"Desarrollo de un sistema de Business Intelligence con software libre y datos abiertos, para la identificación de segmentos de mercado, aptos para la implementación de negocios orientados a la venta y mantenimiento de bicicletas en el perímetro urbano de Quito"

Aldas Barrera, Darwin Vinicio

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magister en Gestión de Sistemas de Información e Inteligencia de Negocios

Ing.Lascano, Jorge Edison Phd.

31 de agosto de 2021

Curiginal

Document Information

Analyzed document	02Proyecto_PRY_SEGDO_BICI_UIO_V2.pdf (D110628361)
Submitted	7/21/2021 3:53:00 AM
Submitted by	Lascano Jorge Edison
Submitter email	jelascano@espe.edu.ec
Similarity	5%
Analysis address	jelascano.espe@analysis.urkund.com

Sources included in the report

SA	**Universidad de las Fuerzas Armadas ESPE / Tesis Cueva Tandazo FINAL Rev. WFuertes 1.0.pdf** Document Tesis Cueva Tandazo FINAL Rev. WFuertes 1.0.pdf (D44954487) Submitted by: mpdiaz@espe.edu.ec Receiver: mpdiaz.espe@analysis.urkund.com	13
W	URL: http://45.238.216.28/bitstream/123456789/6512/1/TUAEXCOMMIS002-2017.pdf Fetched: 4/3/2021 7:21:58 PM	1
SA	**DM - Tesis MGS-XII.docx** Document DM - Tesis MGS-XII.docx (D9745764)	5
W	URL: https://docplayer.es/658320-Escuela-superior-politecnica-de-chimborazo-facultad-de-informatica-y-electronica-escuela-ingenieria-en-sistemas.html Fetched: 5/20/2021 12:05:21 AM	1
SA	**Fernando Rea - Proyecto de investigación.pdf** Document Fernando Rea - Proyecto de investigación.pdf (D38468490)	2
W	URL: https://jossjack.wordpress.com/tag/pentaho/ Fetched: 7/7/2020 10:27:11 PM	1
W	URL: https://dspace.uclv.edu.cu/bitstream/handle/123456789/8405/Casta%C3%B1eda%20Alarc%C3%B3n%2C%20Michel%20Alexander%20.pdf?sequence=1&isAllowed=y Fetched: 12/17/2020 2:55:39 AM	2
SA	**Proyecto Investigacion.doc** Document Proyecto Investigacion.doc (D34459800)	1
W	URL: https://expeditiorepositorio.utadeo.edu.co/bitstream/handle/20.500.12010/8529/Trabajo%20de%20grado.pdf?sequence=1&isAllowed=y Fetched: 12/4/2020 3:56:51 AM	3
W	URL: https://repositorio.unc.edu.pe/bitstream/handle/UNC/3400/TESIS%20-%20GUADA%C3%91A%20JUL%C3%93N%2C%20Britaldo.pdf?sequence=1&isAllowed=y Fetched: 5/22/2021 5:40:08 PM	1
W	URL: https://repositorio.upn.edu.pe/bitstream/handle/11537/178/Melsi%20OcasTerrones.pdf?sequence=1&isAllowed=y Fetched: 12/27/2020 10:26:08 PM	2

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, **"Desarrollo de un sistema de Business Intelligence con software libre y datos abiertos, para la identificación de segmentos de mercado, aptos para la implementación de negocios orientados a la venta y mantenimiento de bicicletas en el perímetro urbano de Quito"** fue realizado por el señor **Aldas Barrera, Darwin Vinicio** el mismo que ha sido revisado y analizado en su totalidad, por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 31 de agosto del 2021.

Ing. Jorge Edison Lascano, PhD.
Director
C.C.: 1710893114

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

RESPONSABILIDAD DE AUTORÍA

Yo **Aldas Barrera, Darwin Vinicio**, con cédula de ciudadanía n°1719279729, declaro que el contenido, ideas y criterios del trabajo de titulación*:* **"Desarrollo de un sistema de Business Intelligence con software libre y datos abiertos, para la identificación de segmentos de mercado, aptos para la implementación de negocios orientados a la venta y mantenimiento de bicicletas en el perímetro urbano de Quito"** es de mí autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 31 de agosto del 2021.

