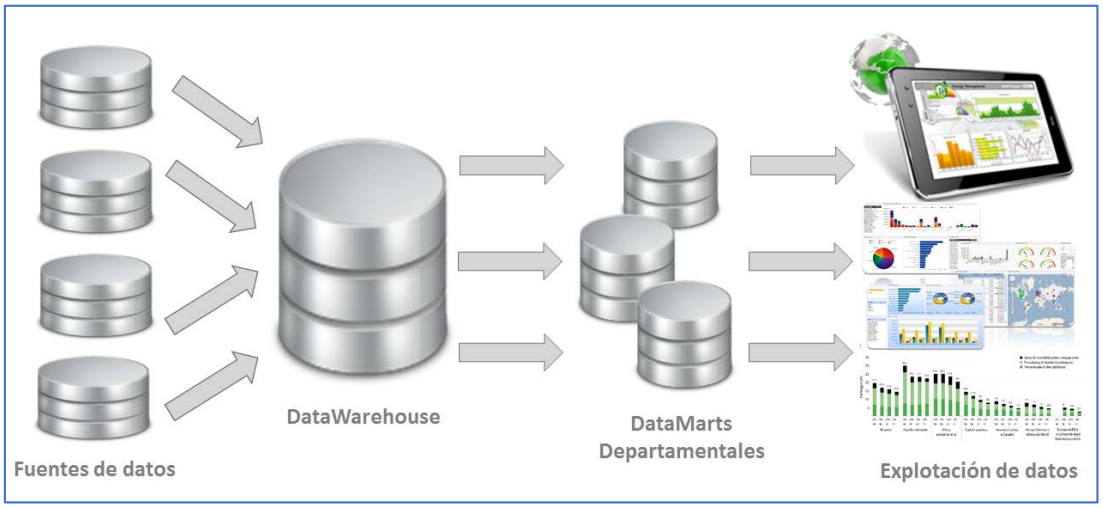


Figura 3. Metodología Inmon

Fuente: (Dertiano, 2015)

2.2.2. Metodología Kimball

Ralph Kimball define un Data Warehouse como: “Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”. (SENA, 2015). El

enfoque Kimball defiende la metodología de trabajo ascendente “Bottom-Up[15]” donde empieza por la construcción de pequeños componentes los mismos que evolucionan a modelos superiores.

La metodología se centra en la construcción de un Data Warehouse, el mismo que está conformado por Data Marts pequeños estructurados en modelo de datos dimensionales de estrella o copo de nieve, el análisis de datos se lo realiza de forma directa sin hacer uso de estructuras intermedias. La metodología se basa en cuatro principios básicos:

- Centrarse en el negocio.
- Construir una infraestructura de información adecuada.
- Crear el Data Warehouse con entregables incrementales en plazos de 6 a 12 meses.
- Entregar la solución completa con un Data Warehouse bien diseñado.

[15] Enfoque de análisis conocida como abajo-arriba, es decir va de menor complejidad a mayor complejidad

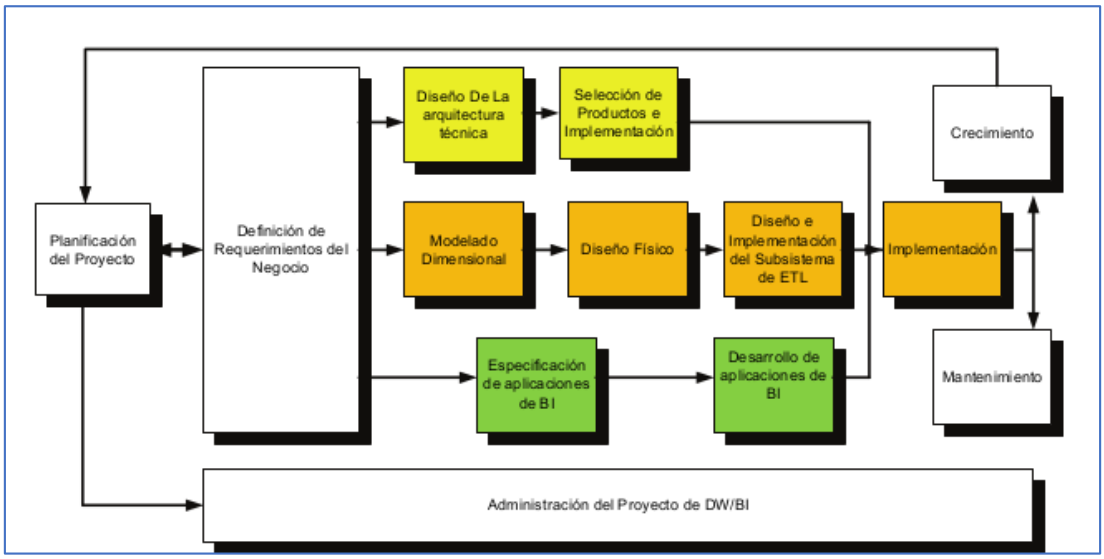


Figura 4. Tareas de la Metodología de Kimball

Fuente: *(Kimball, 2008)*

2.2.3. Metodología Hefesto

HEFESTO es una metodología propia, está fundamentada en la investigación, comparación de otras metodologías, experiencias propias en la construcción de Data Warehouse (almacenes de datos). La metodología se encuentra en constante evolución y ha tomado en cuenta la retroalimentación de quienes han utilizado la metodología. Se comienza con la recolección de las necesidades de los usuarios, luego se identifican los indicadores y perspectivas del análisis que se obtuvo en las entrevistas realizadas a los usuarios, con esta información se construirá el modelo conceptual de datos, se analizarán los OLTP[16] con el fin de determinar la construcción de los indicadores, señalar las correspondencias con las fuentes de datos y de esta manera seleccionar los campos que se utilizarán en cada perspectiva. Realizado las tareas anteriormente descritas se procederá con la construcción del modelo lógico del repositorio de información para

[16] Online Transaction Processing (Procesamiento de Transacciones en Línea)

definir el tipo de esquema que se utilizará, construido el modelo lógico se realiza el proceso ETL[17] (Extracción de datos desde diferentes fuentes de información, Transformación de datos en esta etapa se realiza una limpieza de datos para luego integrar, filtrar y depurar, y por último la Carga de Datos). La metodología se puede resumir en la Figura 5.

Figura 5. Metodología HEFESTO

Fuente: (Darío, 2010)

[17] Extract, Transform and Load (Extraer, Transformar, Cargar)

2.3. Minería de Datos

Existen diferentes definiciones para el término Minería de Datos tales como:

La tarea no trivial de extraer información implícita, previamente desconocida y potencialmente útil de bases de datos. (González, 2016)

Una forma simple de definir Minería de Datos sería que es el análisis y tratamiento de un gran volumen de datos, para extraer patrones de comportamiento e información relevante que servirá como apoyo para la toma de decisiones.

Las principales tareas en la minería de datos son:

- **Agrupación**. Encuentra grupos en los datos, sin necesidad de hacer uso las estructuras observadas en los datos.
- **Aprendizaje de reglas de asociación.** Relaciones que existen entre las variables.
- **Regresión.** Encontrar alguna función que permita realizar el modelo de los datos con el menor margen de error.

Figura 6. Proceso Minería de Datos

2.4. Métodos de Minería de Datos

Los métodos de Minería de Datos ayudan a clasificar la información, la finalidad es formar hipótesis y encontrar información relevante que en muchas ocasiones no es visible a simple vista. Se clasifican en:

- Métodos descriptivos
- Métodos predictivos

2.4.1. Métodos Descriptivos

Los métodos descriptivos ayudan a encontrar patrones y tendencias de datos actuales, permiten formar grupos de datos que no son conocidos con anterioridad, todas las variables son tratadas en el mismo nivel y la información se obtiene de los datos que se tiene en ese momento. Se clasifican en:

- **Clasificación.** Predice un nuevo valor basado en los datos disponibles.
- **Categorización.** Para cada entrada existen una o más correspondencias para determinar la salida.
- **Preferencias.** Obtiene una lista de preferencias a partir de datos previamente ordenados.
- **Regresión**. A cada entrada le corresponde un único valor de salida, la función de regresión debe ser capaz de predecir nuevos datos.

2.4.2. Métodos Predictivos

Los métodos predictivos utilizan el entrenamiento de un modelo con diferentes datos para predecir valores desconocidos o futuros de otras variables, partiendo de los datos ingresados. Las técnicas predictivas también se desarrollan en dos fases:

- Entrenamiento, se construye un modelo con un subconjunto de datos con etiqueta conocida.
- Pruebas del modelo con el resto de datos.

Se clasifican en:

- **Agrupamiento.** Obtiene elementos similares entre sí agrupados en conjuntos.
- **Correlaciones.** Tiene como punto de partida un conjunto de atributos de un elemento, cuya finalidad es determinar si existe dos o más atributos relacionados.
- **Asociación.** Utiliza reglas de asociación.
- **Detección de valores atípicos**. Detecta valores que no son similares de ninguna forma.

2.5. Técnicas de Minería de Datos

Las técnicas de minería de datos son herramientas que se aplican en diferentes áreas las mismas que ayudan al análisis de grandes cantidades de datos, con el fin de encontrar patrones de comportamiento de la información.

2.5.1. Redes Neuronales

Es un método de aprendizaje descriptivo y predictivo, el proceso es automático inspirado en la forma que funciona el cerebro humano aprendiendo del pasado y de experiencias, generando conocimiento para resolver los nuevos problemas. Los nodos se encuentran interconectados a través de sinapsis[18] entre sí para producir una salida, la cual determina el comportamiento de la red, esta técnica detecta y aprende patrones complejos y características en los datos Una ventaja es que procesa información en tiempo real y en paralelo, uno de los principales inconvenientes se presenta el momento

[18] Unión entre dos neuronas

de acceso y comprensión del modelo que se genera ya que se dificulta la extracción de reglas.

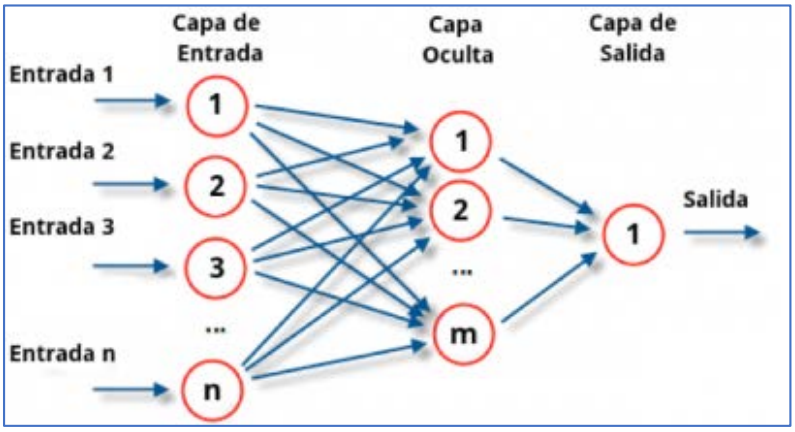


Figura 7. Red Neuronal Multicapa

Fuente: (Alfredo, 2017)

2.5.2. Regresión Lineal

La técnica de Regresión Lineal es un método matemático cuyo estudio es de tipo analítico, se utiliza para establecer relaciones entre los datos, se crea un modelo que relaciona una o varias variables dependientes con un conjunto de variables independientes y una constante.

2.5.2.1. Regresión Lineal Simple

Es la técnica de regresión más simple los datos son modelados usando la línea recta, utiliza dos variables una aleatoria, una variable de respuesta (y) que se encuentra en función lineal de una variable aleatoria (x). La ecuación se representa:

$$y = a + bx$$

donde **y:** valor estimado de la variable dependiente.

x: valor que toma la variable independiente.

a: punto en el que la recta corta al eje y

b: pendiente de la recta

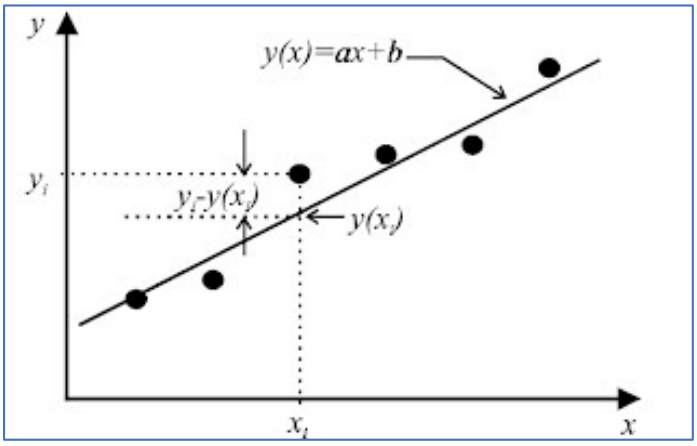


Figura 8. Modelo Regresión lineal Simple

Fuente: (Anónimo, 2017)

2.5.2.2. Regresión Lineal Múltiple

La técnica de regresión lineal múltiple es una técnica estadística extensión de la regresión lineal simple, utiliza más de una variable para la predicción de las variables explicativas. El modelo de regresión lineal múltiple se representa como:

$$y = b_0 + b_1 x_1 + \cdots \ldots \ldots + b_n x_n$$

donde **y:** valor estimado de la variable dependiente.

$\mathbf{x_i}$**:** valores que toman las variables independientes.

$\mathbf{b_i}$**:** coeficientes del modelo, son calculados para minimizar los residuos.

2.5.3. Naive Bayes

La técnica de naive bayes es un método predictivo, considerado como un algoritmo de clasificación, se basa en el Teorema de Bayes, se utiliza para generar modelos de

minería de datos de una forma rápida, que permitan predecir posibles resultados, encuentra asociaciones y relaciones haciendo uso de datos históricos.

2.5.4. Árboles de Decisión

Un árbol de decisión es un conjunto de condiciones que tienen una organización jerárquica, permite establecer la decisión final. Los árboles de decisión son apropiados para procedimientos médicos, legales, comerciales, estratégicos, matemáticos, lógicos, entre otros.

La técnica de árboles de decisión (random forest[19]), utiliza un algoritmo de clasificación, la representación e interpretación es sencilla y fácil, se puede expresar como reglas de decisión, permitiendo analizar decisiones secuenciales usando los resultados y probabilidades asociadas.

Los algoritmos de árboles de decisión más representativos son: el ID3[20] y el C4.5[21] desarrollados por Ross Quinlan. Un árbol de decisión tiene las siguientes características:

- Planteamiento del problema desde diferentes perspectivas.
- Análisis completo de todas las soluciones posibles.
- Proporciona un esquema para estimar el costo del resultado y la probabilidad de uso.
- Escoge las mejores decisiones basándose en la información existente y los mejores supuestos.
- Análisis de las alternativas, los sucesos, las probabilidades y los resultados.

[19] Técnica de minería de datos
[20] Algoritmo Indcution of Decision Trees.
[21] Extensión del algoritmo ID3

2.5.5. Máquina de Soporte Virtual (SVM)

La técnica SVM[22], funciona como un clasificador lineal que separa los datos en dos clases, encuentra la superficie ideal que maximiza el margen entre los vectores de soporte[23]. Si las dos clases no son linealmente separables, la técnica SVM, usa la función de kernel[24] para proyectar los puntos de los datos en un espacio dimensional superior.

2.5.6. Regresión Logística

La Regresión Logística es una técnica de clasificación estadística usada para realizar modelos de minería de datos que tienen resultados binarios (valores entre 0 y 1), es considerada como una extensión de la regresión lineal, esta técnica es flexible ya que toma cualquier tipo de variable como entrada, pueden ser variables continuas o categóricas. Es ampliamente usada para temas médicos y sociales. El algoritmo de regresión logística crea una línea curva, como se observa en la Figura 9, la curva no pasa sobre valor 1 ni bajo el valor 0.

[22] Técnica de minería de datos desarrollado por Vladimir Vapnik y su grupo de colaboradores en los Laboratorio Bell AT&T
[23] Subconjunto de datos que representan los puntos más cercanos de las clases, que definen la posición del hiperplano ideal.
[24] Función matemática que busca la separabilidad lineal.

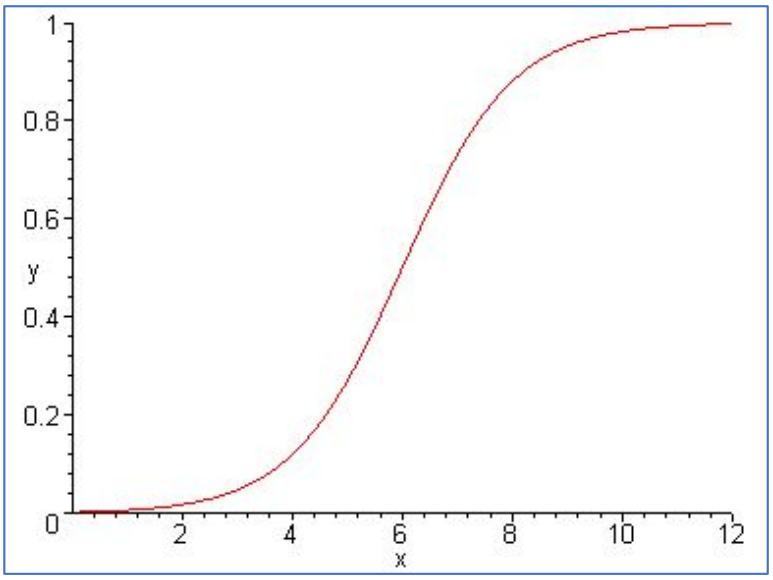


Figura 9. Curva de Regresión Logística

Fuente: (Martínez, 2018)

2.5.7. Algoritmo A priori

El algoritmo Apriori es un algoritmo predictivo de minería de datos pertenece al grupo de reglas de asociación, trabaja con bases de datos transaccionales que almacenan grandes volúmenes de datos, toma como base el conocimiento previo, consiste en hallar, "conjuntos de ítems frecuentes", con los cuales se generan reglas de asociación. Se encuentra constituido por:

- item-sets (conjuntos de valores que se repiten frecuentemente)
- Soporte
- Confianza o confidencia

Comúnmente este algoritmo se aplica en el análisis de transacciones comerciales y en problemas de predicción. El tipo de dato que requiere el algoritmo tanto de entrada y salida son datos categóricos. (Corral, 2014)

2.6. Herramientas para Minería de Datos

Existen muchas herramientas de minería de datos tanto software de código abierto como propietario, se describe algunas de ellas.

2.6.1. Knime

KNIME (Konstanz Information Miner) se encuentra escrito en lenguaje JAVA[25], se basa en Eclipse[26] y puede usar métodos para soportar plugins[27], lo cual permite a los usuarios añadir nuevos componentes de texto, imagen, procesamientos de series de tiempo, adicional se integra con proyectos de código abierto como Lenguaje R[28], Weka[29].

KNIME es una plataforma de código abierto de fácil uso y comprensible para integración de datos, procesamiento, análisis, y exploración. Los flujos de datos son creados de manera visual, permite seleccionar los pasos de análisis o ejecutar todos, analizar los resultados, modelos y vistas interactivas. (Jmacoe, 2018)

2.6.2. RapidMiner

RapidMiner es una herramienta multiplataforma de análisis de datos que nació en el año 2006 de código abierto con licencia GPL[30] desarrollado en java, permite desarrollar procesos de análisis de datos a través de encadenamiento de operadores visualizados en un entorno gráfico, provee un entorno de aprendizaje para el proceso de minería de

[25] Lenguaje de programación.
[26] Plataforma para desarrollar entornos de desarrollo integrados.
[27] Aplicación que añade una nueva funcionalidad al software.
[28] Lenguaje de programación de análisis estadístico.
[29] Software de aprendizaje automático y minería de datos.
[30] Genral Public License (Licencia Pública General).

datos, el software es utilizado para aplicaciones de negocios, investigación, educación, creación de prototipos y desarrollo de aplicaciones.

RapidMiner incluye procedimientos de aprendizaje automático y minería de datos como:

- Carga y transformación de datos (Proceso ETL).
- Pre-procesamiento de datos y visualización.
- Análisis predictivo.
- Evaluación y despliegue.

2.6.3. Weka

Weka (Waikato Environment for Knowledge Analysis - Entorno para Análisis del Conocimiento de la Universidad de Waikato), desarrollado en lenguaje Java, nació en la Universidad de Waikato en el año 1993, es un software de código abierto distribuido bajo licencia GNU-GPL, una plataforma de aprendizaje automático y minería de datos, recoge herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, agrupados en un entorno gráfico que facilita el acceso a las funcionalidades. Contiene herramientas para diferentes tareas:

- Preprocesamiento de Datos
- Algoritmos de Clasificación
- Algoritmos de Segmentación (Clustering)
- Algoritmos de Asociación: Encuentra relaciones de asociación entre variables.
- Selección de Atributos: Weka es capaz de buscar las mejores variables del modelo.

- Visualización

2.6.4. Lenguaje R

R es un lenguaje muy utilizado por los estadistas distribuido de forma gratuita, uso libre y código abierto bajo licencia GNU-GPL, se enfoca en análisis de datos estadísticos que son usados en la minería de datos, maneja grandes volúmenes de datos, es independiente del hardware y software, permite la carga de bibliotecas y paquetes con diferentes funciones.

2.7. Metodología CRISP-DM

La metodología CRISP-DM (siglas en inglés Cross Industry Standard Process for Data Mining). El ciclo de vida para un proyecto de minería de datos consiste en seis fases que se muestran en la siguiente figura.

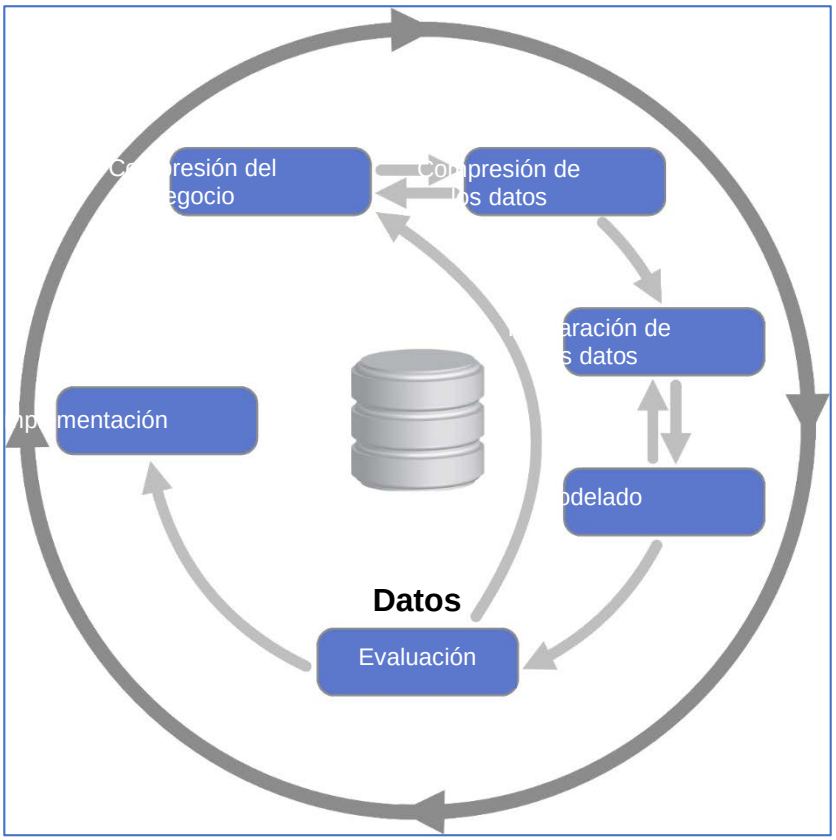


Figura 10: Ciclo de Vida – Metodología CRISP-DM

Fuente: (Román, 2016)

De acuerdo a la Figura 10 se observa que las fases no siguen una secuencia estricta, permitiendo realizar movimientos entre fases hacia adelante y hacia atrás, el resultado de cada fase establece que fase se debe ejecutar después.

El proyecto de minería de datos no finaliza con el despliegue de la solución, la información encontrada a través del proceso y la solución encontrada puede ocasionar nuevas iteraciones para el modelo, los procesos de análisis subsecuentes se favorecerán de las experiencias previas. La metodología CRISP-DM se encuentra estructurada en seis (6) fases:

- Comprensión del negocio

- Comprensión de los datos
- Preparación de los datos
- Modelado
- Evaluación
- Implementación o despliegue

2.7.1. Comprensión del Negocio

Esta es la primera fase de la metodología se puede considerar como la más importante, reúne varias tareas para la comprensión de los objetivos enfocados en una perspectiva empresarial, es necesario entender el problema que se desea resolver. Las tareas que se realiza en esta fase son:

- **Establecer objetivos del negocio:** Esta primera tarea se desarrolla para determinar el problema que se va a resolver, el por qué se necesita utilizar minería de datos y definir los criterios de éxito que pueden ser cualitativo y cuantitativo.
- **Evaluación de la situación:** En esta fase se definen los requisitos del problema y el estado de la situación antes de iniciar el proceso de minería de datos, se deben considerar:
    - ¿Cuál es el conocimiento previo disponible acerca del problema?
    - ¿Se tiene los suficientes datos para resolver el problema?
    - ¿Cuál es la relación coste beneficio de la aplicación de minería de datos?
- **Determinación de los objetivos de minería de datos:** Esta fase representa los objetivos del negocio de acuerdo a las metas para el proyecto de minería de datos.

- **Producción de un plan de proyecto:** Esta es la última tarea de la fase de Comprensión del Negocio, la finalidad es desarrollar el para el proyecto, los pasos a seguir y las técnicas que se emplearan.

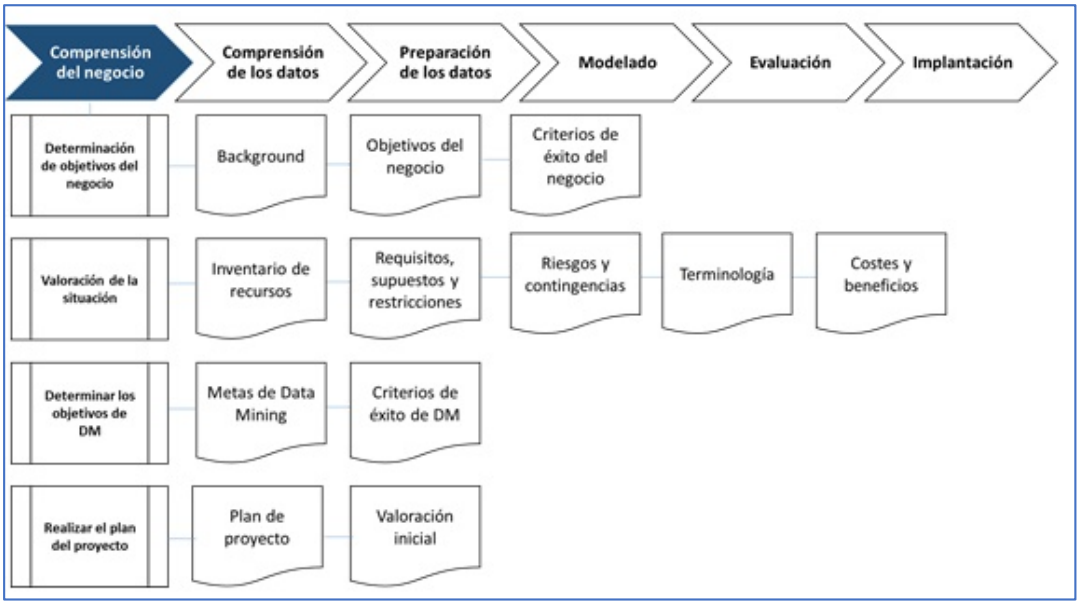


Figura 11. Fase 1. Comprensión del Negocio

Fuente: (Díaz, Smartbase Group, 2016)

2.7.2. Comprensión de los Datos

La segunda fase comprende la recolección inicial de datos, se tiene una primera visualización del problema a resolver, permite identificar la calidad de los datos y establecer relaciones existentes entre ellos que definirán las primeras hipótesis. Las principales tareas de esta fase son:

- **Recolección de datos iniciales:** En esta la primera tarea se preparará los datos para el procesamiento. Tiene como objetivo, crear informes con un conjunto de datos obtenidos, técnicas utilizadas en la recolección de los datos iniciales, problemas y soluciones que se presenten en este proceso.

- **Descripción de los datos:** Una vez que se han adquirido los datos iniciales, deben ser descritos, en este proceso se establece la cantidad de datos como: el número de registro y atributos por registro, identificación, significado de cada atributo y la representación del formato inicial.
- **Exploración de datos:** Esta tarea tiene como fin el descubrir una estructura de forma general para los datos, como resultado se obtiene un informe de esta tarea.
- **Verificación de la calidad de los datos:** Esta tarea realiza las verificaciones de los datos y establece la consistencia de los valores, el volumen y forma de distribución de los valores nulos, se encuentra los valores fuera de rango que producen ruido en el proceso, el objetivo es asegurar que los datos sean completos y corregir los datos si fuese necesario.

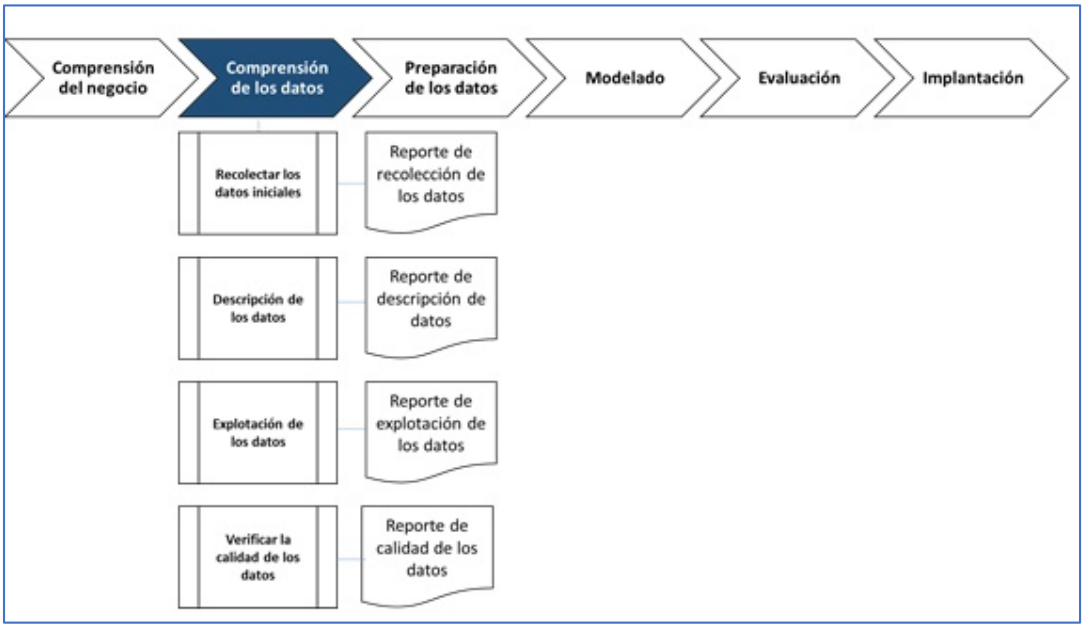


Figura 12. Fase 2: Comprensión de los Datos

Fuente: (Díaz, Smartbase Group, 2016)

2.7.3. Preparación de los Datos

En la fase de preparación de los datos se ejecutan todas las actividades para la construcción de la base de datos final donde se ingresarán los datos iniciales. Las tareas que se incluyen son:

- **Selección de datos:** En esta tarea se selecciona un subconjunto de datos, los mismos que fueron obtenidos en la fase de comprensión de datos, estos serán usados para el análisis, apoyándose en criterios establecidos en las fases anteriores.
- **Limpieza de los datos:** Esta tarea es el complemento de la tarea de selección de datos, se puede decir que esta tarea es la que consume más tiempo y mayor esfuerzo, todo depende de las técnicas de minería de datos elegidas, estas técnicas ayudan a la obtención de datos de calidad, los mismos que serán preparados ser utilizados en la fase de modelamiento. Las técnicas que se usan son:
    - Normalización de los datos
    - Discretización de campos numéricos
    - Tratamiento de valores ausentes o lejanos
    - Disminución del volumen de datos
- **Estructuración de los datos:** Esta tarea realiza operaciones de para la preparación de datos, estas operaciones son creación de nuevos atributos tomando como punto de partida los atributos existentes, transformación de valores para atributos existentes integración de nuevos registros. (Díaz, Smartbase Group, 2016)

- **Integración de los datos:** Implica la creación de nuevos registros o valores, a partir de los datos seleccionados.
- **Formateo de los datos:** Esta tarea consiste en realizar modificaciones sintácticas de los datos sin cambiar su significado, facilitando el uso de alguna técnica de minería de datos.

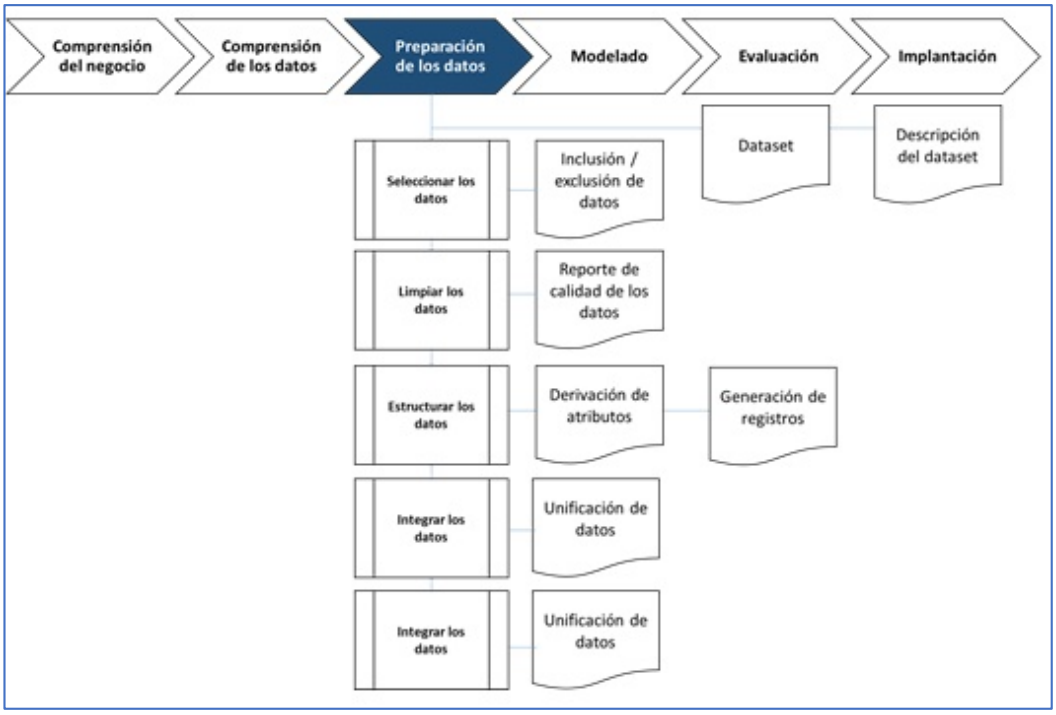


Figura 13. Fase 3. Preparación de los Datos

Fuente: (Díaz, Smartbase Group, 2016)

2.7.4. Modelado

En esta fase, se seleccionan y aplican diversas técnicas de modelado apropiadas para el proyecto de minería de datos, Existen varias técnicas de minería de datos que se pueden aplicar al mismo tipo de problema, muchas veces es necesario regresar a fase de preparación de datos. Las tareas que se incluyen en esta fase son:

- **Selección de la técnica de modelado:** Esta tarea nos ayuda a seleccionar la técnica de minería de datos adecuada para el tipo de problema a resolver, considerando el objetivo específico del proyecto.
- **Generación del plan de prueba:** Construido el modelo, se genera un procedimiento para probar la calidad y validez del modelo. Se toman dos conjuntos de datos, el uno servirá como entrenamiento y el otro de prueba, se construye el modelo tomando como base el conjunto de entrenamiento que mide la eficacia del modelo que se genera con el conjunto de datos de prueba.
- **Construcción del Modelo:** Una vez que la técnica de minería de datos ha sido seleccionada, se ejecuta la herramienta de modelado sobre el conjunto de datos que fueron preparados con anterioridad para generar dependiendo de cada caso un modelo o más. Las técnicas de modelado permiten crear el modelo de acuerdo a ciertas características que son establecidas por parámetros, estos parámetros pueden ser ajustados, la selección de los parámetros es un proceso repetitivo que se basa en los resultados obtenidos, que deben ser interpretados y su rendimiento justificado.

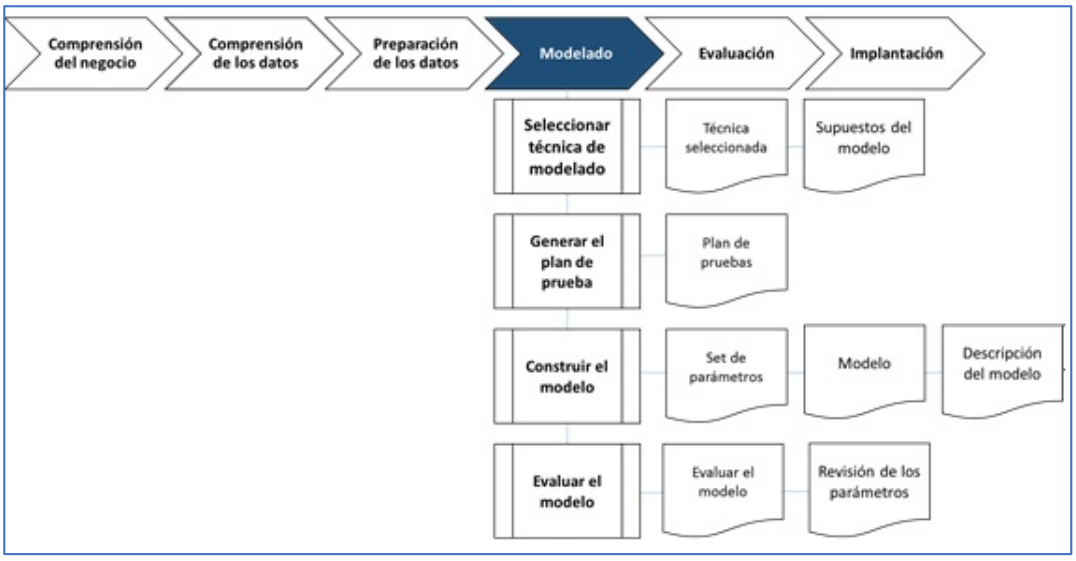


Figura 14. Fase4. Modelado

Fuente: (Díaz, Smartbase Group, 2016)

2.7.5. Evaluación

En esta fase, ya se tiene el modelo o modelos construidos, es importante evaluar y analizar los pasos realizados para crearlo, de esta manera se asegura que el modelo cumpla con los objetivos del negocio, en este paso se puede determinar si hay algo del negocio que no ha sido considerado. Finalizada la fase de evaluación, podríamos tomar una decisión del uso o no de los resultados del proyecto de minería de datos. Las principales tareas en esta fase son:

- **Evaluación de los resultados:** Se evalúa el modelo en relación a los objetivos del negocio, esta fase ayuda a determinar si existe alguna razón del negocio en la cual es modelo es deficiente, se recomienda de ser posible verificar el modelo con datos de un problema real.
- **Revisión del proceso:** Se califica el proceso completo del proyecto de minería de datos, con el fin de identificar elementos a ser mejorados.

- **Determinación de futuras fases:** Si con la ejecución de las fases anteriores se determina que los resultados generados han sido satisfactorios se puede seguir con la siguiente fase de implementación, si los resultados no fueran satisfactorios se puede decidir por ejecutar otra iteración con parámetros diferentes a los que se utilizó anteriormente y empezar desde la fase de preparación de datos o de modelado según sea el caso.

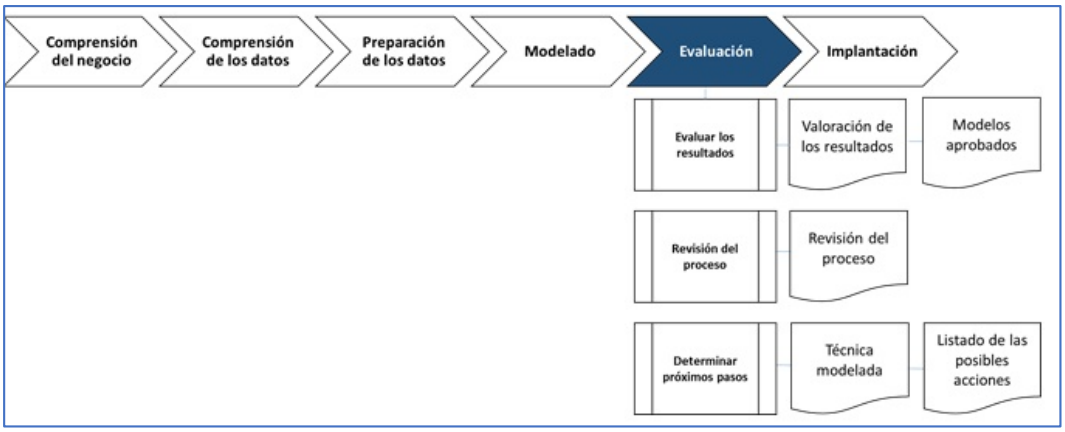


Figura 15. Fase 5. Evaluación

Fuente: (Díaz, Smartbase Group, 2016)

2.7.6. Implementación o Despliegue

Esta es la última fase de la metodología, el conocimiento que se ha adquirido con el modelo construido y validado tiene que ser organizado y presentado de una forma comprensible al usuario. Las tareas que intervienen en esta fase son:

- **Plan de implementación:** Se toma los resultados obtenidos en la fase de evaluación y presenta la estrategia a implementarse. Si se ha reconocido un procedimiento para la creación del modelo, este debe ser documentado para poder realizar la implementación.

- **Monitorización y Mantenimiento:** Preparación de las estrategias de monitorización y mantenimiento que son aplicados en los modelos.
- **Informe Final:** Con esta tarea se finaliza el proyecto de minería de datos, dependiendo de la tarea de plan de implementación, se presenta un resumen del proyecto de lo que se ha logrado en los puntos más relevantes, o la exposición final de los resultados obtenidos en el proyecto de minería de datos.
- **Revisión del proyecto:** En esta tares se evalúa lo qué se realizó bien y lo que necesita ser mejorado.

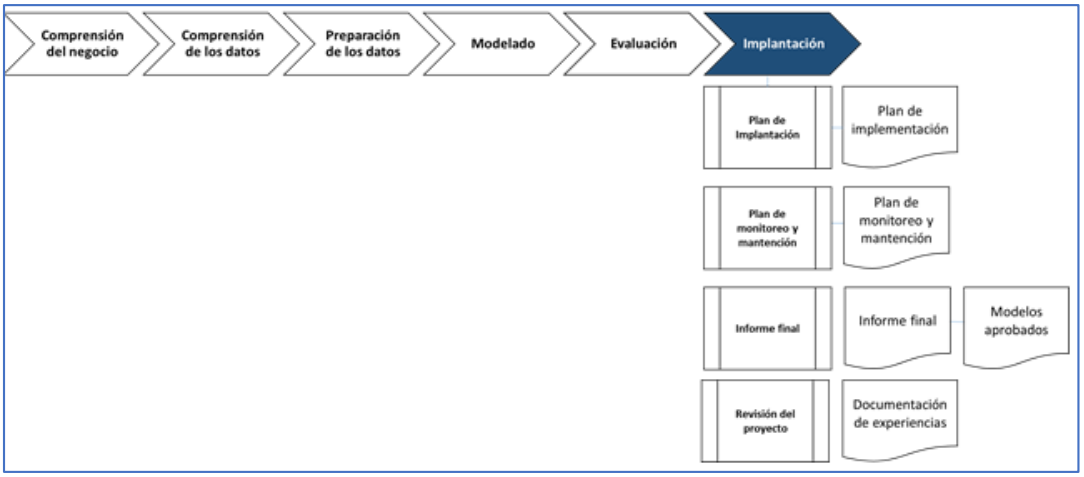


Figura 16. Fase 6. Implementación o Despliegue

Fuente: (Díaz, Smartbase Group, 2016)

CAPÍTULO III. DESARROLLO DE LA SOLUCIÓN

El presente capítulo muestra el desarrollo en dos partes, la primera parte se detalla la Construcción Data Warehouse usando la Metodología de Kimball ya que nos proporciona un conjunto de conocimientos para la construcción de un Data Warehouse haciendo uso de un enfoque ascendente, la versatilidad de esta metodología se basa en el mantenimiento constante y el intercambio de información con los usuarios finales. En la segunda parte se detalla la Construcción del Modelo de Minería de Datos utilizando la Metodología CRISP-DM.

3.1. Construcción Data Warehouse con la Metodología Kimball

3.1.1. Planificación del Proyecto

A continuación, se detalla la planificación que se elaboró para la construcción del Data Warehouse.

La metodología Kimball se escogió en base a la comparativa realizada en la Tabla1, donde se hizo un resumen de tres de las metodologías más relevantes como son: HEFESTO, INMON y KIMBALL, dando como resultado que la metodología que se adapta a nuestro proyecto es KIMBALL.

En la Tabla 4. Se observa la planificación de cada tarea que comprende la metodología que se ha seguido para el proyecto, duración de las tareas desde cuando inician hasta el su finalización y las tareas predecesoras de acuerdo al caso.

Tabla 4.

Planificación del Proyecto

Nro.	Fase	Días
1	Planificación del proyecto	10
2	Definición de Requerimientos del Negocio	13
3	Modelado Dimensional	6
4	Diseño de la arquitectura técnica	6
5	Selección e instalación de herramientas	5
6	Especificación de aplicaciones BI	2
7	Diseño Físico	20
8	Diseño e Implementación ETL	21
9	Exploración de Herramientas BI	15
Total en Días		**98**

3.1.2. Definición de Requerimientos del Negocio

El Ministerio de Trabajo es una institución pública que rige las políticas de trabajo y empleo en el sector privado y público, para nuestro caso nos centramos en el sector privado, para mejorar los tiempos de verificación que las empresas cumplan con la Normativa Legal Vigente. Para el caso de estudio se apoyará a la Dirección de Control e Inspecciones en el área de Inspectoría de Trabajo. De las conversaciones mantenidas con el área funcional se pudo llegar a determinar los siguientes requerimientos:

Tabla 5.

Requerimientos del Negocio

Tema	Requerimiento	Proceso de Negocio	Observación
Planificación inspecciones a empresas de acuerdo a la ubicación	Analizar información histórica en el registro de contratos	Registro de Contratos	Por provincia, por región.
	Analizar información histórica en actas de finiquito.	Registro de Actas de Finiquito	Por provincia, por región.

CONTINÚA

	Analizar información histórica en trámites y boletas únicas	Registro de trámites y boletas únicas	Por provincia, por región
	Analizar los motivos de salida de los trabajadores de las empresas	Registro de Actas de Finiquito	Por motivo de salida
	Analizar estabilidad laboral de los empleados	Registro de Contratos	Por actividad económica
	Analizar los factores de incumplimiento de las empresas	Registro de Contratos, Actas de Finiquito, Trámites y Boletas	Por provincia

3.1.3. Diseño de la arquitectura técnica

El proceso inicia con la recolección de datos, para el caso de estudio los datos se obtienen bases de datos de los sistemas SAITE, SINACOI y SGI, las mismas que pertenecen Ministerio del Trabajo. Antes de realizar el ETL[31] se debe realizar el proceso de limpieza de datos, lo que implica cambios de formatos, validación de los datos, etc.

El proceso de ETL, se realizó con la herramienta Pentaho PDI[32], los datos que se encuentran en la base de datos resultado del ETL serán explotados y presentados en un reporte elaborado en Tableau[33].

[31] Proceso de Extracción, Transformación y Carga.
[32] Pentaho Data Integration, herramienta que permite realizar el proceso de ETL.
[33] Herramienta analítica visual e inteligencia de negocios que ayuda a la toma decisiones.

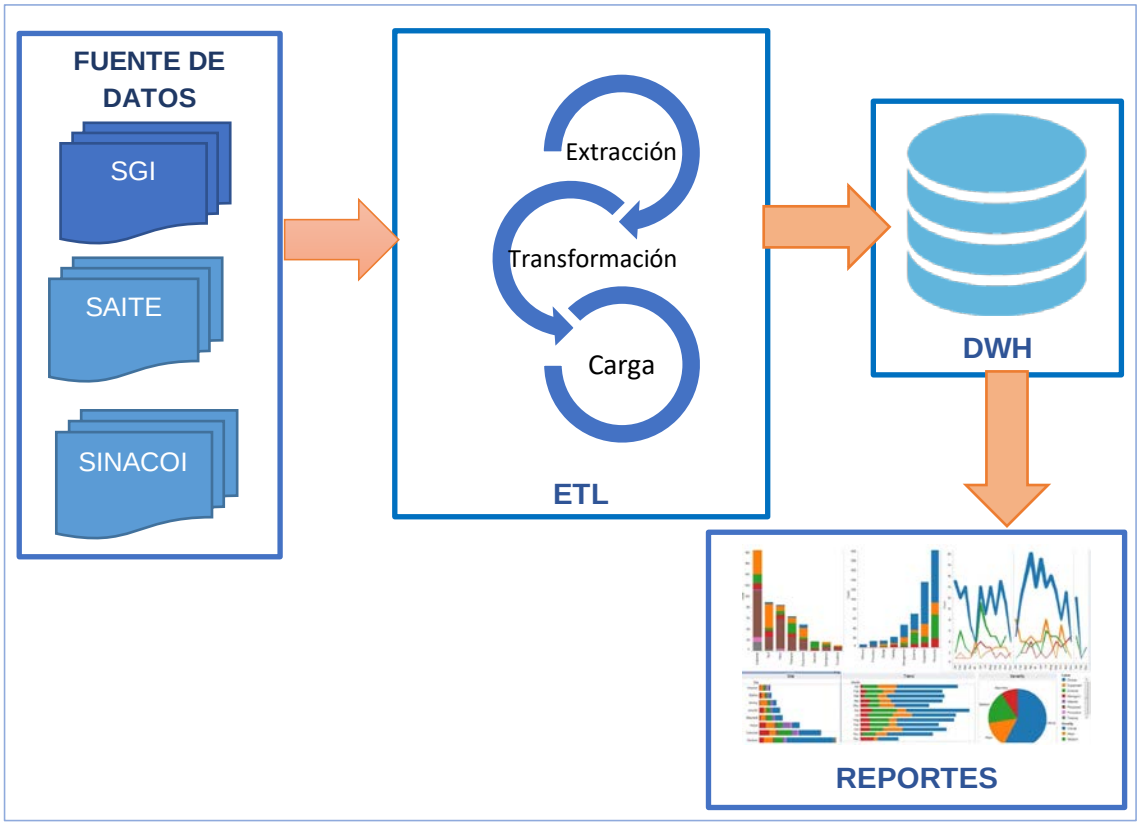


Figura 17. Diseño de la Arquitectura Técnica

3.1.4. Selección e Instalación de Herramientas

En la Figura 16., se muestra las herramientas seleccionadas para construir el Data Warehouse.