Aldas Barrera, Darwin Vinicio

C.C.: 1719279729

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORIZACIÓN DE PUBLICACIÓN

Yo **Aldas Barrera, Darwin Vinicio,** con cédula de ciudadanía n°1719279729, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **"Desarrollo de un sistema de Business Intelligence con software libre y datos abiertos, para la identificación de segmentos de mercado, aptos para la implementación de negocios orientados a la venta y mantenimiento de bicicletas en el perímetro urbano de Quito"** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 31 de agosto del 2021.

Aldas Barrera, Darwin Vinicio

C.C.: 1719279729

Dedicatoria

A mi esposa Carla Regalado, por haberme brindado su apoyo incondicional y su fortaleza, para poder alcanzar mis objetivos, a mi hija Amalia Aldas, quien con su paciencia, amor y comprensión, supo despertar en mí, la capacidad de sobrellevar los momentos más difíciles y enfrentarme a los obstáculos.

Agradecimiento

Principalmente a Dios por bendecirme y guiarme de forma espiritual, por el mejor camino hacia el éxito.

A mi director de tesis el Ing. Jorge Lascano, hombre de grandes conocimientos y excelentes valores.

A mis compañeros de aula, hombres y mujeres de buen corazón, solidarios, luchadores y estudiantes de la sexta promoción de la maestría en gestión de sistemas de información e inteligencia de negocios.

A los docentes de esta prestigiosa universidad, con quienes tuve la oportunidad de interactuar y de quienes pude aprender nuevas cosas en beneficio a mi carrera profesional.

Índice de contenido

Índice de tablas

Índice de gráficos

Resumen

En los últimos años un gran número de personas en nuestro país ha decidido implementar su propio negocio con la finalidad de estabilizar o mejorar su economía. Según (Lasio et al., 2018) en el 2017 alrededor de 3 millones de adultos en el Ecuador, empezaron el proceso de puesta en marcha de un negocio. El éxito de un emprendimiento depende de ciertos factores, que pueden ser determinantes y que en ocasiones son omitidos. La inexistencia de planes de marketing más la falta de interés en el análisis y estudio de segmentos de mercado, han ocasionado decaimiento en la productividad de muchos negocios y en algunos casos su fracaso. El presente proyecto tiene como objetivo el desarrollo de un sistema de BI alineado al giro de negocio de las bicicletas, enfocado al perímetro urbano de Quito. Esta herramienta permitirá al emprendedor, visualizar y analizar información estratégica en un dashboard (panel de control), que servirá de soporte al momento de decidir sobre un segmento geográfico o demográfico al cual atacar. La construcción del sistema se sustenta en la propuesta metodológica de Ralph Kimball, que se caracteriza por su agilidad y dinamismo. Los resultados reflejan, la existencia de áreas geográficas delimitadas por parroquias, con atributos que las vuelve aptas para la instalación de un negocio orientado al mundo de las bicicletas. Según expertos en marketing, la información reflejada en el BI apoya considerablemente al emprendedor a elegir un nicho de mercado. Ellos afirman que la herramienta puede ser el complemento de un estudio más completo, acompañado de otras técnicas de segmentación y recomiendan enriquecer el dashboard con un nuevo concepto denominado Geo BI.

Palabras Clave

- **INTELIGENCIA DE NEGOCIOS**
- **SEGMENTACIÓN DE MERCADOS**
- **MARKETIN**

Abstract

In recent years, a large number of people in our country decided to implement their own business in order to stabilize or improve their economy. According to (Lasio et al., 2018) in 2017, around 3 million adults in Ecuador began the process of starting a business. The success of an entrepreneurship depends on certain factors, these factors can be decisive but sometimes they are omitted. The lack of marketing plans and the lack of interest in the analysis of market segments have caused a decline in the productivity of many businesses and in some cases with failure. The objective of this project is to develop a BI system aligned with the bicycle business, focused on the urban perimeter of Quito. This tool will be able to the entrepreneur, visualize and analyze strategic information in a control panel, which will serve as support when deciding on a geographic or demographic segment to attack. The construction of the system is based on the methodological proposal of Ralph Kimball, which is characterized by its agility and dynamism. The results reflect the existence of geographic areas delimited by parishes, with attributes that make them suitable for the installation of a business oriented to the world of bicycles. According to marketing experts, the information reflected in the BI considerably supports the entrepreneur in choosing a market niche. The experts affirm that the tool can be the complement of a more complete study, accompanied by other segmentation techniques and recommend enriching the dashboard with a new concept called Geo BI.