Figura 18. Herramientas Seleccionadas

Base de Datos

- **Postgres:** Almacena los datos resultado del proceso ETL.

Modelado de Datos

- **Power Designer:** Permite crear el modelo lógico de la base de datos

ETL

- **PDI:** Pentaho Data Integration se encarga de realizar el proceso ETL (Extracción, Transformación y Carga de Datos).

Inteligencia de Negocios

- **Tableau:** Sirve para la explotación de los datos que se ha obtenido en el Data Warehouse a través de dashboards[34] y reportes.

3.1.5. Modelado Dimensional

La tarea de modelado dimensional para una mejor comprensión se dividió en secciones.

En el Anexo A se describe el estándar utilizado para la nomenclatura en nombres de variables, bases de datos, modelos, esquemas, tablas de hechos, dimensiones y campos, que conforman el data warehouse.

3.1.5.1. Estructura de Datos

Para la estructura de datos se ha utilizado el modelo Estrella una de las principales razones es la sencillez de su estructura se puede afirmar que es el modelo más utilizado, la extracción de datos es rápida, la información se encuentra en cada una de las dimensiones.

[34] Representación en forma visual de indicadores que apoya la toma de decisiones.

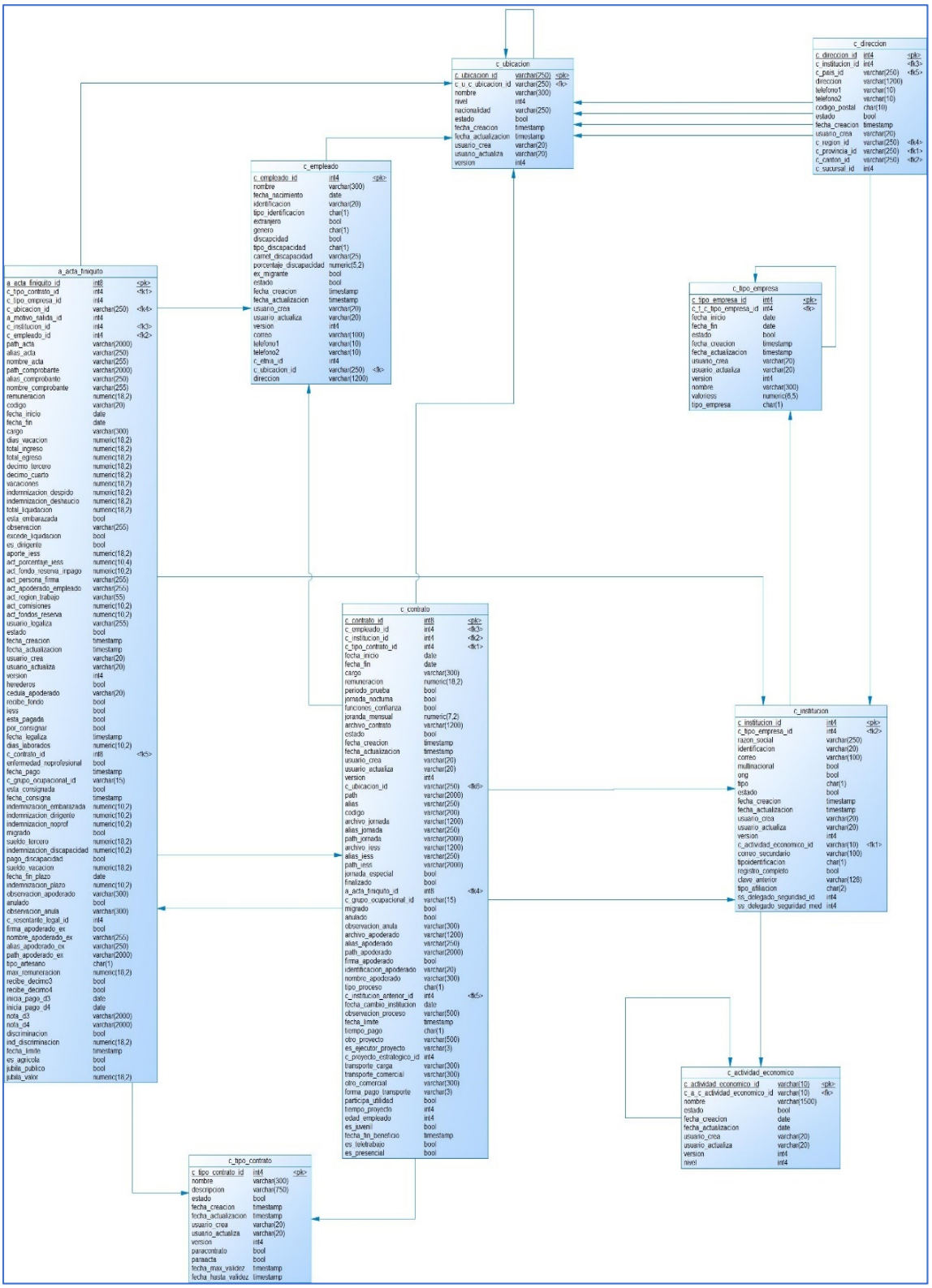


Figura 19. Modelo Entidad Relación Sistema SAITE

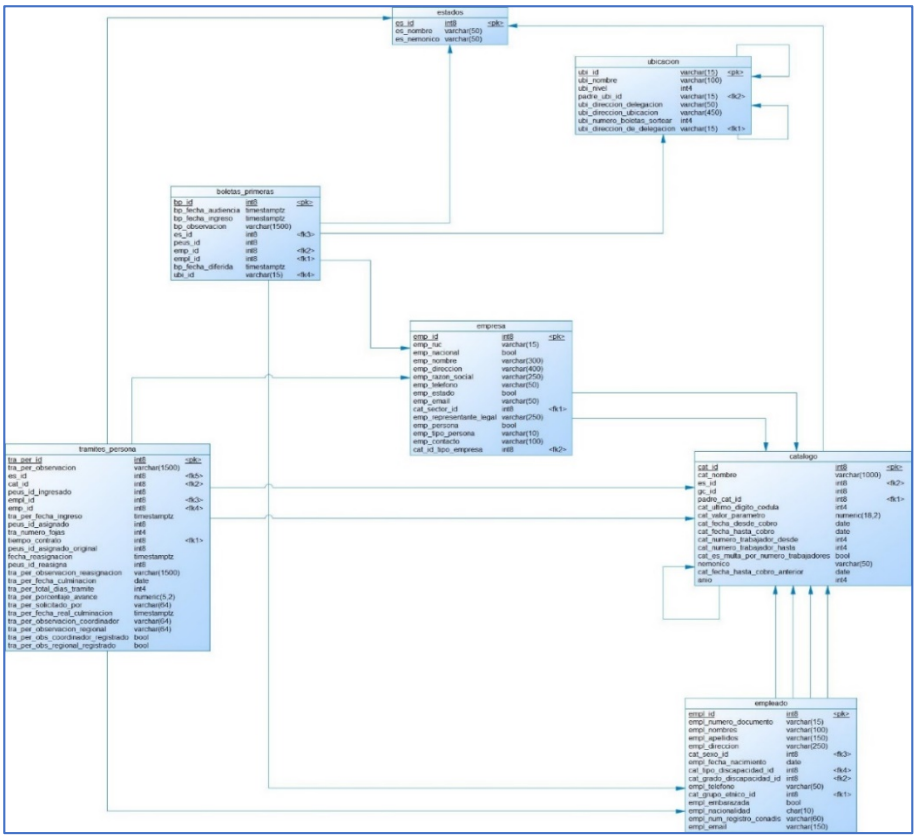


Figura 20. Modelo Entidad Relación Sistema SINACOI

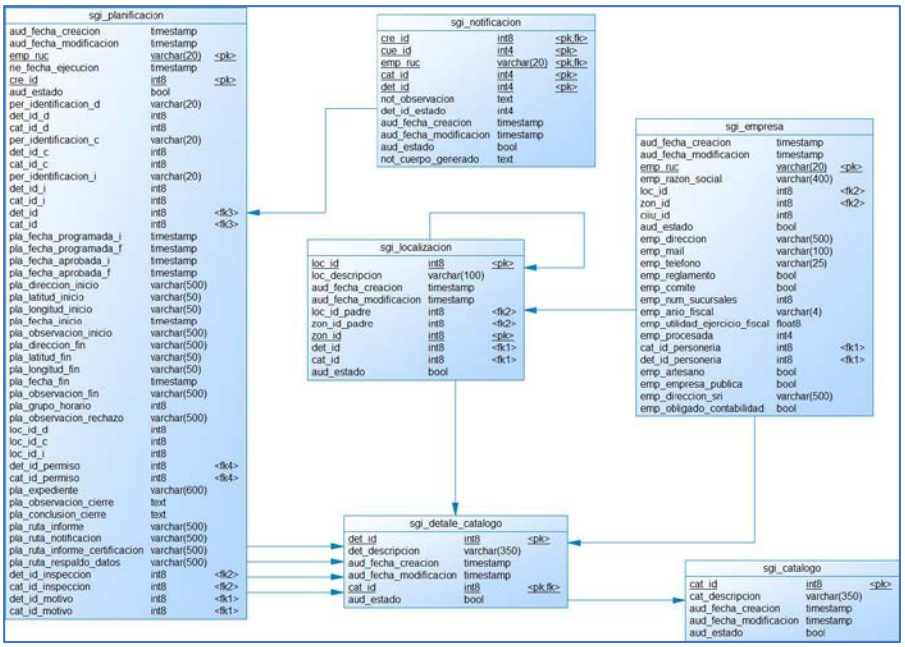


Figura 21. Modelo Entidad Relación Sistema SGI

3.1.5.2. Tablas de Hechos y Dimensiones

El modelado se dividirá en dos partes, tablas de hechos y las dimensiones.

3.1.5.2.1. Tablas de Dimensiones

- **Dimensión Discapacidad** (dim_discapacidad): Contiene información de los tipos de discapacidades registradas.

Tabla 6.

Dimensión Discapacidad

Campo	Tipo	Tamaño	Llave	Descripción
sk_discapacidad	integer		PK	Clave única
ddis_tipo_discapacidad	varchar	100		Tipo de discapacidad (Auditiva, Física, Intelectual, Lenguaje, Psicológica, Visual.)
ddis_fecha_carga	timestamp			Fecha carga de los registros.

- **Dimensión Género (dim_genero):** Contiene información de los tipos de género.

Tabla 7.

Dimensión Género

Campo	Tipo	Tamaño	Llave	Descripción
sk_genero	integer		PK	Clave única del género.
dgen_nombre	varchar	100		Descripción del género.
dgen_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Etnia (dim_etnia):** Contiene la información de los tipos de etnias.

Tabla 8.

Dimensión Etnia

Campo	Tipo	Tamaño	Llave	Descripción
sk_etnia	integer		PK	Clave única de la etnia.
detn_codigo	integer			Código de la etnia.
detn_nombre	varchar	300		Nombre de la etnia.
detn_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Actividad Económica (dim_actividad_economica):** Contiene información de los tipos de actividades económicas de acuerdo al CIIU 4.0[35]

Tabla 9.

Dimensión Actividad Económica

Campo	Tipo	Tamaño	Llave	Descripción
sk_actividad_economica	integer		PK	Clave única de la actividad económica.
dact_cod_l1	varchar	10		Código de la actividad económica nivel uno.
dact_nombre_l1	varchar	1500		Nombre de la actividad económica nivel uno
dact_cod_l2	varchar	10		Código de la actividad económica nivel dos.
dact_nombre_l2	varchar	1500		Nombre de la actividad económica nivel dos
dact_cod_l3	varchar	10		Código de la actividad económica nivel tres.
dact_nombre_l3	varchar	1500		Nombre de la actividad económica nivel tres

CONTINÚA

[35]Clasificación Industrial Internacional Uniforme otorgada por las Naciones Unidas que cataloga las actividades económicas en una serie de categorías y subcategorías, donde cada una posee un código alfanumérico.

dact_cod_l4	varchar	10	Código de la actividad económica nivel cuatro.
dact_nombre_l4	varchar	1500	Nombre de la actividad económica nivel cuatro
dact_cod_l5	varchar	10	Código de la actividad económica nivel cinco.
dact_nombre_l5	varchar	1500	Nombre de la actividad económica nivel cinco
dact_cod_l6	varchar	10	Código de la actividad económica nivel seis.
dact_nombre_l6	varchar	1500	Nombre de la actividad económica nivel seis
dact_cod_l7	varchar	10	Código de la actividad económica nivel siete.
dact_nombre_l7	varchar	1500	Nombre de la actividad económica nivel siete
dact_nivel	varchar	10	Nivel estándar de la actividad económica.
dact_fecha_carga	timestamp		Fecha del proceso de carga de los registros.

- **Dimensión Ubicación (dim_ubicacion):** Contiene información de la ubicación geográfica.

Tabla 10.

Dimensión Ubicación

Campo	Tipo	Tamaño	Llave	Descripción
sk_ubicacion	integer		PK	Clave única de la ubicación.
dubi_cod_pais	varchar	250		Código del país.
dubi_nombre_pais	varchar	300		Nombre del país.
dubi_cod_region	varchar	250		Código de la región
dubi_nombre_region	varchar	300		Nombre de la región
dubi_cod_provincia	varchar	250		Código de la provincia
dubi_nombre_provincia	varchar	300		Nombre de la provincia

CONTINÚA

dubi_nombre_provinica_unificado	varchar	300	Nombre de la provincia unificado
dubi_cod_canton	varchar	250	Código del cantón
dubi_nombre_canton	varchar	300	Nombre del cantón.
dubi_codigo_sinacoi	integer		Código de referencia a las ubicaciones de la tabla ubicación del Sistema SINACOI.
dubi_fecha_carga	timestamp		Fecha del proceso de carga de los registros.

- **Dimensión Contrato (dim_contrato):** Contiene información general de los contratos.

Tabla 11.

Dimensión Contrato

Campo	Tipo	Tamaño	Llave	Descripción
sk_contrato	integer		PK	Clave única del contrato.
dcon_codigo	integer			Código del contrato
dcon_fecha_creacion	timestamp			Fecha de creación del contrato en el sistema.
dcon_fecha_inicio	timestamp			Fecha de inicio del contrato
dcon_fecha_fin	timestamp			Fecha de finalización del contrato.
dcon_cargo	varchar	300		Cargo del empleado del contrato.
dcon_legalizado	integer			Estado que indica si el contrato esta legalizado.
dcon_anulado	integer			Estado que indica si el contrato esta anulado.
dcon_finalizado	integer			Estado que indica si el contrato fue finalizado
dcon_juvenil	integer			Estado que indica si el contrato es juvenil.
dcon_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Tipo de Contrato (dim_tipo_contrato):** contiene información de los tipos de contratos.

Tabla 12.

Dimensión Tipo de Contrato

Campo	Tipo	Tamaño	Llave	Descripción
sk_tipo_contrato	integer		PK	Clave única del tipo de contrato.
dtic_codigo	integer			Código del tipo de contrato
dtic_nombre	varchar	300		Nombre del tipo de contrato.
dtic_descripcion	varchar	300		Descripción del tipo de contrato.
dtic_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Institución (dim_institucion):** Contiene información de las empresas.

Tabla 13.

Dimensión Institución

Campo	Tipo	Tamaño	Llave	Descripción
sk_institucion	integer		PK	Clave única de la empresa.
dins_codigo	integer			Código de institución
dins_identificacion	varchar	20		Identificador de la empresa.
dins_razon_social	varchar	250		Razón social de la empresa
dins_correo	varchar	250		Correo electrónico de la empresa
dins_tipo_identificacion	varchar	100		Tipo de registro: (RUC, Cédula, Pasaporte)
dins_tipo_representante	varchar	100		Tipo de registro: (Natural o Jurídica)

CONTINÚA

dins_tipo_institución	varchar	100	Tipo de registro: (Privada o Pública)
dins_tipo_afiliación	varchar	100	Tipo de registro: (IP, IF, IE)
dins_nivel_actividad	varchar	10	Nivel de actividad a la que pertenece la empresa
dins_fecha_carga	timestamp		Fecha del proceso de carga de los registros.

- **Dimensión Empleado (dim_empleado):** Contiene información de los empleados.

Tabla 14.

Dimensión Empleado

Campo	Tipo	Tamaño	Llave	Descripción
sk_empleado	integer		PK	Clave única del empleado.
demp_codigo	integer			Código de empleado
demp_identificador	varchar	20		Identificador del empleado.
demp_nombre	varchar	300		Nombres completos del empleado.
demp_fec_nacimiento	timestamp			Fecha de nacimiento del empleado.
demp_tipo_identificacion	varchar	300		Tipo de identificación: Ruc, Cédula, Pasaporte
demp_extranjero	integer			Estado indica si el empleado es extranjero.
demp_edad	numeric			Edad del empleado
dins_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Fecha (dim_fecha):** Contiene las fechas de cada una de las transacciones realizadas por las empresas.

Tabla 15.

Dimensión Fecha

Campo	Tipo	Tamaño	Llave	Descripción
sk_fecha	integer		PK	Clave única de la fecha.
dfec_fecha	timestamp			Fecha
dfec_dia	integer			Número que corresponde al día de la fecha.
dfec_nombre_dia	varchar	50		Nombre que corresponde al día de la fecha.
dfec_mes	integer			Número que corresponde al mes de la fecha.
dfec_nombre_mes	varchar	50		Nombre que corresponde al mes de la fecha.
dfec_anio	integer			Número que corresponde al año de la fecha.
dfec_trimestre	integer			Número que corresponde al trimestre de la fecha.
dfec_nombre_trimestre	varchar	50		Nombre que corresponde al trimestre de la fecha.
dfec_semestre	integer			Nombre que corresponde al semestre de la fecha.
dfec_nombre_semestre	varchar	50		Número que corresponde al semestre de la fecha.
dfec_dia_semana	integer			Número del día de la semana.
dfec_dia_anio	integer			Número del día del año.
dfec_semana_anio	integer			Número de la semana del año
dfec_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Tipo de Empresa (dim_tipo_empresa):** Contiene información de los tipos de empresas.

Tabla 16.

Dimensión Tipo Empresa

Campo	Tipo	Tamaño	Llave	Descripción
sk_tipo_empresa	integer		PK	Clave única del tipo de empresa.
dtip_codigo_l1	integer			Código del nivel uno del tipo de empresa

CONTINÚA

dtip_nombre_l1	varchar	300	Nombre del nivel uno del tipo de empresa
dtip_cod_l2	integer		Código del nivel dos del tipo de empresa
dtip_nombre_l2	varchar	300	Nombre del nivel dos del tipo de empresa
dtip_tipo_empresa	varchar	300	Descripción del tipo de empresa
dtip_fecha_carga	timestamp		Fecha del proceso de carga de los registros.

- **Dimensión Motivo de Salida (dim_motivo_salida):** Contiene información de los motivos de salida de los empleados de las empresas.

Tabla 17.

Dimensión Motivo de Salida

Campo	Tipo	Tamaño	Llave	Descripción
sk_motivo_salida	integer		PK	Clave única del motivo de salida.
dmot_codigo	integer			Código del motivo salida
dmot_nombre	varchar	300		Nombre del motivo de salida.
dmot_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Grupo Ocupacional (dim_grupo_ocupacional):** Contiene información de los grupos ocupacionales de acuerdo CIUO08[36].

[36] Clasificación Uniforme Internacional de Ocupaciones preparada por la Organización Internacional del Trabajo.

Tabla 18.

Dimensión Grupo Ocupacional

Campo	Tipo	Tamaño	Llave	Descripción
sk_grupo_ocupacional	integer		PK	Clave única del grupo ocupacional.
dgru_cod_l1	varchar	10		Código del nivel uno del grupo ocupacional.
dgru_nombre_l1	varchar	1500		Nombre del nivel uno del grupo ocupacional.
dgru_cod_l2	varchar	10		Código del nivel dos del grupo ocupacional.
dgru_nombre_l2	varchar	1500		Nombre del nivel dos del grupo ocupacional.
dgru_cod_l3	varchar	10		Código del nivel tres del grupo ocupacional.
dgru_nombre_l3	varchar	1500		Nombre del nivel tres del grupo ocupacional.
dgru_cod_l4	varchar	10		Código del nivel cuatro del grupo ocupacional.
dgru_nombre_l4	varchar	1500		Nombre del nivel cuatro del grupo ocupacional.
dgru_cod_l5	varchar	10		Código del nivel cinco del grupo ocupacional.
dgru_nombre_l5	varchar	1500		Nombre del nivel cinco del grupo ocupacional.
dgru_cod_l6	varchar	10		Código del nivel seis del grupo ocupacional.
dgru_nombre_l6	varchar	1500		Nombre del nivel seis del grupo ocupacional.
dgru _fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Acta de Finiquito (dim_acta_finiquito):** Contiene información de las actas de finiquito.

Tabla 19.

Dimensión Acta de Finiquito

Campo	Tipo	Tamaño	Llave	Descripción
sk_acta_finiquito	integer		PK	Clave única del acta de finiquito.
dact_codigo	varchar	300		Código del acta de finiquito
dact_fecha_creacion	timestamp			Fecha de creación del acta de finiquito
dact_fecha_inicio	timestamp			Fecha de inicio de vigencia del acta de finiquito
dact_fecha_fin	timestamp			Fecha fin de vigencia del acta de finiquito
dact_anulado	integer			Estado que indica si el acta fue anulada.
dact_consignada	integer			Estado que indica si el acta fue consignada.
dact_estado_cancelacion	integer			Estado que indica si el acta fue cancelada.
dact_cargo	varchar	300		Cargo del empledo
dact_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Boletas (dim_boletas):** Contiene información de las boletas.

Tabla 20.

Dimensión Boletas

Campo	Tipo	Tamaño	Llave	Descripción
sk_boletas	integer		PK	Clave única de la boleta.
dbol_codigo	integer			Código de la boleta
dbol_fecha_ingreso	timestamp			Fecha de ingreso de la boleta
dbol_fecha_audiencia	timestamp	100		Fecha de audiencia de la boleta
dbol_estado	varchar	100		Estado de la boleta
dbol_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Tipo Trámite (dim_tipo_tramite):** Contiene información de los tipos de trámite.

Tabla 21.

Dimensión Tipo Trámite

Campo	Tipo	Tamaño	Llave	Descripción
sk_tipo_tramite	integer		PK	Clave única del tipo de trámite.
dtip_nombre	varchar	300		Código del tipo trámite
dtip_nombre	varchar	300		Nombre del tipo de trámite.
dtip_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Dimensión Trámite (dim_tramite):** Contiene información de los trámites.

Tabla 22.

Dimensión Trámite

Campo	Tipo	Tamaño	Llave	Descripción
sk_tramite	integer		PK	Clave única del trámite.
dtra_codigo	integer			Código del trámite
dtra_fecha_ingreso	timestamp			Fecha de ingreso del trámite.
dtra_fecha_culminacion	timestamp	100		Fecha de culminación del trámite.
dtra_solicitante	varchar	300		Identificación del solicitante del trámite.
dtra_estado	varchar	100		Estado en que se encuentra el trámite
dtra_tipo	varchar	1000		Tipo de Trámite
dtra_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

3.1.5.2.2. Tablas de Hechos

- **Tabla de Hechos Contratos (fact_contratos):** Contiene información de los contratos de los empleados registrados por las empresas.

Tabla 23.

Tabla de Hechos Contratos

Campo	Tipo	Tamaño	Llave	Descripción
sk_empleado	integer		FK	Clave foránea del empleado
sk_institucion	integer		FK	Clave foránea de la empresa
sk_ubicacion	integer		FK	Clave foránea de la ubicación
sk_genero	integer		FK	Clave foránea del genero
sk_actividad_economica	integer		FK	Clave foránea de la actividad económica
sk_tipo_contrato	integer		FK	Clave foránea del tipo de contrato
sk_fecha_inicio	integer		FK	Clave foránea de la fecha
sk_fecha_fin	integer		FK	Clave foránea de la fecha
sk_etnia	integer		FK	Clave foránea de la etnia
sk_contrato	integer		FK	Clave foránea del contrato
sk_discapacidad	integer		FK	Clave foránea de la discapacidad
sk_tipo_empresa	integer		FK	Clave foránea de tipo empresa
fcon_numero	integer			Identificador (1)
fcon_sueldo	integer			Sueldo registrado en el contrato.
fcon_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Tabla de Hechos Acta de Finiquito (fact_acta_finiquito):** Contiene información de las actas de finiquito registradas por las empresas.

Tabla 24.

Tabla de Hechos Acta de Finiquito

Campo	Tipo	Tamaño	Llave	Descripción
sk_actividad_economica	integer		FK	Clave foránea de la actividad económica
sk_tipo_contrato	integer		FK	Clave foránea del tipo de contrato
sk_fecha	integer		FK	Clave foránea de la fecha
sk_acta_finiquito	integer		FK	Clave foránea del acta de finiquito
sk_tipo_empresa	integer		FK	Clave foránea del tipo de empresa
sk_contrato	integer		FK	Clave foránea del contrato
sk_motivo_salida	integer		FK	Clave foránea del motivo de salida
sk_ubicacion	integer		FK	Clave foránea de la ubicación
sk_institucion	integer		FK	Clave foránea de la institución
sk_empleado	integer		FK	Clave foránea del empleado
sk_grupo_ocupacional	integer		FK	Clave foránea de grupo ocupacional
ffin_numero	integer			Identificador (1)
ffin_monto	numeric			Valor cancelado.
ffin_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Tabla de Hechos Boletas (fact_boletas):** Contiene información de las boletas registradas por los empleados en contra de las empresas.

Tabla 25.

Tabla de Hechos Boletas

Campo	Tipo	Tamaño	Llave	Descripción
sk_institucion	integer		FK	Clave foránea de la empresa
sk_empleado	integer		FK	Clave foránea del empleado
sk_boleta	integer		FK	Clave foránea de la boleta
sk_fecha	integer		FK	Clave foránea de la fecha
sk_ubicacion	integer		FK	Clave foránea de ubicación
fbol_numero	integer			Identificador (1)
fbol_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Tabla de Hechos Trámite (fact_tramite):** Contiene información de los trámites registrados por los empleados en contra de las empresas.

Tabla 26.

Tabla de Hechos Trámite

Campo	Tipo	Tamaño	Llave	Descripción
sk_institucion	integer		FK	Clave foránea de la empresa
sk_empleado	integer		FK	Clave foránea del empleado
sk_trámite	integer		FK	Clave foránea del trámite
sk_tipo_tramite	integer		FK	Clave foránea de tipo trámite
sk_fecha	integer		FK	Clave foránea de fecha
ftra_numero	integer			Identificador (1)
ftra_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

- **Tabla de Hechos Cumplimiento (fact_cumplimiento):** Contiene información del cumplimiento de factores de cada una de las empresas.

Tabla 27.

Tabla de Hechos Cumplimiento

Campo	Tipo	Tamaño	Llave	Descripción
sk_institucion	integer		FK	Clave foránea de la empresa
sk_tipo_empresa	integer		FK	Clave foránea de tipo empresa
sk_actividad_economica	integer		FK	Clave foránea de actividad económica
fcum_trabajo_menores	integer			Total trabajadores menores de 15 años
fcum_trabajo_juvenil	integer			Total trabajadores bajo modalidad juvenil
fcum_trabajo_discapacidad	integer			Total trabajadores discapacitados
fcum_contratos_registro_atrasados	integer			Total contratos con registro después de 30 días.
fcum_actas_registro_atrasadas	integer			Total actas de finiquito pagadas después de 30 días.
fcum_tramites_boletas	integer			Total de denuncias
fcum_resultado_incumplimiento	varchar	50		Resultado de factores evaluados
fcum_resultado_sgi	varchar	50		Resultado del Sistema SGI
fcum_fecha_carga	timestamp			Fecha del proceso de carga de los registros.

3.1.5.2.3. Contexto y Universo

El modelado de datos es realizado con la herramienta Power Designer, donde se tiene que las tablas de dimensiones tienen un triángulo negro en la parte superior izquierda y las tablas de hechos un cubo de color rojo en la parte superior izquierda. Los contextos ayudan a visualizar la lógica que tiene el Data Warehouse y la explotación del mismo, tienen como eje principal las tablas de hechos y sus dimensiones.

- **Contexto 1. Actas de Finiquito**

En la Figura 22, se observa cómo queda el contexto Actas de Finiquito.

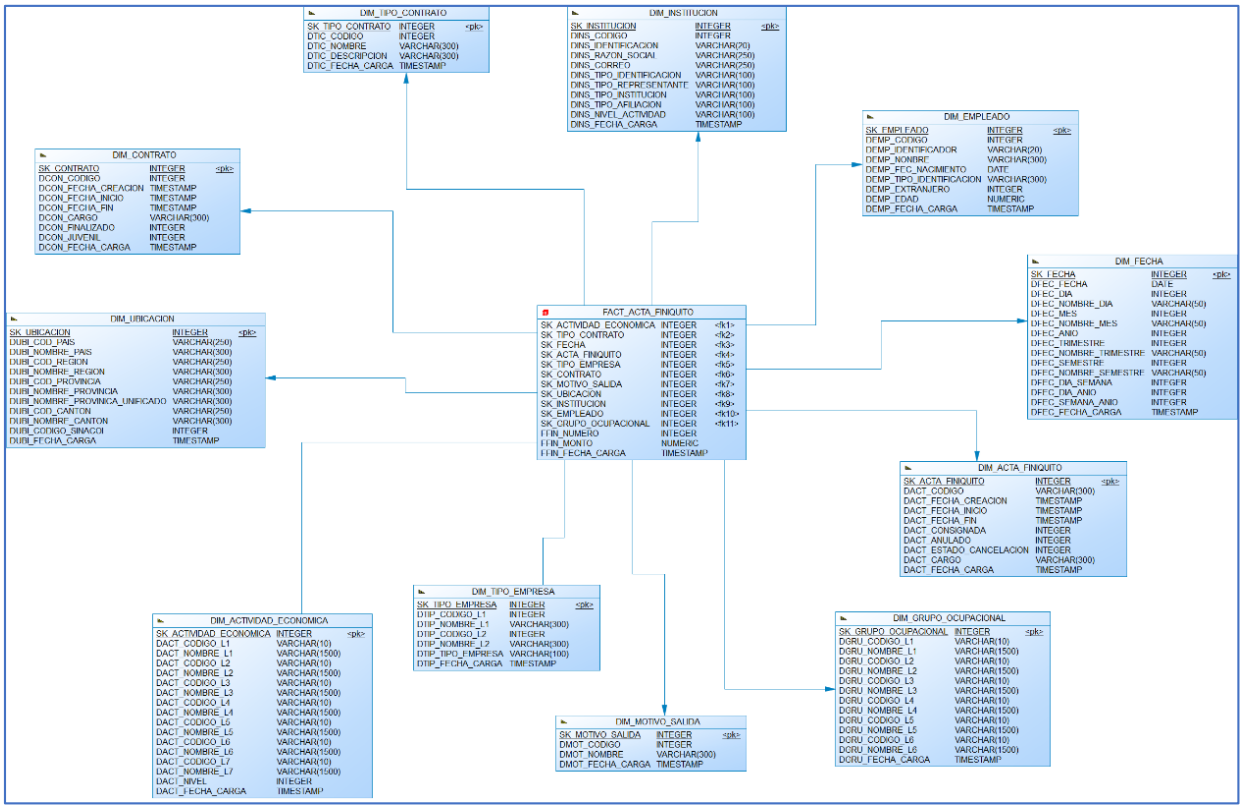


Figura 22. Contexto Actas de Finiquito

- **Contexto 2. Contratos**

En la Figura 23, se observa cómo queda el contexto Contratos

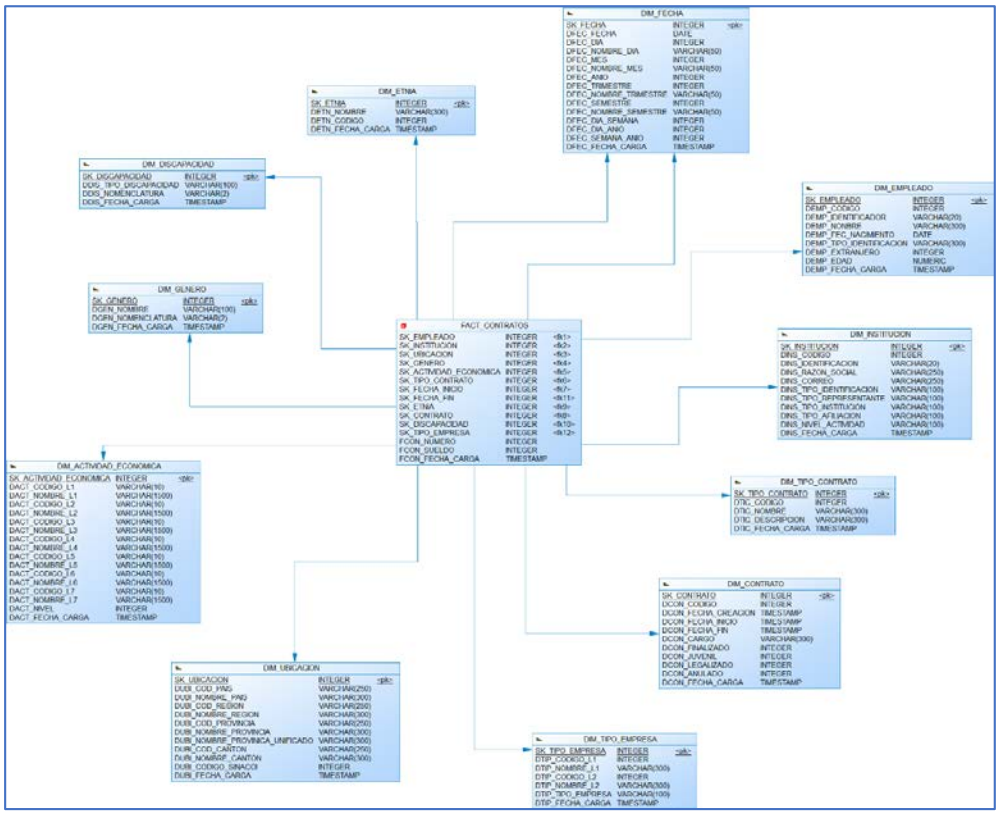


Figura 23. Contexto Contratos

- **Contexto 3. Trámites**

En la Figura 24, se observa cómo queda el contexto Trámites.

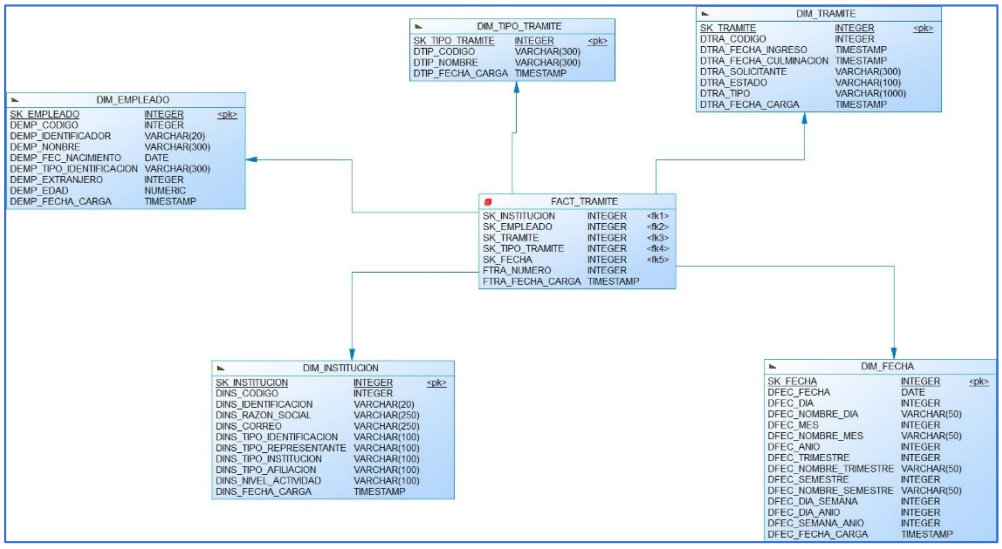


Figura 24. Contexto Trámites

- **Contexto 4. Boletas**

  En la Figura 25, se observa cómo queda el contexto Boletas.

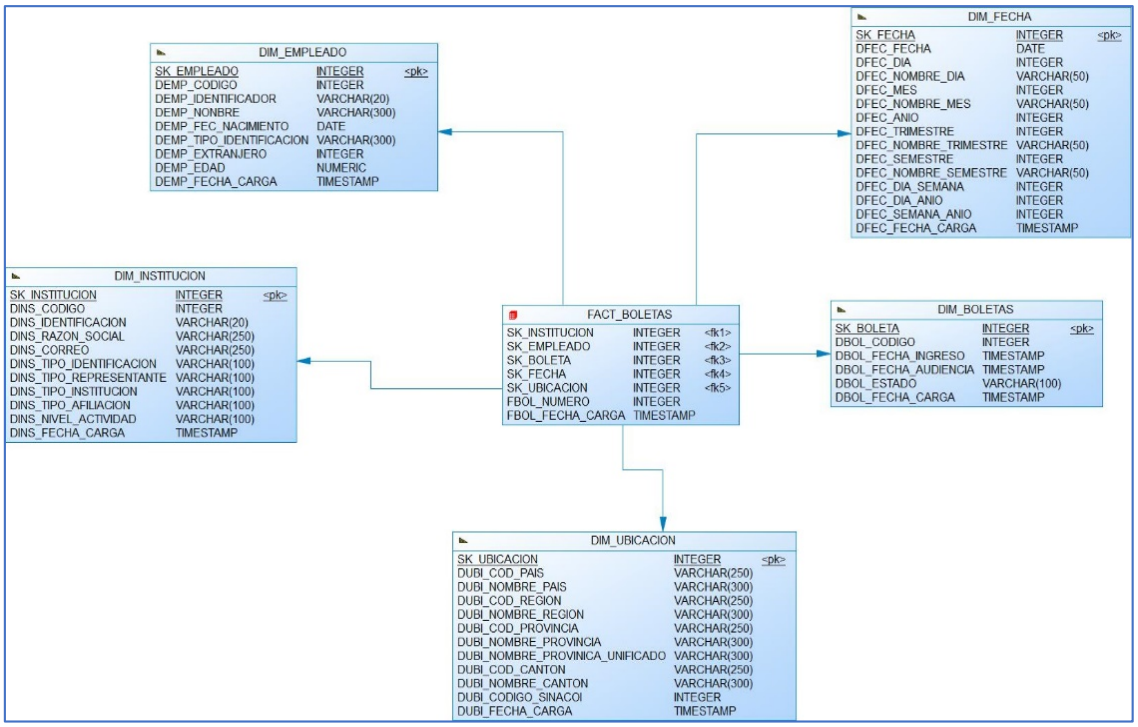


Figura 25. Contexto Boletas

- **Contexto 3. Factores de Cumplimiento**

  En la Figura 26, se observa cómo queda el contexto Factores de Cumplimiento.

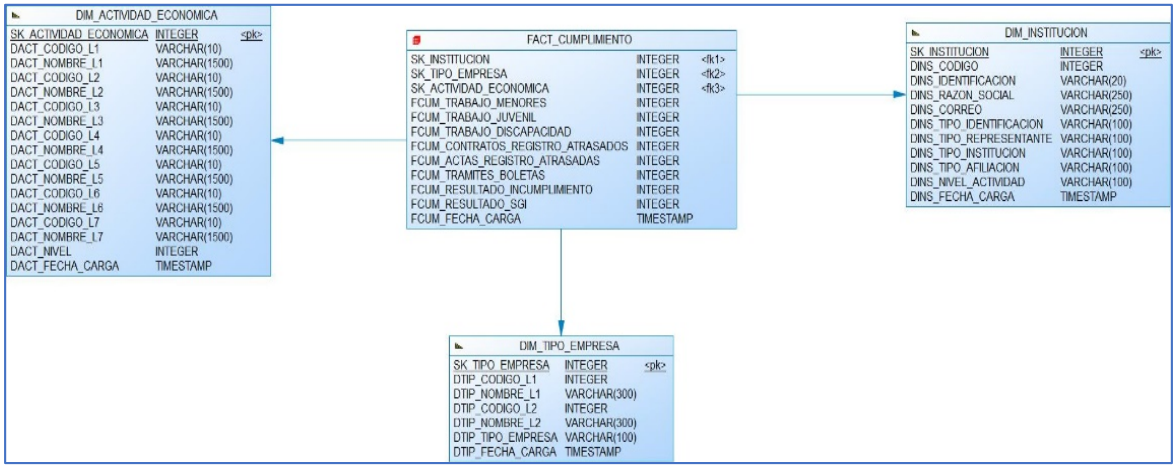


Figura 26. Contexto Factores de Cumplimiento

El universo del Data Warehouse se conforma por todas las tablas de hechos y sus dimensiones. En la Figura 27 se muestra el universo.

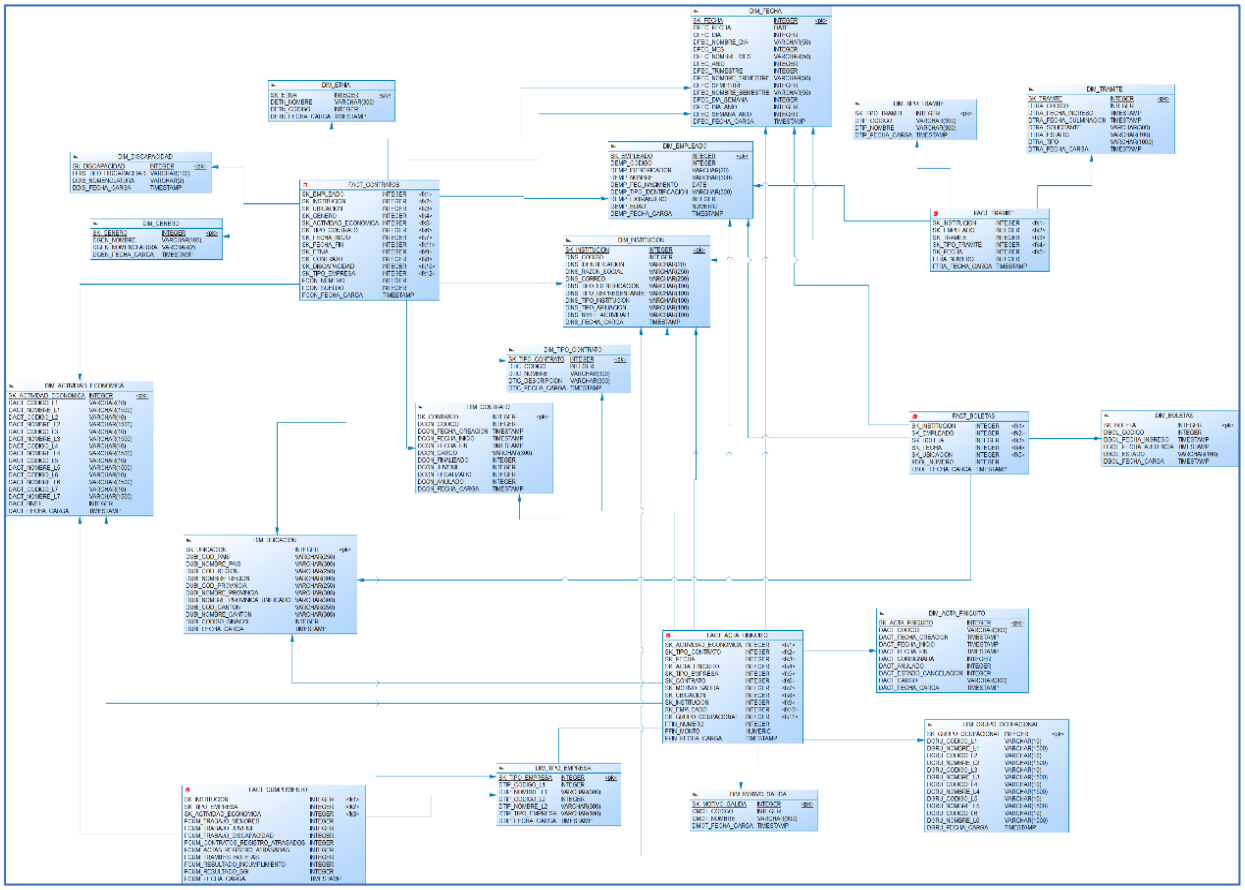


Figura 27. Universo

3.1.6. Especificación de aplicaciones BI

En esta parte se utilizará la herramienta BI Tableau, que se encuentra dentro de las aplicaciones analíticas, estas aplicaciones hacen una exploración de un proceso centrándose en algo específico permitiendo realizar un análisis e interpretación del proceso.

3.1.7. Diseño Físico

El diseño físico nos ayuda a dar soporte a las estructuras que se encuentran en el diseño lógico, se evalúan diferentes elementos:

- **Establecer el tamaño de la base de datos para el Data Warehouse:** Se tomo en cuenta los datos almacenados en el proceso de extracción donde se efectuó la limpieza de datos, con la información obtenida luego de haber realizado el proceso se estimó el tamaño de la base de datos, la misma que se encuentra dividida en esquemas con la finalidad de tener un acceso rápido a los datos, para nuestro caso el tamaño de la base de datos es de 13 GB, el tamaño de la base de datos puede aumentar en un futuro si se realiza un mantenimiento al Data Warehouse o una nueva carga de datos.
- **Configuración del sistema:** No se realizaron configuraciones adicionales a la base de datos, la base se encuentra con las configuraciones de acceso y usuarios privilegiados.
- **Servidores y memoria a utilizar:** El data warehouse no se almacena en un servidor, actualmente se almacena en un computador portátil que hace de servidor, por el tamaño de la base de datos se necesita 8 GB de memoria ram. En un futuro el Ministerio del Trabajo debe tomar la decisión de almacenar el data warehouse en un servidor propio de la institución.
- **Espacio en equipos de trabajo:** No se necesita, puesto que la base de datos se encuentra almacenada en un servidor.

- **Modelo físico en la base de datos:** Se crea la base de datos, para luego crear las tablas con las claves primarias y los otros campos, la base de datos se puede visualizar como se muestra en la Figura 28.

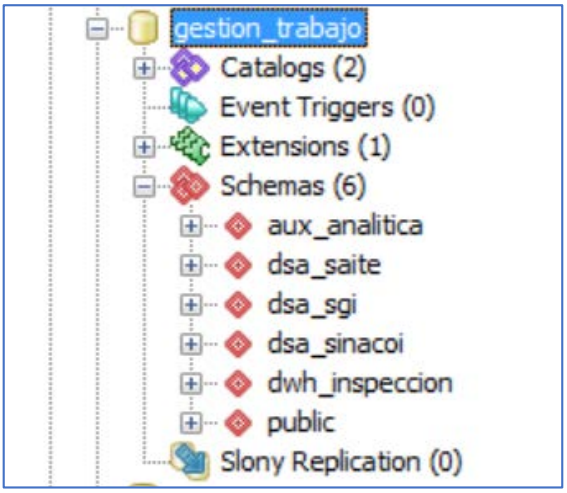


Figura 28. Esquema de la Base de Datos en PostgreSQL

- **Sistema de indexación:** Este sistema se realiza a través de herramienta PDI de Pentaho, para disminuir el coste computacional, ya que si se lo realiza desde la base de datos este coste aumenta.

3.1.8. Diseño e Implementación del ETL

Para una mejor comprensión, esta fase se dividirá en dos partes.

3.1.8.1. Diseño del ETL

Para no afectar la transaccionalidad de los sistemas del Ministerio del Trabajo se crea una copia idéntica de los datos en un esquema denominado DSA.

En esta fase los datos deben ser cargados al data warehouse de forma correcta, por lo que se debe tener en cuenta el orden en que se realizara la carga de datos en la base de datos.

En la Figura 29, se muestra la carga de datos en las tablas de dimensiones.