Key Words

- **BUSINESS INTELLIGENCE**
- **MARKET SEGMENTATION**
- **MARKETING**

Capítulo 1

Introducción

Antecedentes

Actualmente muchas ciudades han experimentado un aumento en el número de vehículos lo cual conduce a un sin número de inconvenientes, la contaminación ambiental y auditiva es parte de la problemática que enfrenta la sociedad actual y que en cierta forma traumatiza su calidad de vida. La bicicleta se considera como una forma de movilidad amigable con el medio ambiente, saludable y con un desarrollo sostenible (Suero, 2010).

En la ciudad de Quito gracias a la gestión municipal se ha logrado incorporar una amplia red de ciclo vías las cuales acompañadas del proyecto Ciclopaseo, han motivado el uso de la bicicleta y la oportunidad de emprender en este giro de negocio, la realización del Ciclopaseo dominical apuntaló al grupo de personas que vendían bienes y servicios a los ciclistas, la venta de bicicletas aumentó y se fortaleció el mercado de la bicicleta de gama superior debido a la emergencia de deportistas salidos del ciclo paseo (Oleas & Albornoz, 2015). Hoy en día la ciudad continúa adaptando ciclo rutas y motivando el transporte no motorizado a los ciudadanos, esto implica nuevas oportunidades de aceptación para aquellos emprendedores dedicados a la venta y mantenimiento de bicicletas, quienes deberán buscar estrategias para crear o fortaleces sus negocios.

La satisfacción del cliente se ha convertido en un factor estratégico para empresas y negocios que buscan su fidelización, sin embargo, no todas las personas tienen preferencias idénticas por lo que difícilmente un producto satisface completamente a todos. La segmentación de mercado tiene como objetivo la clasificación de los consumidores (personas o empresas) en diferentes grupos según

ciertas características, necesidades o deseos. El objetivo de todo negocio es mejorar su rentabilidad buscando métodos que le permitan orientar sus productos o servicios a las personas ideales en sitios estratégicos (Camilleri, 2018). Un adecuado estudio de segmentación de mercado permite a los gerentes de marketing dividir la demanda total en segmentos relativamente homogéneos identificados por variables geográficas, demográficas, psicográficas o de comportamiento (Tynan, 1987).

Segmentación Demográfica: La segmentación demográfica implica dividir el mercado en grupos que son identificables en términos de datos físicos y fácticos. Las variables demográficas pueden incluir; género edad, ingresos, ocupación, estado civil, tamaño de la familia, raza, religión y nacionalidad (Tynan, 1987).

Segmentación Geográfica: La segmentación geográfica implica la selección de mercados potenciales de acuerdo con su ubicación. Este enfoque de segmentación puede considerar variables como el clima, el terreno, la naturaleza, recursos, densidad de población, entre otras variables geográficas (Tynan, 1987).

Segmentación Psicográfica: La segmentación psicográfica podría usarse para segmentar los mercados según la personalidad, rasgos, valores, motivos, intereses y estilos de vida. Las variables psicográficas se utilizan cuando los comportamientos de compra se correlacionan con la personalidad o estilos de vida de los consumidores (Tynan, 1987).

Adicionalmente, hoy en día los avances tecnológicos han permitido que las empresas optimicen el flujo de sus procesos mediante el uso de herramientas tecnológicas generando un gran impacto en sus objetivos estratégicos.

(Azma & Mostafapour, 2012) encontraron en su investigación que las organizaciones buscan estrategias comerciales y tecnológicas en este nuevo mundo competitivo. Consideran que la tecnología es una buena opción para mejorar los

procesos y optimizar el tiempo y los recursos. La inteligencia de negocios comprende un conjunto de estrategias y herramientas tecnológicas que son usadas por las empresas para analizar datos del negocios y tomar decisiones oportunas (Gómez & Bautista, 2010). En los últimos diez años esta herramienta se ha convertido en una buena referencia para la automatización y transformación empresarial, pues su lógica se basa en la obtención del dato, su procesamiento para transformarlo en información y el conocimiento que puede generar a partir de su análisis. Hoy en día muchas empresas se están familiarizando con el concepto de inteligencia de negocios debido a sus grandes ventajas al momento de su implementación.(Fan et al., 2015) hacen un estudio en el cual analizan el uso de inteligencia empresarial combinada con big data y minería de datos como una nueva forma de crear estrategias de marketing. Ellos exponen en su investigación la identificación de métodos y aplicaciones basados en cinco importantes perspectivas de marketing: personas, producto, lugar, precio y promoción. Los datos recolectados en grandes cantidades permitirían la construcción de sistemas de inteligencia de negocios que den soporte a la toma de decisiones. El trabajo realizado por (Fan et al., 2015) presenta una propuesta metodológica basada en ideas innovadoras. En base a este estudio, la implementación práctica de esta metodología podría otorgar valor agregado a las estrategias de marketing de las organizaciones, para así mejorar su productividad.

En Ecuador, varios negocios inician su implementación mediante la ejecución de planes de marketing y segmentación de mercado ejecutados de manera tradicional, es así como técnicas de recolección de datos manuales o encuestas ayudan a identificar un lugar físico estratégico para la instalación de una sucursal o a su vez enfocar un producto o servicio a las personas adecuadas. De igual manera, la interpretación de esta información se la realiza de manera manual, mediante la construcción de reportes

específicos en base a los datos recolectados que posiblemente se representa en una herramienta ofimática como excel o pdf's. Este proceso requiere del planteamiento de un objetivo, la generación de una muestra y el diseño de un formulario. De acuerdo a (Lasio et al., 2018), varios negocios en el Ecuador se implementan sin un estudio de mercado apropiado, y en muchos casos este proceso no se ejecuta, a pesar de ser actividades primordiales para un emprendimiento. (RACINES, 2016) por ejemplo, desarrolla un plan de marketing en el cual identifica un segmento de mercado basado en la recolección de datos mediante encuestas. Los tipos de variables incluidos en este estudio se basan en criterios geográficos, demográficos y psicográficos que tienen como objetivo ayudar a identificar las estrategias de marketing para impulsar un negocio de preparación de jugos y batidos en la ciudad de Quito. El estudio realizado fue exitoso puesto que se pudo identificar factores de gran valoración al momento de establecer estrategias para impulsar un negocio. Mediante el análisis y la interpretación de los datos recolectados se pudo conocer que la calidad del producto, el servicio brindado, la agilidad en la entrega y el grado de satisfacción de los clientes juegan un papel muy importante al momento de generar ganancias. Gracias al estudio realizado se pudo establecer estrategias de marketing orientadas a la difusión del producto rescatando asequibilidad en sus precios ya que el conjunto de clientes es diverso.

Problema de la Investigación

Contexto del Problema

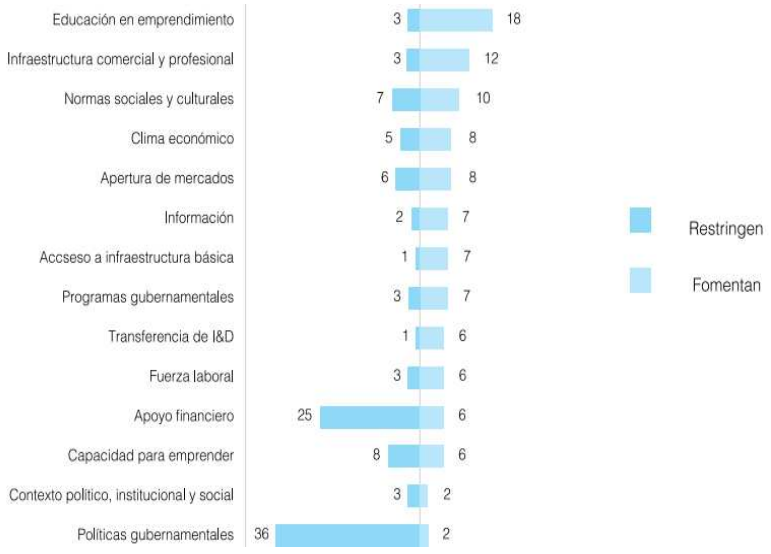
El GEM (Global Enterpreneurship Monitor) destacado por su investigación internacional acerca de las actitudes, actividades y características del emprendimiento, en su última publicación GEM ECUADOR (2017) refleja varios factores que fomentan o restringen la creación y el desarrollo de los emprendedores en el Ecuador.