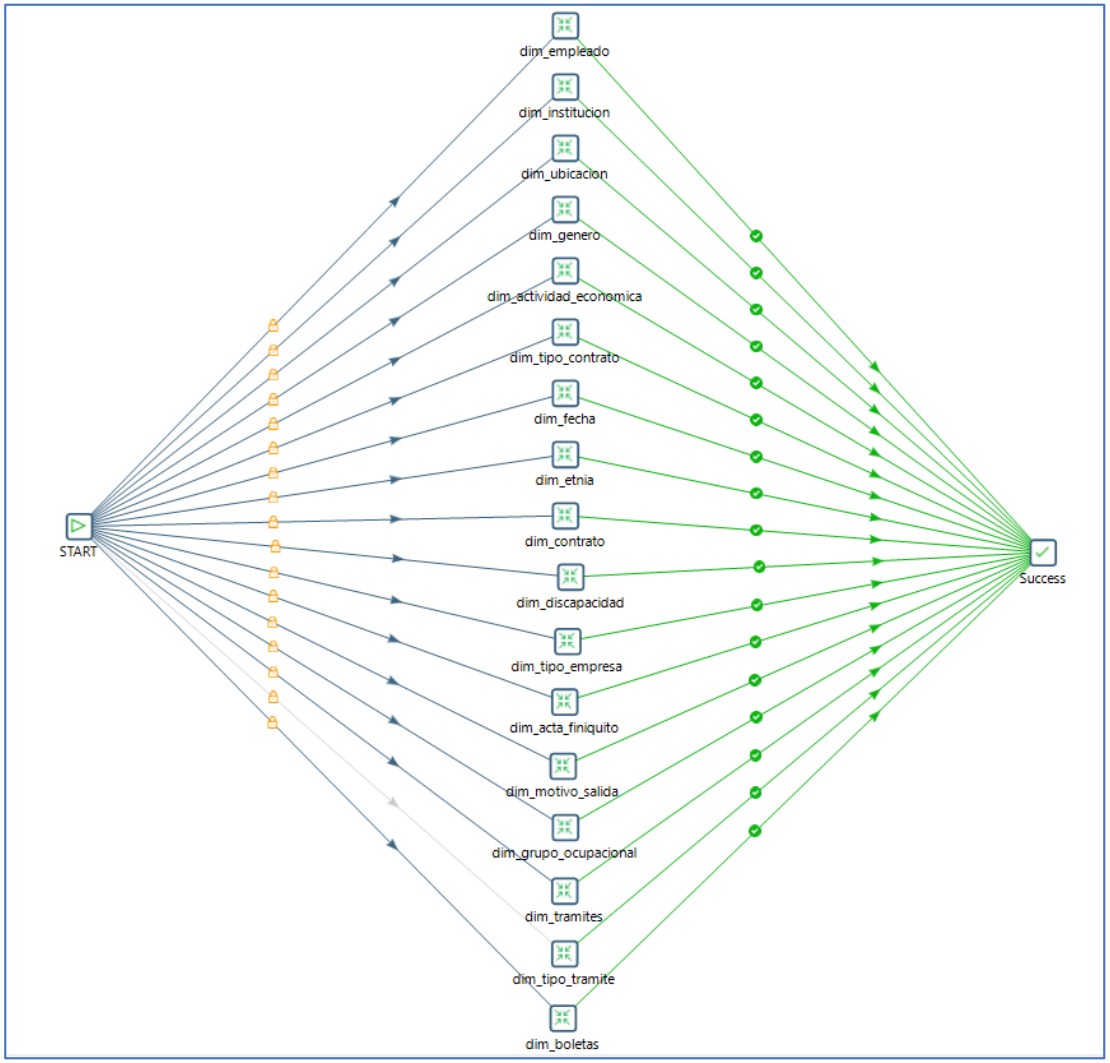


Figura 29. Carga de Datos en las Tablas de Dimensiones

Una vez finalizada la carga de datos en las tablas de dimensiones se continua con la carga de datos en las tablas de hechos, como se muestra en la Figura 30.

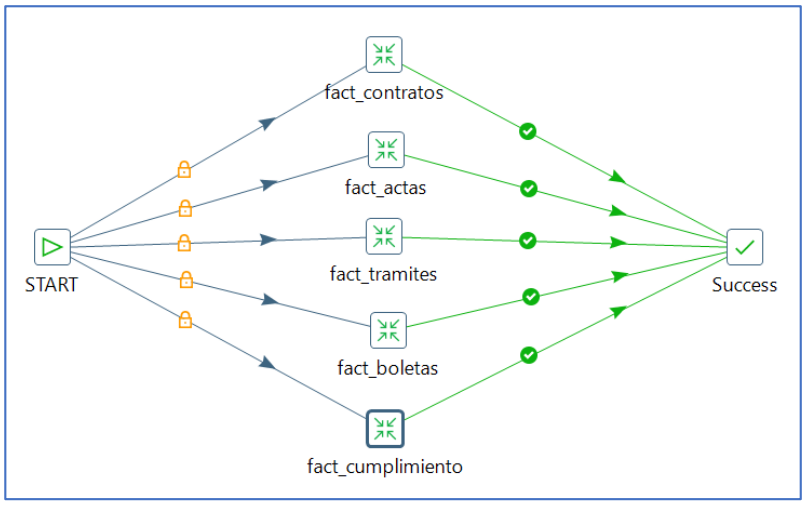


Figura 30. Carga de Datos en las Tablas de Hechos

3.1.8.2. Implementación del ETL

En esta sección se detalla el proceso para transformación de los datos del esquema DSA al esquema DWH. Cabe indicar que el esquema DSA es una copia de las bases de datos de los sistemas transaccionales SAITE, SINACOI y SGI, la única variante es que se agregó un nuevo campo a cada una de las tablas denominado fecha_carga.

De acuerdo al análisis de datos realizado por los expertos del negocio, se tomó los datos necesarios que nos sirven para realizar la transformación, la misma que será usada en la Sección 3.2. “Construcción Modelo de Minería de Datos aplicando la Metodología CRISP-DM”.

El proceso inicia con la carga de datos del esquema DSA al esquema DWH de las siguientes tablas.

Carga de Datos en Tablas de Dimensiones

- **Empleado:** En esta tabla se cargan los datos de los empleados que registran las empresas. En la Figura 31 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Cálculo de la edad del empleado
    - Reemplazo tipo_identificación (R –> Ruc, C–> Cédula, P –> Pasaporte)
    - Uso de variables dicotómicas (true –> 1, false–> 0)
    - Transformación a letras mayúsculas (nombre, tipo_identificacion)

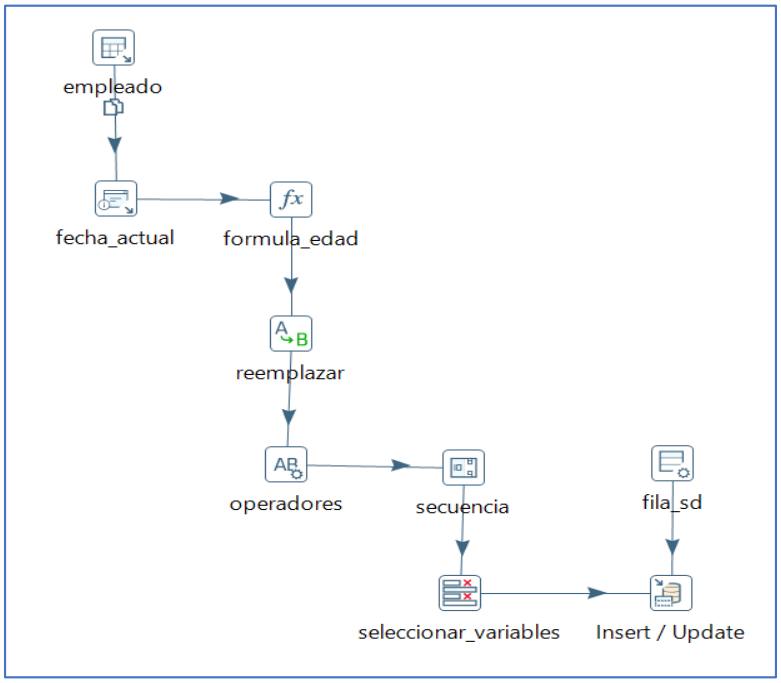
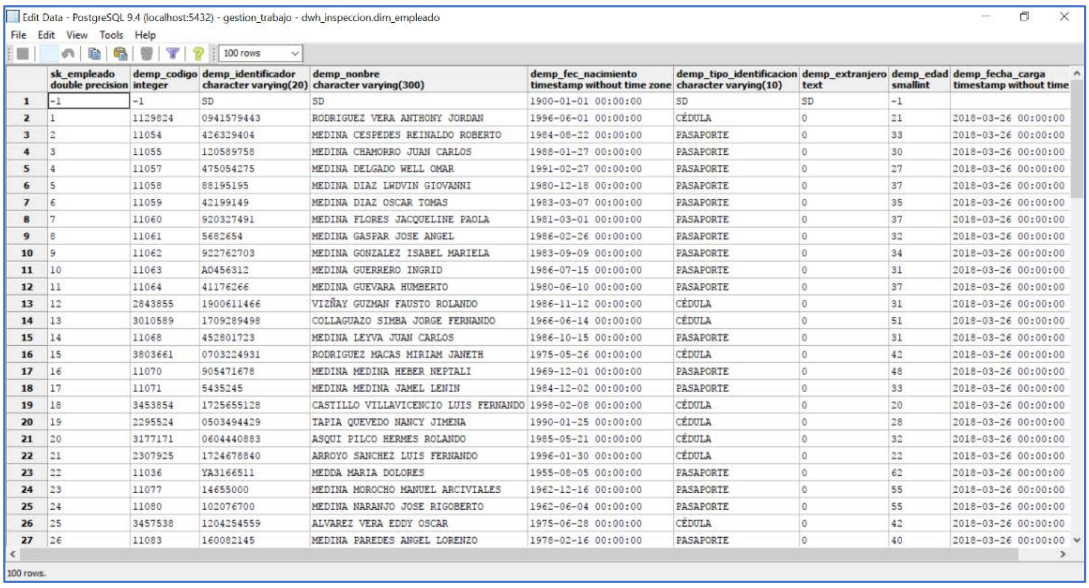


Figura 31. Proceso Carga de Datos Dimensión Empleado

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 32.



	sk_empleado double precision	demp_codigo integer	demp_identificador character varying(20)	demp_nonbre character varying(300)	demp_fec_nacimiento timestamp without time zone	demp_tipo_identificacion character varying(10)	demp_extranjero text	demp_edad smallint	demp_fecha_carga timestamp without time
1	-1	-1	SD	SD	1900-01-01 00:00:00	SD	SD	-1	
2	1	1129824	0941579443	RODRIGUEZ VERA ANTHONY JORDAN	1996-06-01 00:00:00	CÉDULA	0	21	2018-03-26 00:00:00
3	2	11054	426329404	MEDINA CESPEDES REINALDO ROBERTO	1984-08-22 00:00:00	PASAPORTE	0	33	2018-03-26 00:00:00
4	3	11055	120589758	MEDINA CHAMORRO JUAN CARLOS	1988-01-27 00:00:00	PASAPORTE	0	30	2018-03-26 00:00:00
5	4	11057	475054275	MEDINA DELGADO WELL OMAR	1991-02-27 00:00:00	PASAPORTE	0	27	2018-03-26 00:00:00
6	5	11058	88195195	MEDINA DIAZ LWDVIN GIOVANNI	1980-12-18 00:00:00	PASAPORTE	0	37	2018-03-26 00:00:00
7	6	11059	42199149	MEDINA DIAZ OSCAR TOMAS	1983-03-07 00:00:00	PASAPORTE	0	35	2018-03-26 00:00:00
8	7	11060	920327491	MEDINA FLORES JACQUELINE PAOLA	1981-03-01 00:00:00	PASAPORTE	0	37	2018-03-26 00:00:00
9	8	11061	5682654	MEDINA GASPAR JOSE ANGEL	1986-02-26 00:00:00	PASAPORTE	0	32	2018-03-26 00:00:00
10	9	11062	922762703	MEDINA GONZALEZ ISABEL MARIELA	1983-09-09 00:00:00	PASAPORTE	0	34	2018-03-26 00:00:00
11	10	11063	A0456312	MEDINA GUERRERO INGRID	1986-07-15 00:00:00	PASAPORTE	0	31	2018-03-26 00:00:00
12	11	11064	41176266	MEDINA GUEVARA HUMBERTO	1980-06-10 00:00:00	PASAPORTE	0	37	2018-03-26 00:00:00
13	12	2843855	1900611466	VIZÑAY GUZMAN FAUSTO ROLANDO	1986-11-12 00:00:00	CÉDULA	0	31	2018-03-26 00:00:00
14	13	3010589	1709289498	COLLAGUAZO SIMBA JORGE FERNANDO	1966-06-14 00:00:00	CÉDULA	0	51	2018-03-26 00:00:00
15	14	11068	452801723	MEDINA LEYVA JUAN CARLOS	1986-10-15 00:00:00	PASAPORTE	0	31	2018-03-26 00:00:00
16	15	3803661	0703224931	RODRIGUEZ MACAS MIRIAM JANETH	1975-05-26 00:00:00	CÉDULA	0	42	2018-03-26 00:00:00
17	16	11070	905471678	MEDINA MEDINA HEBER NEPTALI	1969-12-01 00:00:00	PASAPORTE	0	48	2018-03-26 00:00:00
18	17	11071	5435245	MEDINA MEDINA JAMEL LENIN	1984-12-02 00:00:00	PASAPORTE	0	33	2018-03-26 00:00:00
19	18	3453854	1725655128	CASTILLO VILLAVICENCIO LUIS FERNANDO	1998-02-08 00:00:00	CÉDULA	0	20	2018-03-26 00:00:00
20	19	2295524	0503494429	TAPIA QUEVEDO NANCY JIMENA	1990-01-25 00:00:00	CÉDULA	0	28	2018-03-26 00:00:00
21	20	3177171	0604440883	ASQUI PILCO HERMES ROLANDO	1985-05-21 00:00:00	CÉDULA	0	32	2018-03-26 00:00:00
22	21	2307925	1724678840	ARROYO SANCHEZ LUIS FERNANDO	1996-01-30 00:00:00	CÉDULA	0	22	2018-03-26 00:00:00
23	22	11036	YA3166511	MEDDA MARIA DOLORES	1955-08-05 00:00:00	PASAPORTE	0	62	2018-03-26 00:00:00
24	23	11077	14655000	MEDINA MOROCHO MANUEL ARCIVIALES	1962-12-16 00:00:00	PASAPORTE	0	55	2018-03-26 00:00:00
25	24	11080	102076700	MEDINA NARANJO JOSE RIGOBERTO	1962-06-04 00:00:00	PASAPORTE	0	55	2018-03-26 00:00:00
26	25	3457538	1204254559	ALVAREZ VERA EDDY OSCAR	1975-06-28 00:00:00	CÉDULA	0	42	2018-03-26 00:00:00
27	26	11083	160082145	MEDINA PAREDES ANGEL LORENZO	1978-02-16 00:00:00	PASAPORTE	0	40	2018-03-26 00:00:00

100 rows.

Figura 32. Datos en la Tabla Dimensión Empleado

- **Institución:** En esta tabla se cargan los datos de las empresas que se registran los sistemas. En la Figura 33 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo tipo (N –> Natural, J –> Juídico)
    - Reemplazo tipo_identificación (R –> Ruc, C–> Cédula, P –> Pasaporte)
    - Reemplazo del tipo_afiliación (IP –> ISPOL, IF–> ISFA, IE –> IESS)
    - Reemplazo del tipo_empresa (F –> FINANCIERA, P–> PÚBLICA, E –> ESPECIAL, R –> PRIVADA, A –> ARTESANAL, M –> EMPRESA PÚBLICA)
    - Reemplazo valores nulos por “SD” en las variables tipo_afiliacion, tipo_afiliacion.

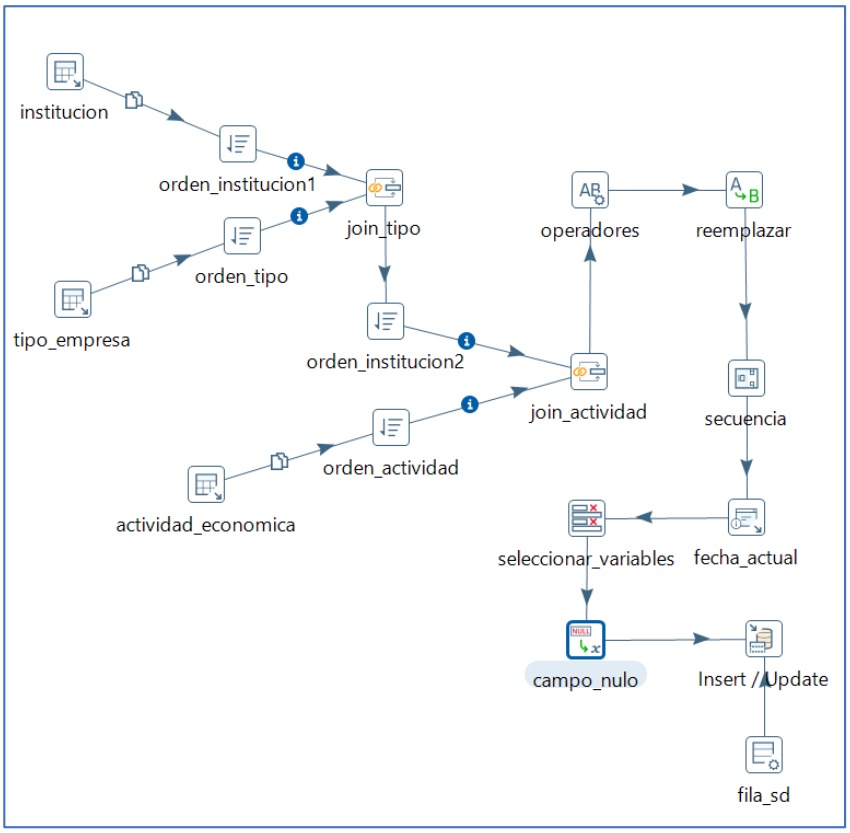
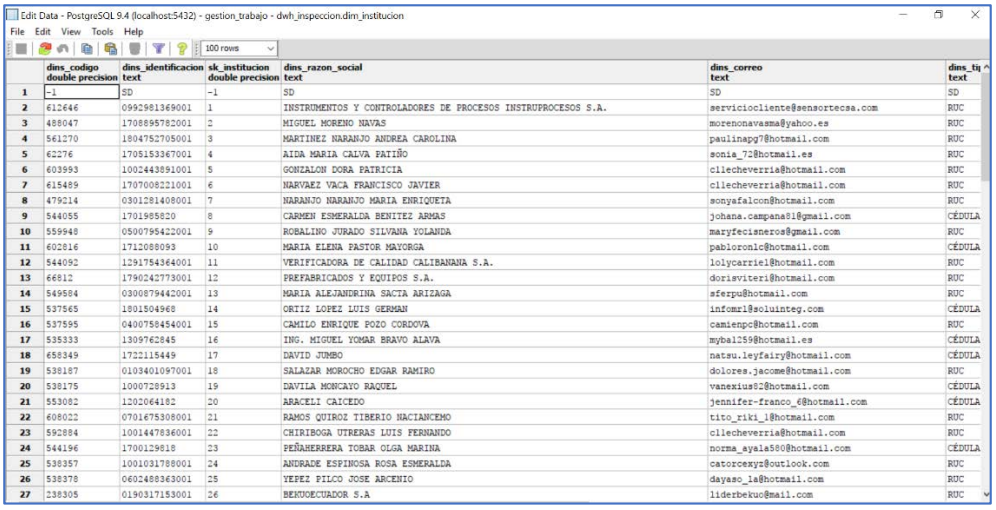


Figura 33. Proceso Carga de Datos Dimensión Institución

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 34.



Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_institucion

File Edit View Tools Help

100 rows

	dins_codigo double precision	dins_identificacion text	sk_institucion double precision	dins_razon_social text	dins_correo text	dins_tip text
1	-1	SD	-1	SD	SD	SD
2	612646	0992981369001	1	INSTRUMENTOS Y CONTROLADORES DE PROCESOS INSTRUPROCESOS S.A.	serviciocliente@sensortecsa.com	RUC
3	488047	1708895782001	2	MIGUEL MORENO NAVAS	morenonavasma@yahoo.es	RUC
4	561270	1804752705001	3	MARTINEZ NARANJO ANDREA CAROLINA	paulinapg7@hotmail.com	RUC
5	62276	1705153367001	4	AIDA MARIA CALVA PATIÑO	sonia_72@hotmail.es	RUC
6	603993	1002443891001	5	GONZALON DORA PATRICIA	cllecheverria@hotmail.com	RUC
7	615489	1707008221001	6	NARVAEZ VACA FRANCISCO JAVIER	cllecheverria@hotmail.com	RUC
8	479214	0301281408001	7	NARANJO NARANJO MARIA ENRIQUETA	sonyafalcon@hotmail.com	RUC
9	544055	1701985820	8	CARMEN ESMERALDA BENITEZ ARMAS	johana.campana81@gmail.com	CÉDULA
10	559948	0500795422001	9	ROBALINO JURADO SILVANA YOLANDA	maryfecisneros@gmail.com	RUC
11	602816	1712088093	10	MARIA ELENA PASTOR MAYORGA	pabloronlc@hotmail.com	CÉDULA
12	544092	1291754364001	11	VERIFICADORA DE CALIDAD CALIBANANA S.A.	lolycarrie1@hotmail.com	RUC
13	66812	1790242773001	12	PREFABRICADOS Y EQUIPOS S.A.	dorisviteri@hotmail.com	RUC
14	549584	0300879442001	13	MARIA ALEJANDRINA SACTA ARIZAGA	sferpu@hotmail.com	RUC
15	537565	1801504968	14	ORTIZ LOPEZ LUIS GERMAN	infomr1@soluinteg.com	CÉDULA
16	537595	0400758454001	15	CAMILO ENRIQUE POZO CORDOVA	camienpc@hotmail.com	RUC
17	535333	1309762845	16	ING. MIGUEL YOMAR BRAVO ALAVA	myba1259@hotmail.es	CÉDULA
18	658349	1722115449	17	DAVID JUMBO	natsu.leyfairy@hotmail.com	CÉDULA
19	538187	0103401097001	18	SALAZAR MOROCHO EDGAR RAMIRO	dolores.jacome@hotmail.com	RUC
20	538175	1000728913	19	DAVILA MONCAYO RAQUEL	vanexius82@hotmail.com	CÉDULA
21	553082	1202064182	20	ARACELI CAICEDO	jennifer-franco_6@hotmail.com	CÉDULA
22	608022	0701675308001	21	RAMOS QUIROZ TIBERIO HACIANCEMO	tito_riki_1@hotmail.com	RUC
23	592884	1001447836001	22	CHIRIBOGA UTRERAS LUIS FERNANDO	cllecheverria@hotmail.com	RUC
24	544196	1700129818	23	PEÑAHERRERA TOBAR OLGA MARINA	norma_ayala580@hotmail.com	CÉDULA
25	538357	1001031788001	24	ANDRADE ESPINOSA ROSA ESMERALDA	catorcexyz@outlook.com	RUC
26	538378	0602488363001	25	YEPEZ PILCO JOSE ARCENIO	dayasc_1a@hotmail.com	RUC
27	238305	0190317153001	26	BEKUOECUADOR S.A	liderbekuo@mail.com	RUC

Figura 34. Datos en la Tabla Dimensión Institución

- **Ubicación:** En esta tabla se cargan los datos de las ubicaciones geográficas de los sistemas. En la Figura 35 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Aplanar la tabla a cuatro niveles (País, Región, Provincia y Cantón).
    - Unificación de nombres de provincia**.**

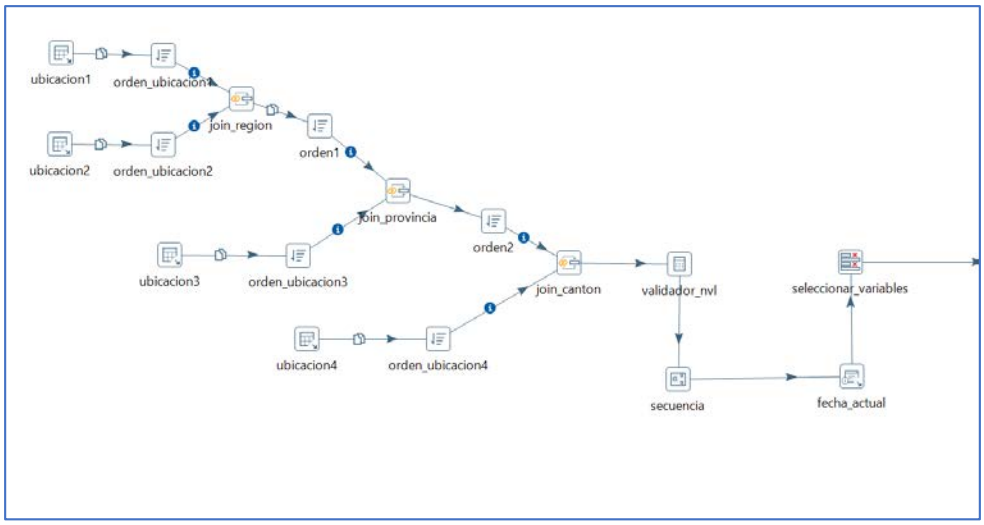
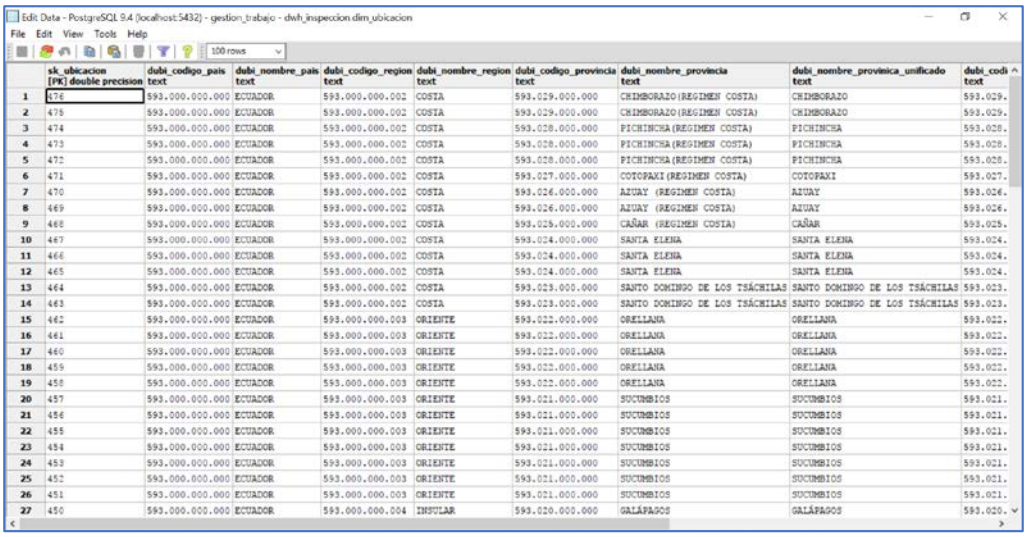


Figura 35. Proceso Carga de Datos Dimensión Ubicación

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 36.



Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_ubicacion

File Edit View Tools Help

100 rows

	sk_ubicacion [PK] double precision	dubi_codigo_pais text	dubi_nombre_pais text	dubi_codigo_region text	dubi_nombre_region text	dubi_codigo_provincia text	dubi_nombre_provincia text	dubi_nombre_provincia_unificado text	dubi_codi text
1	476	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.029.000.000	CHIMBORAZO(REGIMEN COSTA)	CHIMBORAZO	593.029.
2	475	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.029.000.000	CHIMBORAZO(REGIMEN COSTA)	CHIMBORAZO	593.029.
3	474	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.028.000.000	PICHINCHA(REGIMEN COSTA)	PICHINCHA	593.028.
4	473	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.028.000.000	PICHINCHA(REGIMEN COSTA)	PICHINCHA	593.028.
5	472	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.028.000.000	PICHINCHA(REGIMEN COSTA)	PICHINCHA	593.028.
6	471	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.027.000.000	COTOPAXI(REGIMEN COSTA)	COTOPAXI	593.027.
7	470	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.026.000.000	AZUAY (REGIMEN COSTA)	AZUAY	593.026.
8	469	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.026.000.000	AZUAY (REGIMEN COSTA)	AZUAY	593.026.
9	468	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.025.000.000	CAÑAR (REGIMEN COSTA)	CAÑAR	593.025.
10	467	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.024.000.000	SANTA ELENA	SANTA ELENA	593.024.
11	466	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.024.000.000	SANTA ELENA	SANTA ELENA	593.024.
12	465	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.024.000.000	SANTA ELENA	SANTA ELENA	593.024.
13	464	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.023.000.000	SANTO DOMINGO DE LOS TSÁCHILAS	SANTO DOMINGO DE LOS TSÁCHILAS	593.023.
14	463	593.000.000.000	ECUADOR	593.000.000.002	COSTA	593.023.000.000	SANTO DOMINGO DE LOS TSÁCHILAS	SANTO DOMINGO DE LOS TSÁCHILAS	593.023.
15	462	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.022.000.000	ORELLANA	ORELLANA	593.022.
16	461	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.022.000.000	ORELLANA	ORELLANA	593.022.
17	460	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.022.000.000	ORELLANA	ORELLANA	593.022.
18	459	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.022.000.000	ORELLANA	ORELLANA	593.022.
19	458	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.022.000.000	ORELLANA	ORELLANA	593.022.
20	457	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.021.000.000	SUCUMBIOS	SUCUMBIOS	593.021.
21	456	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.021.000.000	SUCUMBIOS	SUCUMBIOS	593.021.
22	455	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.021.000.000	SUCUMBIOS	SUCUMBIOS	593.021.
23	454	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.021.000.000	SUCUMBIOS	SUCUMBIOS	593.021.
24	453	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.021.000.000	SUCUMBIOS	SUCUMBIOS	593.021.
25	452	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.021.000.000	SUCUMBIOS	SUCUMBIOS	593.021.
26	451	593.000.000.000	ECUADOR	593.000.000.003	ORIENTE	593.021.000.000	SUCUMBIOS	SUCUMBIOS	593.021.
27	450	593.000.000.000	ECUADOR	593.000.000.004	INSULAR	593.020.000.000	GALÁPAGOS	GALÁPAGOS	593.020.

Figura 36. Datos en la Tabla Dimensión Ubicación

- **Género:** En esta tabla se cargan los datos de los géneros de los empleados. En la Figura 37 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo genero (M –> Masculino, F –> Femenino)
    - Transformación a letras mayúsculas (genero, nomenclatura_genero)

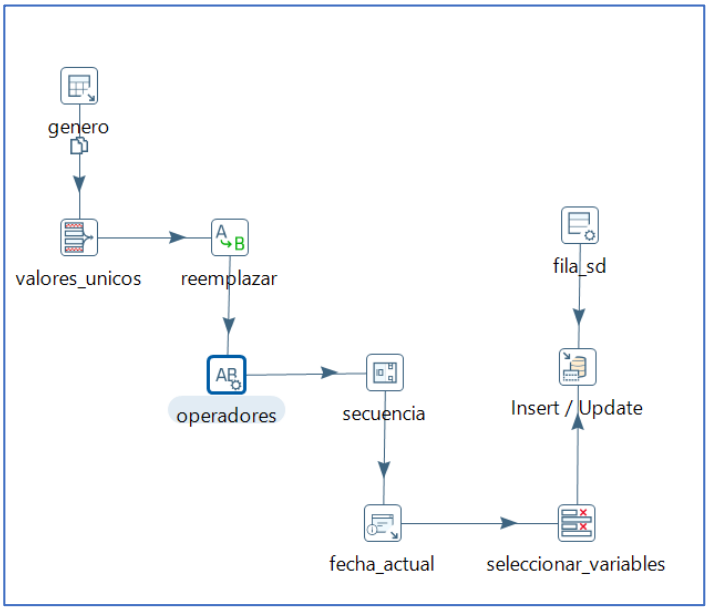


Figura 37. Proceso Carga de Datos Dimensión Género

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 38.

Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_genero

File Edit View Tools Help

No limit

	sk_genero double precision	dgen_nombre text	dgen_nomenclatura text	dgen_fecha_carga timestamp without time zone
1	-1	SD	SD	
2	1	MASCULINO	M	2018-03-26 14:30:55.597
3	2	FEMENINO	F	2018-03-26 14:30:55.597

Figura 38. Datos en la Tabla Dimensión Género

- **Actividad Económica:** En esta tabla se cargan los datos la Actividad Económica de las empresas. En la Figura 39 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Aplanar la tabla a siete niveles.
    - Transformación a letras mayúsculas los nombres de los siete niveles.

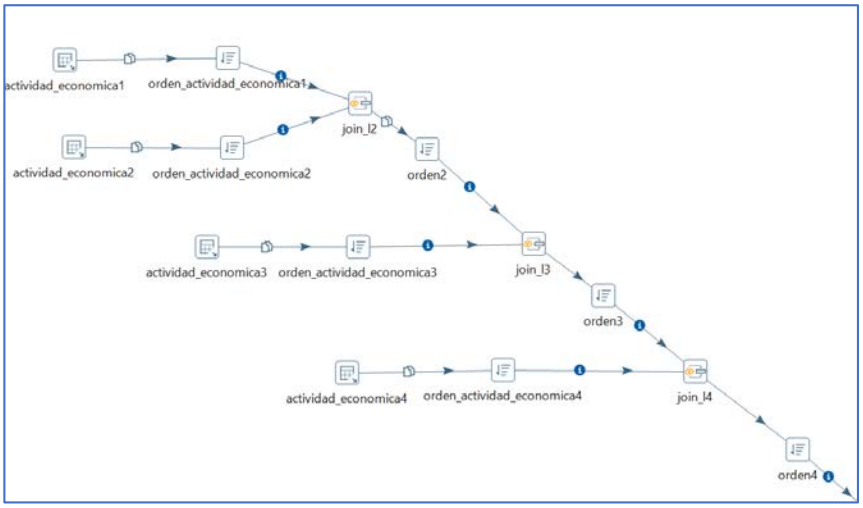
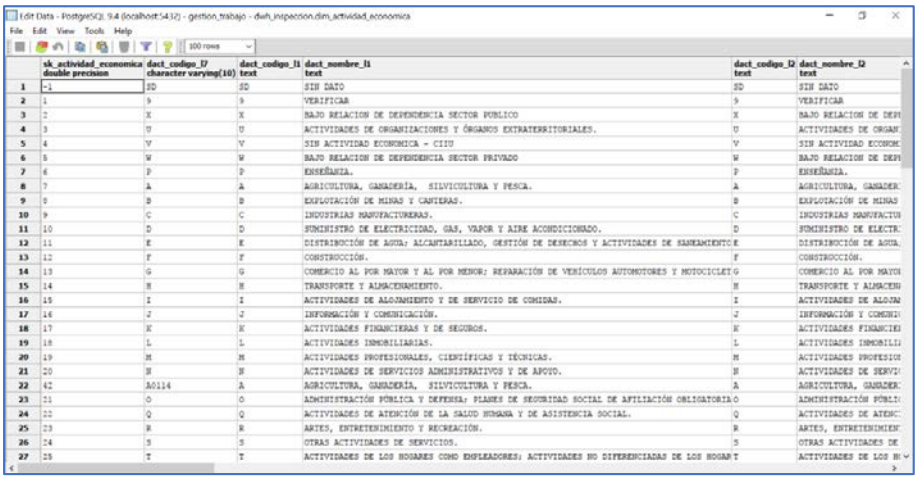


Figura 39. Proceso Carga de Datos Dimensión Actividad Económica

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 40.



	sk_actividad_economica double precision	dact_codigo_l7 character varying(10)	dact_codigo_l1 text	dact_nombre_l1 text	dact_codigo_l2 text	dact_nombre_l2 text
1	-1	SD	SD	SIN DATO	SD	SIN DATO
2	1	9	9	VERIFICAR	9	VERIFICAR
3	2	X	X	BAJO RELACION DE DEPENDENCIA SECTOR PUBLICO	X	BAJO RELACION DE DEPI
4	3	U	U	ACTIVIDADES DE ORGANIZACIONES Y ÓRGANOS EXTRATERRITORIALES.	U	ACTIVIDADES DE ORGAN
5	4	V	V	SIN ACTIVIDAD ECONOMICA - CIIU	V	SIN ACTIVIDAD ECONOM
6	5	W	W	BAJO RELACION DE DEPENDENCIA SECTOR PRIVADO	W	BAJO RELACION DE DEPI
7	6	P	P	ENSEÑANZA.	P	ENSEÑANZA.
8	7	A	A	AGRICULTURA, GANADERÍA, SILVICULTURA Y PESCA.	A	AGRICULTURA, GANADER
9	8	B	B	EXPLOTACIÓN DE MINAS Y CANTERAS.	B	EXPLOTACIÓN DE MINAS
10	9	C	C	INDUSTRIAS MANUFACTURERAS.	C	INDUSTRIAS MANUFACTU
11	10	D	D	SUMINISTRO DE ELECTRICIDAD, GAS, VAPOR Y AIRE ACONDICIONADO.	D	SUMINISTRO DE ELECTR
12	11	E	E	DISTRIBUCIÓN DE AGUA; ALCANTARILLADO, GESTIÓN DE DESECHOS Y ACTIVIDADES DE SANEAMIENTO	E	DISTRIBUCIÓN DE AGUA,
13	12	F	F	CONSTRUCCIÓN.	F	CONSTRUCCIÓN.
14	13	G	G	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACIÓN DE VEHÍCULOS AUTOMOTORES Y MOTOCICLET	G	COMERCIO AL POR MAYO
15	14	H	H	TRANSPORTE Y ALMACENAMIENTO.	H	TRANSPORTE Y ALMACEN
16	15	I	I	ACTIVIDADES DE ALOJAMIENTO Y DE SERVICIO DE COMIDAS.	I	ACTIVIDADES DE ALOJAI
17	16	J	J	INFORMACIÓN Y COMUNICACIÓN.	J	INFORMACIÓN Y COMUNI
18	17	K	K	ACTIVIDADES FINANCIERAS Y DE SEGUROS.	K	ACTIVIDADES FINANCIEI
19	18	L	L	ACTIVIDADES INMOBILIARIAS.	L	ACTIVIDADES INMOBILI
20	19	M	M	ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICAS.	M	ACTIVIDADES PROFESIO
21	20	N	N	ACTIVIDADES DE SERVICIOS ADMINISTRATIVOS Y DE APOYO.	N	ACTIVIDADES DE SERVI
22	42	A0114	A	AGRICULTURA, GANADERÍA, SILVICULTURA Y PESCA.	A	AGRICULTURA, GANADER
23	21	O	O	ADMINISTRACIÓN PÚBLICA Y DEFENSA; PLANES DE SEGURIDAD SOCIAL DE AFILIACIÓN OBLIGATORIA	O	ADMINISTRACIÓN PÚBLI
24	22	Q	Q	ACTIVIDADES DE ATENCIÓN DE LA SALUD HUMANA Y DE ASISTENCIA SOCIAL.	Q	ACTIVIDADES DE ATENC
25	23	R	R	ARTES, ENTRETENIMIENTO Y RECREACIÓN.	R	ARTES, ENTRETENIMIEN
26	24	S	S	OTRAS ACTIVIDADES DE SERVICIOS.	S	OTRAS ACTIVIDADES DE
27	25	T	T	ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES; ACTIVIDADES NO DIFERENCIADAS DE LOS HOGAR	T	ACTIVIDADES DE LOS H

Figura 40. Datos en la Tabla Dimensión Actividad Económica

- **Tipo Contrato:** En esta tabla se cargan los datos los tipos de contratos que registran las empresas. En la Figura 41 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - o Reemplazo valores nulos por “SD” en las variables descripción.
    - o Transformación a letras mayúsculas (nombre y descripción).

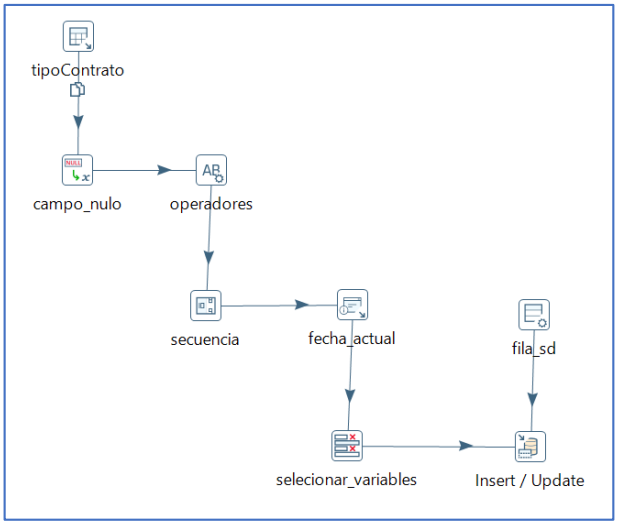
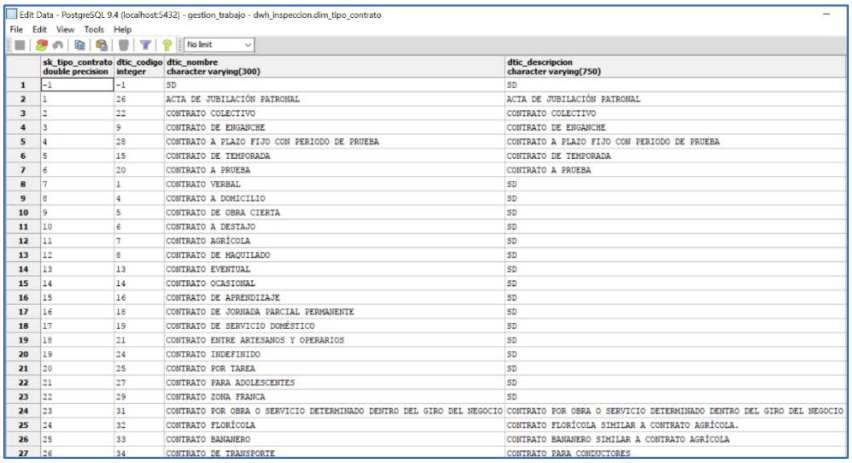


Figura 41. Proceso Carga de Datos Dimensión Tipo Contrato

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 42.



Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_tipo_contrato

File Edit View Tools Help

No limit

	sk_tipo_contrato double precision	dtic_codigo integer	dtic_nombre character varying(300)	dtic_descripcion character varying(750)
1	-1	-1	SD	SD
2	1	26	ACTA DE JUBILACIÓN PATRONAL	ACTA DE JUBILACIÓN PATRONAL
3	2	22	CONTRATO COLECTIVO	CONTRATO COLECTIVO
4	3	9	CONTRATO DE ENGANCHE	CONTRATO DE ENGANCHE
5	4	28	CONTRATO A PLAZO FIJO CON PERIODO DE PRUEBA	CONTRATO A PLAZO FIJO CON PERIODO DE PRUEBA
6	5	15	CONTRATO DE TEMPORADA	CONTRATO DE TEMPORADA
7	6	20	CONTRATO A PRUEBA	CONTRATO A PRUEBA
8	7	1	CONTRATO VERBAL	SD
9	8	4	CONTRATO A DOMICILIO	SD
10	9	5	CONTRATO DE OBRA CIERTA	SD
11	10	6	CONTRATO A DESTAJO	SD
12	11	7	CONTRATO AGRÍCOLA	SD
13	12	8	CONTRATO DE MAQUILADO	SD
14	13	13	CONTRATO EVENTUAL	SD
15	14	14	CONTRATO OCASIONAL	SD
16	15	16	CONTRATO DE APRENDIZAJE	SD
17	16	18	CONTRATO DE JORNADA PARCIAL PERMANENTE	SD
18	17	19	CONTRATO DE SERVICIO DOMÉSTICO	SD
19	18	21	CONTRATO ENTRE ARTESANOS Y OPERARIOS	SD
20	19	24	CONTRATO INDEFINIDO	SD
21	20	25	CONTRATO POR TAREA	SD
22	21	27	CONTRATO PARA ADOLESCENTES	SD
23	22	29	CONTRATO ZONA FRANCA	SD
24	23	31	CONTRATO POR OBRA O SERVICIO DETERMINADO DENTRO DEL GIRO DEL NEGOCIO	CONTRATO POR OBRA O SERVICIO DETERMINADO DENTRO DEL GIRO DEL NEGOCIO
25	24	32	CONTRATO FLORÍCOLA	CONTRATO FLORÍCOLA SIMILAR A CONTRATO AGRÍCOLA.
26	25	33	CONTRATO BANANERO	CONTRATO BANANERO SIMILAR A CONTRATO AGRÍCOLA
27	26	34	CONTRATO DE TRANSPORTE	CONTRATO PARA CONDUCTORES

Figura 42. Datos en la Tabla Dimensión Tipo Contrato

- **Fecha:** En esta tabla se cargan las fechas comprendidas entre el año 2009 hasta el año 2038. En la Figura 43 se muestra el proceso ETL. Esta tabla requirió de la siguiente transformación.
    - Transformación para obtener datos de semestre, trimestre, día de la semana, nombre del día, día del mes, día del año.

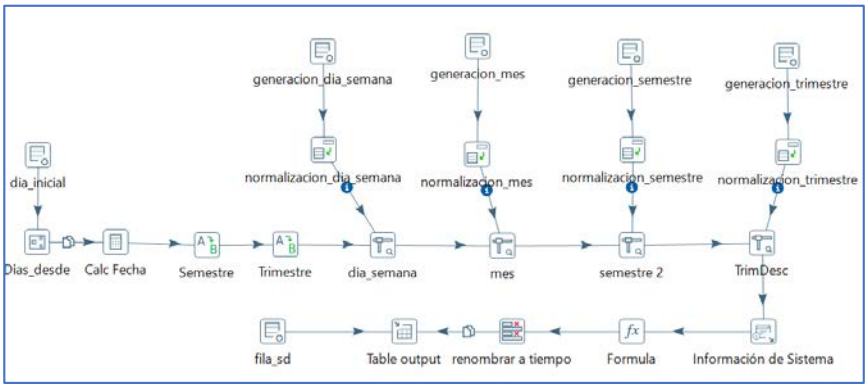
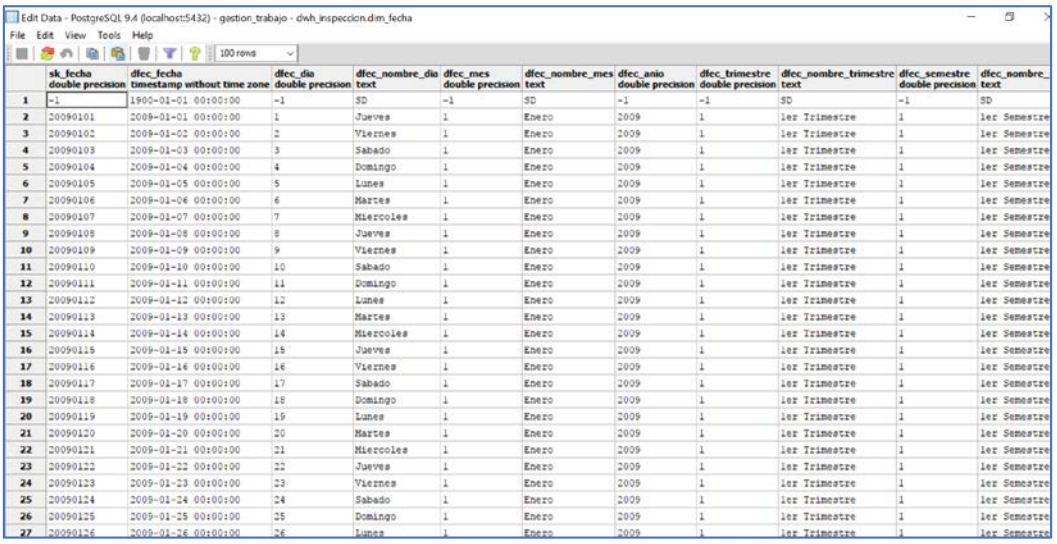


Figura 43. Proceso Carga de Datos Dimensión Fecha

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 44.



	sk_fecha double precision	dfec_fecha timestamp without time zone	dfec_dia double precision	dfec_nombre_dia text	dfec_mes double precision	dfec_nombre_mes text	dfec_anio double precision	dfec_trimestre double precision	dfec_nombre_trimestre text	dfec_semestre double precision	dfec_nombre_ text
1	-1	1900-01-01 00:00:00	-1	SD	-1	SD	-1	-1	SD	-1	SD
2	20090101	2009-01-01 00:00:00	1	Jueves	1	Enero	2009	1	1er Trimestre	1	1er Semestre
3	20090102	2009-01-02 00:00:00	2	Viernes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
4	20090103	2009-01-03 00:00:00	3	Sabado	1	Enero	2009	1	1er Trimestre	1	1er Semestre
5	20090104	2009-01-04 00:00:00	4	Domingo	1	Enero	2009	1	1er Trimestre	1	1er Semestre
6	20090105	2009-01-05 00:00:00	5	Lunes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
7	20090106	2009-01-06 00:00:00	6	Martes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
8	20090107	2009-01-07 00:00:00	7	Miercoles	1	Enero	2009	1	1er Trimestre	1	1er Semestre
9	20090108	2009-01-08 00:00:00	8	Jueves	1	Enero	2009	1	1er Trimestre	1	1er Semestre
10	20090109	2009-01-09 00:00:00	9	Viernes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
11	20090110	2009-01-10 00:00:00	10	Sabado	1	Enero	2009	1	1er Trimestre	1	1er Semestre
12	20090111	2009-01-11 00:00:00	11	Domingo	1	Enero	2009	1	1er Trimestre	1	1er Semestre
13	20090112	2009-01-12 00:00:00	12	Lunes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
14	20090113	2009-01-13 00:00:00	13	Martes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
15	20090114	2009-01-14 00:00:00	14	Miercoles	1	Enero	2009	1	1er Trimestre	1	1er Semestre
16	20090115	2009-01-15 00:00:00	15	Jueves	1	Enero	2009	1	1er Trimestre	1	1er Semestre
17	20090116	2009-01-16 00:00:00	16	Viernes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
18	20090117	2009-01-17 00:00:00	17	Sabado	1	Enero	2009	1	1er Trimestre	1	1er Semestre
19	20090118	2009-01-18 00:00:00	18	Domingo	1	Enero	2009	1	1er Trimestre	1	1er Semestre
20	20090119	2009-01-19 00:00:00	19	Lunes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
21	20090120	2009-01-20 00:00:00	20	Martes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
22	20090121	2009-01-21 00:00:00	21	Miercoles	1	Enero	2009	1	1er Trimestre	1	1er Semestre
23	20090122	2009-01-22 00:00:00	22	Jueves	1	Enero	2009	1	1er Trimestre	1	1er Semestre
24	20090123	2009-01-23 00:00:00	23	Viernes	1	Enero	2009	1	1er Trimestre	1	1er Semestre
25	20090124	2009-01-24 00:00:00	24	Sabado	1	Enero	2009	1	1er Trimestre	1	1er Semestre
26	20090125	2009-01-25 00:00:00	25	Domingo	1	Enero	2009	1	1er Trimestre	1	1er Semestre
27	20090126	2009-01-26 00:00:00	26	Lunes	1	Enero	2009	1	1er Trimestre	1	1er Semestre

Figura 44. Datos en la Tabla Dimensión Fecha

- **Etnia:** En esta tabla se cargan los datos de las etnias de los empleados. En la Figura 45 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Transformación a letras mayúsculas (nombre).

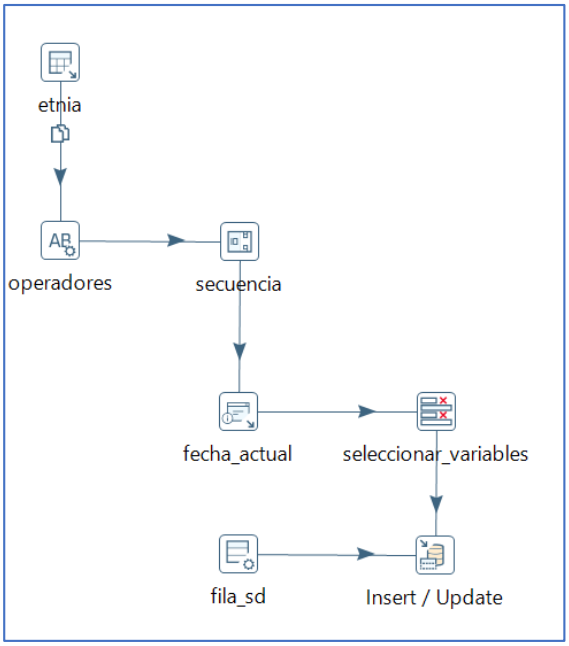
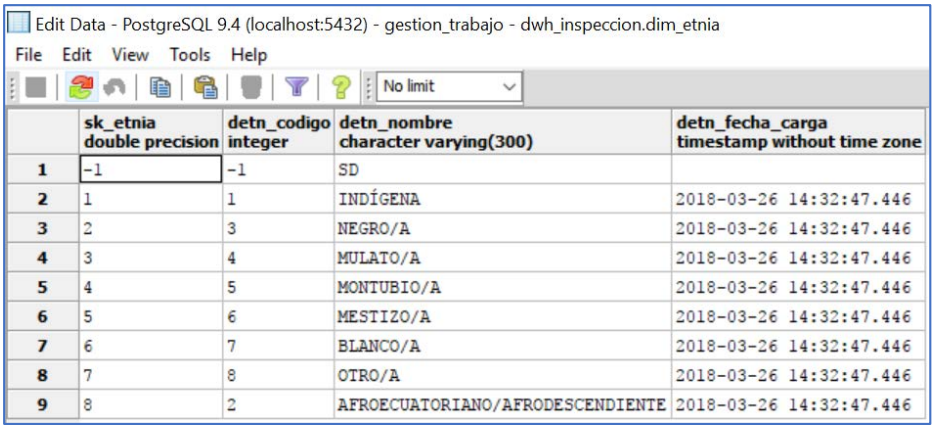


Figura 45. Proceso Carga de Datos Dimensión Etnia

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 46.



	sk_etnia double precision	detn_codigo integer	detn_nombre character varying(300)	detn_fecha_carga timestamp without time zone
1	-1	-1	SD	
2	1	1	INDÍGENA	2018-03-26 14:32:47.446
3	2	3	NEGRO/A	2018-03-26 14:32:47.446
4	3	4	MULATO/A	2018-03-26 14:32:47.446
5	4	5	MONTUBIO/A	2018-03-26 14:32:47.446
6	5	6	MESTIZO/A	2018-03-26 14:32:47.446
7	6	7	BLANCO/A	2018-03-26 14:32:47.446
8	7	8	OTRO/A	2018-03-26 14:32:47.446
9	8	2	AFROECUATORIANO/AFRODESCENDIENTE	2018-03-26 14:32:47.446

Figura 46. Datos en la Tabla Dimensión Etnia

- **Contrato:** En esta tabla se cargan los datos de los contratos registrados por las empresas. En la Figura 47 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo finalizado (true –> 1, false –> 0)
    - Reemplazo es_juvenil (true –> 1, false –> 0)
    - Transformación a letras mayúsculas (cargo).
    - Reemplazo valores nulos por “0” en la variable es_juvenil.

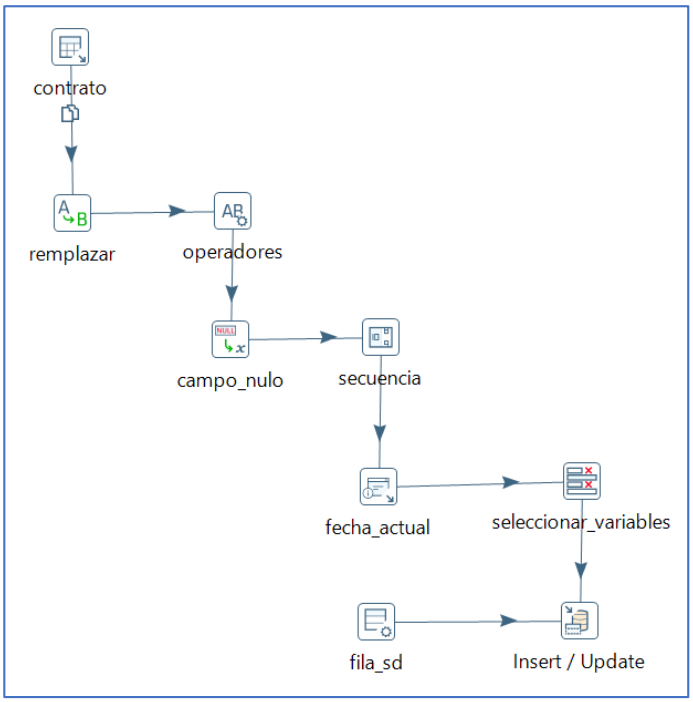
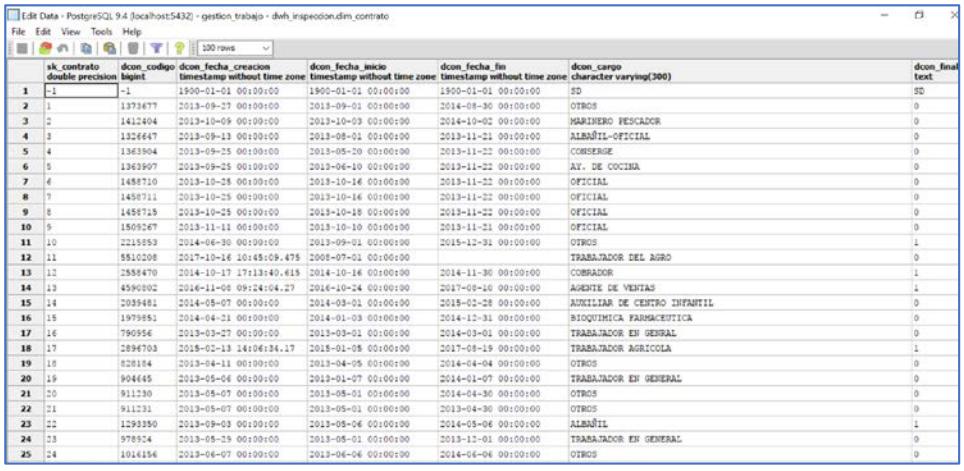


Figura 47. Proceso Carga de Datos Dimensión Contrato

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 48.



	sk_contrato double precision	dcon_codigo bigint	dcon_fecha_creacion timestamp without time zone	dcon_fecha_inicio timestamp without time zone	dcon_fecha_fin timestamp without time zone	dcon_cargo character varying(300)	dcon_final text
1	-1	-1	1900-01-01 00:00:00	1900-01-01 00:00:00	1900-01-01 00:00:00	SD	SD
2	1	1373677	2013-09-27 00:00:00	2013-09-01 00:00:00	2014-08-30 00:00:00	OTROS	0
3	2	1412404	2013-10-09 00:00:00	2013-10-03 00:00:00	2014-10-02 00:00:00	MARINERO PESCADOR	0
4	3	1326647	2013-09-13 00:00:00	2013-08-01 00:00:00	2013-11-21 00:00:00	ALBAÑIL-OFICIAL	0
5	4	1363904	2013-09-25 00:00:00	2013-05-20 00:00:00	2013-11-22 00:00:00	CONSERGE	0
6	5	1363907	2013-09-25 00:00:00	2013-06-10 00:00:00	2013-11-22 00:00:00	AY. DE COCINA	0
7	6	1458710	2013-10-25 00:00:00	2013-10-16 00:00:00	2013-11-22 00:00:00	OFICIAL	0
8	7	1458711	2013-10-25 00:00:00	2013-10-16 00:00:00	2013-11-22 00:00:00	OFICIAL	0
9	8	1458715	2013-10-25 00:00:00	2013-10-18 00:00:00	2013-11-22 00:00:00	OFICIAL	0
10	9	1509267	2013-11-11 00:00:00	2013-10-10 00:00:00	2013-11-21 00:00:00	OFICIAL	0
11	10	2215853	2014-06-30 00:00:00	2013-09-01 00:00:00	2015-12-31 00:00:00	OTROS	1
12	11	5510208	2017-10-16 10:45:09.475	2008-07-01 00:00:00		TRABAJADOR DEL AGRO	0
13	12	2558470	2014-10-17 17:13:40.615	2014-10-16 00:00:00	2014-11-30 00:00:00	COBRADOR	1
14	13	4590002	2016-11-08 09:24:04.27	2016-10-24 00:00:00	2017-08-16 00:00:00	AGENTE DE VENTAS	1
15	14	2039481	2014-05-07 00:00:00	2014-03-01 00:00:00	2015-02-28 00:00:00	AUXILIAR DE CENTRO INFANTIL	0
16	15	1979851	2014-04-21 00:00:00	2014-01-03 00:00:00	2014-12-31 00:00:00	BIOQUIMICA FARMACEUTICA	0
17	16	790956	2013-03-27 00:00:00	2013-03-01 00:00:00	2014-03-01 00:00:00	TRABAJADOR EN GENRAL	0
18	17	2896703	2015-02-13 14:06:34.17	2015-01-05 00:00:00	2017-08-19 00:00:00	TRABAJADOR AGRICOLA	1
19	18	828184	2013-04-11 00:00:00	2013-04-05 00:00:00	2014-04-04 00:00:00	OTROS	0
20	19	904445	2013-05-06 00:00:00	2013-01-07 00:00:00	2014-01-07 00:00:00	TRABAJADOR EN GENERAL	0
21	20	911230	2013-05-07 00:00:00	2013-05-01 00:00:00	2014-04-30 00:00:00	OTROS	0
22	21	911231	2013-05-07 00:00:00	2013-05-01 00:00:00	2013-04-30 00:00:00	OTROS	0
23	22	1293350	2013-09-03 00:00:00	2013-05-06 00:00:00	2014-05-06 00:00:00	ALBAÑIL	1
24	23	978924	2013-05-29 00:00:00	2013-05-01 00:00:00	2013-12-01 00:00:00	TRABAJADOR EN GENERAL	0
25	24	1016156	2013-06-07 00:00:00	2013-06-06 00:00:00	2014-06-06 00:00:00	OTROS	0

Figura 48. Datos en la Tabla Dimensión Contrato

- **Discapacidad:** En esta tabla se cargan los datos de los tipos de discapacidad de los empleados. En la Figura 49 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo tipo_discapacidad (A –> AUDITIVA, F –> FISICA, I –> INTELECTUAL, L –> LENGUAJE, P –> PSICOLÓGICA, V –> VISUAL)
    - Transformación a letras mayúsculas (tipo_discapacidad, nomenclatura_discapacidad).

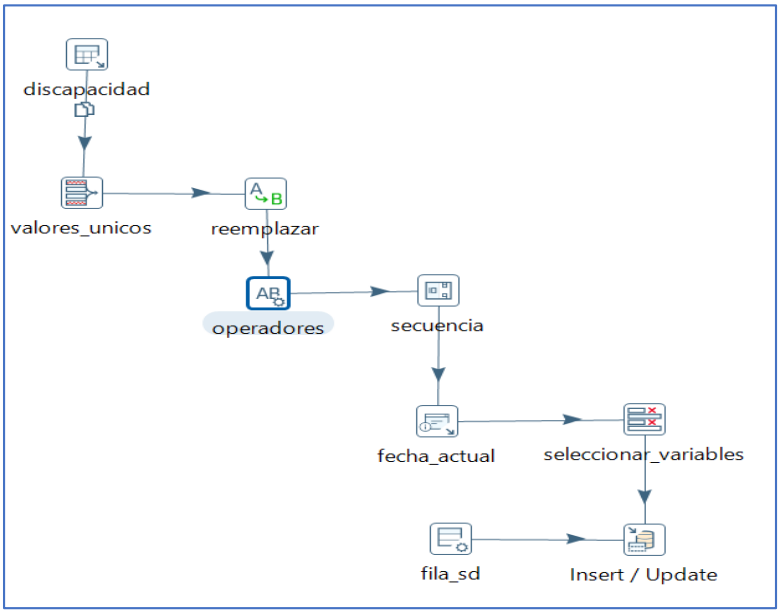


Figura 49. Proceso Carga de Datos Dimensión Discapacidad

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 50.

Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_discapacidad

File Edit View Tools Help

No limit

	sk_discapacidad double precision	ddis_tipo_discapacidad text	ddis_nomenclatura text	ddis_fecha_carga timestamp without time zone
1	-1	SD	SD	
2	1	VISUAL	V	2018-03-26 14:36:03.403
3	2	INTELECTUAL	I	2018-03-26 14:36:03.403
4	3	AUDITIVA	A	2018-03-26 14:36:03.403
5	4	PSICOLÓGICA	P	2018-03-26 14:36:03.403
6	5	FÍSICA	F	2018-03-26 14:36:03.403
7	6	LENGUAJE	L	2018-03-26 14:36:03.403

Figura 50. Datos en la Tabla Dimensión Discapacidad

- **Tipo Empresa:** En esta tabla se cargan los datos de los tipos de empresa. En la Figura 51 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Aplanar la tabla a dos niveles
    - Reemplazo tipo_empresa (F –> FINANCIERA, P–> PÚBLICA, E –> ESPECIAL, R –> PRIVADA, A –> ARTESANAL, M –> EMPRESA PÚBLICA)
    - Transformación a letras mayúsculas los nombres de los dos niveles.

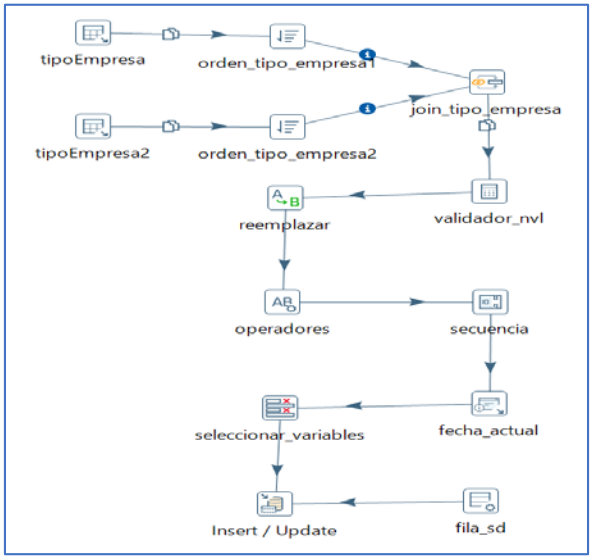
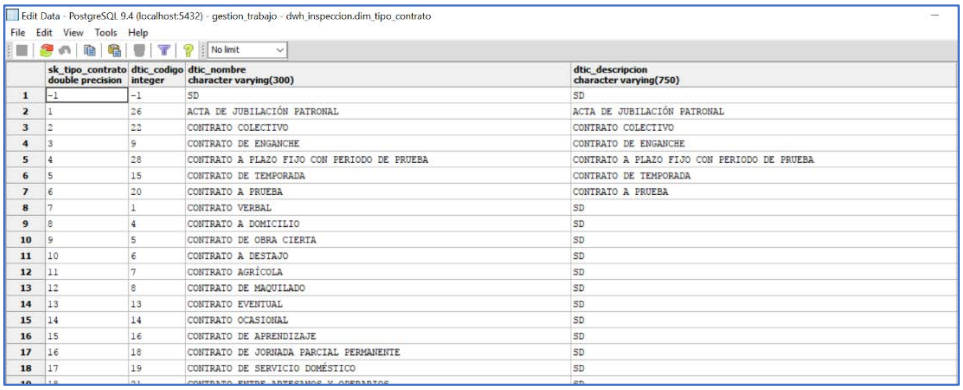


Figura 51. Proceso Carga de Datos Dimensión Tipo Contrato

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 52.



	sk_tipo_contrato double precision	dtic_codigo integer	dtic_nombre character varying(300)	dtic_descripcion character varying(750)
1	-1	-1	SD	SD
2	1	26	ACTA DE JUBILACIÓN PATRONAL	ACTA DE JUBILACIÓN PATRONAL
3	2	22	CONTRATO COLECTIVO	CONTRATO COLECTIVO
4	3	9	CONTRATO DE ENGANCHE	CONTRATO DE ENGANCHE
5	4	28	CONTRATO A PLAZO FIJO CON PERIODO DE PRUEBA	CONTRATO A PLAZO FIJO CON PERIODO DE PRUEBA
6	5	15	CONTRATO DE TEMPORADA	CONTRATO DE TEMPORADA
7	6	20	CONTRATO A PRUEBA	CONTRATO A PRUEBA
8	7	1	CONTRATO VERBAL	SD
9	8	4	CONTRATO A DOMICILIO	SD
10	9	5	CONTRATO DE OBRA CIERTA	SD
11	10	6	CONTRATO A DESTAJO	SD
12	11	7	CONTRATO AGRÍCOLA	SD
13	12	8	CONTRATO DE MAQUILADO	SD
14	13	13	CONTRATO EVENTUAL	SD
15	14	14	CONTRATO OCASIONAL	SD
16	15	16	CONTRATO DE APRENDIZAJE	SD
17	16	18	CONTRATO DE JORNADA PARCIAL PERMANENTE	SD
18	17	19	CONTRATO DE SERVICIO DOMÉSTICO	SD

Figura 52. Datos en la Tabla Dimensión Tipo Contrato

- **Acta de Finiquito:** En esta tabla se cargan los datos de las actas de finiquito que registran las empresas. En la Figura 53 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo esta_consignada (true –> 1, false –> 0)
    - Reemplazo esta_pagada (true –> 1, false –> 0)
    - Transformación a letras mayúsculas (cargo).
    - Reemplazo valores nulos por “SD” en la variable cargo.

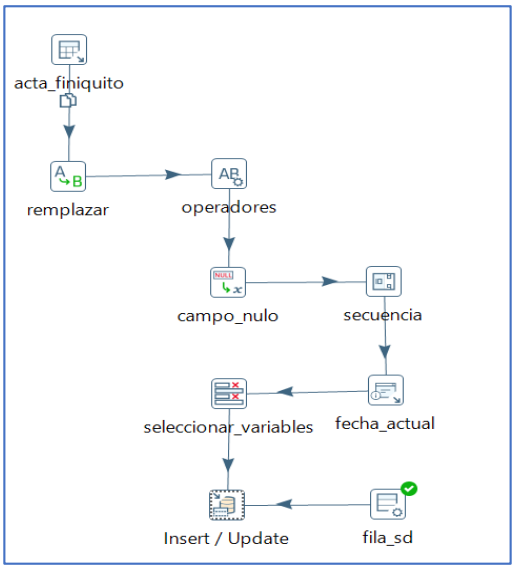
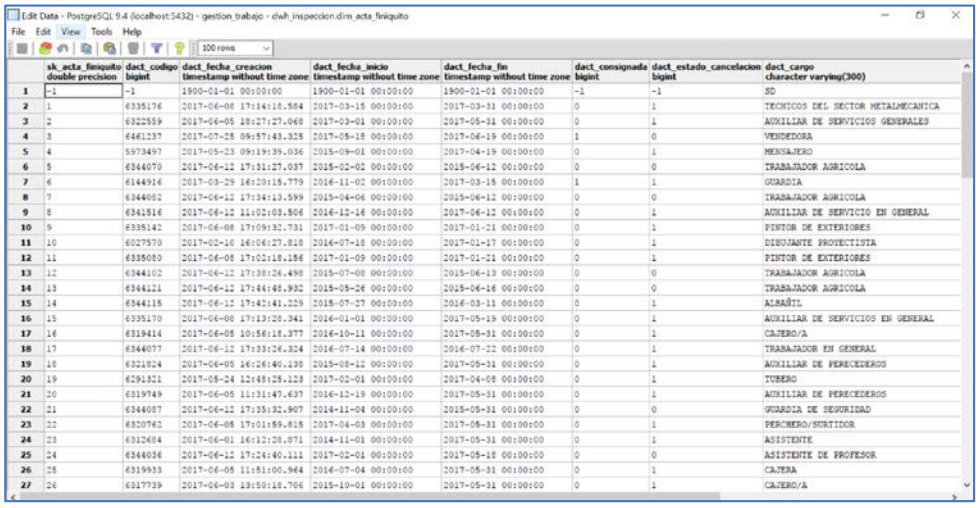


Figura 53. Proceso Carga de Datos Dimensión Acta de Finiquito

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 54.



	sk_acta_finiquito double precision	dact_codigo bigint	dact_fecha_creacion timestamp without time zone	dact_fecha_inicio timestamp without time zone	dact_fecha_fin timestamp without time zone	dact_consignada bigint	dact_estado_cancelacion bigint	dact_cargo character varying(300)
1	-1	-1	1900-01-01 00:00:00	1900-01-01 00:00:00	1900-01-01 00:00:00	-1	-1	SD
2	1	6335176	2017-06-08 17:14:18.584	2017-03-15 00:00:00	2017-03-31 00:00:00	0	1	TECNICOS DEL SECTOR METALMECANICA
3	2	6322559	2017-06-05 18:27:27.060	2017-03-01 00:00:00	2017-05-31 00:00:00	0	1	AUXILIAR DE SERVICIOS GENERALES
4	3	6461237	2017-07-25 09:57:43.325	2017-05-18 00:00:00	2017-06-19 00:00:00	1	0	VENDEDORA
5	4	5973497	2017-05-23 09:19:39.036	2015-09-01 00:00:00	2017-04-19 00:00:00	0	1	MENSAJERO
6	5	6344070	2017-06-12 17:31:27.037	2015-02-02 00:00:00	2015-06-12 00:00:00	0	0	TRABAJADOR AGRICOLA
7	6	6144916	2017-03-29 16:20:15.779	2016-11-02 00:00:00	2017-03-15 00:00:00	1	1	GUARDIA
8	7	6344082	2017-06-12 17:34:13.599	2015-04-06 00:00:00	2015-06-12 00:00:00	0	0	TRABAJADOR AGRICOLA
9	8	6341516	2017-06-12 11:02:03.506	2016-12-16 00:00:00	2017-06-12 00:00:00	0	1	AUXILIAR DE SERVICIO EN GENERAL
10	9	6335142	2017-06-08 17:09:32.731	2017-01-09 00:00:00	2017-01-21 00:00:00	0	1	PINTOR DE EXTERIORES
11	10	6027570	2017-02-10 16:06:27.818	2016-07-18 00:00:00	2017-01-17 00:00:00	0	1	DIBUJANTE PROYECTISTA
12	11	6335080	2017-06-08 17:02:18.156	2017-01-09 00:00:00	2017-01-21 00:00:00	0	1	PINTOR DE EXTERIORES
13	12	6344102	2017-06-12 17:38:26.498	2015-07-08 00:00:00	2015-06-13 00:00:00	0	0	TRABAJADOR AGRICOLA
14	13	6344121	2017-06-12 17:46:48.932	2015-05-26 00:00:00	2015-06-16 00:00:00	0	0	TRABAJADOR AGRICOLA
15	14	6344115	2017-06-12 17:42:41.229	2015-07-27 00:00:00	2016-03-11 00:00:00	0	1	ALBAÑIL
16	15	6335170	2017-06-08 17:13:28.341	2016-01-01 00:00:00	2017-05-19 00:00:00	0	1	AUXILIAR DE SERVICIOS EN GENERAL
17	16	6319414	2017-06-05 10:56:18.377	2016-10-11 00:00:00	2017-05-31 00:00:00	0	1	CAJERO/A
18	17	6344077	2017-06-12 17:33:26.324	2016-07-14 00:00:00	2016-07-22 00:00:00	0	1	TRABAJADOR EN GENERAL
19	18	6321824	2017-06-05 16:26:40.138	2015-08-12 00:00:00	2017-05-31 00:00:00	0	1	AUXILIAR DE PERECEDEROS
20	19	6291321	2017-05-24 12:48:25.123	2017-02-01 00:00:00	2017-04-08 00:00:00	0	1	TUBERO
21	20	6319749	2017-06-05 11:31:47.637	2016-12-19 00:00:00	2017-05-31 00:00:00	0	1	AUXILIAR DE PERECEDEROS
22	21	6344087	2017-06-12 17:35:32.907	2014-11-04 00:00:00	2015-05-31 00:00:00	0	0	GUARDIA DE SEGURIDAD
23	22	6320762	2017-06-05 17:01:59.615	2017-04-03 00:00:00	2017-05-31 00:00:00	0	1	PERCHERO/SURTIDOR
24	23	6312684	2017-06-01 16:12:28.871	2014-11-01 00:00:00	2017-05-31 00:00:00	0	1	ASISTENTE
25	24	6344036	2017-06-12 17:26:40.111	2017-02-01 00:00:00	2017-05-18 00:00:00	0	0	ASISTENTE DE PROFESOR
26	25	6319933	2017-06-05 11:51:00.964	2016-07-04 00:00:00	2017-05-31 00:00:00	0	1	CAJERA
27	26	6317739	2017-06-03 13:50:18.706	2015-10-01 00:00:00	2017-05-31 00:00:00	0	1	CAJERO/A

Figura 54. Datos en la Tabla Dimensión Tipo Contrato

- **Motivo Salida:** En esta tabla se cargan los datos de los motivos de salida de los empleados. En la Figura 55 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Transformación a letras mayúsculas (nombre).

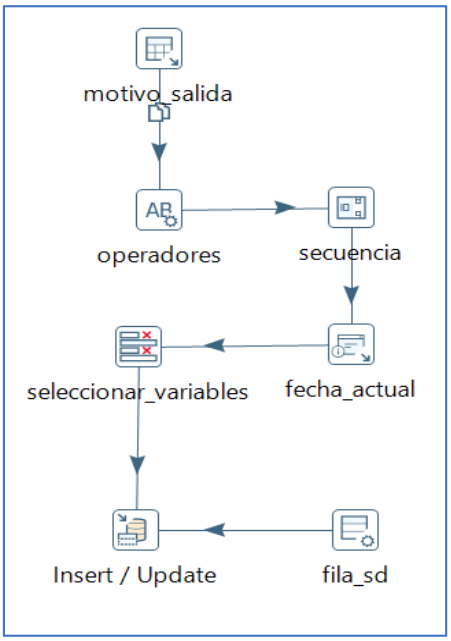


Figura 55. Proceso Carga de Datos Dimensión Motivo de Salida

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 56.

Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_motivo_salida

File Edit View Tools Help

No limit

	sk_motivo_salida double precision	dmot_codigo integer	dmot_nombre character varying(300)
1	-1	-1	SD
2	1	2	POR ACUERDO DE LAS PARTES.
3	2	3	POR LA CONCLUSIÓN DE LA OBRA, PERÍODO DE LABOR O SERVICIOS OBJETO DEL CONTRATO.
4	3	4	POR MUERTE O INCAPACIDAD DEL EMPLEADOR O EXTINCIÓN DE LA PERSONA JURÍDICA CONTRATANTE.
5	4	5	POR MUERTE DEL TRABAJADOR O INCAPACIDAD PERMANENTE Y TOTAL PARA EL TRABAJO.
6	5	6	POR CASO FORTUITO O FUERZA MAYOR QUE IMPOSIBILITEN EL TRABAJO, COMO INCENDIO, TERREMOTO, TEMPESTAD, EXPLOSIÓN, PLAGAS DEL CA
7	6	7	POR VOLUNTAD DEL EMPLEADOR PREVIO VISTO BUENO.
8	7	8	POR VOLUNTAD DEL TRABAJADOR PREVIO VISTO BUENO.
9	8	9	POR DESAHUCIO.
10	9	10	POR DESPIDO INTEMPESTIVO.
11	10	14	POR TERMINACIÓN DEL CONTRATO ANTES DEL PLAZO CONVENIDO.
12	11	12	POR TERMINACIÓN DENTRO DEL PERIODO DE PRUEBA
13	12	11	POR SER ASUMIDOS POR OTRO EMPLEADOR EN APLICACIÓN DEL MANDATO CONSTITUYENTE NO 8.
14	13	15	POR LAS CAUSAS LEGALMENTE PREVISTAS EN EL CONTRATO

Figura 56. Datos en la Tabla Dimensión Motivo de Salida

- **Grupo Ocupacional:** En esta tabla se cargan los datos de los grupos ocupacionales a los que pertenecen los empleados. En la Figura 57 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Aplanar la tabla a seis niveles
    - Transformación a letras mayúsculas los nombres de los seis niveles.

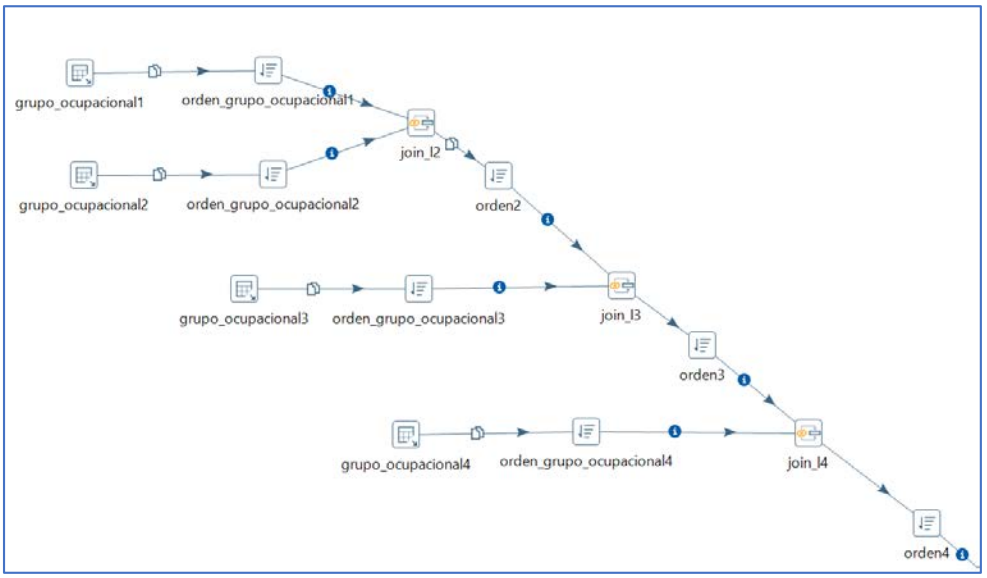


Figura 57. Proceso Carga de Datos Dimensión Grupo Ocupacional

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 58.

Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_grupo_ocupacional

File Edit View Tools Help

100 rows

	sk_grupo_ocupacional double precision	dgru_codigo_l6 character varying(15)	dgru_codigo_l1 text	dgru_nombre_l1 text	dgru_codigo_l2 text	dgru_nombre_l2 text
1	-1	SD	SD	SIN DATO	SD	SIN DATO
2	1	0	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
3	2	3	3	TÉCNICOS Y PROFESIONALES DEL NIVEL MEDIO	3	TÉCNICOS Y PROFESIONALES DEL N
4	3	4	4	PERSONAL DE APOYO ADMINISTRATIVO	4	PERSONAL DE APOYO ADMINISTRATI
5	4	9	9	OCUPACIONES ELEMENTALES	9	OCUPACIONES ELEMENTALES
6	5	5	5	TRABAJADORES DE LOS SERVICIOS Y VENDEDORES DE COMERCIOS Y MERCADOS	5	TRABAJADORES DE LOS SERVICIOS
7	6	6	6	AGRICULTORES Y TRABAJADORES CALIFICADOS AGROPECUARIOS, FORESTALES Y PESQUEROS	6	AGRICULTORES Y TRABAJADORES CA
8	7	1	1	DIRECTORES Y GERENTES	1	DIRECTORES Y GERENTES
9	8	7	7	OFICIALES, OPERARIOS Y ARTESANOS DE ARTES MECÁNICAS Y DE OTROS OFICIOS	7	OFICIALES, OPERARIOS Y ARTESAN
10	9	2	2	PROFESIONALES CIENTÍFICOS E INTELECTUALES	2	PROFESIONALES CIENTÍFICOS E IN
11	10	8	8	OPERADORES DE INSTALACIONES Y MÁQUINAS Y ENSAMBLADORES	8	OPERADORES DE INSTALACIONES Y
12	11	02	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
13	12	01	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
14	13	03	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
15	14	2131.03.16	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
16	15	2131.03.17	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
17	16	2131.01.32	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
18	17	2131.01.31	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
19	18	011	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
20	19	0110	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
21	20	0110.01	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
22	21	0110.02	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
23	22	0110.03	0	OCUPACIONES MILITARES	0	OCUPACIONES MILITARES
24	23	0110.01.10	0	OCUPACIONES MILITARES	01	OFICIALES DE LAS FUERZAS ARMAD
25	24	0110.01.09	0	OCUPACIONES MILITARES	01	OFICIALES DE LAS FUERZAS ARMAD
26	25	0110.01.08	0	OCUPACIONES MILITARES	01	OFICIALES DE LAS FUERZAS ARMAD
27	26	0110.01.07	0	OCUPACIONES MILITARES	01	OFICIALES DE LAS FUERZAS ARMAD

Figura 58. Datos en la Tabla Dimensión Grupo Ocupacional

- **Trámites:** En esta tabla se cargan los datos de los trámites que ingresan los empleados. En la Figura 59 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.

- Obtener de los nombres de estado y tipo de trámite, realizando el cruce de la tabla catálogo.
- Transformación a letras mayúsculas (cat_nombre, es_nombre, tra_per_solicitado_por).
- Reemplazo valores nulos por “SD” en las variables dtra_estado, dtra_solicitante, dtra_tipo.

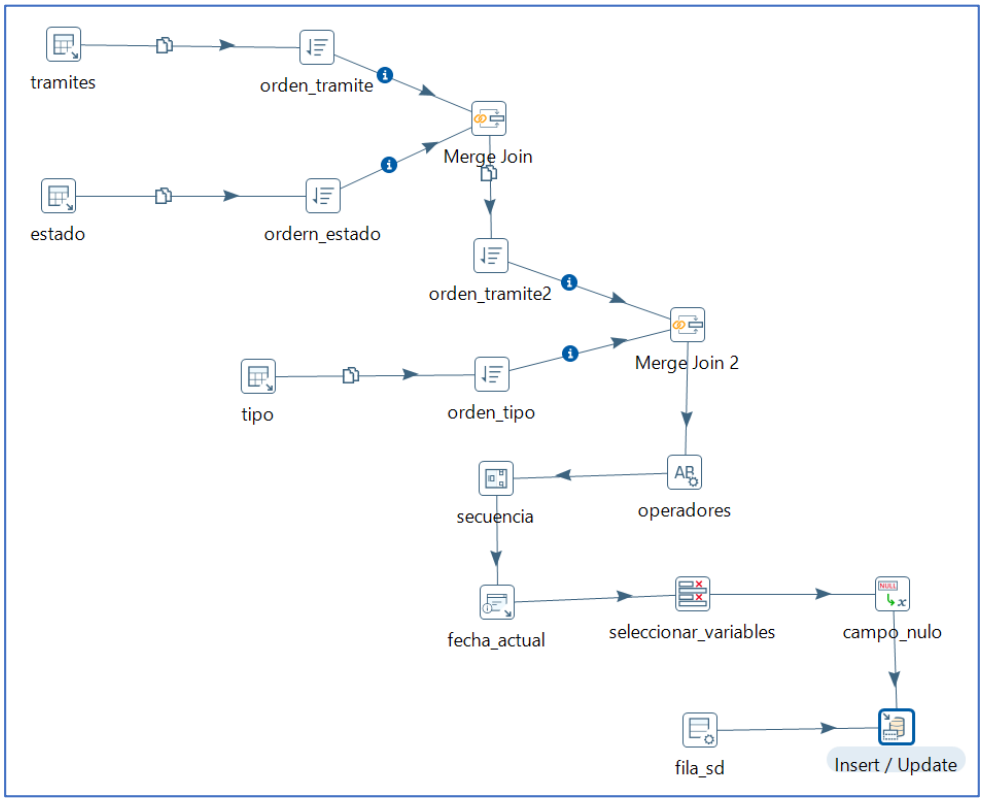
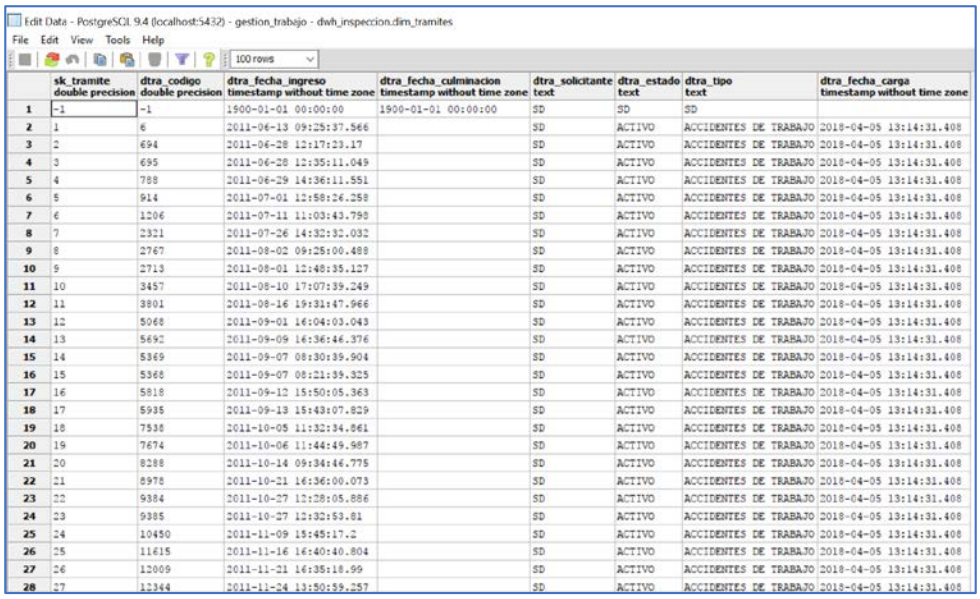


Figura 59. Proceso Carga de Datos Dimensión Trámites

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 60.


Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_tramites

File Edit View Tools Help

100 rows

	sk_tramite double precision	dtra_codigo double precision	dtra_fecha_ingreso timestamp without time zone	dtra_fecha_culminacion timestamp without time zone	dtra_solicitante text	dtra_estado text	dtra_tipo text	dtra_fecha_carga timestamp without time zone
1	-1	-1	1900-01-01 00:00:00	1900-01-01 00:00:00	SD	SD	SD	
2	1	6	2011-06-13 09:25:37.566		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
3	2	694	2011-06-28 12:17:23.17		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
4	3	695	2011-06-28 12:35:11.049		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
5	4	788	2011-06-29 14:36:11.551		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
6	5	914	2011-07-01 12:58:26.258		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
7	6	1206	2011-07-11 11:03:43.798		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
8	7	2321	2011-07-26 14:32:32.032		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
9	8	2767	2011-08-02 09:25:00.488		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
10	9	2713	2011-08-01 12:48:35.127		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
11	10	3457	2011-08-10 17:07:39.249		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
12	11	3801	2011-08-16 19:31:47.966		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
13	12	5068	2011-09-01 16:04:03.043		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
14	13	5692	2011-09-09 16:36:46.376		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
15	14	5369	2011-09-07 08:30:39.904		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
16	15	5368	2011-09-07 08:21:39.325		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
17	16	5818	2011-09-12 15:50:05.363		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
18	17	5935	2011-09-13 15:43:07.829		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
19	18	7538	2011-10-05 11:32:34.861		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
20	19	7674	2011-10-06 11:44:49.987		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
21	20	8288	2011-10-14 09:34:46.775		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
22	21	8978	2011-10-21 16:36:00.073		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
23	22	9384	2011-10-27 12:28:05.886		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
24	23	9385	2011-10-27 12:32:53.81		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
25	24	10450	2011-11-09 15:45:17.2		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
26	25	11615	2011-11-16 16:40:40.804		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
27	26	12009	2011-11-21 16:35:18.99		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408
28	27	12344	2011-11-24 13:50:59.257		SD	ACTIVO	ACCIDENTES DE TRABAJO	2018-04-05 13:14:31.408

Figura 60. Datos en la Tabla Dimensión Grupo Ocupacional

- **Boletas:** En esta tabla se cargan los datos de las boletas que ingresan los empleados. En la Figura 61 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Obtener el nombre del estado, realizando el cruce de la tabla catálogo.
    - Transformación a letras mayúsculas (es_nombre).
    - Reemplazo valores nulos por “SD” en la variable dbol_estado.

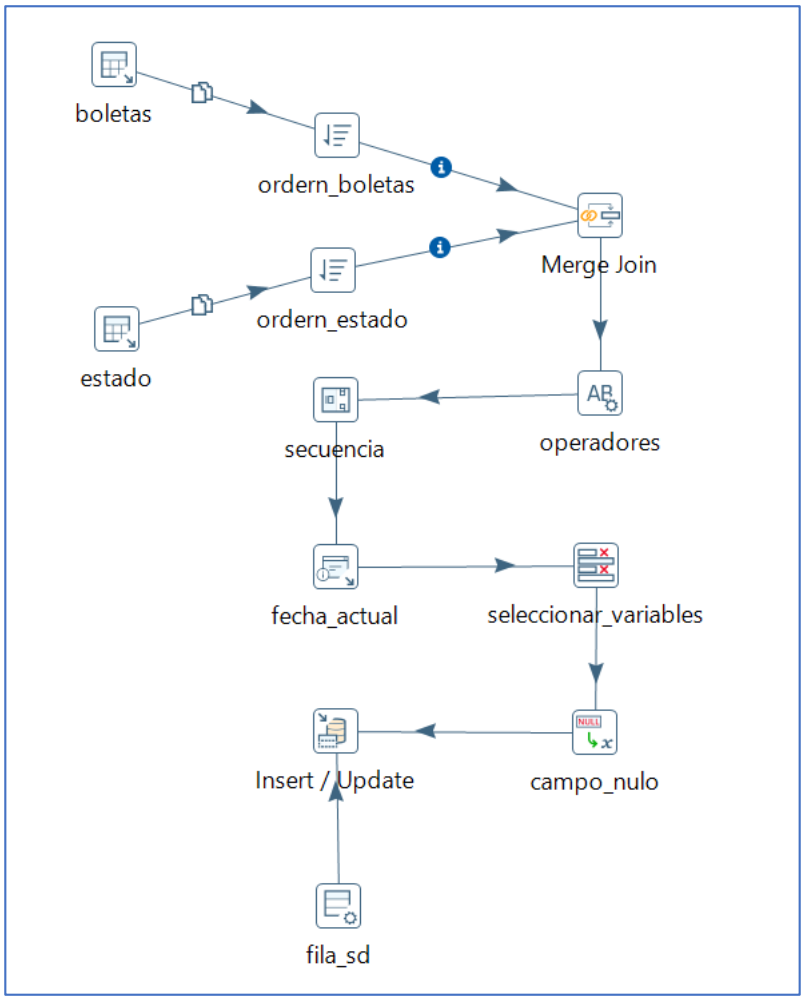
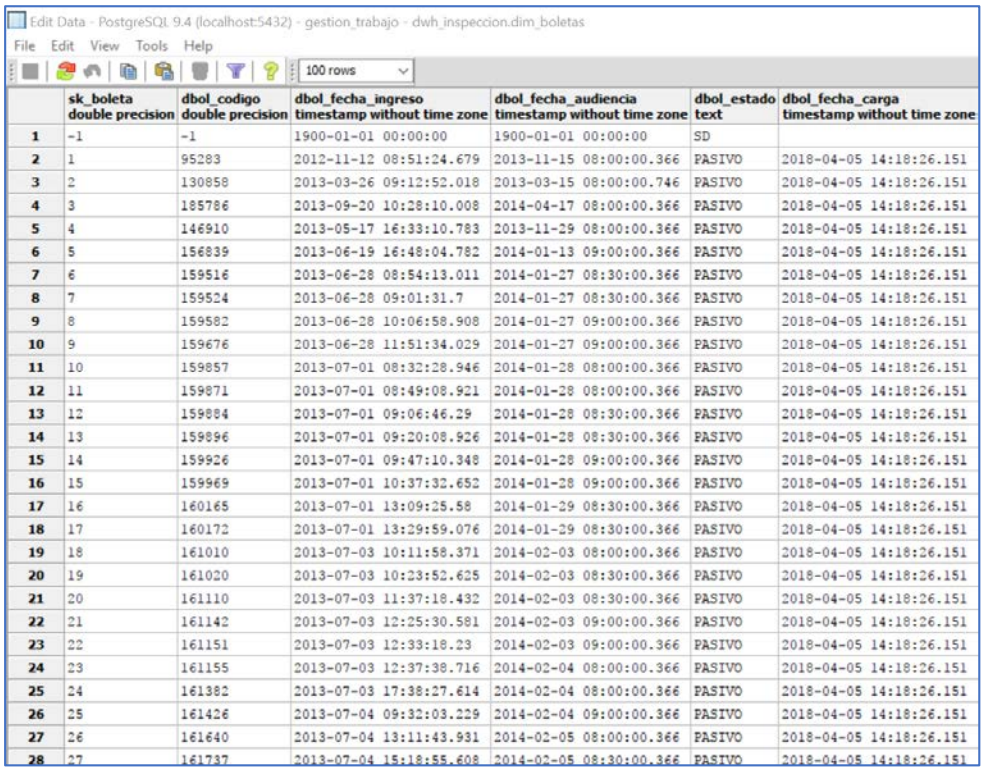


Figura 61. Proceso Carga de Datos Dimensión Boletas

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 62.



Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.dim_boletas

File Edit View Tools Help

100 rows

	sk_boleta double precision	dbol_codigo double precision	dbol_fecha_ingreso timestamp without time zone	dbol_fecha_audiencia timestamp without time zone	dbol_estado text	dbol_fecha_carga timestamp without time zone
1	-1	-1	1900-01-01 00:00:00	1900-01-01 00:00:00	SD	
2	1	95283	2012-11-12 08:51:24.679	2013-11-15 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
3	2	130858	2013-03-26 09:12:52.018	2013-03-15 08:00:00.746	PASIVO	2018-04-05 14:18:26.151
4	3	185786	2013-09-20 10:28:10.008	2014-04-17 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
5	4	146910	2013-05-17 16:33:10.783	2013-11-29 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
6	5	156839	2013-06-19 16:48:04.782	2014-01-13 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
7	6	159516	2013-06-28 08:54:13.011	2014-01-27 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
8	7	159524	2013-06-28 09:01:31.7	2014-01-27 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
9	8	159582	2013-06-28 10:06:58.908	2014-01-27 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
10	9	159676	2013-06-28 11:51:34.029	2014-01-27 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
11	10	159857	2013-07-01 08:32:28.946	2014-01-28 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
12	11	159871	2013-07-01 08:49:08.921	2014-01-28 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
13	12	159884	2013-07-01 09:06:46.29	2014-01-28 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
14	13	159896	2013-07-01 09:20:08.926	2014-01-28 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
15	14	159926	2013-07-01 09:47:10.348	2014-01-28 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
16	15	159969	2013-07-01 10:37:32.652	2014-01-28 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
17	16	160165	2013-07-01 13:09:25.58	2014-01-29 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
18	17	160172	2013-07-01 13:29:59.076	2014-01-29 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
19	18	161010	2013-07-03 10:11:58.371	2014-02-03 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
20	19	161020	2013-07-03 10:23:52.625	2014-02-03 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
21	20	161110	2013-07-03 11:37:18.432	2014-02-03 08:30:00.366	PASIVO	2018-04-05 14:18:26.151
22	21	161142	2013-07-03 12:25:30.581	2014-02-03 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
23	22	161151	2013-07-03 12:33:18.23	2014-02-03 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
24	23	161155	2013-07-03 12:37:38.716	2014-02-04 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
25	24	161382	2013-07-03 17:38:27.614	2014-02-04 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
26	25	161426	2013-07-04 09:32:03.229	2014-02-04 09:00:00.366	PASIVO	2018-04-05 14:18:26.151
27	26	161640	2013-07-04 13:11:43.931	2014-02-05 08:00:00.366	PASIVO	2018-04-05 14:18:26.151
28	27	161737	2013-07-04 15:18:55.608	2014-02-05 08:30:00.366	PASIVO	2018-04-05 14:18:26.151

Figura 62. Datos en la Tabla Dimensión Boletas

Carga de Datos en Tablas de Hechos

- **Contratos:** En esta tabla se cargan datos que nos permiten realizar la evaluación de los indicadores de los contratos registrados por las empresas. La tabla de hechos Contratos se relaciona con las tablas de dimensiones Empleado, Institución, Ubicación, Género, Actividad Económica, Tipo Contrato, Etnia, Contratos, Discapacidad, Tipo Empresa y con las medidas Número, Sueldo. En la Figura 61 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo valores nulos por “-1” en todos los identificadores de la tabla.
    - Cálculo del identificador de la fecha utilizando la fórmula:

      Año X10000 + Mes X 100 + Día

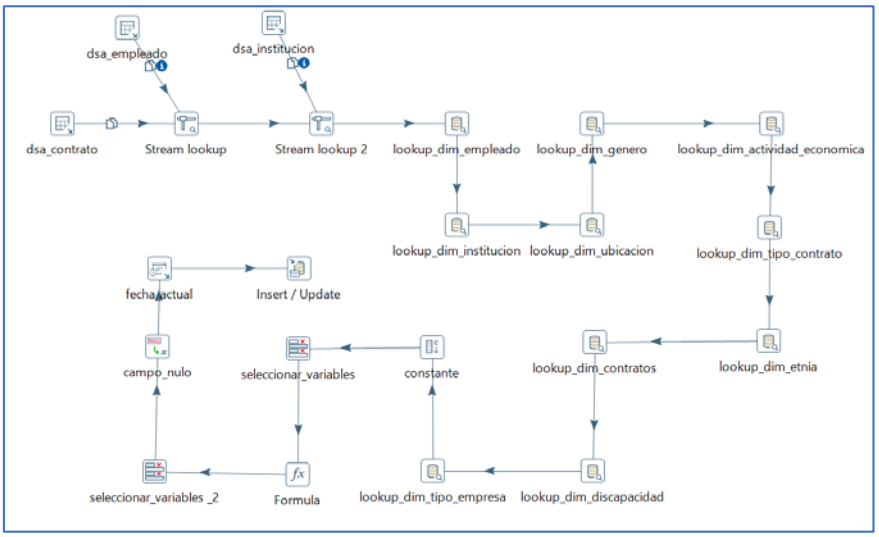
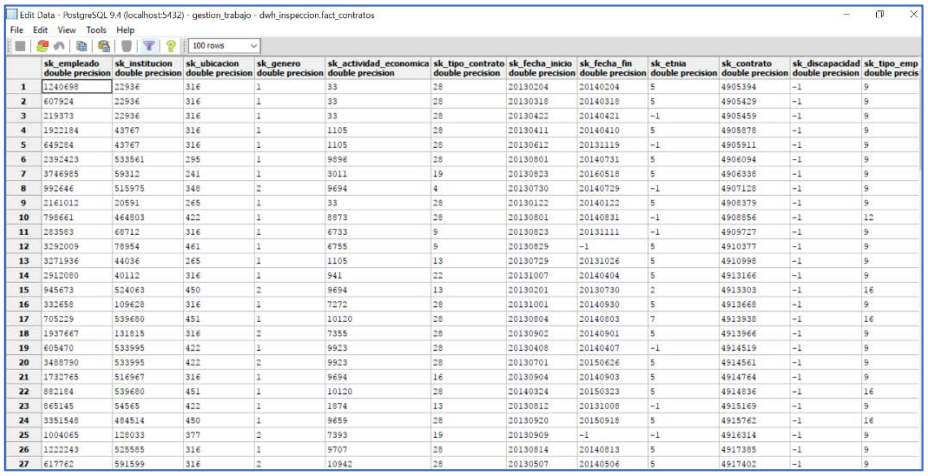


Figura 63. Proceso Carga de Datos Tabla de Hechos Contratos

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 64.



Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.fact_contratos

File Edit View Tools Help

100 rows

	sk_empleado double precision	sk_institucion double precision	sk_ubicacion double precision	sk_genero double precision	sk_actividad_economica double precision	sk_tipo_contrato double precision	sk_fecha_inicio double precision	sk_fecha_fin double precision	sk_etnia double precision	sk_contrato double precision	sk_discapacidad double precision	sk_tipo_emp double precis
1	1240698	22936	316	1	33	28	20130204	20140204	5	4905394	-1	9
2	607924	22936	316	1	33	28	20130318	20140318	5	4905429	-1	9
3	219373	22936	316	1	33	28	20130422	20140421	-1	4905459	-1	9
4	1922184	43767	316	1	1105	28	20130411	20140410	5	4905878	-1	9
5	649284	43767	316	1	1105	28	20130612	20131119	-1	4905911	-1	9
6	2392423	533561	295	1	9896	28	20130801	20140731	5	4906094	-1	9
7	3746985	59312	241	1	3011	19	20130823	20160518	5	4906338	-1	9
8	992646	515975	348	2	9694	4	20130730	20140729	-1	4907128	-1	9
9	2161012	20591	265	1	33	28	20130122	20140122	5	4908379	-1	9
10	798661	464803	422	1	8873	28	20130801	20140831	-1	4908856	-1	12
11	283583	68712	316	1	6733	9	20130823	20131111	-1	4909727	-1	9
12	3292009	78954	461	1	6755	9	20130829	-1	5	4910377	-1	9
13	3271936	44036	265	1	1105	13	20130729	20131026	5	4910998	-1	9
14	2912080	40112	316	1	941	22	20131007	20140404	5	4913166	-1	9
15	945673	524063	450	2	9694	13	20130201	20130730	2	4913303	-1	16
16	332658	109628	316	1	7272	28	20131001	20140930	5	4913668	-1	9
17	705229	539680	451	1	10120	28	20130804	20140803	7	4913938	-1	16
18	1937667	131815	316	2	7355	28	20130902	20140901	5	4913966	-1	9
19	605470	533995	422	1	9923	28	20130408	20140407	-1	4914519	-1	9
20	3488790	533995	422	2	9923	28	20130701	20150626	5	4914561	-1	9
21	1732765	516967	316	1	9694	16	20130904	20140903	5	4914764	-1	9
22	882184	539680	451	1	10120	28	20140324	20150323	5	4914836	-1	16
23	865145	54565	422	1	1074	13	20130812	20131008	-1	4915169	-1	9
24	3351548	484514	450	1	9659	28	20130920	20150918	5	4915762	-1	16
25	1004065	128033	377	2	7393	19	20130909	-1	-1	4916314	-1	9
26	1222243	525585	316	1	9707	28	20130814	20140813	5	4917385	-1	9
27	617762	591599	316	2	10942	28	20130507	20140506	5	4917402	-1	9

Figura 64. *Datos en la Tabla de Hechos Contratos*

- **Actas de Finiquito:** En esta tabla se cargan datos que nos permiten realizar la evaluación de los indicadores de las actas de finiquito registradas por las empresas. La tabla de hechos Actas de Finiquito se relaciona con las tablas de dimensiones Empleado, Institución, Ubicación, Motivo Salida, Actividad Económica, Tipo Contrato, Grupo Ocupacional, Contratos, Acta Finiquito, Tipo

Empresa y con las medidas Número, Monto Liquidado. En la Figura 63 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.

- Reemplazo valores nulos por “-1” en todos los identificadores de la tabla.
- Cálculo del identificador de la fecha utilizando la fórmula:

  Año X10000 + Mes X 100 + Día

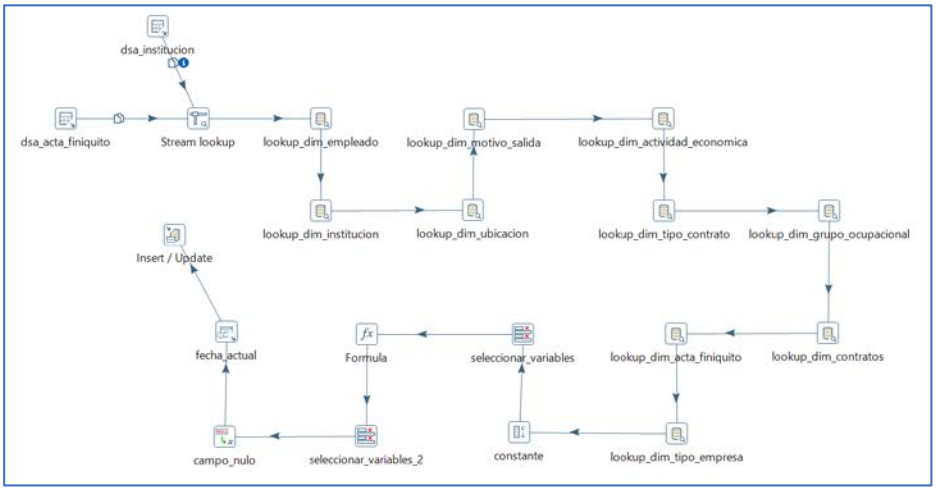


Figura 65. Proceso Carga de Datos Tabla de Hechos Actas de Finiquito

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 66.

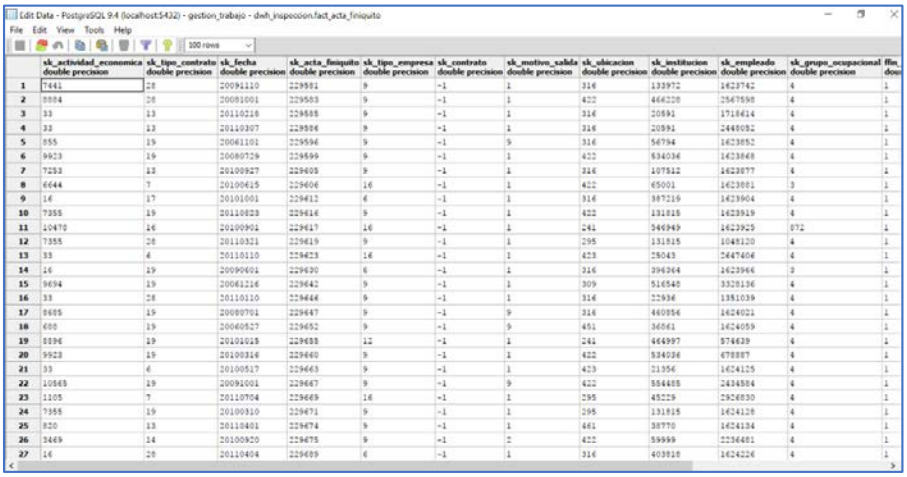


Figura 66. Datos en la Tabla de Hechos Actas de Finiquito

.

- **Trámites:** En esta tabla se cargan datos que nos permiten realizar la evaluación de los indicadores de los trámites registrados por los empleados. La tabla de hechos Trámites se relaciona con las tablas de dimensiones Empleado, Institución, Trámite, Tipo Trámite y con la medida Número. En la Figura 67 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo valores nulos por "-1" en todos los identificadores de la tabla.
    - Cálculo del identificador de la fecha utilizando la fórmula:

      Año X10000 + Mes X 100 + Día

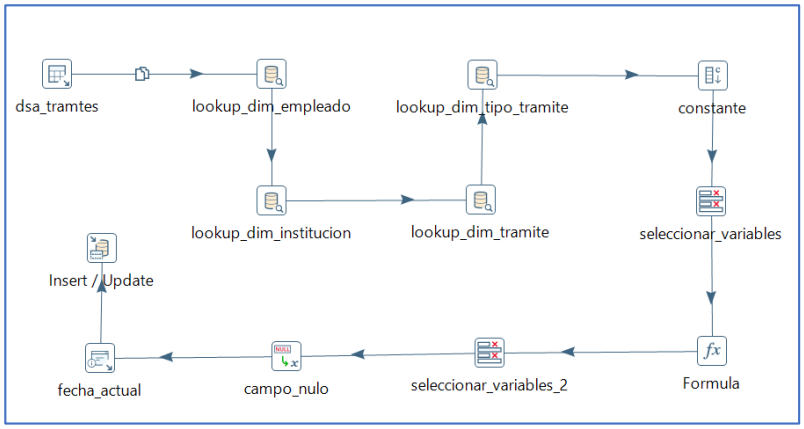
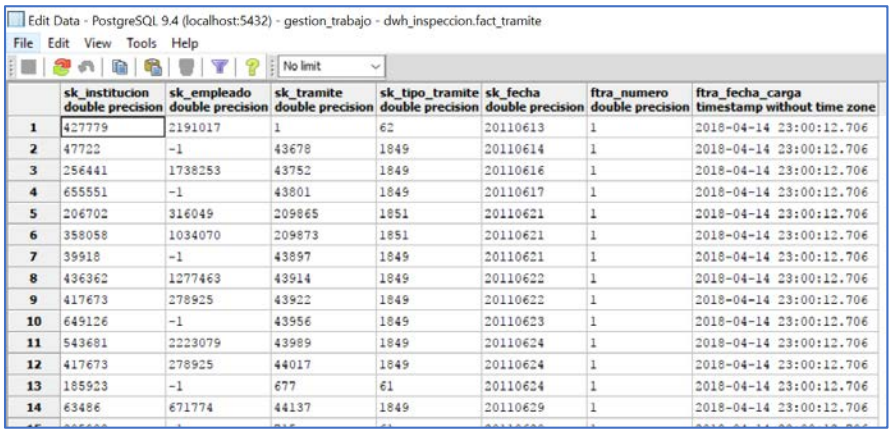


Figura 67. Proceso Carga de Datos Tabla de Hechos Trámites

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 68.



	sk_institucion double precision	sk_empleado double precision	sk_tramite double precision	sk_tipo_tramite double precision	sk_fecha double precision	ftra_numero double precision	ftra_fecha_carga timestamp without time zone
1	427779	2191017	1	62	20110613	1	2018-04-14 23:00:12.706
2	47722	-1	43678	1849	20110614	1	2018-04-14 23:00:12.706
3	256441	1738253	43752	1849	20110616	1	2018-04-14 23:00:12.706
4	655551	-1	43801	1849	20110617	1	2018-04-14 23:00:12.706
5	206702	316049	209865	1851	20110621	1	2018-04-14 23:00:12.706
6	358058	1034070	209873	1851	20110621	1	2018-04-14 23:00:12.706
7	39918	-1	43897	1849	20110621	1	2018-04-14 23:00:12.706
8	436362	1277463	43914	1849	20110622	1	2018-04-14 23:00:12.706
9	417673	278925	43922	1849	20110622	1	2018-04-14 23:00:12.706
10	649126	-1	43956	1849	20110623	1	2018-04-14 23:00:12.706
11	543681	2223079	43989	1849	20110624	1	2018-04-14 23:00:12.706
12	417673	278925	44017	1849	20110624	1	2018-04-14 23:00:12.706
13	185923	-1	677	61	20110624	1	2018-04-14 23:00:12.706
14	63486	671774	44137	1849	20110629	1	2018-04-14 23:00:12.706

***Figura 68*.** Datos en la Tabla de Hechos Trámites

- **Boletas:** En esta tabla se cargan datos que nos permiten realizar la evaluación de los indicadores de las boletas registradas por los empleados. La tabla de hechos Boletas se relaciona con las tablas de dimensiones Empleado, Institución, Boletas, Ubicación y con la medida Número. En la Figura 69 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
    - Reemplazo valores nulos por “-1” en todos los identificadores de la tabla.
    - Cálculo del identificador de la fecha utilizando la fórmula:

      Año X10000 + Mes X 100 + Día

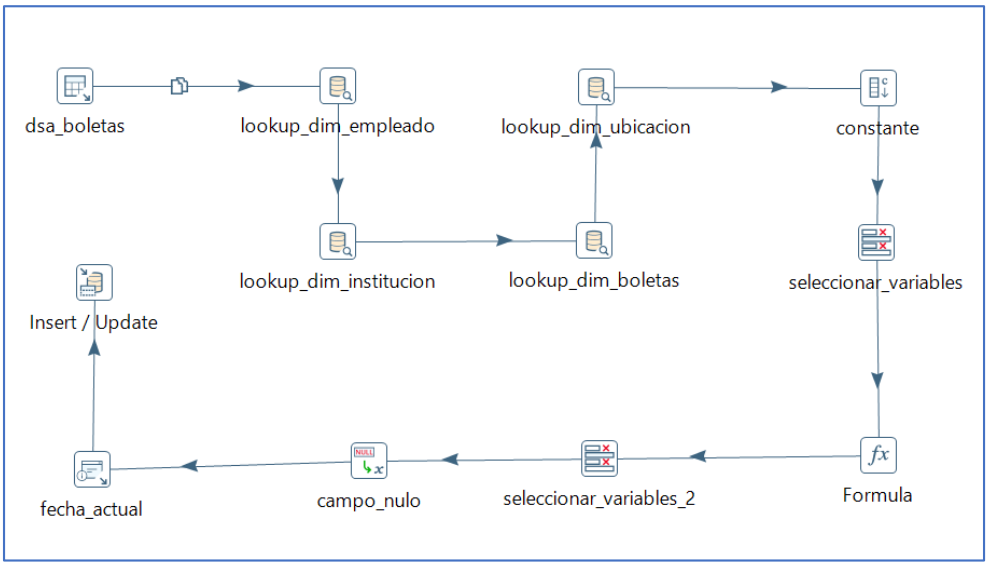


Figura 69. Proceso Carga de Datos Tabla de Hechos Boletas

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 70.

Edit Data - PostgreSQL 9.4 (localhost:5432) - gestion_trabajo - dwh_inspeccion.fact_boletas

File Edit View Tools Help

No limit

	sk_institucion double precision	sk_empleado double precision	sk_boleta double precision	sk_fecha double precision	sk_ubicacion double precision	fbol_numero double precision	fbol_fecha_carga timestamp without time zone
1	38659	2515112	906	20120807	305	1	2018-04-14 23:20:23.4
2	556123	1506663	917	20120820	422	1	2018-04-14 23:20:23.4
3	214303	3180256	921	20120802	422	1	2018-04-14 23:20:23.4
4	33318	2268887	1171	20121005	367	1	2018-04-14 23:20:23.4
5	605444	3766466	1234	20121114	430	1	2018-04-14 23:20:23.4
6	416805	-1	1238	20121119	397	1	2018-04-14 23:20:23.4
7	50775	1599141	1259	20121212	241	1	2018-04-14 23:20:23.4
8	535120	3652109	368966	20121204	422	1	2018-04-14 23:20:23.4
9	40158	2870928	1312	20130118	316	1	2018-04-14 23:20:23.4
10	22907	2920983	1334	20130204	422	1	2018-04-14 23:20:23.4
11	105192	-1	1388	20130319	379	1	2018-04-14 23:20:23.4
12	155198	620164	1406	20130318	423	1	2018-04-14 23:20:23.4
13	153977	3179133	1412	20130314	422	1	2018-04-14 23:20:23.4
14	166746	1992447	369011	20131003	341	1	2018-04-14 23:20:23.4
15	31455	3226926	1440	20130222	423	1	2018-04-14 23:20:23.4
16	21964	3411489	2	20130326	423	1	2018-04-14 23:20:23.4
17	485880	62749	369031	20130927	367	1	2018-04-14 23:20:23.4
18	497475	835318	1495	20131009	316	1	2018-04-14 23:20:23.4
19	109390	3491948	1577	20130411	379	1	2018-04-14 23:20:23.4
20	620295	-1	369050	20131004	309	1	2018-04-14 23:20:23.4
21	74926	412046	1606	20130409	316	1	2018-04-14 23:20:23.4
22	326476	-1	1645	20130416	309	1	2018-04-14 23:20:23.4
23	656739	1784679	1683	20130924	422	1	2018-04-14 23:20:23.4
24	17154	113665	1692	20130422	423	1	2018-04-14 23:20:23.4
25	78994	-1	369069	20130423	380	1	2018-04-14 23:20:23.4
26	565857	331731	1739	20130422	316	1	2018-04-14 23:20:23.4
27	111654	908730	1740	20130425	241	1	2018-04-14 23:20:23.4
28	539716	2682729	1795	20131016	422	1	2018-04-14 23:20:23.4

1012 rows.

Figura 70. Datos en la Tabla de Hechos Boletas

- **Factores Cumplimiento:** En esta tabla se cargan datos que nos permiten realizar la evaluación de los indicadores de los factores de cumplimiento de las empresas. La tabla de hechos Factores Cumplimiento se relaciona con las tablas de dimensiones Institución, Tipo Empresa, Actividad Económica y con las medidas Trabajo Menores, Trabajo Juvenil, Trabajo Discapacidad, Contratos Registrados Atrasados, Contratos Registradas Atrasadas, Trámites, Boletas, Resultado Incumplimiento, Resultado SGI. En la Figura 71 se muestra el proceso ETL. Esta tabla requirió de las siguientes transformaciones.
  - Reemplazo valores nulos por “-1” en todos los identificadores de la tabla.

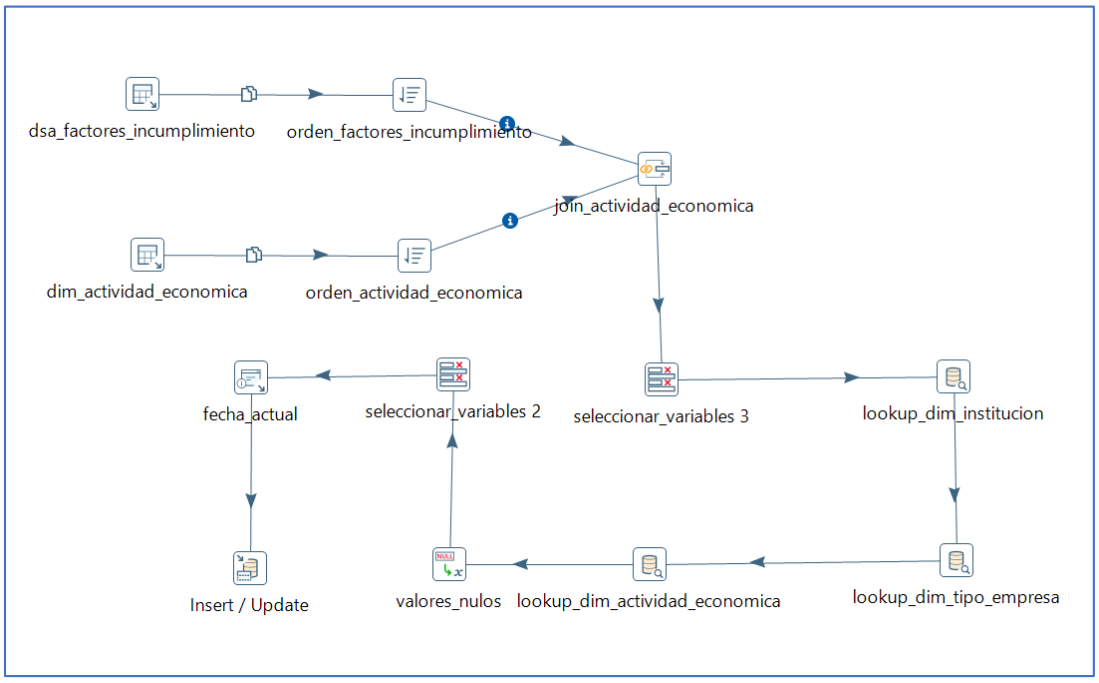
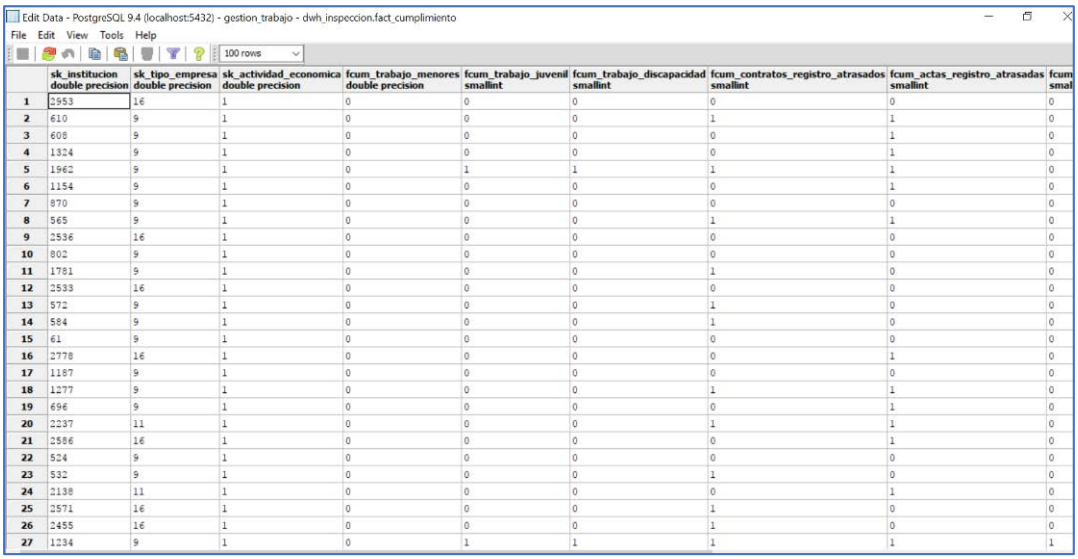


Figura 71. Proceso Carga de Datos Tabla de Hechos Factores Cumplimiento

Finalmente se realiza la carga al esquema DWH, donde se cargan los datos como se observa en la Figura 72.



	sk_institucion double precision	sk_tipo_empresa double precision	sk_actividad_economica double precision	fcum_trabajo_menores double precision	fcum_trabajo_juvenil smallint	fcum_trabajo_discapacidad smallint	fcum_contratos_registro_atrasados smallint	fcum_actas_registro_atrasadas smallint	fcum smal
1	2953	16	1	0	0	0	0	0	0
2	610	9	1	0	0	0	1	1	0
3	608	9	1	0	0	0	0	1	0
4	1324	9	1	0	0	0	0	1	0
5	1962	9	1	0	1	1	1	1	0
6	1154	9	1	0	0	0	0	1	0
7	870	9	1	0	0	0	0	0	0
8	565	9	1	0	0	0	1	1	0
9	2536	16	1	0	0	0	0	0	0
10	802	9	1	0	0	0	0	0	0
11	1781	9	1	0	0	0	1	0	0
12	2533	16	1	0	0	0	0	0	0
13	572	9	1	0	0	0	1	0	0
14	584	9	1	0	0	0	1	0	0
15	61	9	1	0	0	0	0	0	0
16	2778	16	1	0	0	0	0	1	0
17	1187	9	1	0	0	0	0	0	0
18	1277	9	1	0	0	0	1	1	0
19	696	9	1	0	0	0	0	1	0
20	2237	11	1	0	0	0	1	1	0
21	2586	16	1	0	0	0	0	1	0
22	524	9	1	0	0	0	0	0	0
23	532	9	1	0	0	0	1	0	0
24	2138	11	1	0	0	0	0	1	0
25	2571	16	1	0	0	0	1	0	0
26	2455	16	1	0	0	0	1	0	0
27	1234	9	1	0	1	1	1	1	1

Figura 72. Datos en la Tabla de Hechos Factores Cumplimiento

3.1.9. Exploración de Herramientas BI

En esta fase se va a explorar, explotar y visualizar los datos obtenidos en el Data Warehouse con la herramienta de BI Tableau, se expondrán ejemplos de reportes elaborados con Tableau.

3.1.10. Mantenimiento y Crecimiento del Data Warehouse

La base de datos y el modelo lógico del Data Warehouse se encuentran preparados para crecer en un futuro, es decir, para la base de datos se puede aumentar el tablespace y para el Data Warehouse puede incluirse nuevas dimensiones y tablas de hechos, siempre y cuando tengan relación con el proyecto planteado en el presente documento.

El mantenimiento no resulta caro debido a que los datos se encuentran en la base de datos y la exploración, explotación y visualización se puede realizar desde cualquier herramienta de BI solo conectándose a la base de datos, lo que se recomendaría es realizar un respaldo de la base de datos.

Cabe aclarar que esta fase el Ministerio del Trabajo es quién debe tomar la decisión de continuidad.

3.2. Construcción Modelo de Minería de Datos aplicando la Metodología CRISP-DM

En esta parte se utilizará la Metodología CRISP-DM para la construcción del Modelo de Minería de Datos, se aplicará las fases que comprenden la metodología.

La metodología CRISP-DM se escogió en base a la comparativa realizada en la Tabla 2, donde se hizo un resumen de dos de las metodologías más relevantes como son: CRISP-DM y SEMMA, dando como resultado que la metodología que se adapta a nuestro proyecto es CRISP-DM, adicional se tomó en cuenta el resultado de la encuesta realizada sobre el uso de metodologías usadas para proyectos de minería de datos, como se muestra en la Figura 1.

3.2.1. Comprensión del negocio

En esta primera fase de la minería de datos se determinarán los objetivos y requerimientos del proyecto.

3.2.1.1. Determinar los Objetivos del Negocio

El objetivo para el proyecto es realizar predicciones fiables a partir de los datos que se tiene de los Sistemas SAITE, SINACOI y SGI. El objetivo es determinar si una empresa está cumpliendo con la Normativa Legal Vigente (Código del Trabajo, Ley de Justicia Laboral y Acuerdos Ministeriales), la mismas que amparan al trabajador, a través de las predicciones se podrá establecer si una empresa debe ser inspeccionada o no, esto

servirá como apoyo al área de Inspectoría de Trabajo para que realice una planificación de las empresas a ser inspeccionadas de esta manera se optimizaran el tiempo que dedican los inspectores de trabajo en esta tarea.

Adicionalmente se obtiene Reglas de Asociación para determinar la tendencia que siguen las empresas inspeccionadas, en lo referente a la contratación, en relación al género, edad del empleado y a la actividad económica a la que pertenecen las empresas.

Contexto

Para la ejecución del presente proyecto se cuenta con un Data Warehouse de las tres bases de datos de los sistemas transaccionales SINACOI, SAITE y SGI, con los datos obtenidos en el data warehouse se realizó un estudio del comportamiento de las empresas en el cumplimiento de la Normativa Legal Vigente, así como la relación que existe entre el género y la edad de un empleado con la actividad económica de la empresa el momento de contratar recurso humano, tomando en cuenta si empresa es inspeccionada.

Objetivos del negocio

El propósito del negocio es la predicción de datos para las nuevas empresas que se registran en los sistemas del Ministerio del Trabajo, con la finalidad de realizar una predicción fiable partiendo de los datos que se tienen de las empresas registradas y encontrar la relación que existe entre los datos, en este proyecto se definieron los siguientes objetivos:

- Determinar un modelo predictivo que ayude a establecer si una empresa debe ser inspeccionada o no inspeccionada.
- Determinar los patrones que siguen las empresas inspeccionadas en lo referente a contratación y salida de personal.

Esto permitirá a las autoridades del ministerio específicamente en el área de Inspectoría de Trabajo, detectar las empresas que tiene inconvenientes en el cumplimiento de la Normativa Legal Vigente, con lo cual se podría identificar por qué las empresas no cumplen con sus obligaciones, una posible causa podría ser el poco conocimiento en el uso de los sistemas por parte de las empresas o falta de socialización por parte del Ministerio, lo cual permitirá al Ministerio del Trabajo mejorar los servicios ofrecidos a los usuarios externos.

Criterios de éxito del negocio

Se consideran como criterios de éxito:

- Predecir si nuevas empresas que registran información en los sistemas del Ministerio del Trabajo deben ser inspeccionadas o no, teniendo una alta fiabilidad.
- Detectar patrones de comportamiento en la contratación y salida de personal en las empresas inspeccionadas, con el fin de apoyar la toma de decisiones por parte de las autoridades del Ministerio.

3.2.1.2. Evaluación de la Situación

Se tiene un data warehouse, con información detallada de trámites, contratos y actas de finiquito de las empresas desde el año 2012 hasta enero del 2018, por lo que se puede

aseverar que se dispone de gran cantidad de datos que servirán para encontrar el modelo y resolver el problema de minería de datos.

Inventario de recursos

El software con el que se dispone para realizar la minería de datos es Knime que provee un entorno visual, integra diferentes módulos para aprendizaje automático y tareas de minería de datos, se trabajara con una base de datos Postgres en donde se almacenan los datos.

El hardware con el que contamos para realizar el proyecto son dos computadores portátiles con las siguientes características:

Tabla 28.

Características Hardware

Características	Computador Portátil 1	Computador Portátil 2
Marca	Toshiba	ASUS
Modelo	SATELLITE S55T	UX-303
Procesador	INTEL CORE I7 4710 HQ 2.50 GHZ	INTEL CORE I7 6500U HQ 2.5GHZ
Memoria Ram	16GB	12 GB
Disco Duro	1TB	512GB
Sistema Operativo	WINDOWS 10	WINDOWS 10

La fuente de dato que se tiene es un data warehouse que contienen información de los trámites, contratos de trabajo y actas de finiquito desde el año 2012 hasta enero del 2018.

El recurso humano con el que se cuenta para realizar el proyecto es el siguiente:

Tabla 29.

Detalle Recurso Humano

Cargo	Recurso Humano
Expertos en Minería de Datos	Ing. Edison Ayala Ing. Ana Logacho
Expertos en el negocio	Director de Control e Inspecciones Coordinador de Inspectoría de Trabajo de Pichincha Inspectores de Trabajo

Requisitos, Supuestos y Restricciones

- No se usan datos reales de las empresas debido a cuestiones legales, se utilizan datos ficticios (datos alterados por un factor).
- No se usan datos reales de los empleados debido a cuestiones legales, se utilizan ficticios.

Costes y Beneficios

La información que se usa en el proyecto no ocasiona un coste adicional al Ministerio del Trabajo, ya que la información pertenece al Ministerio a partir del instante que el responsable de ingresar la información a los sistemas transaccionales lo hace.

Este proyecto no genera beneficios económicos para el Ministerio, pero disminuye el tiempo que la Inspectoría de Trabajo se toma para realizar la planificación de inspecciones a las empresas que no cumplen con la Normativa Legal Vigente.

3.2.1.3. Determinar los Objetivos de la Minería de Datos

Los objetivos de la minería de datos son:

- Clasificar las empresas que deben ser inspeccionadas y no inspeccionadas, esto es determinado por el incumplimiento en el registro de contratos y actas de finiquito, y de las denuncias que los empleados registran en contra de las empresas.
- Buscar la relación que existe entre las variables relacionadas con la contratación y salida de personal en las empresas inspeccionadas.

Criterio de éxito de la minería de datos

Para la minería de datos se considera como criterios de éxito:

- Realizar predicciones sobre nuevas empresas con un porcentaje alto de fiabilidad, se definió el porcentaje en un 75%. La fiabilidad la determinará los algoritmos que se empleen para determinar el modelo de la minería de datos.
- Obtener reglas de asociación relevantes para las autoridades del Ministerio de Trabajo.

Estos temas se abordarán en la fase de evaluación de la metodología.

3.2.1.4. Generación plan de proyecto

El plan del proyecto se muestra en la siguiente tabla, donde se muestra el tiempo que se estima para cada fase de la metodología.

Tabla 30.

Plan del Proyecto

Nro.	Fase	Días
1	Comprensión del Negocio	10
2	Comprensión de los datos	15
3	Preparación de los datos	30
4	Modelado	20
5	Evaluación	20
6	Implementación o Despliegue	10
Total, en Días		**105**

Evaluación inicial de herramientas y técnicas

La herramienta que se ha escogido para nuestro proyecto de minería de datos es Knime, para la elección de esta herramienta se consideró el reporte del Cuadrante Mágico de Gartner para Plataformas de Ciencia de Datos y Aprendizaje Automático el último estudio es de enero del 2018, el estudio fue realizado a 16 herramientas como se muestra en la Figura 73 Knime se encuentra en el cuadrante de Líderes.

Figura 73. Cuadrante Mágico de Gartner para Plataformas de Ciencia de Datos y Aprendizaje Automático

Fuente: (Carlie Idoine, Peter Krensky, Erick Brethenoux, Jim Hare, Svetlana Sicular, Shubhangi Vashisth, 2018)

Las principales ventajas que ofrece esta herramienta son:

- Enfoque de código abierto permite a los usuarios y organizaciones minimizar sus costos de software de minería de datos sin comprometer la calidad.

- Ofrece una gran cantidad de algoritmos de apoyo tanto para principiantes como para expertos en minería de datos.
- Integración con otras herramientas y plataformas, como R, Python, Spark, H2O.ai, Weka, DL4J y Keras.
- Facilidad de uso por la interfaz de usuario y los ejemplos que proporciona.
- Automatiza la creación e implementación de modelos haciendo uso de la metodología CRISP.
- Enfoques automatizados para la calidad de los datos y la generación de características.
- Puede activar el reciclaje de modelos y admite la actualización y sincronización automática de datos.

Las técnicas de minería de datos que se han escogido para encontrar el modelo predictivo son técnicas de clasificación, estas técnicas usan registros históricos para la generación de un modelo de minería de datos, el cuál puede predecir el comportamiento futuro, en nuestro caso las técnicas de clasificación de minería datos se adaptan ya que se tiene datos desde el año 2012, el tiempo de respuesta de las técnicas seleccionadas es aceptable para el volumen de datos que estamos manejando, las técnicas seleccionadas son:

- Árboles de Decisión
- Regresión Logística
- Redes Neuronales

Se realizó la Tabla donde se evalúan diferentes criterios para la selección de las técnicas de minería de datos, se ha tomado como referencia dos estudios comparativos, *(Chintan Shah, Comparison of Data Mining Classification Algorithms for Breast Cancer Prediction, 2013)* y *(Pandey, 2016)*

Tabla 31.

Selección de Técnica de Minería de Datos

Criterio	Árboles de Decisión	Regresión Logística	Redes Neuronales
Fácil comprensión	Si	Si	Si
Uso de la técnica	Amplio	Amplio	Amplio
Preparación de datos	Poca	Poca	Poca
Tipos de Variables de entrada	Numéricos y Categóricos	Dummy[37]	Numéricos
Limite en el número de variables de entrada	No	No	No
Manejo del ruido y datos faltates	Si	Si	Si
Manejo de grandes volúmenes de datos	Si	Si	Si
Coste y Tiempo de Procesamiento	Depende de los datos	Depende de los datos	Alto
Precisión del clasificador	Alta	Depende de los datos	Alto
Facilidad de Interpretación	Si	Si	Si

La técnica de minería de datos que se ha escogido para encontrar reglas de asociación (buscar hechos comunes que suceden en un conjunto de datos), en nuestro caso de estudio como estamos utilizando la herramienta Knime se utilizara el algoritmo A

[37] Variables que pueden tomar el valor de 0 o 1, indican ausencia o presencia de una cualidad o atributo.

priori que es el que se encuentra disponible en la herramienta. (Katherine González, 2016)

3.2.2. Comprensión de los datos

En esta segunda fase se realiza la recolección inicial de datos con la finalidad de tener una idea inicial del problema, familiarizarnos con los datos y determinar la calidad de los mismos, adicional se identificará las relaciones existentes para establecer las primeras suposiciones.

3.2.2.1. Recolectar los Datos Iniciales

Los datos que se utilizan en el proyecto son referentes a empresas y empleados que incluyen información de tipo personal como nombre de la empresa, ruc, dirección, cédula de identidad, nombre de empleado, etc., por lo que por cuestiones legales no se utilizaran datos reales que se encuentran en las bases de datos del Ministerio del Trabajo.

El Data Warehouse construido con las bases de datos de los sistemas transaccionales SAITE, SINACOI y SGI propiedad del Ministerio del Trabajo, es la única fuente de extracción de datos, de este repositorio se tomará la información necesaria para dar cumplimiento a los objetivos planteados, los datos se encuentran comprendidos entre el año 2012 hasta enero del 2018.

3.2.2.2. Descripción de los Datos

Los datos se almacenan en el Data Warehouse, la descripción de cada una de las tablas que se relacionan entre sí, así como los campos de cada una de las tablas, se explica en la Sección 3.1.3.2. del presente capítulo.

3.2.2.3. Exploración de los Datos

Concluida la descripción de los datos, se continua con la exploración de lo datos, lo cual nos ayudara a determinar la consistencia de los datos.

En las provincias donde se concentran la mayor cantidad de registro de empresas son Pichincha, seguida por Guayas, Azuay, etc., como se puede observar en la Figura 74. El total de empresas que poseen contratos en estado vigente son 327805.

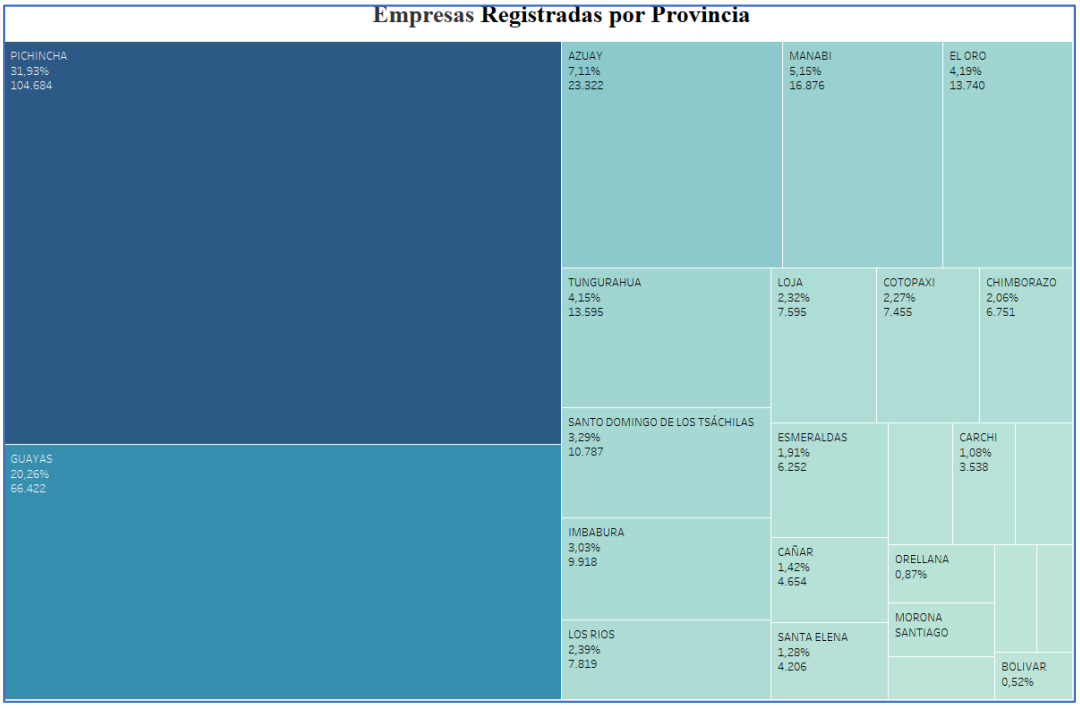


Figura 74. Empresas Registradas por Provincia

Del total de empresas que poseen contratos en estado vigente se tiene que 138803 no han sido inspeccionadas. Como se muestra en Figura 75.

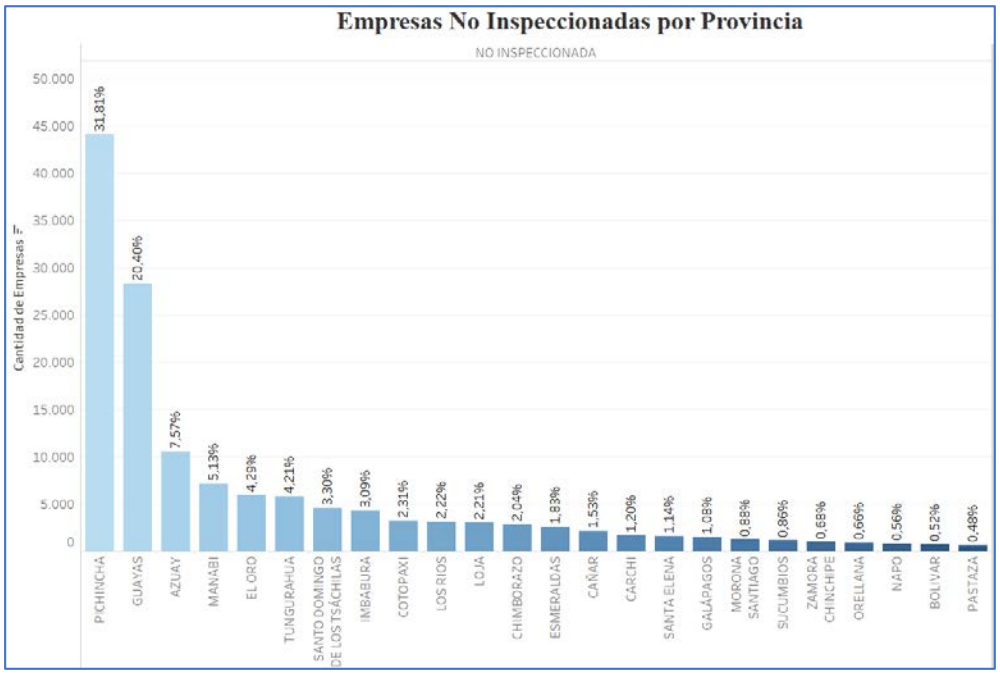


Figura 75. Empresas No Inspeccionadas por Provincia

Del total de empresas que poseen contratos en estado vigente se tiene que 189002 han sido inspeccionadas. Como se muestra en Figura 76.

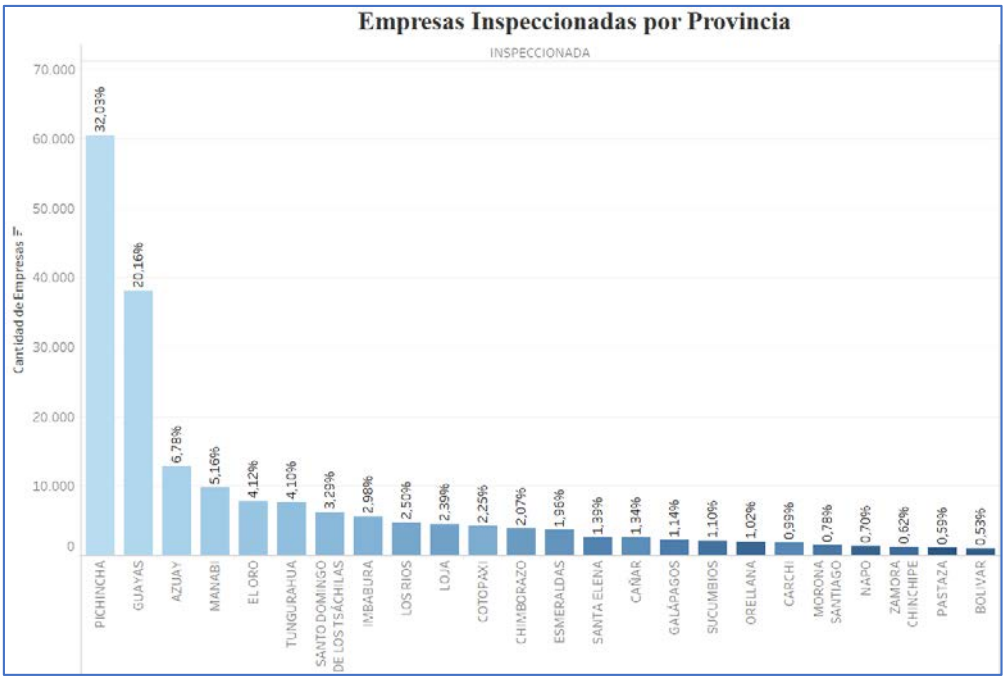


Figura 76. Empresas No Inspeccionadas por Provincia

En la Figura 77 se muestra la distribución de los empleados en relación con el género y la actividad económica, se observa que en el sector de la construcción contratan más empleados de género masculino y existen muy pocas contrataciones de género femenino.

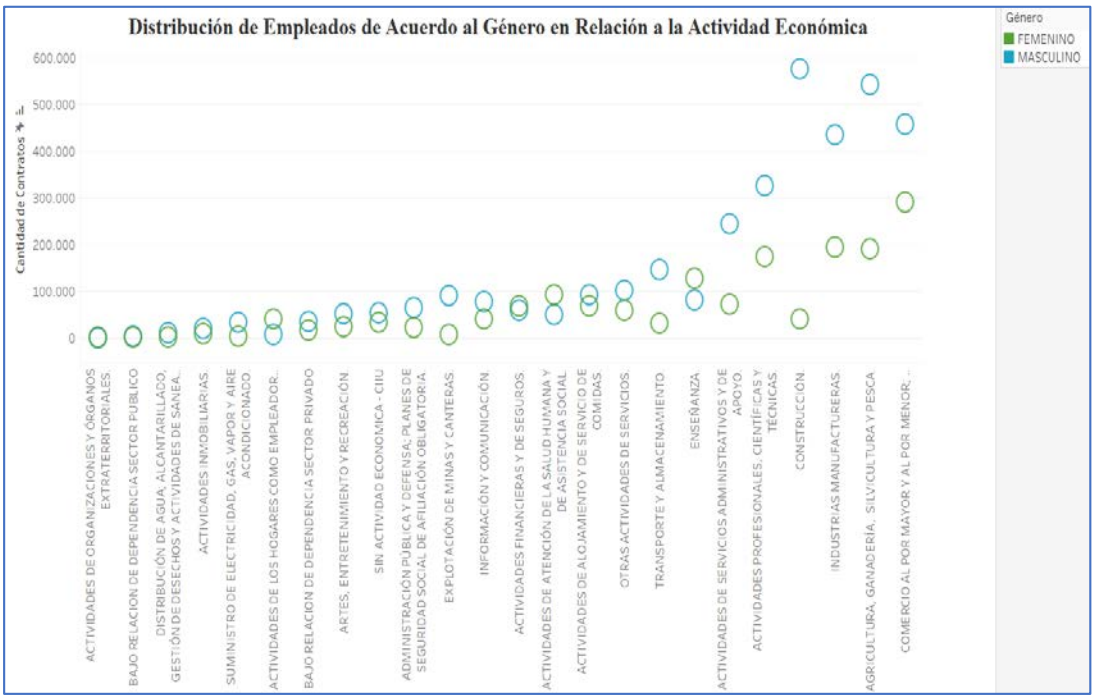


Figura 77. Distribución de Empleados de Acuerdo al Género en Relación a la Actividad Económica de la Empresa

En lo que se refiere a contratos vigentes por trabajo juvenil, se visualiza en la Figura 78, que los datos no se encuentran muy dispersos en 22 provincias, lo que se observa es que dos provincias Pichincha y Guayas tienen el mayor número de contratos con modalidad juvenil (edad del empleado comprendida entre 18 y 26 años), esto se debe a que en estas dos provincias es donde se concentran el poder económico del Ecuador.

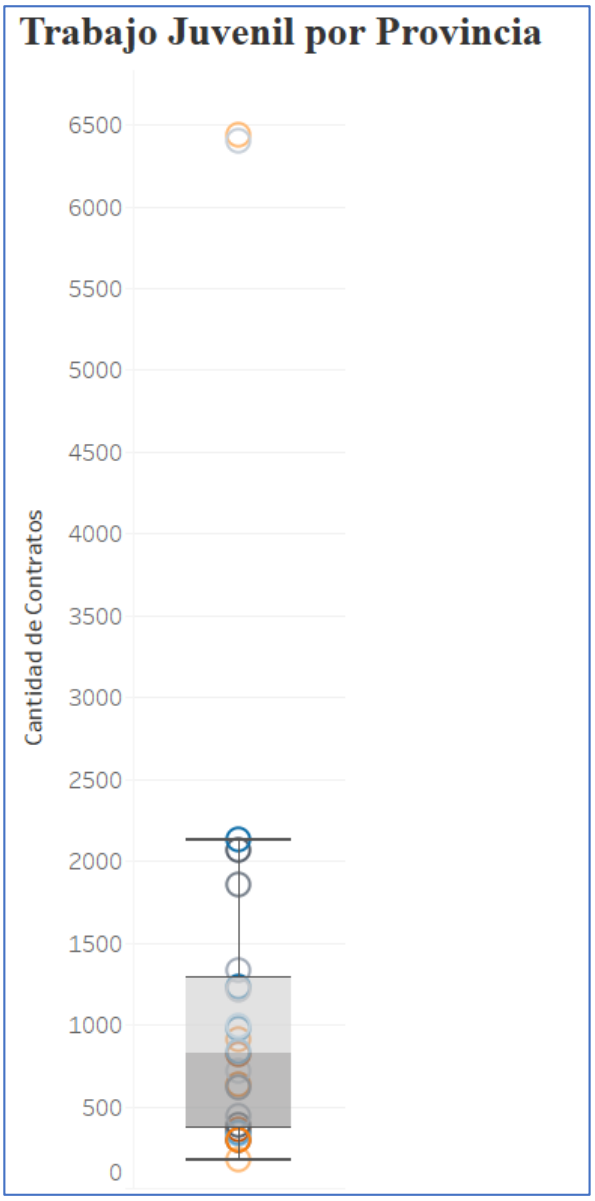


Figura 78. Trabajo Juvenil por Provincia

En la Figura 79, se visualiza que el rango de edad que las empresas prefieren contratar empleados es de 18 a 26 años, es decir los empleadores han acogido el trabajo juvenil como primera opción, seguido por el rango de edad de 26 a 35 años, como tercera opción se tiene el rango es de 35 a 45 años, en cuarto lugar se encuentran los trabajadores que tiene más de 46 años con lo que se comprueba en la realidad que vive

el país que las empresas prefieren contratar poco a las personas que se encuentran en este rango de edad, también se observa que las empresas tienen poco personal contratado en las edades de 15 a 18 años, y por último se tiene que el trabajo infantil es muy poco en comparación con los otros rangos de edades, esto se debe a los programas implementados por el gobierno y al apoyo que ha recibido por parte de la empresa privada.

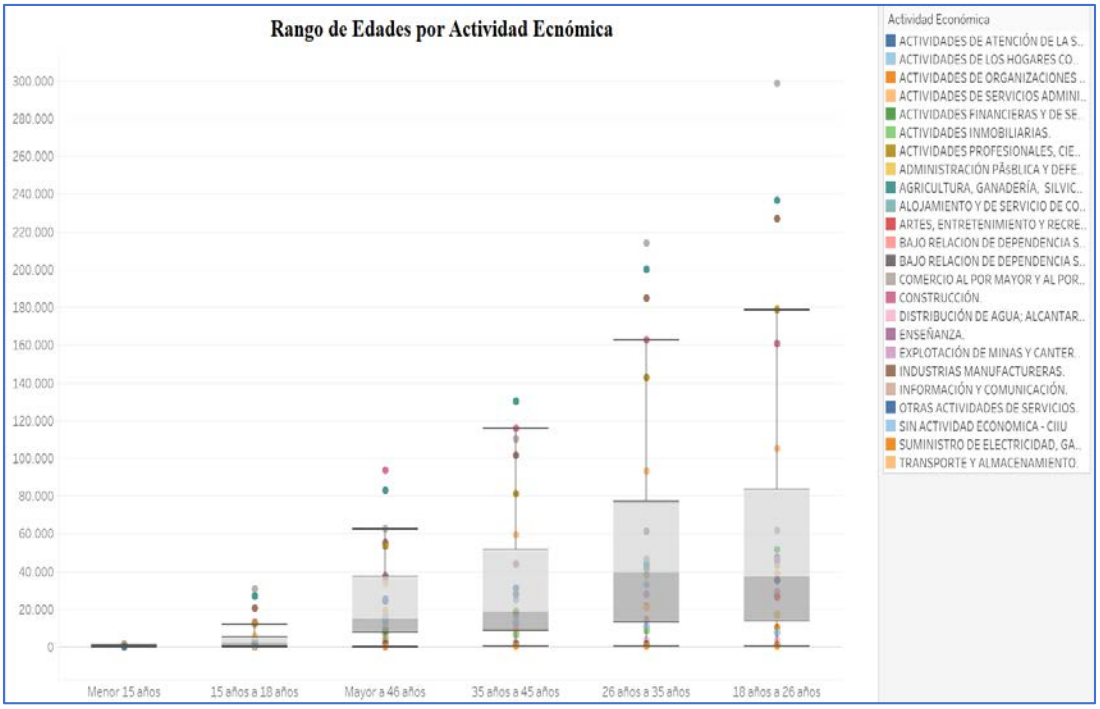


Figura 79. Rango de Edades por Actividad Económica

Los contratos que más se registran son los de tipo indefinido como se muestra en la Figura 80, a continuación, se encuentran los contratos de plazo fijo, este tipo de contrato actualmente ya no existen ya que se eliminó este tipo de contratación por la expedición de la **“LEY ORGÁNICA PARA LA JUSTICIA LABORAL Y RECONOCIMIENTO DEL TRABAJO EN EL HOGAR”,** la fecha de expedición fue el 15 de abril del 2015, donde indica que a partir del 1 de enero del 2016 ya no existe el tipo de contratación a plazo fijo.

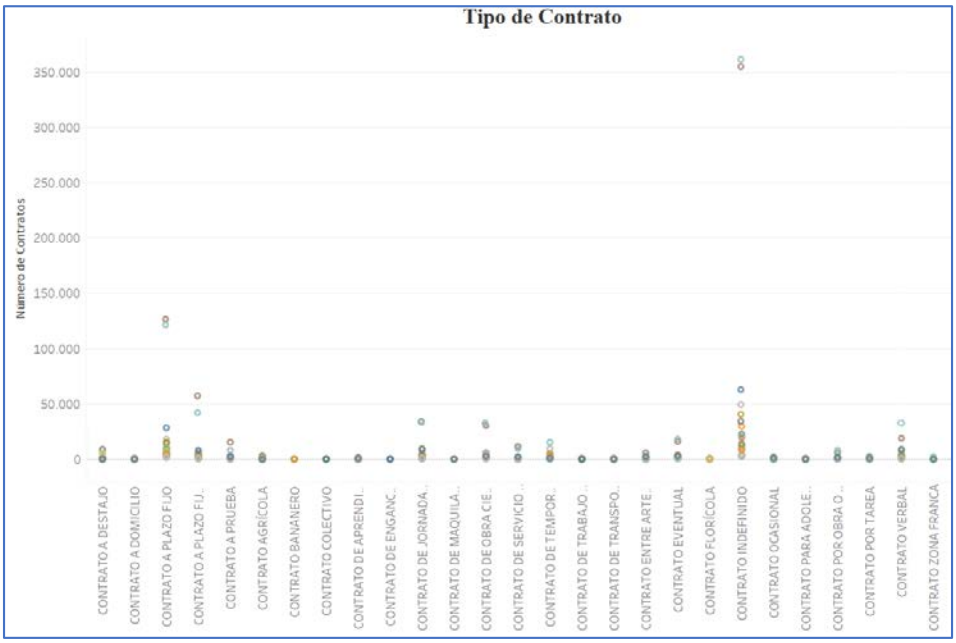


Figura 80. Tipo de Contratos por Provincia

En la Figura 81., se observa que las provincias de Pichincha y Guayas son las que más registran contratos, y las que menos registran son las provincias de Bolívar y Pastaza.

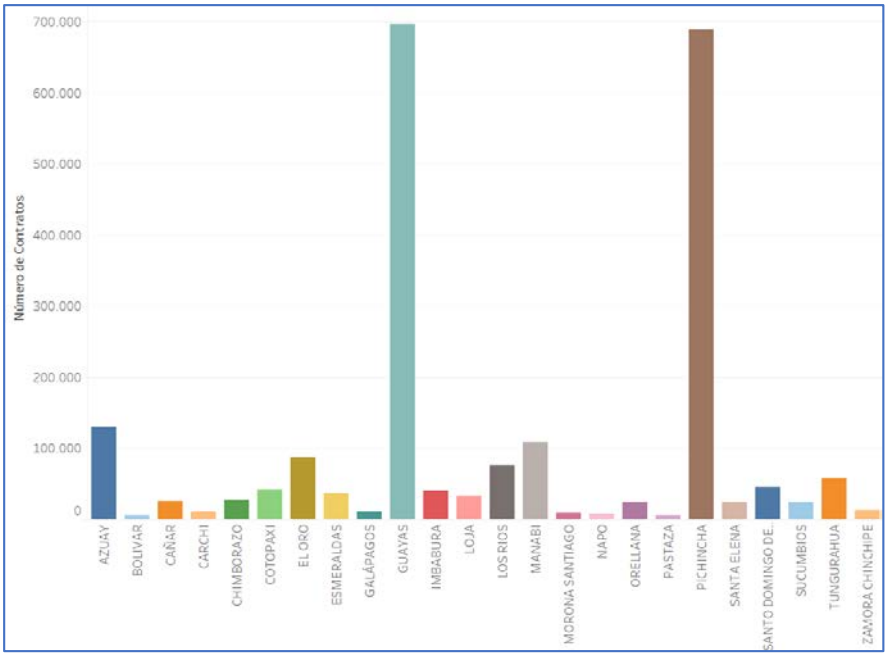


Figura 81. Distribución de Contratos por Provincia

Las actas de finiquito que más se registran son en los empleados en el rango de edad entre los 18 y 26 años, y los rangos que menos actas finiquito se registran son de 15 a 18 años, como se muestra en la Figura 82., adicional se observa el motivo de salida en donde mayor cantidad de actas se registran son “Por Acuerdo de las Partes”.

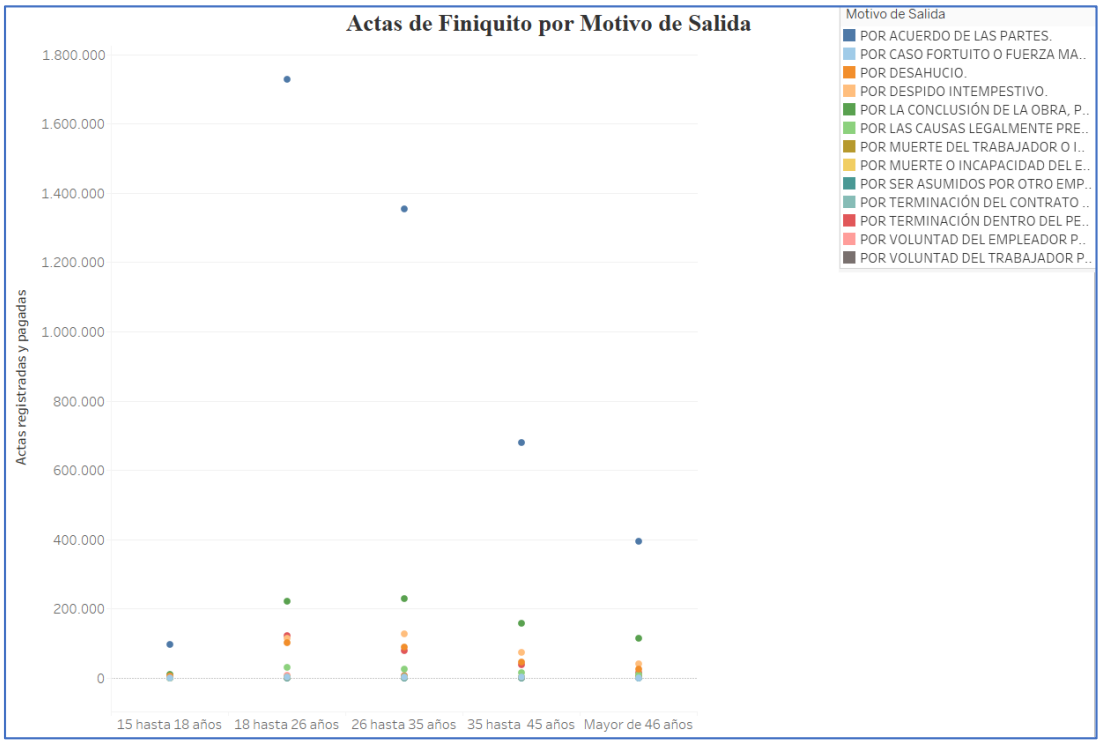


Figura 82. Actas Registradas por Motivo de Salida y Rango de Edad

En la Figura 83., se muestra que las provincias donde más se registran actas de finiquito son Pichincha con el 32.16% y Guayas con el 30.66%, la provincia donde menos actas de finiquito se registran es Bolívar con un 0.24%

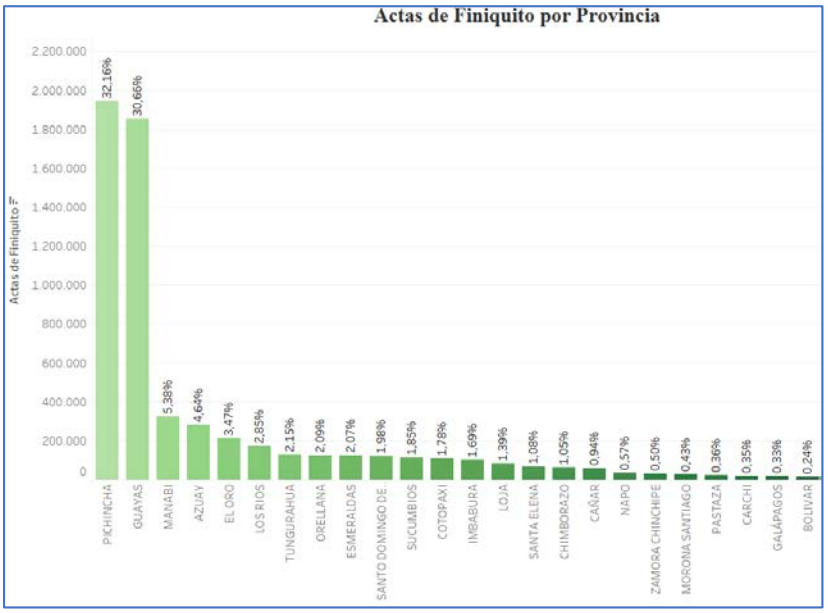


Figura 83. Actas de Finiquito Registrada por Provincia

Los trámites que tienen mayor cantidad de registro en el Ministerio son los Desahucios por terminación laboral y Vistos Buenos como se observa en la Figura 84.

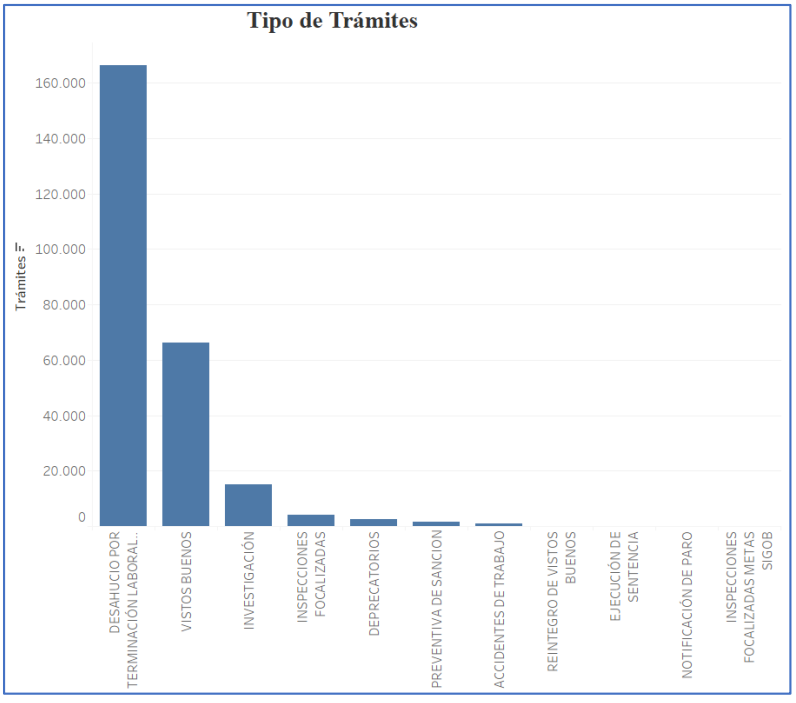


Figura 84. Tipo de Trámites

De la Figura 85, se tiene que la Región Sierra es donde los empleados registran más boletas en contra de las empresas, esta región tiene el 55.33%, luego se tiene la Región Costa con un 36.49%, y muy por debajo de las dos regiones anteriormente descritas se tiene a la Región Oriente con el 7.75% y por último se tiene a la Región Insular con un 0.42%.

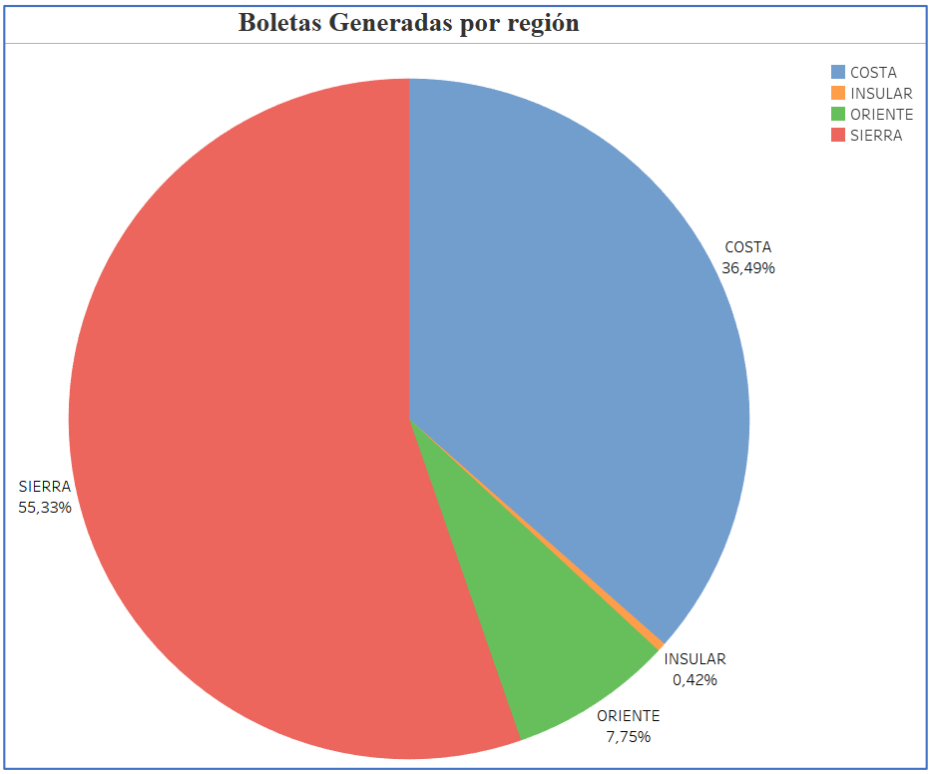


Figura 85. Boletas Generadas por Región

3.2.2.4. Verificar la Calidad de los Datos

Una vez que se ha realizado la exploración de los datos hemos podido confirmar que los datos son consistentes, y nos permitirán obtener los resultados para cumplir con los objetivos planteados en el proyecto. Los datos son de calidad por lo siguiente:

- Los datos no contienen errores, ya que se sometieron a un proceso ETL, es decir se realizó una limpieza de datos, estandarización de datos, completitud de datos, etc.
- Los datos al ser ingresados cuentan con validaciones que evitan el ingreso de datos erróneos eliminando el ruido en el conjunto de datos.
- Los valores nulos fueron procesados cuando se realizó la construcción del data warehouse.
- El modelo de base de datos cumple con las reglas de normalización de datos, es decir, evita la redundancia, todas las tablas se encuentran relacionadas, no existe valores duplicados, los valores en las diferentes columnas son congruentes en relación al tipo de dato, logrando la integridad de los datos.

3.2.3. Preparación de los datos

La fase de preparación de datos nos ayuda a dar el formato y adecuar los datos que serán usados en las técnicas de minería de datos escogidas.

3.2.3.1. Selección de los datos

En esta sección se seleccionan los campos necesarios de la base de datos que nos permiten cumplir los objetivos del negocio planteados en el proyecto. Los campos escogidos del Data Warehouse son los siguientes:

- **Dimensión Género (dim_genero)**
    - dgen_nombre
- **Dimensión Empleado (dim_empleado)**

    - sk_empleado
    - demp_edad
- **Dimensión Institución (dim_institucion)**
    - sk_institucion
- **Dimensión Actividad Económica (dim_actividad_economica)**
    - sk_actividad_economica
    - dact_nombre_l1
- **Dimensión Grupo Ocupacional (dim_grupo_ocupacional)**
    - dgru_nombre_l1
- **Dimensión Tipo Contrato (dim_tipo_contrato)**
    - dtic_nombre
- **Dimensión Tipo Empresa (dim_tipo_empresa)**
    - dtip_nombre_l1
- **Dimensión Contrato (dim_contrato)**
    - sk_contrato
    - dcon_finalizado
    - dcon_juvenil
    - dcon_legalizado
    - dcon_anulado
    - dcon_fecha_creacion
- **Dimensión Actas de Finiquito (dim_acta_finiquito)**
    - sk_acta_finiquito

  - dact_consignada
  - dact_anulada
  - dact_pagada

- **Dimensión Trámites (dim_tramite)**
  - sk_tramite.
- **Dimensión Boletas (dim_boletas)**
  - sk_boleta
- **Tabla de Hechos Cumplimiento (fact_cumplimiento)**
  - fcum_trabajo_menores
  - fcum_trabajo_juvenil
  - fcum_trabajo_discapacidad
  - fcum_contratos_registro_atrasados
  - fcum_actas_registro_atrasadas
  - fcum_tramites_boletas
  - fcum_resultado_incumplimiento
  - fcum_resultado_sgi

Los campos escogidos son los que tienen relación con los objetivos de la minería de datos planteados en el proyecto.

3.2.3.2. Limpieza de los Datos

En el Data Warehouse se tiene la información, los datos se encuentran limpios, ya que en la Sección 3.1.8.2. se realizó el proceso de limpieza de datos.

3.2.3.3. Construcción de los Datos

En esta fase se utiliza la herramienta Pentaho (PDI), la cual nos permite realizar la transformación de los campos necesarios para el desarrollo de la analítica mismos que son almacenados en la Base Datos, almacenados en el esquema “aux_analitica”.

Atributos Derivados

Los atributos derivados se muestran en la Tabla 31, donde se especifican los campos y las transformaciones realizadas.

Tabla 32.

Transformación de Campos

Campo	Transformación	Tipo de Dato
total_contratos	Conteo cantidad de contratos que tiene la empresa en todos sus estados.	bigint
total_contratos_nulos	Conteo cantidad de contratos que tiene la empresa en estado nulo.	bigint
total_contratos_periodo_prueba	Conteo cantidad de contratos que tiene la empresa con el tipo período de prueba.	bigint
total_contratos_finalizados	Conteo cantidad de contratos que tiene la empresa en estado finalizado.	bigint
total_contratos_vigentes	Conteo cantidad de contratos que tiene la empresa en estado vigente	bigint
total_contratos_legalizados	Conteo cantidad de contratos que tiene la empresa en estado legalizado	bigint
total_actas	Conteo cantidad de actas de finiquito que tiene la empresa en todos sus estados.	bigint
total_actas_anuladas	Conteo cantidad de actas de finiquito que tiene la empresa en estado nulo.	bigint
total_actas_registradas	Conteo cantidad de actas de finiquito que tiene la empresa en estado registrado.	bigint
total_actas_pagadas	Conteo cantidad de actas de finiquito que tiene la empresa en estado pagado.	bigint
total_actas_consignadas_ministerio	Conteo cantidad de actas de finiquito que tiene la empresa en estado consignado en el ministerio.	bigint
total_tramites	Conteo cantidad de tramites ingresados al Ministerio por empresa.	bigint
total_boletas	Conteo cantidad de boletas ingresadas al Ministerio por empresa.	bigint

CONTINÚA

rango_edad	Generación de rangos de edades: • Menores de 15años (<=15) • Entre 15 y 18 años (15>x<=18) • Entre 18 y 26 años (18>x<=26) • Entre 26 y 35 años (26>x<=35) • Entre 35 y 45 años (35>x<=45) • Mayores de 46 años (>=46)	text

Como punto adicional se realizó el pivoteo[38] de los siguientes campos:

Tabla 33.

Pivoteo de Campos

Campo	Columna de Salida	Tipo de Dato
dact_nombe_l1	• actividad_economica_verificar • actividad_economica_dependencia_publico • actividad_economica_organzaciones • actividad_economica_sin_actividad • actividad_economica_dependencia_privado • actividad_economica_ensenancia • actividad_economica_agricultura • actividad_economica_explotacion • actividad_economica_industrias • actividad_economica_electricas • actividad_economica_agua • actividad_economica_construccion • actividad_economica_comercio • actividad_economica_transporte • actividad_economica_alojamiento • actividad_economica_informacion • actividad_economica_financiera • actividad_economica_inmobiliares • actividad_economica_profesionales • actividad_economica_administrativos • actividad_economica_seguridad • actividad_economica_salud • actividad_economica_artes • actividad_economica_otros_servicios • actividad_economica_hogar • actividad_economica_no_definido	double precision
dgu_nombe_l1	• grupo_ocupacional_militar	double precision

[38] Convertir valores únicos de una columna en varias columnas de salida

CONTINÚA

	• grupo_ocupacional_tecnico • grupo_ocupacional_personal_apoyo • grupo_ocupacional_elementales • grupo_ocupacional_servicios • grupo_ocupacional_agropecuario • grupo_ocupacional_directores • grupo_ocupacional_oficiles • grupo_ocupacional_profesionales • grupo_ocupacional_operadores • grupo_ocupacional_no_definido	
dtic_nombe	• tipo_contrato_acta_jubilacion • tipo_contrato_colectivo • tipo_contrato_enganche • tipo_contrato_fijo • tipo_contrato_temporada • tipo_contrato_prueba • tipo_contrato_verbal • tipo_contrato_domicilio • tipo_contrato_obra_cierta • tipo_contrato_destajo • tipo_contrato_agricola • tipo_contrato_maquilado • tipo_contrato_eventual • tipo_contrato_ocasional • tipo_contrato_aprendizaje • tipo_contrato_jornada_parcial • tipo_contrato_servicio_domestico • tipo_contrato_artesano • tipo_contrato_indefinido • tipo_contrato_tarea • tipo_contrato_adolescente • tipo_contrato_franca • tipo_contrato_determinado • tipo_contrato_floricola • tipo_contrato_bananero • tipo_contrato_transporte • tipo_contrato_juvenil • tipo_contrato_plazo_fijo • tipo_contrato_no_definido	double precision
dtip_nombe_l1	• tipo_empresa_especial_12 • tipo_empresa_especial_17 • tipo_empresa_financiera • tipo_empresa_publica • tipo_empresa_atesanal	double precision

CONTINÚA 

	• tipo_empresa_privada • tipo_empresa_publica_1 • tipo_empresa_publica_2 • tipo_empresa_no_definido double precision	

3.2.3.4. Integración de los Datos

No es necesario la integración de datos ya que los sistemas transaccionales del Ministerio del Trabajo utilizados comparten información entre sí, cada sistema hace una copia de los datos requeridos hacia su propia base de datos por lo que se tendrían los mismos datos.

3.2.3.5. Formateo de los Datos

El proceso de formateo de datos se describió en la Sección 3.1.8.2. del presente documento.

3.2.4. Modelado

En la fase de modelado se elige las técnicas de minería de datos adecuadas para cumplir con los objetivos de la minería de datos, luego se genera el plan de prueba, a continuación, se aplica las técnicas de minería de datos escogidas sobre los datos para construir el modelo y finalmente se evalúa el modelo para determinar si cumple con los criterios de éxito.

3.2.4.1. Selección Técnica de Modelado

Se construyeron tres modelos, basados en las técnicas de minería de datos descritas en la sección 3.2.1.4., las técnicas que se escogieron son:

- Árboles de Decisión

- Regresión Logística
- Redes Neuronales
- Reglas de Asociación (Algoritmo Apriori)

3.2.4.2. Generación del diseño de Pruebas

En la sección de generación del diseño de pruebas, se verifica que el modelo generado sea válido, las técnicas de minería de datos seleccionadas ejecutan una entrada, proceso y salida que se detallan a continuación

Entrada:

Dataset[39] con información necesaria para la generación de los modelos, la información es desde el año 2012 hasta enero 2018, esta información se obtuvo del Data Warehouse

Proceso:

La técnica que se utilizó para la validación del modelo es cross-validation[40], que nos permite tener un determinado número de validaciones del conjunto de datos para el entrenamiento y un determinado número de validaciones del conjunto de datos para pruebas, esto nos ayuda a validar la exactitud del modelo.

Como se está utilizando la herramienta Knime hacemos uso del componente:

- **X-Patitioner:** Permite determinar el número de iteraciones de validación cruzada que se deben realizar. En la validación cruzada en k , la muestra

[39] Conjunto de Datos
[40] Validación cruzada

original se divide en k submuestras de igual tamaño. De las k submuestras, una sola submuestra se retiene como los datos de validación para probar el modelo, y las submuestras k - 1 restantes se usan como datos de entrenamiento. Esto significa que, si establece el número de validaciones como 10, tendrá un 10% utilizado para la validación y un 90% para el entrenamiento en cada iteración. (Knime, Open for innovation - Knime, 2017)

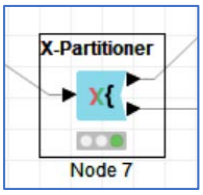


Figura 86. Componente X-Partitioner

Fuente: (Knime, Open for innovation-Knime, 2017)

Para nuestro caso se configura el componente X-Partitioner para que realice 30 iteracciones con un muestreo aleatorio del dataset seleccionado.

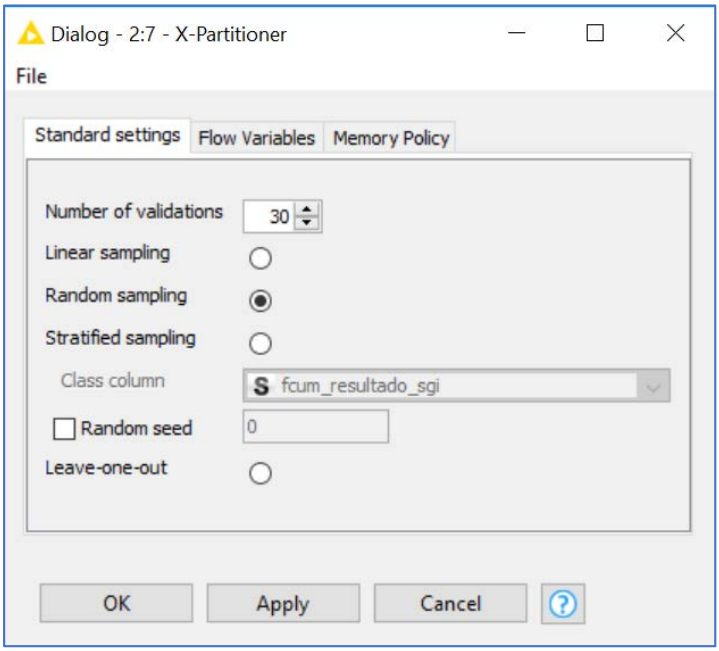


Figura 87. Configuración X-Partitioner

Salida

En la herramienta Knime para la salida del proceso se tiene:

- **X-Agregator:** Agrega el resultado de la validación cruzada. (Knime, Open for innovation-Knime, 2017)

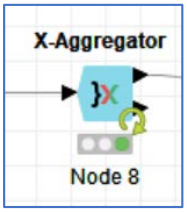


Figura 88. Componente X-Agregator

Fuente: (Knime, Open for innovation-Knime, 2017)

Agrega el resultado de cada una de las iteracciones realizadas por el X-Partitioner.

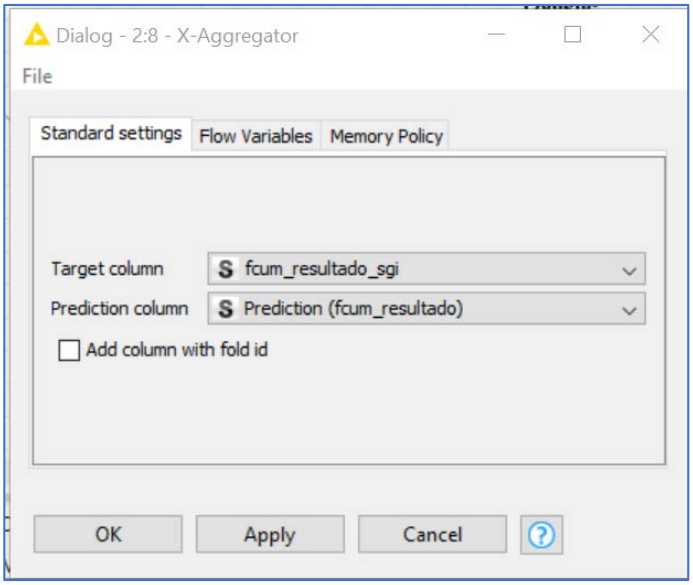


Figura 89. Configuración X-Partitioner

3.2.4.3. Construcción del Modelo

Se construyó diferentes modelos para cumplir con los objetivos planteados, los cuales consisten en la generación del modelo predictivo orientado a predecir si una empresa

debe ser inspeccionada o no inspeccionada. De acuerdo a la herramienta escogida para la aplicación de los modelos se realizan las diferentes configuraciones de los parámetros establecidas para cada modelo, en nuestro caso en la herramienta Knime.

La construcción de los modelos se realizó en base las variables más relevantes, se inicia con cada una de las técnicas seleccionadas: árboles de decisión, regresión logística, redes neuronales y reglas de asociación.

- **Árboles de Decisión**

Con la aplicación de la técnica de árboles de decisión, nos ayuda a identificar las variables que intervienen en la decisión de si una empresa debe ser inspeccionada o no inspeccionada, se define como el atributo clase al campo fcum_resultado_sgi (variable a predecir). La Figura 90 muestra el modelo de árbol de decisión en Knime.

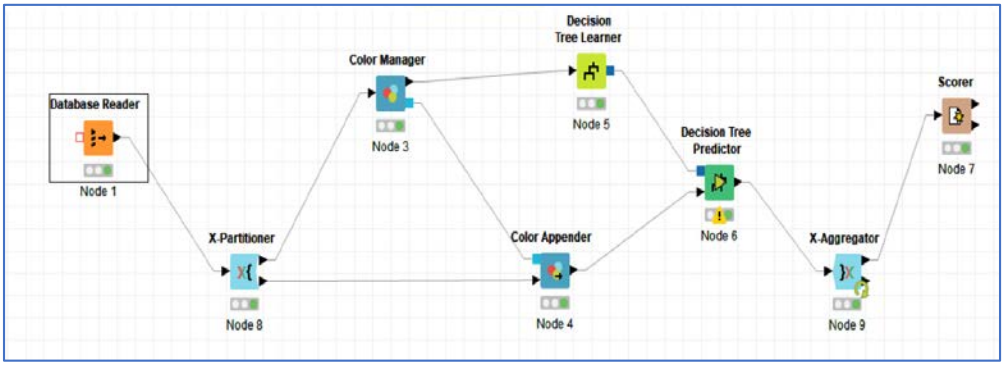


Figura 90. Árbol de Decisión

La técnica de árboles de decisión nos permite visualizar la clasificación de los datos con la ayuda de un esquema gráfico, el cual permite conocer de mejor manera el modelo predictivo.

Para la generación del modelo predictivo del árbol de decisión se utilizó los siguientes nodos:

- Un nodo de carga de datos que nos permite obtener todas las variables que están involucradas en el proceso de la minería, los datos son obtenidos del esquema “aux_analitica”.
- Un nodo de partición de validación cruzada el cual posee una configuración de 30 particiones, esto debido a que se tiene 663297 registros con 91 variables, esto nos permite dividir los datos de la siguiente manera en cada una de las iteraciones:
    - 641188 para el entrenamiento del modelo.
    - 22109 para la validación del modelo propuesto.
- Un nodo de configuración de color para las variables de inspeccionada y no inspeccionada.
- Un nodo de aprendizaje de árbol de decisión, con los siguientes parámetros:
    - Selección de la clase: Para nuestro caso de estudio la variable seleccionada es fcum_resultado_sgi.
    - Medida de calidad: Se selecciona GINI[41] (coeficiente de impureza) la probabilidad de obtener dos registros de la misma clase.
    - Método de poda: No se utiliza ningún método de poda
    - Registros mimos por nodo: 10000
    - Nivel de profundidad del árbol: 10000

[41] La función es la decisión dividida en el árbol de decisión. (Knime, Open for Innovation Knime, 2015)

En la Figura 91 se presenta las configuraciones que se realizaron en el nodo de aprendizaje del árbol de decisión.

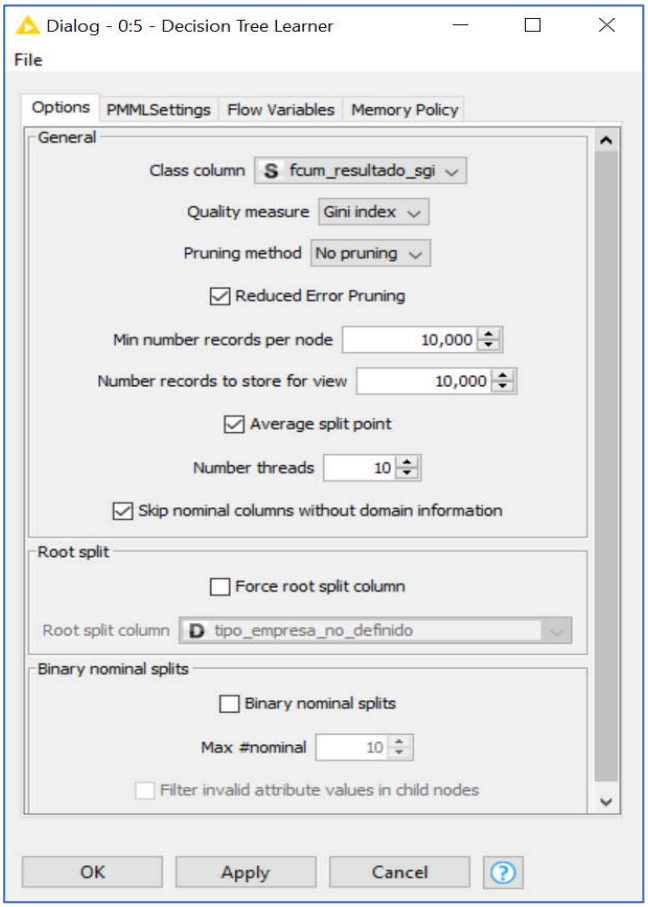


Figura 91. Configuración Nodo de Aprendizaje

- Un nodo de predicción de árbol de decisión.
- Un nodo que agrupa todas las iteraciones realizadas por el nodo de partición de validación cruzada.
- Un nodo de recopilación obtención de datos para la generación de estadísticas.
- Un nodo para la generación de la curva ROC.

Con esta configuración el modelo obtuvo una exactitud del 76.119% y un error del 23.904%.

En la Figura 92, de acuerdo a atributo de mayor peso que son las actas de finiquito registradas con retraso (fcum_actas_registro_atrasadas).

Figura 92. Árbol de Decisión Parte Superior

Como se observa en la Figura 93 y Figura 94, la ramificación izquierda nos indica que la distribución de los datos está determinada por las variables actividad_economica_otros_servicios, seguida por la variable fcum_contratos_registro_atrasados, esta variable es determinante para establecer si una empresa es inspeccionada o no inspeccionada ya que según el Acuerdo 309 emitido por el Ministerio del Trabajo en el año 2017, indica que los contratos de trabajo deben ser registrados máximo en 30 días que el empleado empieza la relación laboral, a continuación se tiene las variables total_contratos_periodo_prueba, total_actas, total_boletas y por último la variable total_contratos3_legalizados, todas

estas variables en conjunto permiten determinar si la empresa es inspeccionada o no inspeccionada.

Figura 93. Parte Izquierda Árbol de Decisión – Primera Parte

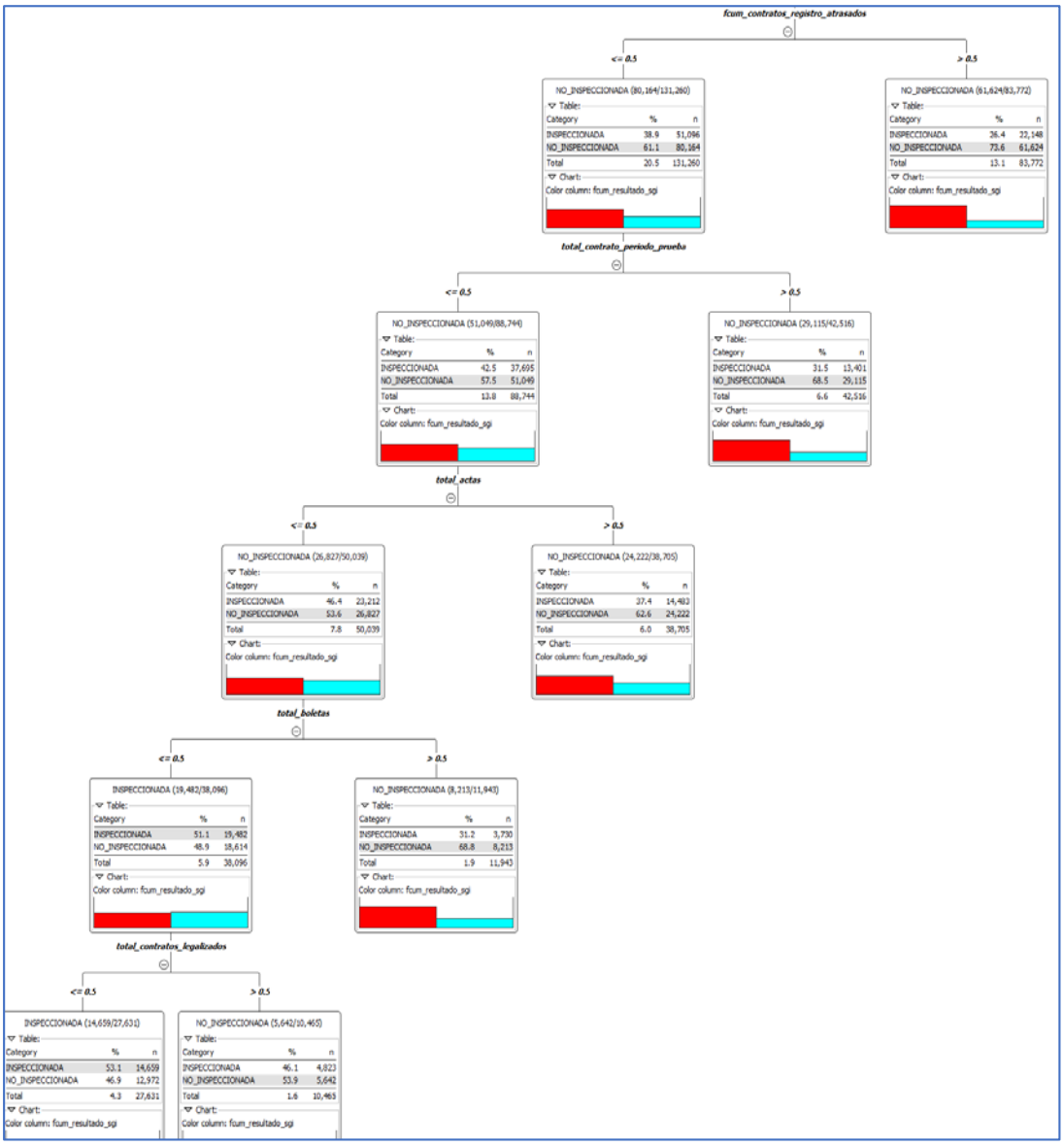


Figura 94. Parte Izquierda Árbol de Decisión – Segunda Parte

En la Figura 95 se tiene la ramificación derecha del árbol decisión, en el cual se nota que la variable fcum_contratos_registro_atrasados, nuevamente es determinante para la toma decisión.

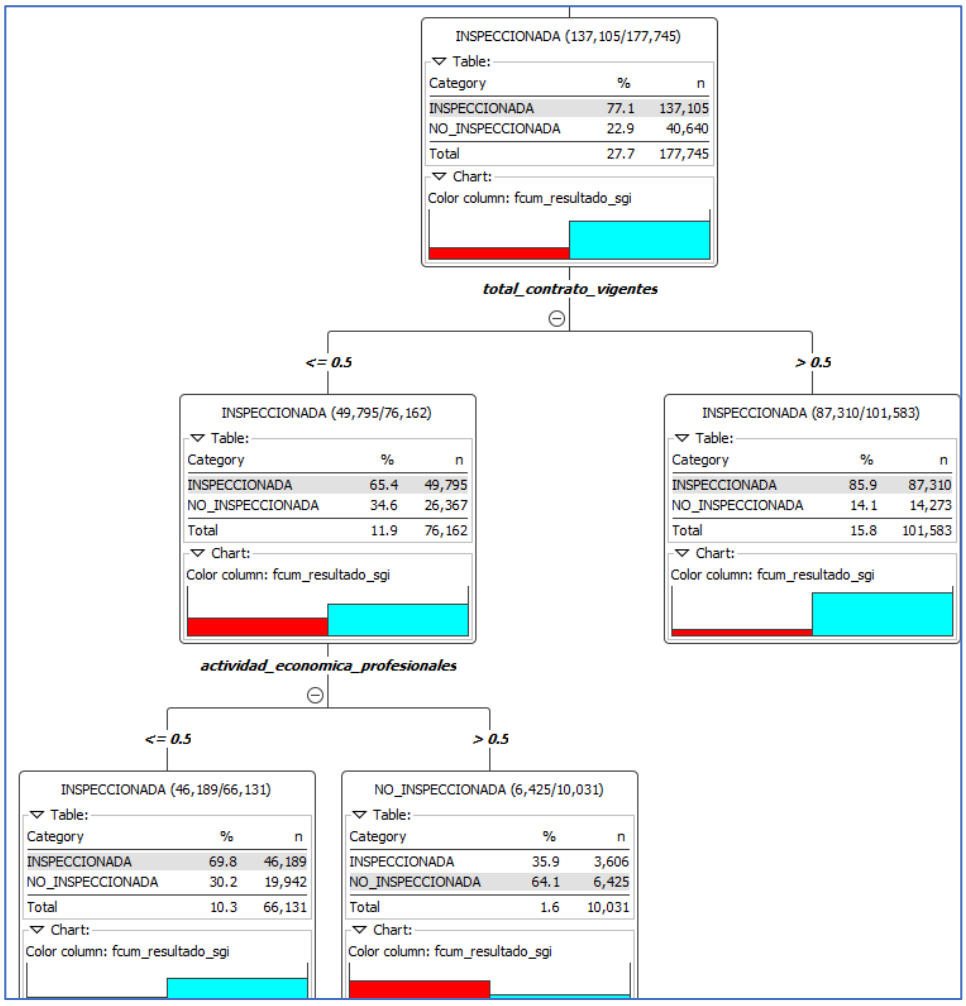


Figura 95. Parte Derecha Árbol de Decisión

- **Regresión Logística**

En la Figura 96 se muestra el modelo de regresión logística en Knime.

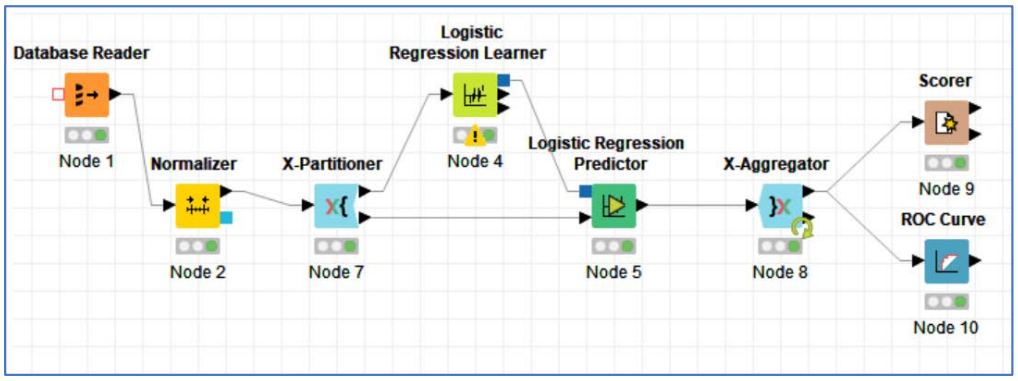


Figura 96. Regresión Logística

Para la generación del algoritmo se utilizó los siguientes nodos:

- Un nodo de carga de datos que nos permite obtener todas las variables que están involucradas en el proceso de la minería, los datos son obtenidos del esquema “aux_analitica”.
- Un nodo de normalización de datos entre 0 y 1.
- Un nodo de partición de validación cruzada el cual posee una configuración de 30 particiones, esto debido a que se tiene 663297 registros con 91 variables, esto nos permite dividir los datos de la siguiente manera en cada una de las iteraciones:
    - 641188 para el entrenamiento del modelo.
    - 22109 para la validación del modelo propuesto.
- Un nodo de aprendizaje de regresión logística, con los siguientes parámetros:
    - Selección de la clase: Para nuestro caso de estudio la variable seleccionada es fcum_resultado_sgi.
    - Categoría de referencia: Seleccionamos “INSPECCIONADA”.
    - Seleccionamos las 91 características disponibles.

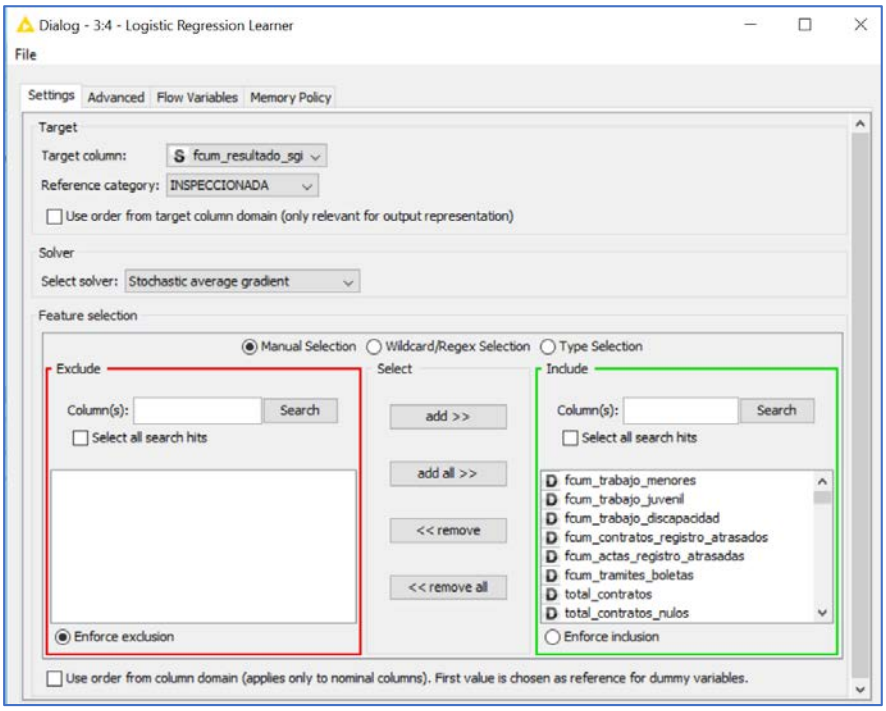


Figura 97. Configuración de Ajustes Iniciales

- Número máximo de iteraciones: 100.
- Un épsilon de 1.0E-5
- Una tasa de aprendizaje de fijo con un tamaño de cada uno de los pasos de 0.1
- Una constante de regulación de GAUS con una varianza de 0.1

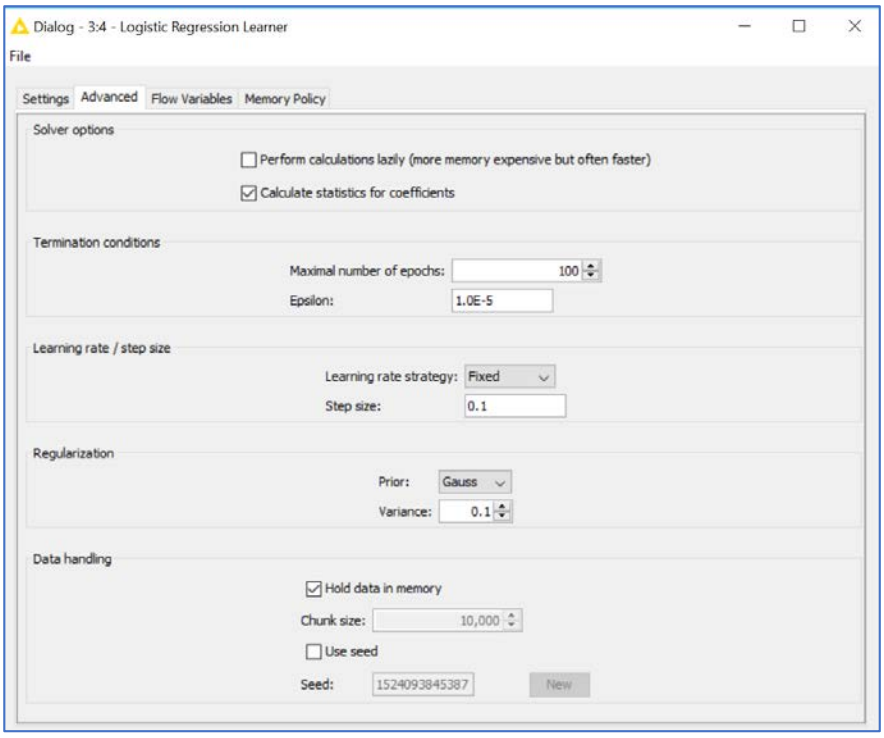


Figura 98. Configuración de Ajustes Avanzada

- Un nodo de predicción de regresión logística.
- Un nodo que agrupa todas las iteraciones realizadas por el nodo de partición de validación cruzada.
- Un nodo de recopilación obtención de datos para la generación de estadísticas
- Un nodo para la generación de la curva ROC.

Con esta configuración el modelo obtuvo una exactitud del 75.531% y un error del 24.469%. Se obtienen la siguiente tabla con los coeficientes de cada una de las variables

Coefficients and Statistics - 0:4 - Logistic Regression Learner

File Hilite Navigation View

Table "Coefficients and Statistics" - Rows: 92 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	Logit	Variable	Coeff.	Std. Err.	z-score	P>|z|
Row1	NO_INSPECCIONADA	fcum_trabajo_menores	-0.517	0.052	-9.922	0
Row2	NO_INSPECCIONADA	fcum_trabajo_juvenil	-1.205	0.076	-15.802	0
Row3	NO_INSPECCIONADA	fcum_trabajo_discap...	-1.581	0.041	-38.754	0
Row4	NO_INSPECCIONADA	fcum_contratos_regi...	0.223	0.009	23.556	0
Row5	NO_INSPECCIONADA	fcum_actas_registro...	-1.912	0.007	-256.776	0
Row6	NO_INSPECCIONADA	fcum_tramites_boletas	0.139	0.018	7.834	0
Row7	NO_INSPECCIONADA	total_contratos	-0.095	0.316	-0.302	0.763
Row8	NO_INSPECCIONADA	total_contratos_nulos	-0.047	0.316	-0.149	0.882
Row9	NO_INSPECCIONADA	total_contrato_perio...	-0.082	0.316	-0.261	0.794
Row10	NO_INSPECCIONADA	total_contratos_finali...	-0.074	0.316	-0.233	0.816
Row11	NO_INSPECCIONADA	total_contrato_vigen...	-0.098	0.316	-0.31	0.757
Row12	NO_INSPECCIONADA	total_contratos_legal...	-0.077	0.316	-0.243	0.808
Row13	NO_INSPECCIONADA	total_actas	-0.186	0.315	-0.592	0.554
Row14	NO_INSPECCIONADA	total_actas_anuladas	-0.06	0.316	-0.189	0.85
Row15	NO_INSPECCIONADA	total_actas_registradas	-0.164	0.315	-0.521	0.602
Row16	NO_INSPECCIONADA	total_actas_pagadas	-0.099	0.315	-0.315	0.752
Row17	NO_INSPECCIONADA	total_actas_consigna...	-0.018	0.316	-0.056	0.956
Row18	NO_INSPECCIONADA	total_tramites	-0.119	0.315	-0.376	0.707
Row19	NO_INSPECCIONADA	total_boletas	-0.038	0.313	-0.12	0.904
Row20	NO_INSPECCIONADA	actividad_economica...	-0.14	0.066	-2.111	0.035
Row21	NO_INSPECCIONADA	actividad_economica...	-0.856	0.086	-9.978	0
Row22	NO_INSPECCIONADA	actividad_economica...	-0.784	0.155	-5.044	0
Row23	NO_INSPECCIONADA	actividad_economica...	-0.647	0.067	-9.702	0
Row24	NO_INSPECCIONADA	actividad_economica...	-0.801	0.067	-11.902	0
Row25	NO_INSPECCIONADA	actividad_economica...	-0.119	0.069	-1.724	0.085
Row26	NO_INSPECCIONADA	actividad_economica...	0.428	0.066	6.481	0
Row27	NO_INSPECCIONADA	actividad_economica...	0.311	0.08	3.897	0
Row28	NO_INSPECCIONADA	actividad_economica...	0.042	0.066	0.635	0.526
Row29	NO_INSPECCIONADA	actividad_economica...	0.405	0.097	4.161	0
Row30	NO_INSPECCIONADA	actividad_economica...	0.005	0.095	0.049	0.961
Row31	NO_INSPECCIONADA	actividad_economica...	0.395	0.066	5.956	0
Row32	NO_INSPECCIONADA	actividad_economica...	-0.025	0.065	-0.381	0.703
Row33	NO_INSPECCIONADA	actividad_economica...	0.027	0.065	0.41	0.682
Row34	NO_INSPECCIONADA	actividad_economica...	0.286	0.066	4.313	0
Row35	NO_INSPECCIONADA	actividad_economica...	0.742	0.064	11.544	0
Row36	NO_INSPECCIONADA	actividad_economica...	0.654	0.074	8.853	0
Row37	NO_INSPECCIONADA	actividad_economica...	0.564	0.07	8.044	0
Row38	NO_INSPECCIONADA	actividad_economica...	0.267	0.065	4.134	0
Row39	NO_INSPECCIONADA	actividad_economica...	0.179	0.067	2.652	0.008
Row40	NO_INSPECCIONADA	actividad_economica...	-0.1	0.086	-1.162	0.245
Row41	NO_INSPECCIONADA	actividad_economica...	0.137	0.067	2.042	0.041
Row42	NO_INSPECCIONADA	actividad_economica...	0.32	0.07	4.591	0
Row43	NO_INSPECCIONADA	actividad_economica...	-1.021	0.065	-15.65	0
Row44	NO_INSPECCIONADA	actividad_economica...	-0.269	0.065	-4.148	0
Row45	NO_INSPECCIONADA	grupo_ocupacional_...	-0.13	0.283	-0.46	0.645
Row46	NO_INSPECCIONADA	grupo_ocupacional_t...	-0.969	0.296	-3.276	0.001
Row47	NO_INSPECCIONADA	grupo_ocupacional_p...	-2.508	0.221	-11.326	0
Row48	NO_INSPECCIONADA	grupo_ocupacional_e...	-2.057	0.143	-14.395	0
Row49	NO_INSPECCIONADA	grupo_ocupacional_s...	-0.64	0.218	-2.933	0.003
Row50	NO_INSPECCIONADA	grupo_ocupacional_a...	0.978	0.206	4.743	0
Row51	NO_INSPECCIONADA	grupo_ocupacional_d...	-0.882	0.283	-3.118	0.002
Row52	NO_INSPECCIONADA	grupo_ocupacional_o...	0.941	0.243	3.867	0
Row53	NO_INSPECCIONADA	grupo_ocupacional_p...	-1.664	0.289	-5.769	0
Row54	NO_INSPECCIONADA	grupo_ocupacional_o...	-0.07	0.247	-0.285	0.776
Row55	NO_INSPECCIONADA	tipo_contrato_acta_j...	0.136	0.309	0.438	0.661
Row56	NO_INSPECCIONADA	tipo_contrato_colectivo	0.105	0.297	0.353	0.724
Row57	NO_INSPECCIONADA	tipo_contrato_engan...	0.004	0.277	0.015	0.988
Row58	NO_INSPECCIONADA	tipo_contrato_fijo	-0.16	0.014	-11.522	0
Row59	NO_INSPECCIONADA	tipo_contrato_tempo...	-0.178	0.035	-5.074	0
Row60	NO_INSPECCIONADA	tipo_contrato_prueba	-0.311	0.02	-15.781	0
Row61	NO_INSPECCIONADA	tipo_contrato_verbal	-0.1	0.015	-6.834	0
Row62	NO_INSPECCIONADA	tipo_contrato_domicilio	-0.145	0.061	-2.365	0.018
Row63	NO_INSPECCIONADA	tipo_contrato_obra_...	0.302	0.033	9.159	0
Row64	NO_INSPECCIONADA	tipo_contrato_destajo	0.093	0.08	1.168	0.243
Row65	NO_INSPECCIONADA	tipo_contrato_agricola	-0.025	0.058	-0.43	0.667
Row66	NO_INSPECCIONADA	tipo_contrato_maquil...	-0.123	0.284	-0.434	0.664
Row67	NO_INSPECCIONADA	tipo_contrato_eventual	-0.238	0.023	-10.212	0
Row68	NO_INSPECCIONADA	tipo_contrato_ocasional	-0.25	0.05	-5.015	0
Row69	NO_INSPECCIONADA	tipo_contrato_apren...	-0.186	0.088	-2.121	0.034
Row70	NO_INSPECCIONADA	tipo_contrato_jornad...	-0.233	0.011	-20.926	0
Row71	NO_INSPECCIONADA	tipo_contrato_servici...	-0.122	0.015	-7.957	0
Row72	NO_INSPECCIONADA	tipo_contrato_artesano	0.045	0.03	1.488	0.137
Row73	NO_INSPECCIONADA	tipo_contrato_indefin...	-0.165	0.01	-17.01	0
Row74	NO_INSPECCIONADA	tipo_contrato_tarea	-0.259	0.073	-3.532	0
Row75	NO_INSPECCIONADA	tipo_contrato_adoles...	0.048	0.089	0.538	0.591
Row76	NO_INSPECCIONADA	tipo_contrato_franca	-0.142	0.282	-0.502	0.615
Row77	NO_INSPECCIONADA	tipo_contrato_deter...	0.115	0.068	1.689	0.091
Row78	NO_INSPECCIONADA	tipo_contrato_floricola	-0.191	0.245	-0.78	0.436
Row79	NO_INSPECCIONADA	tipo_contrato_banan...	-0.036	0.259	-0.14	0.888
Row80	NO_INSPECCIONADA	tipo_contrato_transp...	-0.088	0.057	-1.549	0.121
Row81	NO_INSPECCIONADA	tipo_contrato_juvenil	-0.182	0.249	-0.732	0.464
Row82	NO_INSPECCIONADA	tipo_contrato_plazo_...	-0.29	0.01	-28.465	0
Row83	NO_INSPECCIONADA	tipo_contrato_no_de...	-0.03	0.313	-0.095	0.924
Row84	NO_INSPECCIONADA	tipo_empresa_especi...	0.089	0.141	0.63	0.528
Row85	NO_INSPECCIONADA	tipo_empresa_especi...	0.255	0.2	1.277	0.202
Row86	NO_INSPECCIONADA	tipo_empresa_financi...	0.023	0.129	0.179	0.858
Row87	NO_INSPECCIONADA	tipo_empresa_publica	0.09	0.06	1.508	0.132
Row88	NO_INSPECCIONADA	tipo_empresa_atesanal	0.118	0.062	1.895	0.058
Row89	NO_INSPECCIONADA	tipo_empresa_privada	0.164	0.059	2.767	0.006
Row90	NO_INSPECCIONADA	tipo_empresa_public...	0.235	0.073	3.216	0.001
Row91	NO_INSPECCIONADA	tipo_empresa_public...	0.362	0.078	4.632	0
Row92	NO_INSPECCIONADA	Constant	0.727	0.087	8.35	0

Figura 99: Coeficientes de Variables

- **Redes Neuronales**

En la Figura 100 se muestra el modelo de redes neuronales en Knime.

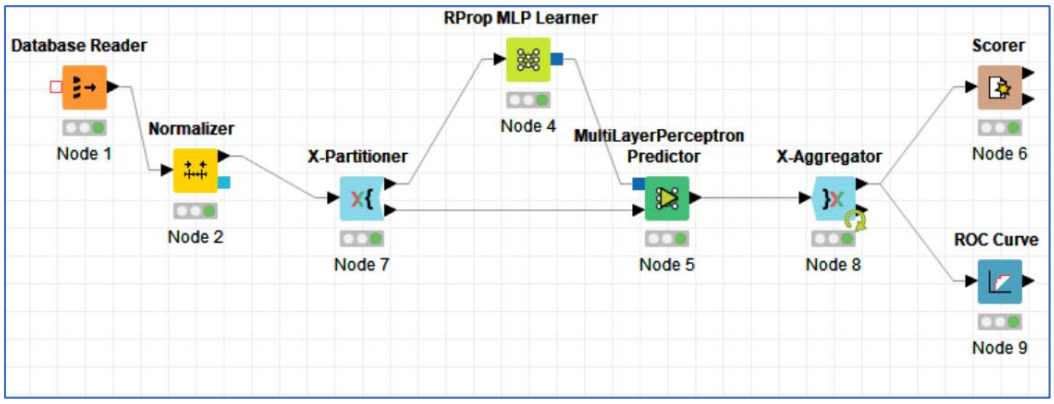


Figura 100. Redes Neuronales

Para la generación del algoritmo se utilizó los siguientes nodos:

- Un nodo de carga de datos que nos permite obtener todas las variables que están involucradas en el proceso de la minería, los datos son obtenidos del esquema “aux_analitica”.
- Un nodo de normalización de datos entre 0 y 1.
- Un nodo de partición de validación cruzada el cual posee una configuración de 30 particiones, esto debido a que se tiene 663297 registros con 91 variables, esto nos permite dividir los datos de la siguiente manera en cada una de las iteraciones:
    - 641188 para el entrenamiento del modelo.
    - 22109 para la validación del modelo propuesto.
- Un nodo de aprendizaje de red neuronal, con los siguientes parámetros:
    - Número máximo de iteraciones: 50
    - Número de capas ocultas: 5

- Número de neuronas ocultas por capa: 10
- Selección de la clase: Para nuestro caso de estudio la variable seleccionada es fcum_resultado_sgi.

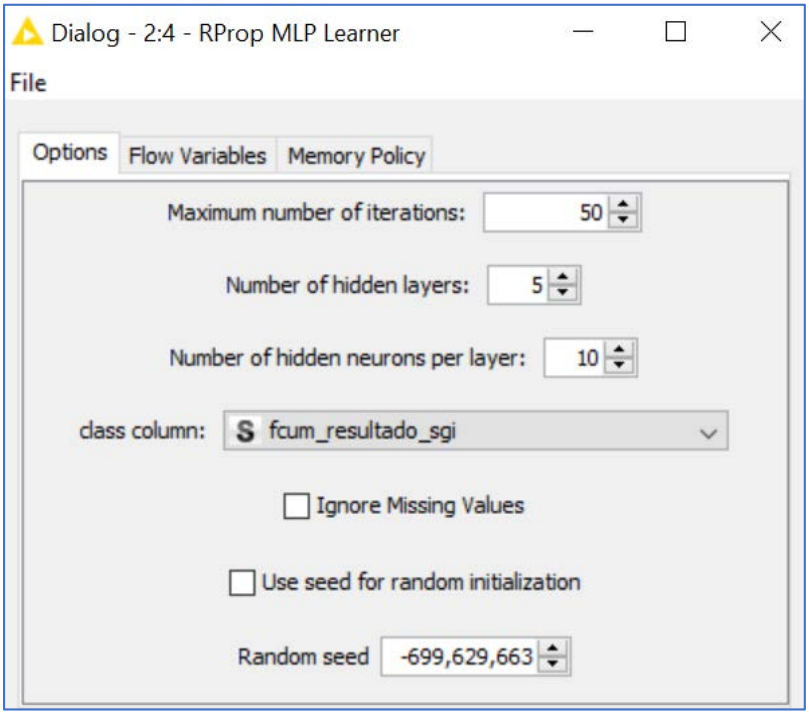


Figura 101. Configuración Red Neuronal

- Un nodo de predicción de red neuronal.
- Un nodo que agrupa todas las iteraciones realizadas por el nodo de partición de validación cruzada.
- Un nodo de recopilación obtención de datos para la generación de estadísticas
- Un nodo para la generación de la curva ROC.

Con esta configuración el modelo obtuvo una exactitud del 76.784% y un error del 23.216%.

- **Reglas de Asociación**

En las Figuras 102, 103, 104 y 105 se muestra el modelo de reglas de asociación en Knime.

Para la implementación de la técnica de reglas de asociación se utilizó el algoritmo apriori, cada una de las iteraciones realizadas fueron implementadas con un plugin de R de acuerdo al Anexo B, la herramienta Knime permite instalar plugins que facilita la interacción con otros aplicativos como Lenguaje R. Cada uno de los algoritmos han sido configurados para obtener reglas de asociación con un soporte mínimo de 0.15 y confianza de 0.5, pero en el modelo donde interviene la actividad económica con esta configuración no se obtuvo resultados, por lo que se cambio el soporte a 0.05 con la finalidad de encontrar reglas.

En la construcción de los modelos para cada regla de asociación se utilizaron los siguientes nodos de Lenguaje R:

- Nodo de obtención de datos.
- Nodo de aplicación del algoritmo.
- Nodo de conversión de lista de reglas para ser almacenadas en una tabla.
- Nodo de visualización de la tabla generada.

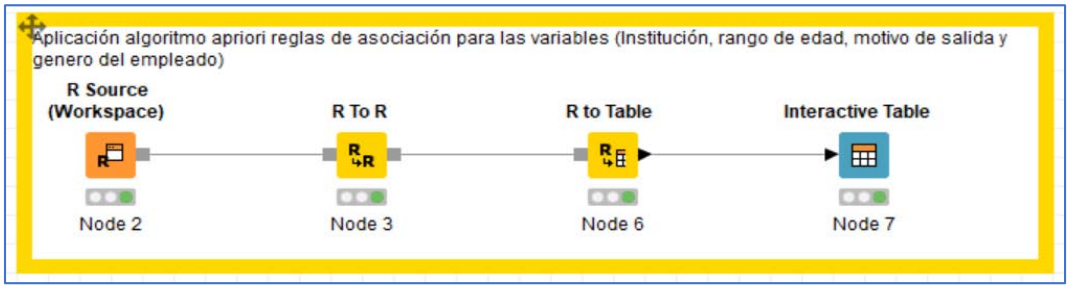


Figura 102. Regla de Asociación 1

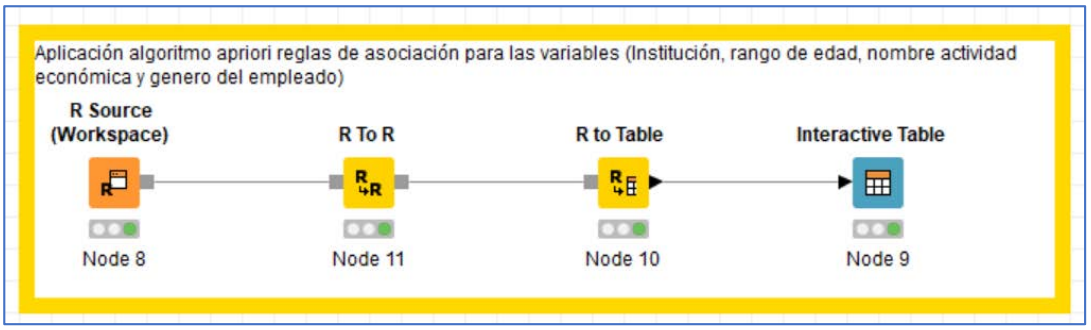


Figura 103. Regla de Asociación 2

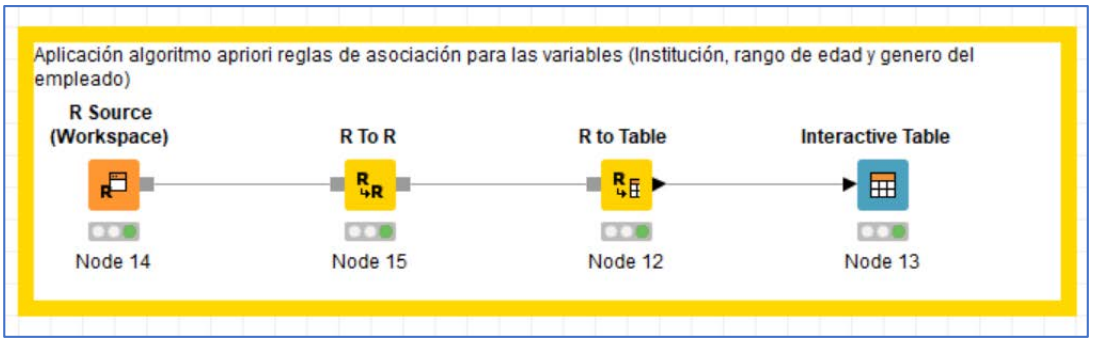


Figura 104. Regla de Asociación 3

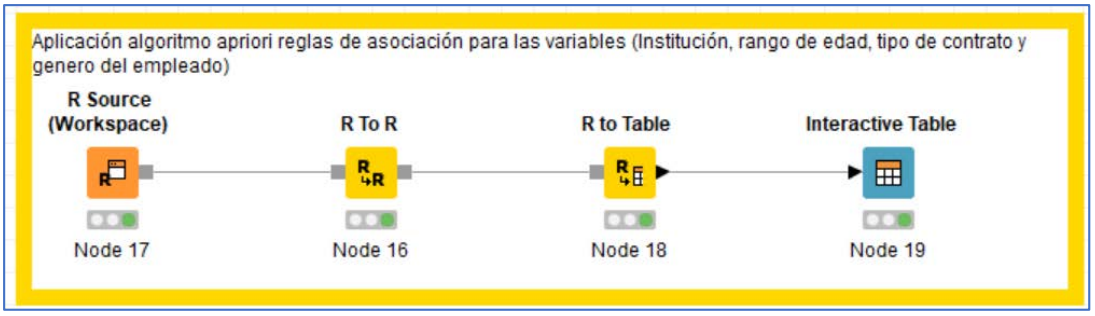


Figura 105. Regla de Asociación 4

3.2.5. Evaluación del Modelo

En esta fase se evalúan los modelos generados en la fase de modelado, la evaluación se realiza para decidir si los objetivos tanto del negocio como de minería de datos han sido cumplidos para seguir con la fase de despliegue.

3.2.5.1. Evaluación de Resultados

En la Tabla 34 se muestra el cumplimiento de los objetivos del negocio una vez aplicadas las técnicas de minería de datos seleccionadas.

Tabla 34.

Cumplimiento de los Objetivos de Negocio

Objetivos del Negocio	Arboles de Decisión	Regresión Logística	Redes Neuronales	Reglas de Asociación
Determinar un modelo predictivo que ayude a establecer si una empresa debe ser inspeccionada o no inspeccionada.	Si	Si	Si	No
Determinar los patrones que siguen las empresas inspeccionadas en lo referente a contratación y salida de personal.	Si	No	No	Si

En la Tabla 35 se muestra el cumplimiento de los objetivos de la minería de datos una vez aplicadas la técnica de minería de datos seleccionadas.

Tabla 35.

Cumplimiento de los Objetivos de Minería de Datos

Objetivos de la Minería de Datos	Arboles de Decisión	Regresión Logística	Redes Neuronales	Reglas de Asociación
Clasificar las empresas que deben ser inspeccionadas y no inspeccionadas, esto es determinado por el incumplimiento en el registro de contratos y actas de finiquito, y de las denuncias que los empleados registran en contra de las empresas	SI	SI	SI	NO
Buscar la relación que existe entre las variables relacionadas con la contratación y salida de personal en las empresas inspeccionadas.	SI	NO	NO	SI

De acuerdo al resultado de la tabla de cumplimiento de objetivos del negocio y la tabla de cumplimiento de objetivos de la minería de datos se tiene que la técnica de árboles de decisión cumple en forma total con los objetivos del negocio y de la minería de datos, las otras técnicas tienen un cumplimiento parcial.

3.2.6. Despliegue

La fase de despliegue es la última fase de la metodología CRISP-DM, en esta fase se explica el proceso que permite a los encargados del Ministerio del Trabajo la toma de decisiones a través del modelo de minería de datos encontrado, además se presenta los resultados obtenidos para que las autoridades puedan entender de una forma fácil. También esta fase tiene como objetivo crear un método para dar mantenimiento a los modelos en un futuro.

3.2.6.1. Plan de Despliegue

Para poder realizar el despliegue de los modelos es necesario lo siguiente:

- Acceso a la base de datos gestion_trabajo del datawarehouse.
- Acceso al esquema de base de datos “aux_analítica” propiedad del Ministerio de Trabajo
- Instalación de la herramienta de minería de datos Knime versión 3.5.1 o superior.
- Memoria RAM disponible de mínimo de 8GB.

Una vez que se ha realizado los pasos anteriores se puede realizar:

- Actualizar la base de datos gestión_trabajo con los datos actualizados a la fecha que institución estime según sus necesidades.
- Ejecutar los ETL's para que se actualice el Data Warehouse.
- Actualizar la información extraída del Data Warehouse en las tablas del esquema “aux_analitica”.
- Ejecutar los modelos generados.
- Verificar los algoritmos que presenten menor tasa de error, llevar una bitácora de los datos.

3.2.6.2. Plan de Monitoreo y Mantenimiento

El monitoreo y mantenimiento del modelo debe considerar la frecuencia con que se actualiza la información en los sistemas transaccionales SAITE, SINACOI y SGI, esta información se actualiza a cada momento por lo que el Ministerio debe decidir la periodicidad con la que se actualizara la información en el Data Warehouse de donde se extrae la información para el esquema “aux_analitica”. El proceso de minería de datos es recomendable realizarlo mensualmente ya que las empresas contratan personal a inicios de mes, despiden personal a finales de mes lo cual ocasiona que los empleados que no estén de acuerdo con el motivo de salida presenten denuncias en el Ministerio del Trabajo, sin embargo, esto puede variar de acuerdo a las necesidades de obtener información para la toma de decisiones por parte de la institución.

3.2.6.3. Informe Final

El informe final presenta el informe en forma resumida los resultados obtenidos y lo más relevante del proyecto de minería de datos, en el Capítulo 4., se realiza la descripción y resultados obtenidos una vez que se ejecutan los modelos.

CAPÍTULO IV. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

En este capítulo se presenta los resultados obtenidos en la ejecución de los modelos, estos resultados son analizados e interpretados en base al criterio de los funcionarios que son los expertos del negocio.

Para tener un modelo que cumpla con los objetivos del negocio y de la minería de datos, se realizaron varias pruebas, las mismas que tuvieron diferentes parametrizaciones. A continuación, se analiza los resultados obtenidos en cada uno de los modelos.

4.1. Empresas Inspeccionadas o No Inspeccionadas

Para determinar si una empresa debe ser inspeccionada o no inspeccionada se generó tres modelos de los cuáles se eligió el modelo más óptimo, las técnicas de minería de datos seleccionadas en el Capítulo 3, fueron aplicadas en cada modelo.

4.1.1. Primer Modelo

En el primer modelo se obtiene un error muy bajo en los tres algoritmos, esto se debe a que tomo variables identificadas por los expertos del negocio, a las cuáles se les otorgo un peso y al aplicar una fórmula de cálculo, los expertos identificaron 6 variables con información relevante y consistente para la evaluación de la empresa en el incumplimiento de sus obligaciones laborales, estas se encontraban ponderadas según un nivel de importancia.

Tabla 36.

Matriz de Ponderación de Variables

Número	Factores	Ponderación
1.	%Trabajadores menores de 15 años	16%
2.	% Trabajadores modalidad juvenil	7%
3.	%Trabajadores Discapacitados	13%
4.	% Contratos ingresados después de 30 días	23%
5.	Actas ingresadas después de 30 días	28%
6.	Trámites y Boletas	13%
Total		100%

La fórmula de cálculo para determinar si una empresa es inspeccionada o no inspeccionada es la siguiente:

$$P(incumplimiento) = \frac{Número\ de\ factores\ incumplidos}{Total\ de\ factores\ evaluados}$$

De acuerdo al resultado obtenido se clasifica de acuerdo a la Tabla 36.

Tabla 37.

Porcentaje Incumplimiento

Probabilidad	P(incumplimiento)
Inspeccionada	Mayor al 50% (>0.50)
No Inspeccionada	Menor o igual al 50% (<=0.50)

Para aplicar las técnicas de minería de datos seleccionadas se consideró la “Ley de Paretto[42]”, es decir, el 80% de los datos para entrenamiento y el 20% de los datos para

[42] El 80% de las consecuencias proviene del 20% de las causas.

la validación del modelo. Los resultados del error obtenido de la matriz de confusión en las tres técnicas se describen en la Tabla 37, donde se observa que los datos se encuentran clasificados al 99.98%, lo cual no es usual que en la primera aplicación de las técnicas el modelo sea tan eficiente. Este modelo quedó descartado debido a que se tomó la variable a predecir era el resultado de la variable que obtuvo al aplicar la fórmula de P(incumplimiento).

Tabla 38.

Error Técnicas Seleccionadas Primer Modelo

Técnica	Error
Árboles de Decisión	0.016%
Regresión Logística	0.017%
Redes Neuronales	0.014%

4.1.2. Segundo Modelo

Al descartar el primer modelo, se vio la necesidad de generar un segundo modelo donde se debió incorporar al data warehouse la variable del sistema SGI que indica si una empresa fue inspeccionada o no inspeccionada, realizada la actualización del data warehouse a las variables citadas en el primer modelo (Tabla 35), se agregan nuevas variables extraídas del data warehouse que también son de relevancia. Las nuevas variables que se agregan al modelo son:

- sk_tipo_empresa
- sk_actividad_economica
- sk_tipo_contrato
- sk_ubicacion

- sk_grupo_ocupacional
- fcum_total_contratos
- fcum_contratos_nulos
- fcum_contratos_periodo_prueba
- fcum_contratos_finalizados
- fcum_contratos_vigentes
- fcum_contratos_legalizados
- fcum_total_actas
- fcum_actas_anuladas
- fcum_actas_registradas
- fcum_actas_pagadas
- fcum_actas_consignadas_empresa
- fcum_actas_consignadas_ministerio
- fcum_trabajo_menores, fcum_trabajo_juvenil
- fcum_trabajo_discapacidad
- fcum_contratos_registro_atrasados
- fcum_actas_registro_atrasadas
- fcum_tramites_boletas
- fcum_total_tramites
- fcum_total_boletas
- fcum_resultado_sgi

Los tiempos de ejecución para 1579921 registros en cada una de las técnicas de minería de datos seleccionadas se presentan en la Tabla 38

Tabla 39.

Tiempo de Ejecución Técnicas Segundo Modelo

Técnica	Tiempo de Ejecución
Árboles de Decisión	5 minutos
Regresión Logística	5 minutos
Redes Neuronales	15 minutos

Para aplicar las técnicas de minería de datos seleccionadas se consideró la “Ley de Paretto”, es decir, el 80% de los datos para entrenamiento y el 20% de los datos para la validación del modelo. Los resultados del error obtenido de la matriz de confusión en las tres técnicas se describen en la Tabla 37, donde se observa que los datos se encuentran en promedio clasificados al 68%, la técnica que menor error tiene es el de árboles de decisión.

Tabla 40.

Error Técnicas Seleccionadas Primer Modelo

Técnica	Error
Árboles de Decisión	31.48%
Regresión Logística	32.62%
Redes Neuronales	32.60%

Este modelo quedó descartado debido a que presenta un alto índice de error debido a que algunas de las variables agregadas al modelo consideradas relevantes no aportaron en la eficacia esperada en el mismo.

4.1.3. Tercer Modelo

Para llegar al tercer modelo tuvimos que realizar varios ajustes al segundo modelo como eliminación de variables, incorporación de nuevas variables, pivoteo de campos lo cual conlleva a la obtención de variables dicotómicas (actividad económica, tipo contrato, tipo empresa), validación cruzada, cabe indicar que se mantienen las variables del primer modelo (Tabla 35), este nuevo modelo generado es el que mejor se adapta a los objetivos planteados.

Los tiempos de ejecución para 663297 registros en cada una de las técnicas de minería de datos seleccionadas se presentan en la Tabla 40. La técnica de minería de datos que más tiempo tarda en ejecutarse es redes neuronales en comparación a las técnicas de árboles de decisión y regresión logística, estas dos últimas técnicas mencionadas tienen tiempo de ejecución similar o igual dependiendo del número de registros.

Tabla 41.

Tiempo de Ejecución Técnicas Tercer Modelo

Técnica	Tiempo de Ejecución
Árboles de Decisión	2 horas
Regresión Logística	2 horas
Redes Neuronales	6 horas

Se tomarán algunos indicadores que nos permitirán determinar cuál de las técnicas de minería de datos es la más confiable, los indicadores son:

- **Matriz de confusión**, identifica la forma correcta e incorrecta la clasificación de los registros.
- **Exactitud del Modelo (Accuracy)**, porcentaje de valores clasificados de forma correcta por el modelo. Este valor se obtiene aplicando la siguiente fórmula:

$$Exactitud = \frac{VIC + VNIC}{Total}$$

Donde

VIC: Valores inspeccionados clasificados correctamente

VNIC: Valores no inspeccionados clasificados correctamente

Total: Cantidad de registros totales

- **Tasa de Error**, porcentaje de valores clasificados de forma incorrecta por el modelo. Este valor se obtiene aplicando la siguiente fórmula:

$$Tasa\ Error = \frac{VII + VNII}{Total}$$

Donde

VII: Valores inspeccionados clasificados incorrectamente

VNII: Valores no inspeccionados clasificados incorrectamente

Total: Cantidad de registros totales

- **Indicador Kappa**[43]
- **Tasa de verdaderos positivos** (Sensitivity), valores inspeccionados clasificados correctamente, que se calcula con la fórmula:

$$Sensibilidad = \frac{VIC}{Total\ Positivo}$$

Donde

VIC: Valores inspeccionados clasificados correctamente

- **Tasa de verdaderos negativos** (Specifity), valores inspeccionados clasificados incorrectamente, que se calcula con la fórmula:

$$Especificidad = \frac{VNIC}{Total\ Negativos}$$

Donde

VNIC: Valores no inspeccionados clasificados correctamente

- **Precisión Positiva,** valores positivos inspeccionados clasificados correctamente, que se calcula con la fórmula:

$$Precision = \frac{VIC}{Total\ Clasificados\ Positivos}$$

Donde

[43] Medida estadística que se utiliza en escalas nominales, compara la concordancia observada en un conjunto de datos, respecto a las que podrían ocurrir por azar. (Samiuc, 2011)

VIC: Valores inspeccionados clasificados correctamente

- **Precisión Negativa,** valores negativos inspeccionados clasificados correctamente, que se calcula con la fórmula:

$$Precision = \frac{VNIC}{Total\ Clasificados\ Negativos}$$

Donde

VNIC: Valores no inspeccionados clasificados correctamente

- **Curva ROC**

4.1.3.1. Arboles de Decisión

- **Matriz de Confusión**

A continuación, se presenta la matriz de confusión en la Figura 106, se observa:

- 163734 registros han sido clasificados como inspeccionadas de forma correcta por el modelo.
- 341162 registros han sido clasificados como no inspeccionadas de forma correcta por el modelo.
- 101122 registros han sido clasificados como no inspeccionadas de forma incorrecta por el modelo.
- 57279 registros han sido clasificados como inspeccionadas de forma incorrecta por el modelo.

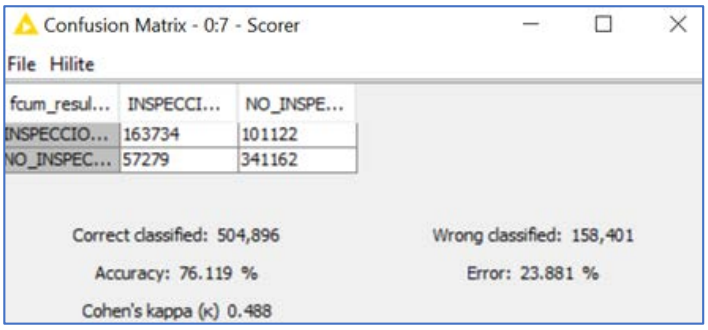


Figura 106. Matriz de Confusión Árboles de Decisión

- **Exactitud del Modelo**

En la Figura 106 se tiene el valor Accuracy que es 76.119%, que representa el porcentaje de valores clasificados de forma correcta por el modelo.

- **Tasa de Error**

En la Figura 106 se tiene el valor Error que es 23.881%, que representa el porcentaje de valores clasificados de forma incorrecta por el modelo.

- **Indicador Kappa**

En la Figura 106 se tiene el valor Cohens kappa (indicador kappa) que es 0.488, nos indica que la mayoría de predicciones son correctas.

- **Tasa de Verdaderos Positivos**

En la Figura 107 se tiene el valor de Sensitivity es 0.618 representa el 61.8% de los valores inspeccionados clasificados correctamente.

Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePo...	FalsePo...	TrueNe...	FalseNegatives	Recall	Precision	Sensitivity	Specifity	F-measure	Accuracy	Cohen's kappa
INSPECCION...	163734	57279	341162	101122	0.618	0.741	0.618	0.856	0.674	?	?
NO_INSPECC...	341162	101122	163734	57279	0.856	0.771	0.856	0.618	0.812	?	?
Overall	?	?	?	?	?	?	?	?	?	0.761	0.488

Figura 107. Tabla de Estadísticas de Precisión Árboles de Decisión

- **Tasa de Verdaderos Negativos**

En Figura la 107 se tiene el valor de Specifity es 0.856 representa el 85.6% de los valores inspeccionados clasificados incorrectamente.

- **Precisión Positiva**

En la Figura 107 se tiene el valor de Precision es 0.741 representa el 74.1% de los valores positivos inspeccionados clasificados correctamente.

- **Precisión Negativa**

En Figura 107 se tiene el valor de Precision es 0.771 representa el 77.1% de los valores negativos no inspeccionados clasificados correctamente.

- **Curva ROC**

En Figura 108 se tiene que la probabilidad que una empresa sea inspeccionada es de 0.7962 lo cual representa el 79.62% del total de empresas. De acuerdo a la Figura 108 se observa que más del 50% de empresas tienen probabilidades de ser inspeccionadas.

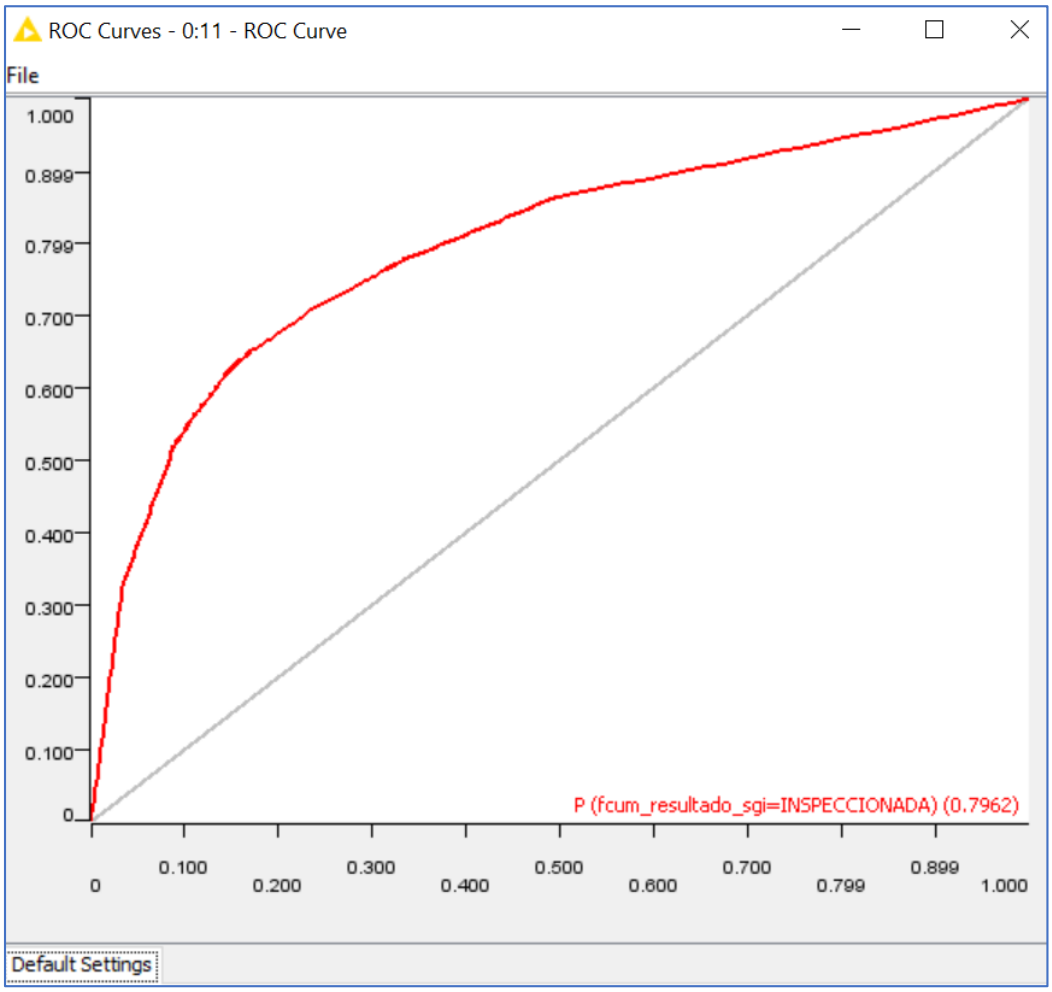


Figura 108. Curva ROC Árboles de Decisión

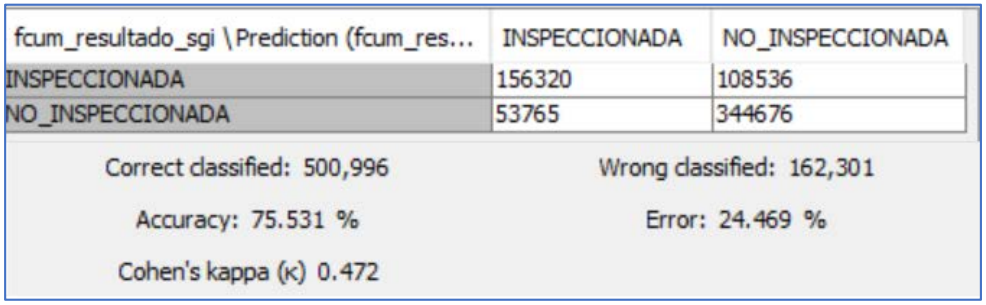
4.1.3.2. Regresión Logística

- **Matriz de Confusión**

A continuación, se presenta la matriz de confusión en la Figura 99, se observa:

- 156320 registros han sido clasificados como inspeccionadas de forma correcta por el modelo.
- 344676 registros han sido clasificados como no inspeccionadas de forma correcta por el modelo.

- 108536 registros han sido clasificados como no inspeccionadas de forma incorrecta por el modelo.
- 53765 registros han sido clasificados como inspeccionadas de forma incorrecta por el modelo.



fcum_resultado_sgi \ Prediction (fcum_res...	INSPECCIONADA	NO_INSPECCIONADA
INSPECCIONADA	156320	108536
NO_INSPECCIONADA	53765	344676

Correct classified: 500,996 Wrong classified: 162,301

Accuracy: 75.531 % Error: 24.469 %

Cohen's kappa (κ) 0.472

Figura 109. Matriz de Confusión Árboles de Decisión

- **Exactitud del Modelo**

  En la Figura 109 se tiene el valor Accuracy que es 75.531%, que representa el porcentaje de valores clasificados de forma correcta por el modelo.

- **Tasa de Error**

  En la Figura 109 se tiene el valor Error que es 24.469%, que representa el porcentaje de valores clasificados de forma incorrecta por el modelo.

- **Indicador Kappa**

  En la Figura 109 se tiene el valor Cohens kappa (indicador kappa) que es 0.472, que nos indica que la mayoría de predicciones son correctas.

- **Tasa de Verdaderos Positivos**

  En la Figura 110 se tiene el valor de Sensitivity que es 0.59 representa el 59 % de los valores inspeccionados clasificados correctamente.

Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specifity	F-measure	Accuracy	Cohen's kappa
INSPECCION...	156320	53765	344676	108536	0.59	0.744	0.59	0.865	0.658	?	?
NO_INSPECC...	344676	108536	156320	53765	0.865	0.761	0.865	0.59	0.809	?	?
Overall	?	?	?	?	?	?	?	?	?	0.755	0.472

***Figura 110*.** Tabla de Estadísticas de Precisión Árboles de Decisión

- **Tasa de Verdaderos Negativos**

  En la Figura 110 se tiene el valor de Specifity que es 0.865 representa el 86.5% de los valores inspeccionados clasificados incorrectamente.

- **Precisión Positiva**

  En la Figura 110 se tiene el valor de Precision que es 0.744 representa el 74.4% de los valores positivos inspeccionados clasificados correctamente.

- **Precisión Negativa**

  En la Figura 110 se tiene el valor de Precision que es 0.761 representa el 76.1% de los valores negativos no inspeccionados clasificados correctamente.

- **Curva ROC**

  En la Figura 111 se tiene que la probabilidad que una empresa sea inspeccionada es de 0.7928 lo cual representa el 79.28% del total de empresas. De acuerdo a la Figura 111 se observa que más del 50% de empresas tienen probabilidades de ser inspeccionadas.

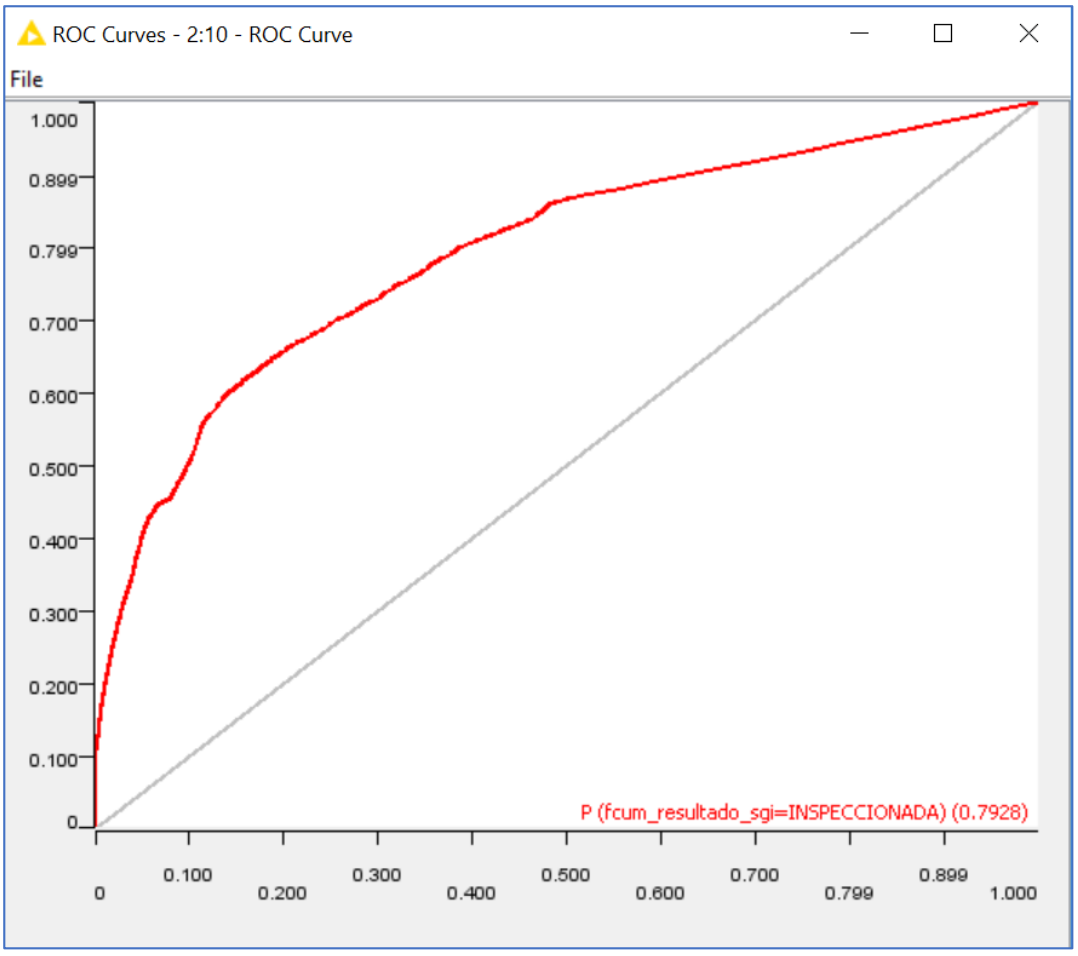


Figura 111. Curva ROC Árboles de Decisión

4.1.3.3. Redes Neuronales

- **Matriz de Confusión**

A continuación, se presenta la matriz de confusión en la Figura 112, se observa:

- 162755 registros han sido clasificados como inspeccionadas de forma correcta por el modelo.
- 346552 registros han sido clasificados como no inspeccionadas de forma correcta por el modelo.

- 102101 registros han sido clasificados como no inspeccionadas de forma incorrecta por el modelo.
- 51889 registros han sido clasificados como inspeccionadas de forma incorrecta por el modelo.

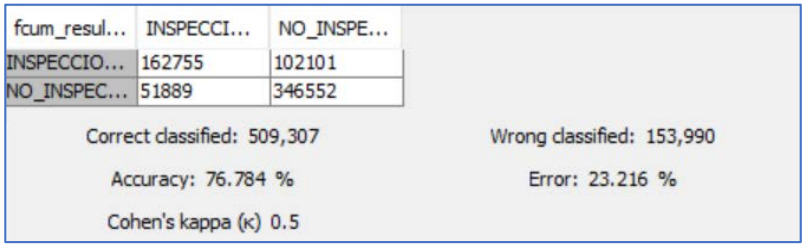


Figura 112. Matriz de Confusión Redes Neuronales

- **Exactitud del Modelo**

  En la Figura 112 se tiene el valor Accuracy que es 76.784%, que representa el porcentaje de valores clasificados de forma correcta por el modelo.

- **Tasa de Error**

  En la Figura 112 se tiene el valor Error que es 23.216%, que representa el porcentaje de valores clasificados de forma incorrecta por el modelo.

- **Indicador Kappa**

  En la Figura 112 se tiene el valor Cohens kappa (indicador kappa) que es 0. 5, que nos indica que la mayoría de predicciones son correctas

- **Tasa de Verdaderos Positivos**

  En la Figura 113 se tiene el valor de Sensitivity que es 0.615 representa el 61.5 % de los valores inspeccionados clasificados correctamente.

Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen'...
INSPECCION...	162755	51889	346552	102101	0.615	0.758	0.615	0.87	0.679	?	?
NO_INSPECC...	346552	102101	162755	51889	0.87	0.772	0.87	0.615	0.818	?	?
Overall	?	?	?	?	?	?	?	?	?	0.768	0.5

Figura 113. Tabla de Estadísticas de Precisión Redes Neuronales

- **Tasa de Verdaderos Negativos**

En la Figura 113 se tiene el valor de Specifity que es 0.87 representa el 87% de los valores inspeccionados clasificados incorrectamente.

- **Precisión Positiva**

En la Figura 113 se tiene el valor de Precision que es 0.758 representa el 75.8% de los valores positivos inspeccionados clasificados correctamente.

- **Precisión Negativa**

En la Figura 113 se tiene el valor de Precision que es 0.772 representa el 77.2% de los valores negativos no inspeccionados clasificados correctamente.

- **Curva ROC**

En la Figura 114 se tiene que la probabilidad que una empresa sea inspeccionada es de 0.8075 lo cual representa el 80.75% del total de empresas. De acuerdo a la Figura 104 se observa que más del 50% de empresas tienen probabilidades de ser inspeccionadas.

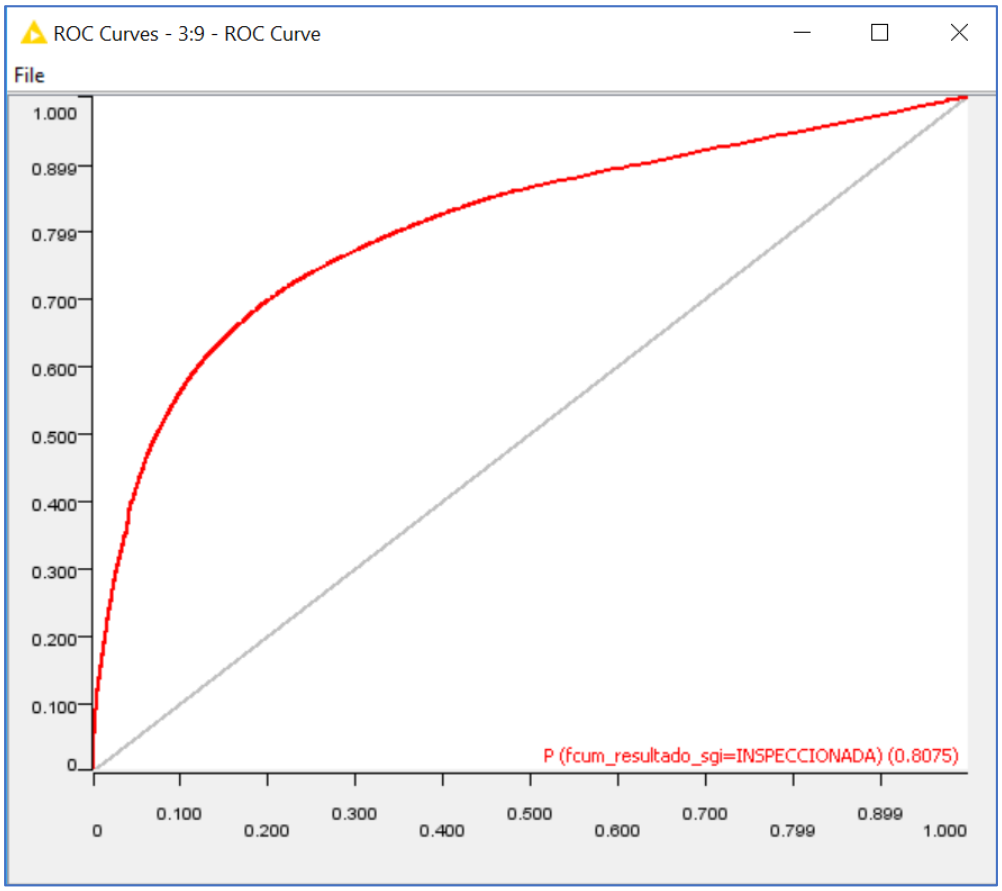


Figura 114. Curva ROC Redes Neuronales

En la Tabla 42, se tiene el resumen de los indicadores más representativos que nos permiten establecer la técnica que es mejor clasifica los, en donde se observa que el modelo que mejor clasifica los datos es el de redes neuronales, seguido por el modelo de árboles de decisión y por último el modelo de regresión logística.

Tabla 42.

Resumen de Indicadores más representativos

Técnica	Accuracy	Error	Indicador Kappa	ROC
Árboles de Decisión	76.119%	23.881%	0.488	0.7962
Regresión Logística	75. 531%	24.469%	0.472	0.7928
Redes Neuronales	76.784%	23.216%	0.5	0.8075

4.2. Reglas de Asociación

Para aplicar la técnica de reglas de asociación se obtuvieron datos almacenados en el data warehouse, estas variables están relacionadas con la empresa, contratos y el empleado, los datos obtenidos son:

- Identificador único de la empresa.
- Actividad económica que desempeña la empresa a su nivel más alto.
- Tipo de contrato generado al empleado
- Edad del empleado al momento de la generación del contrato, la cual se clasifica de la siguiente manera:
    - Menores de 15 años “<15”
    - Entre 15 y 18 años "15>x<=18"
    - Entre 18 y 26 años "18>x<=26"
    - Entre 26 y 35 años "26>x<=35"
    - Entre 35 y 45 años "35>x<=45"
    - Mayores de 46 años ">=46"

Se realizo varias implementaciones donde se obtuvieron 4 modelos, en cada uno de los modelos se relacionaron distintas variables las cuáles son a criterio de los expertos del negocio las más relevantes.

4.2.1. Primer Modelo

En el primer modelo se relacionaron las variables institución, rango de edad, motivo de salida y género del empleado, con la aplicación de esta técnica se obtiene que:

- Datos analizados 306404
- Reglas obtenidas 727

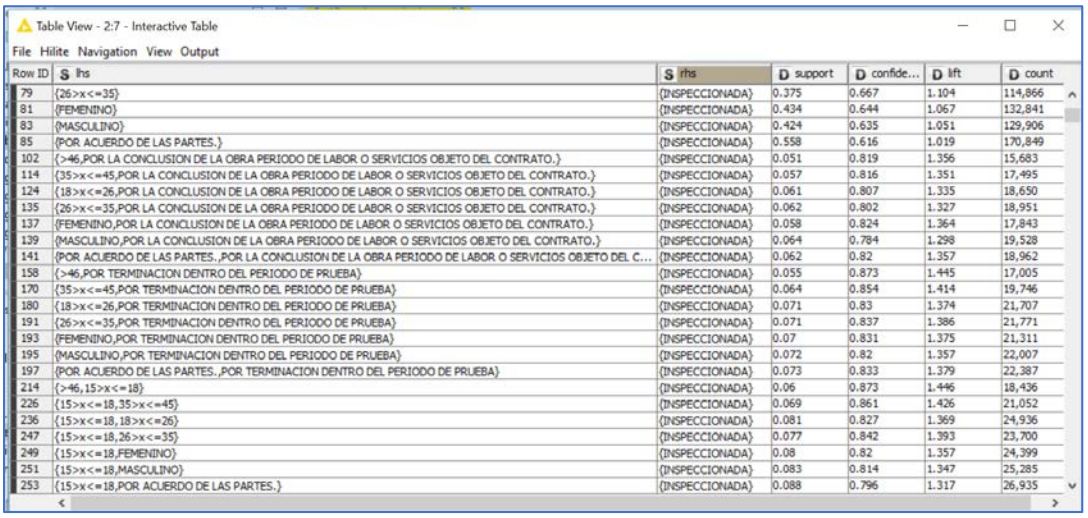


Table View - 2:7 - Interactive Table

File Hilite Navigation View Output

Row ID	lhs	rhs	support	confide...	lift	count
79	{26>x<=35}	{INSPECCIONADA}	0.375	0.667	1.104	114,866
81	{FEMENINO}	{INSPECCIONADA}	0.434	0.644	1.067	132,841
83	{MASCULINO}	{INSPECCIONADA}	0.424	0.635	1.051	129,906
85	{POR ACUERDO DE LAS PARTES.}	{INSPECCIONADA}	0.558	0.616	1.019	170,849
102	{>46,POR LA CONCLUSION DE LA OBRA PERIODO DE LABOR O SERVICIOS OBJETO DEL CONTRATO.}	{INSPECCIONADA}	0.051	0.819	1.356	15,683
114	{35>x<=45,POR LA CONCLUSION DE LA OBRA PERIODO DE LABOR O SERVICIOS OBJETO DEL CONTRATO.}	{INSPECCIONADA}	0.057	0.816	1.351	17,495
124	{18>x<=26,POR LA CONCLUSION DE LA OBRA PERIODO DE LABOR O SERVICIOS OBJETO DEL CONTRATO.}	{INSPECCIONADA}	0.061	0.807	1.335	18,650
135	{26>x<=35,POR LA CONCLUSION DE LA OBRA PERIODO DE LABOR O SERVICIOS OBJETO DEL CONTRATO.}	{INSPECCIONADA}	0.062	0.802	1.327	18,951
137	{FEMENINO,POR LA CONCLUSION DE LA OBRA PERIODO DE LABOR O SERVICIOS OBJETO DEL CONTRATO.}	{INSPECCIONADA}	0.058	0.824	1.364	17,843
139	{MASCULINO,POR LA CONCLUSION DE LA OBRA PERIODO DE LABOR O SERVICIOS OBJETO DEL CONTRATO.}	{INSPECCIONADA}	0.064	0.784	1.298	19,528
141	{POR ACUERDO DE LAS PARTES.,POR LA CONCLUSION DE LA OBRA PERIODO DE LABOR O SERVICIOS OBJETO DEL C...	{INSPECCIONADA}	0.062	0.82	1.357	18,962
158	{>46,POR TERMINACION DENTRO DEL PERIODO DE PRUEBA}	{INSPECCIONADA}	0.055	0.873	1.445	17,005
170	{35>x<=45,POR TERMINACION DENTRO DEL PERIODO DE PRUEBA}	{INSPECCIONADA}	0.064	0.854	1.414	19,746
180	{18>x<=26,POR TERMINACION DENTRO DEL PERIODO DE PRUEBA}	{INSPECCIONADA}	0.071	0.83	1.374	21,707
191	{26>x<=35,POR TERMINACION DENTRO DEL PERIODO DE PRUEBA}	{INSPECCIONADA}	0.071	0.837	1.386	21,771
193	{FEMENINO,POR TERMINACION DENTRO DEL PERIODO DE PRUEBA}	{INSPECCIONADA}	0.07	0.831	1.375	21,311
195	{MASCULINO,POR TERMINACION DENTRO DEL PERIODO DE PRUEBA}	{INSPECCIONADA}	0.072	0.82	1.357	22,007
197	{POR ACUERDO DE LAS PARTES.,POR TERMINACION DENTRO DEL PERIODO DE PRUEBA}	{INSPECCIONADA}	0.073	0.833	1.379	22,387
214	{>46,15>x<=18}	{INSPECCIONADA}	0.06	0.873	1.446	18,436
226	{15>x<=18,35>x<=45}	{INSPECCIONADA}	0.069	0.861	1.426	21,052
236	{15>x<=18,18>x<=26}	{INSPECCIONADA}	0.081	0.827	1.369	24,936
247	{15>x<=18,26>x<=35}	{INSPECCIONADA}	0.077	0.842	1.393	23,700
249	{15>x<=18,FEMENINO}	{INSPECCIONADA}	0.08	0.82	1.357	24,399
251	{15>x<=18,MASCULINO}	{INSPECCIONADA}	0.083	0.814	1.347	25,285
253	{15>x<=18,POR ACUERDO DE LAS PARTES.}	{INSPECCIONADA}	0.088	0.796	1.317	26,935

Figura 115. Reglas de Asociación del Primer Modelo

Las reglas más relevantes encontradas a criterio del experto del negocio en este modelo son:

1. **{FEMENINO, POR ACUERDO DE LAS PARTES}.**

   Para esta regla se tiene un soporte del 40.4% que representa 123871 registros de los cuales el 65.7% son inspeccionadas.

2. **{MASCULINO, POR ACUERDO DE LAS PARTES.}**

   Para esta regla se tiene un soporte del 40.2% que representa 123044 registros de los cuales el 64.5% son inspeccionadas.

3. **{26>x<=35, POR ACUERDO DE LAS PARTES}**

   Para esta regla se tiene un soporte del 35.7% que representa 10484 registros de los cuales el 67.8% son inspeccionadas.

4. **{18>x<=26, POR ACUERDO DE LAS PARTES.}**

   Para esta regla se tiene un soporte del 35.2% que representa 107970 registros de los cuales el 67.9% son inspeccionadas.

4.2.2. Segundo Modelo

En el segundo modelo se relacionaron las variables institución, rango de edad, nombre actividad económica y género del empleado, con la aplicación de esta técnica se obtiene que:

- Datos analizados 296790
- Reglas obtenidas 378

Table View - 2:9 - Interactive Table (141 x 6)

File Hilite Navigation View Output

Ro...	lhs	rhs	support	confide...	lift	count
248	{18>x<=26,MASCULINO}	{INSPECCIONADA}	0.266	0.667	1.225	78,846
245	{18>x<=26,FEMENINO}	{INSPECCIONADA}	0.261	0.69	1.268	77,608
230	{35>x<=45,MASCULINO}	{INSPECCIONADA}	0.259	0.649	1.193	76,772
267	{FEMENINO,MASCULINO}	{INSPECCIONADA}	0.256	0.687	1.263	75,853
227	{35>x<=45,FEMENINO}	{INSPECCIONADA}	0.247	0.672	1.236	73,371
200	{>=46,MASCULINO}	{INSPECCIONADA}	0.243	0.633	1.164	72,034
242	{18>x<=26,26>x<=35}	{INSPECCIONADA}	0.233	0.721	1.325	69,214
197	{>=46,FEMENINO}	{INSPECCIONADA}	0.231	0.647	1.19	68,657
583	{26>x<=35,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.22	0.727	1.337	65,379
224	{26>x<=35,35>x<=45}	{INSPECCIONADA}	0.216	0.725	1.332	64,254
570	{18>x<=26,26>x<=35,MASCULINO}	{INSPECCIONADA}	0.216	0.729	1.34	64,200
574	{18>x<=26,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.215	0.734	1.349	63,677
566	{18>x<=26,26>x<=35,FEMENINO}	{INSPECCIONADA}	0.209	0.753	1.384	61,913
211	{18>x<=26,35>x<=45}	{INSPECCIONADA}	0.203	0.739	1.358	60,340
554	{26>x<=35,35>x<=45,MASCULINO}	{INSPECCIONADA}	0.202	0.734	1.349	60,079
558	{35>x<=45,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.2	0.734	1.349	59,289
194	{>=46,26>x<=35}	{INSPECCIONADA}	0.193	0.727	1.336	57,167
550	{26>x<=35,35>x<=45,FEMENINO}	{INSPECCIONADA}	0.192	0.758	1.393	57,051
839	{18>x<=26,26>x<=35,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.192	0.766	1.409	56,899
533	{18>x<=26,35>x<=45,MASCULINO}	{INSPECCIONADA}	0.191	0.747	1.374	56,797
169	{>=46,35>x<=45}	{INSPECCIONADA}	0.186	0.721	1.326	55,138
518	{>=46,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.184	0.723	1.328	54,716
529	{18>x<=26,35>x<=45,FEMENINO}	{INSPECCIONADA}	0.184	0.769	1.415	54,618
514	{>=46,26>x<=35,MASCULINO}	{INSPECCIONADA}	0.182	0.736	1.353	53,997
525	{18>x<=26,26>x<=35,35>x<=45}	{INSPECCIONADA}	0.181	0.776	1.427	53,629
181	{>=46,18>x<=26}	{INSPECCIONADA}	0.18	0.741	1.362	53,304

Figura 116. Reglas de Asociación del Segundo Modelo

Las reglas más relevantes encontradas a criterio del experto del negocio en este modelo son:

1. **{CONTRATO INDEFINIDO, MASCULINO}**

Para esta regla se tiene un soporte del 23.4% que representa 69320 registros de los cuales el 63.7% son inspeccionadas.

2. **{CONTRATO INDEFINIDO, FEMENINO}**

   Para esta regla se tiene un soporte del 21.4% que representa 63495 registros de los cuales el 67.2% son inspeccionadas.

3. **{CONTRATO A PLAZO FIJO, MASCULINO}**

   Para esta regla se tiene un soporte del 20.7% que representa 61540 registros de los cuales el 68.5% son inspeccionadas.

4. **{CONTRATO A PLAZO FIJO, FEMENINO}**

   Para esta regla se tiene un soporte del 18.3% que representa 54401 registros de los cuales el 73.7% son inspeccionadas.

4.3.3. Tercer Modelo

En el tercer modelo se relacionaron las variables institución, rango de edad y género del empleado, con la aplicación de esta técnica se obtiene que:

- Datos analizados 296790
- Reglas obtenidas 378

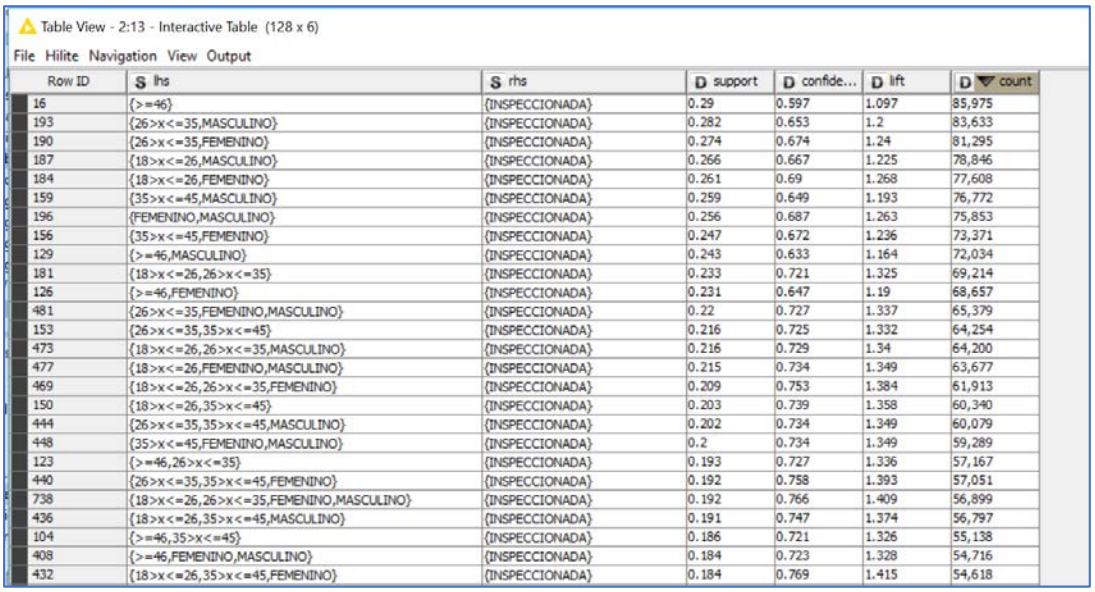


Table View - 2:13 - Interactive Table (128 x 6)

File Hilite Navigation View Output

Row ID	lhs	rhs	support	confide...	lift	count
16	{>=46}	{INSPECCIONADA}	0.29	0.597	1.097	85,975
193	{26>x<=35,MASCULINO}	{INSPECCIONADA}	0.282	0.653	1.2	83,633
190	{26>x<=35,FEMENINO}	{INSPECCIONADA}	0.274	0.674	1.24	81,295
187	{18>x<=26,MASCULINO}	{INSPECCIONADA}	0.266	0.667	1.225	78,846
184	{18>x<=26,FEMENINO}	{INSPECCIONADA}	0.261	0.69	1.268	77,608
159	{35>x<=45,MASCULINO}	{INSPECCIONADA}	0.259	0.649	1.193	76,772
196	{FEMENINO,MASCULINO}	{INSPECCIONADA}	0.256	0.687	1.263	75,853
156	{35>x<=45,FEMENINO}	{INSPECCIONADA}	0.247	0.672	1.236	73,371
129	{>=46,MASCULINO}	{INSPECCIONADA}	0.243	0.633	1.164	72,034
181	{18>x<=26,26>x<=35}	{INSPECCIONADA}	0.233	0.721	1.325	69,214
126	{>=46,FEMENINO}	{INSPECCIONADA}	0.231	0.647	1.19	68,657
481	{26>x<=35,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.22	0.727	1.337	65,379
153	{26>x<=35,35>x<=45}	{INSPECCIONADA}	0.216	0.725	1.332	64,254
473	{18>x<=26,26>x<=35,MASCULINO}	{INSPECCIONADA}	0.216	0.729	1.34	64,200
477	{18>x<=26,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.215	0.734	1.349	63,677
469	{18>x<=26,26>x<=35,FEMENINO}	{INSPECCIONADA}	0.209	0.753	1.384	61,913
150	{18>x<=26,35>x<=45}	{INSPECCIONADA}	0.203	0.739	1.358	60,340
444	{26>x<=35,35>x<=45,MASCULINO}	{INSPECCIONADA}	0.202	0.734	1.349	60,079
448	{35>x<=45,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.2	0.734	1.349	59,289
123	{>=46,26>x<=35}	{INSPECCIONADA}	0.193	0.727	1.336	57,167
440	{26>x<=35,35>x<=45,FEMENINO}	{INSPECCIONADA}	0.192	0.758	1.393	57,051
738	{18>x<=26,26>x<=35,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.192	0.766	1.409	56,899
436	{18>x<=26,35>x<=45,MASCULINO}	{INSPECCIONADA}	0.191	0.747	1.374	56,797
104	{>=46,35>x<=45}	{INSPECCIONADA}	0.186	0.721	1.326	55,138
408	{>=46,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.184	0.723	1.328	54,716
432	{18>x<=26,35>x<=45,FEMENINO}	{INSPECCIONADA}	0.184	0.769	1.415	54,618

Figura 117. Reglas de Asociación del Tercer Modelo

Las reglas más relevantes encontradas a criterio del experto del negocio en este modelo son:

1. **{26>x<=35, MASCULINO}**

   Para esta regla se tiene un soporte del 28.2% que representa 83633 registros de los cuales el 65.3% son inspeccionadas.

2. **{26>x<=35, FEMENINO}**

   Para esta regla se tiene un soporte del 27.4% que representa 81295 registros de los cuales el 67.4% son inspeccionadas.

3. **{18>x<=26, MASCULINO}**

   Para esta regla se tiene un soporte del 26.6% que representa 78846 registros de los cuales el 66.7% son inspeccionadas.

4. **{18>x<=26, FEMENINO}**

Para esta regla se tiene un soporte del 26.1% que representa 77608 registros de los cuales el 69% son inspeccionadas.

4.3.4. Cuarto Modelo

En el tercer modelo se relacionaron las variables institución, rango de edad, tipo de contrato y género del empleado, con la aplicación de esta técnica se obtiene que:

- Datos analizados 296790
- Reglas obtenidas 777

Table View - 2:19 - Interactive Table (768 x 6)

File Hilite Navigation View Output

Row ID	lhs	rhs	support	confide...	lift	count
560	{26>x<=35,35>x<=45}	{INSPECCIONADA}	0.217	0.724	1.331	64,545
1898	{18>x<=26,26>x<=35,MASCULINO}	{INSPECCIONADA}	0.217	0.729	1.34	64,359
1902	{18>x<=26,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.215	0.734	1.349	63,729
488	{CONTRATO INDEFINIDO,FEMENINO}	{INSPECCIONADA}	0.214	0.672	1.235	63,495
1894	{18>x<=26,26>x<=35,FEMENINO}	{INSPECCIONADA}	0.209	0.752	1.383	62,098
429	{CONTRATO A PLAZO FIJO,MASCULINO}	{INSPECCIONADA}	0.207	0.685	1.259	61,540
557	{18>x<=26,35>x<=45}	{INSPECCIONADA}	0.204	0.739	1.358	60,493
485	{26>x<=35,CONTRATO INDEFINIDO}	{INSPECCIONADA}	0.203	0.693	1.275	60,387
1869	{26>x<=35,35>x<=45,MASCULINO}	{INSPECCIONADA}	0.203	0.733	1.348	60,285
1873	{35>x<=45,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.2	0.734	1.349	59,380
530	{>=46,26>x<=35}	{INSPECCIONADA}	0.193	0.727	1.336	57,287
1865	{26>x<=35,35>x<=45,FEMENINO}	{INSPECCIONADA}	0.193	0.757	1.392	57,273
482	{18>x<=26,CONTRATO INDEFINIDO}	{INSPECCIONADA}	0.193	0.707	1.299	57,218
3862	{18>x<=26,26>x<=35,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.192	0.766	1.408	57,013
1861	{18>x<=26,35>x<=45,MASCULINO}	{INSPECCIONADA}	0.192	0.747	1.373	56,925
1749	{26>x<=35,CONTRATO INDEFINIDO,MASCULINO}	{INSPECCIONADA}	0.188	0.711	1.307	55,774
466	{35>x<=45,CONTRATO INDEFINIDO}	{INSPECCIONADA}	0.188	0.695	1.278	55,698
511	{>=46,35>x<=45}	{INSPECCIONADA}	0.186	0.721	1.325	55,349
1833	{>=46,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.185	0.723	1.328	54,762
1857	{18>x<=26,35>x<=45,FEMENINO}	{INSPECCIONADA}	0.184	0.769	1.414	54,751
427	{CONTRATO A PLAZO FIJO,FEMENINO}	{INSPECCIONADA}	0.183	0.737	1.355	54,401
1829	{>=46,26>x<=35,MASCULINO}	{INSPECCIONADA}	0.182	0.736	1.353	54,098
1853	{18>x<=26,26>x<=35,35>x<=45}	{INSPECCIONADA}	0.181	0.776	1.426	53,852
424	{26>x<=35,CONTRATO A PLAZO FIJO}	{INSPECCIONADA}	0.181	0.738	1.357	53,782
1753	{CONTRATO INDEFINIDO,FEMENINO,MASCULINO}	{INSPECCIONADA}	0.181	0.726	1.334	53,615
527	{>=46,18>x<=26}	{INSPECCIONADA}	0.18	0.741	1.362	53,402

Figura 118. Reglas de Asociación del Tercer Modelo

Las reglas más relevantes encontradas a criterio del experto del negocio en este modelo son:

1. **{COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACION DE VEHICULOS AUTOMOTORES Y MOTOCICLETAS, MASCULINO}**

   Para esta regla se tiene un soporte del 7.1% que representa 21077 registros de los cuales el 58.9% son inspeccionadas.

2. **{>=46 AÑOS, ACTIVIDADES PROFESIONALES CIENTIFICAS Y TECNICAS, FEMENINO}**

   Para esta regla se tiene un soporte del 3.1% que representa 9142 registros de los cuales el 69.5% son inspeccionadas.

3. **{26 AÑOS >x<=35 AÑOS, ACTIVIDADES PROFESIONALES CIENTIFICAS Y TECNICAS, MASCULINO}**

   Para esta regla se tiene un soporte del 7.3% que representa 10899 registros de los cuales el 73% son inspeccionadas.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- La construcción de un Data Warehouse ayuda a mejorar el tratamiento de los datos para convertirlos en información, ya que los datos que se encuentran limpios e integrados, son utilizados para realizar explotación de datos con cualquier herramienta BI que sirven como apoyo a las autoridades del Ministerio de Trabajo en la toma de decisiones, adicional estos datos son utilizados para realizar el proyecto de minería de datos y de esta forma se evitan que los datos vayan al proyecto con ruido.
- La aplicación de reglas de asociación nos permitió encontrar patrones y tendencias en los datos analizados, las reglas encontradas permitirán a las autoridades del Ministerio identificar a que parte de la población económicamente activa es la que necesita mayor atención, entre las reglas se evidenció que el tipo de contratación predominante en el Ecuador son los contratos indefinidos para el género masculino y femenino, adicional se encontró que el rango de edad predominante en todos los análisis es de 26 a 35 años.
- El caso de estudio propuesto demuestra que es posible la aplicación de técnicas de minería de datos en la administración integral del trabajo y empleo de las empresas ecuatorianas, ya que la mayoría de estudios que existen han sido orientados a otros sectores como salud y financiero, para el sector laboral existen muy pocas investigaciones.
- De acuerdo a los resultados de los indicadores, se tiene que la técnica de minería que mejor clasifica los datos para nuestro caso de estudio es la técnica de Redes Neuronales, pese a que el tiempo de ejecución es mayor que las otras técnicas.
- Las Redes Neuronales son comúnmente utilizadas para el reconocimiento de patrones y clasificación, son de gran de utilidad en la predicción de datos económicos y financieros. La estructura más utilizada es el Perceptrón Multicapa.

5.2. Recomendaciones

- Se recomienda al Ministerio de Trabajo utilizar el repositorio que existe para realizar reportería, esto ayudara a generar reportes de una forma rápida y confiable.
- Para realizar proyectos de minería de datos se recomienda desarrollar y construir un data warehouse, ya que los datos pasan por una fase de limpieza de datos, y estos datos limpios pueden ser utilizados como datos de entrada para el proyecto de minería de datos.

BIBLIOGRAFÍA

Alfredo. (2 de Abril de 2017). *Time of Software*. Obtenido de http://timeofsoftware.com/descubriendo-informacion/

Anónimo. (01 de 01 de 2017). *Numerictron*. Obtenido de https://sites.google.com/site/numerictron/unidad-4/4-3-regresion-por-minimos-cuadrados-lineal-y-cuadratica

Bustamante, P. (22 de Mayo de 2015). *Péres Bustamante & Ponce*. Obtenido de http://www.pbplaw.com/registro-actas-finiquito-informacion-laboral-saite/

Bustamante, P. (22 de Mayo de 2015). *Péres Bustamante & Ponce*. Obtenido de http://www.pbplaw.com/registro-actas-finiquito-informacion-laboral-saite/

Carlie Idoine, Peter Krensky, Erick Brethenoux, Jim Hare, Svetlana Sicular, Shubhangi Vashisth. (22 de Febrero de 2018). *Gartner*. Obtenido de https://www.gartner.com/doc/reprints?id=1-4RMUF0K&ct=180222&st=sb

Chintan Shah, A. J. (2013). Comparison of Data Minin Classification Algorithmas for Breast Cancer Prediction. *IEEE-31661.* Tiruchengode, India.

Chintan Shah, A. J. (2013). Comparison of Data Mining Classification Algorithms for Breast Cancer Prediction. *IEE*, 4.

Col, A. (1 de Enero de 2018). *Scribd.* Obtenido de https://es.scribd.com/doc/55887136/Trabajo-grupos-ocupacionales-1

Corral, C. R. (1 de Abril de 2014). *Algoritmo Apriori*. Obtenido de https://prezi.com/q6w3v8uxsnlg/algoritmo-apiori/

Darío, B. R. (2010). *Hefesto.* Córdoba, Argentina.

Dertiano, V. (9 de Marzo de 2015). *Arquitectura BI*. Obtenido de http://blog.mirai-advisory.com/arquitectura-bi-parte-ii-el-enfoque-de-william-h-inmon/

Díaz, M. V. (11 de Mayo de 2016). *Smartbase Group*. Obtenido de http://smartbasegroup.com/metodologia-crisp-dm-parte-i/

Díaz, M. V. (9 de Junio de 2016). *Smartbase Group*. Obtenido de http://smartbasegroup.com/metodologia-crisp-dm-parte-2/

Díaz, M. V. (13 de 06 de 2016). *Smartbase Group*. Obtenido de http://smartbasegroup.com/metodologia-crisp-dm-final/

Educar. (7 de Septiembre de 2017). *Educar plus.com*. Obtenido de http://educarplus.com/2017/09/codigo-trabajo-2017-obligaciones-beneficios-empleadores-empleados.html

González, J. A. (2016). *Minería de Datos.* Obtenido de https://ccc.inaoep.mx/~jagonzalez/AI/Sesion13_Data_Mining.pdf

Grández, M. (1 de Enero de 2017). *Aplicación de Minería de Datos para Determinar Patrones de Consumo Futuro en Clientes de una Distribuidora de Suplemento Nutricionales.* Obtenido de http://repositorio.usil.edu.pe/bitstream/USIL/2763/1/2017_Granda_Aplicacion-de-mineria-datos.pdf

Inmon, W. (2005). *Building Data Warehouse* (Vol. Fourth Edition). USA: John Wiley & Sons, Inc.

Jmacoe. (19 de Febrero de 2018). *El rincón de Jmacoe*. Obtenido de http://blog.jmacoe.com/gestion_ti/base_de_datos/5-mejores-software-mineria-datos-codigo-libre-abierto/

Katherine González, O. L. (2016). Conectando Sociedades. *Par-Knime: Conjunto de plugins para extraer reglas de asociación cuantitativas en Kinime*, (pág. 9). La Habana.

Kimball, R. (2008). *The Data Warehouse Lifecycle Toolkit* (Vol. Segunda Edición). New York. Obtenido de Kimball et al., The Data Warehouse Lifecycle Toolkit. 2nd Edition. New York,.

Knime. (27 de Agosto de 2015). *Open for Innovation Knime*. Obtenido de https://www.knime.com/forum/knime-general/the-distinguish-between-gini-index-and-gain-ratio

Knime. (11 de Abril de 2017). *Open for innovation - Knime*. Obtenido de https://www.knime.com/forum/knime-general/x-partitioner-partitions-size

Knime. (1 de Enero de 2017). *Open for innovation-Knime*. Obtenido de https://www.knime.com/features

Legal, E. (26 de Abril de 2015). *EcuadorLegalOnline*. Obtenido de http://www.ecuadorlegalonline.com/laboral/ley-de-justicia-laboral/

Martínez, B. (01 de Enero de 2018). *Universidada Mayor*. Obtenido de http://patoral.umayor.cl/patoral/?p=1012

Pandey, S. (2016). Data Mining Techniques for Medical Data: A Review. *Scopes*, 11.

Pérez, J. (1 de Enero de 2008). *Definición.de.* Obtenido de https://definicion.de/sueldo/

Pérez, J. (1 de Enero de 2012). *Definición.de.* Obtenido de https://definicion.de/discapacidad/

Pérez, J. (1 de Enero de 2017). *Definición.de*. Obtenido de https://definicion.de/actividad-economica/

Piatetsky, G. (1 de Octubre de 2014). *Kdnuggets*. Obtenido de https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html

Román, J. V. (8 de Agosto de 2016). *Singular data&analytics*. Obtenido de https://data.sngular.team/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science

Samiuc. (17 de Noviembre de 2011). *Samiuc*. Obtenido de https://samiuc.es/index.php/estadisticas-con-variables-binarias/medidas-de-concordancia/kappa-de-cohen.html

SENA. (5 de Noviembre de 2015). *Servicio Nacional de Aprendizaje*. Obtenido de https://senaintro.blackboard.com/bbcswebdav/pid47553315-dt-content-rid14706754_4/institution/217213_tvirtual/OAAPs/OAAP2/aa4/oa2aa4/utilidades/descarga

Significados. (1 de Enero de 2018). *Signifcados*. Obtenido de https://www.significados.com/consignar/

Trabajo, M. d. (1 de Enero de 2014). *Ministerio del Trabajo.* Obtenido de http://www.trabajo.gob.ec/wp-content/uploads/2014/08/banco-de-preguntas-contratos.pdf

Trabajo, M. d. (1 de Diciembre de 2016). *Ministerio de Trabajo*. Obtenido de http://www.trabajo.gob.ec/ministerio-del-trabajo-optimiza-procesos-de-inspeccion-laboral-a-traves-del-sistema-inspector-integral-2-0/

Trabajo, M. d. (1 de Enero de 2017). *Ministerio de Trabajo*. Obtenido de http://www.trabajo.gob.ec/httpwww-trabajo-gob-ecwp-contentuploads201603web_reformas-06-pngntrato-juvenil/

Vargas, K. (1 de Enero de 2016). *Mercadotecnia Electrónica*. Obtenido de https://sites.google.com/site/kenyavargasmedrano2016/5-3-1-almacenes-de-datos

Venemedia. (19 de Septiembre de 2011). *Definicion.De*. Obtenido de http://conceptodefinicion.de/genero/

Vinueza, M. (1 de Enero de 2018). *Humanium*. Obtenido de https://www.humanium.org/es/trabajo-infantil/

WordReference. (1 de Enero de 2018). *WordReference.com*. Obtenido de http://www.wordreference.com/definicion/empleado

GLOSARIO

Contrato de trabajo: Contrato individual de trabajo es un convenio en virtud del cual una persona se compromete para con otra u otras a prestar sus servicios lícitos y personales, bajo su dependencia, por una remuneración fijada en el convenio, la ley, el contrato colectivo o la costumbre. (Trabajo M. d., 2014).

Contrato Expreso y Tácito: El contrato es expreso cuando el empleador y el trabajador acuerden las condiciones, sea de palabra o reduciéndolas a escrito. A falta de estipulación expresa, se considera tácita toda relación de trabajo entre empleador y trabajador. (Trabajo M. d., 2014).

Contrato a prueba: En todo contrato de aquellos a los que se refiere el Art. 14, cuando se celebre por primera vez, podrá señalarse un tiempo de prueba, de duración máxima de noventa días. Vencido este plazo, automáticamente se entenderá que continúa en vigencia por el tiempo que faltare para completar el año. Tal contrato no podrá celebrarse sino una sola vez entre las mismas partes. Durante el plazo de prueba, cualquiera de las partes lo puede dar por terminado libremente. (Trabajo M. d., 2014)

Contrato Por obra cierta: Cuando el trabajador toma a su cargo la ejecución de una labor determinada por una remuneración que comprende la totalidad de la misma, sin tomar en consideración el tiempo que se invierta en ejecutarla. (Trabajo M. d., 2014)

Contrato Por tarea: El trabajador se compromete a ejecutar una determinada cantidad de obra o trabajo en la jornada o en un período de tiempo previamente establecido. Se entiende concluida la jornada o período de tiempo, por el hecho de cumplirse la tarea. (Trabajo M. d., 2014)

Contrato Eventuales: Aquellos que se realizan para satisfacer exigencias circunstanciales del empleador, tales como reemplazo de personal que se encuentra ausente por vacaciones, licencia, enfermedad, maternidad y situaciones similares; en cuyo caso, en el contrato deberá puntualizarse las exigencias circunstanciales que motivan la contratación, el nombre o nombres de los reemplazados y el plazo de duración

de la misma. También se podrán celebrar contratos eventuales para atender una mayor demanda de producción o servicios en actividades habituales del empleador, en cuyo caso el contrato no podrá tener una duración mayor de ciento ochenta días continuos o discontinuos dentro de un lapso de trescientos sesenta y cinco días. Si la circunstancia o requerimiento de los servicios del trabajador se repite por más de dos períodos anuales, el contrato se convertirá en contrato de temporada. El sueldo o salario que se pague en los contratos eventuales, tendrá un incremento del 35% del valor hora del salario básico del sector al que corresponda el trabajador. (Trabajo M. d., 2014)

Contrato Ocasionales: Aquellos cuyo objeto es la atención de necesidades emergentes o extraordinarias, no vinculadas con la actividad habitual del empleador, y cuya duración no excederá de treinta días en un año. El sueldo o salario que se pague en los contratos ocasionales, tendrá un incremento del 35% del valor hora del salario básico del sector al que corresponda el trabajador. (Trabajo M. d., 2014)

Contrato De temporada: Aquellos que en razón de la costumbre o de la contratación colectiva, se han venido celebrando entre una empresa o empleador y un trabajador o grupo de trabajadores, para que realicen trabajos cíclicos o periódicos, en razón de la naturaleza discontinua de sus labores, gozando estos contratos de estabilidad, entendida, como el derecho de los trabajadores a ser llamados a prestar sus servicios en cada temporada que se requieran. Se configurará el despido intempestivo si no lo fueren. (Trabajo M. d., 2014).

Acta finiquito: Documento legal mediante el cual se formaliza el pago de valores correspondientes a la liquidación, que se produce el trabajador y el empleador dan por terminada la relación laboral. (Trabajo M. d., 2014)

Desahucio: Es un aviso mediante el cual una de las partes da a conocer a la otra su voluntad de dar por terminado el contrato. (Trabajo M. d., 2014)

Despido intempestivo: Cuando el empleador despide al trabajador, terminando sin causa ni justificación alguna la relación laboral. Al producirse el despido intempestivo el

empleador tiene la obligación de pagarle las indemnizaciones económicas determinadas en la ley. (Trabajo M. d., 2014)

SAITE: El Sistema de Administración Integral de Trabajo y Empleo es una herramienta informática que sirve para (Bustamante, Péres Bustamante & Ponce, 2015)

- **Llevar una base de datos de la información de los trabajadores**: herramienta para que el empleador pueda cumplir su obligación de mantener un registro con la información de sus trabajadores. Los empleadores que utilicen este sistema no estarán obligados a llevar otro tipo de registro. (Bustamante, Péres Bustamante & Ponce, 2015)n
- **Registro de actas de finiquito:** utilizado para que el empleador cumpla con su obligación de elaborar y registrar el acta de finiquito y la constancia de su pago. Para ello registrará la información solicitada por el aplicativo y cargará el acta de finiquito firmada por las partes. (Bustamante, Péres Bustamante & Ponce, 2015)

SINACOI: Sistema Nacional de Control de Inspectores, sistema del Ministerio del Trabajo que registra trámites ingresados a la institución.

CÓDIGO DE TRABAJO: Es un documento jurídico para legislar la actividad laboral, que regula los derechos y obligaciones de patronos y trabajadores. (Educar, 2017)

Boletas: Trámite ingresado al Ministerio del Trabajo por parte del empleado para denunciar el no cumplimiento de las obligaciones por parte del empleador.

SGI: El Sistema “INSPECTOR INTEGRAL 2.0”, permite **r**ealizar las visitas a las empresas anticipando avisos y evitando sanciones innecesarias, los 148 inspectores del trabajo de todo el país accedan a este sistema a través de una tablet que contendrá información importante para su labor, optimizando tiempos y resultados. (Trabajo M. d., Ministerio de Trabajo, 2016)

Consignación: Destinar algo de valor para el pago, la garantía o el depósito de alguna obligación legal. (Significados, 2018)

Actividad económica: Las actividades son aquellas acciones o procesos que llevan a cabo los individuos o las organizaciones. El adjetivo económico, por su parte, alude a lo vinculado a la economía. (Pérez, Definición.de, 2017)

Género: Se refiere a la identidad sexual de los seres vivos, la distinción que se hace entre Femenino y Masculino. (Venemedia, 2011)

Grupo ocupacional: Son categorías que permiten organizar a los servidores en razón a su formación, capacitación o experiencia reconocida. (Col, 2018)

Discapacidad: Una discapacidad es una condición que hace que una persona sea considerada como discapacitada. Esto quiere decir que el sujeto en cuestión tendrá dificultades para desarrollar tareas cotidianas y corrientes que, al resto de los individuos, no les resultan complicadas. El origen de una discapacidad suele ser algún trastorno en las facultades físicas o mentales. (Pérez, Definición.de, 2012)

Sueldo: Remuneración regular asignada por el desempeño de un cargo o servicio profesional. (Pérez, Definición.de, 2008)

Empleado: Persona que desempeña un cargo o trabajo y que a cambio de ello recibe un sueldo. (WordReference, 2018)

Ley de justicia laboral: Ley que reforma aspectos de Código de trabajo y la ley de seguridad social. (Legal, 2015)

Empleador: Persona natural o jurídica que, por su cuenta, contrata bajo su responsabilidad a un trabajador. (Trabajo M. d., Ministerio de Trabajo, 2017)

Trabajador juvenil: Persona entre 18 y 26 años que presta sus servicios. (Trabajo M. d., Ministerio de Trabajo, 2017)

Trabajo infantil: El trabajo infantil se refiere a cualquier trabajo o actividad que priva a los niños de su infancia. En efecto, se trata de actividades que son perjudiciales para su salud física y mental, por lo cual impiden su adecuado desarrollo. (Vinueza, 2018)