Entre los factores del estudio se destaca la capacidad de emprender como una limitación basada en la poca visión global, el desconocimiento del mercado, el desconocimiento de aspectos técnicos y legales e incluso debilidades en la gestión de los nuevos negocios que limitan su crecimiento y sostenibilidad en el tiempo.

Según (Lasio et al., 2018), la capacidad de emprender es un problema asociado a condiciones marco como acceso a información disponible, educación y entrenamiento, evidenciando su incidencia más allá de las habilidades innatas del emprendedor. En el 2017, expertos afirmaron que este factor es crítico para sostener los negocios, la Figura 1 representa de manera visual los factores que influyen en el desarrollo de los emprendedores.

Figura 1

Factores que fomentan o restringen el emprendimiento



Nota: Global Enterpreneurship Monitor (GEM) 2017

La competitividad del mercado actual exige un personal, no solo ávido y lleno de ganas de emprender, sino también con un criterio lo suficientemente capacitado en términos de emprendimiento para poder enfrentarse a un mundo lleno de competencia. La segmentación de mercado forma parte de procesos estratégicos alineados al marketing y planes de negocio, que ayudan a dar un rumbo fijo a la creación de un negocio y su desarrollo económico en un ambiente adecuado.

(Balcazar, 2015) en su estudio y análisis de las PYMES en el Ecuador hace referencia a la importancia de desarrollar un adecuado plan de negocio y el valor que otorga para definir las metas y los objetivos de la empresa. El desconocimiento del mercado se ha convertido en un factor clave que influye en el fracaso de una empresa. El no saber identificar geográfica o demográficamente un nicho al cual atacar, ocasiona la generación de estrategias que no cubren la necesidad del cliente, ni se adaptan al sitio de su ubicación. Bajo este preámbulo se generan las siguientes preguntas ¿Existen herramientas capaces de ayudar a identificar geográficamente un sitio estratégico para la ubicación física de un negocio?, ¿Existen herramientas capaces de ayudar a identificar un grupo de personas como potenciales clientes para el consumo de un producto o servicio?, ¿Se puede disminuir el tiempo y los recursos necesarios para analizar e identificar un segmento de mercado?, ¿Se puede identificar segmentos de mercado en base al análisis de datos abiertos del perímetro urbano de Quito?

Planteamiento del problema

Con el propósito de mantener una estabilidad económica, muchas personas emprenden negocios para generar nuevas fuentes de ingreso, es así que gracias a la creación de proyectos de movilidad alternativos como el CicloPaseo dominical en Quito, a partir del 2003 se han generado nuevas propuestas de negocios alineados al mercado de las bicicletas debido al incremento de ciclistas y su gran demanda (Oleas &

Albornoz, 2015). Uno de los factores que influyen en el éxito al momento de desarrollar un emprendimiento es el análisis y evaluación del mercado. A pesar de su importancia, un gran número de emprendedores omiten este proceso, ya sea por desconocimiento, por complejidad, o posiblemente porque no tienen las herramientas necesarias.

El desconocimiento del mercado puede ocasionar que las empresas o negocios se instalen físicamente en sitios donde no cubren la necesidad del cliente, o que sus productos o servicios se orienten a personas que no son de interés en la perspectiva del negocio. Esto se refleja a corto o mediano plazo en pérdidas en relación a inversión de tiempo y gastos en recursos innecesarios, lo cual lleva al fracaso y en ciertos casos el cierre de sus negocios (Balcazar, 2015).

En el Ecuador, son mínimos los casos de emprendedores que logran identificar un adecuado segmento de mercado para sus negocios. Quienes lo logran, han realizado el proceso de forma tradicional mediante el uso de encuestas o han contratado el servicio de empresas especializadas con el riesgo de invertir en ocasiones una cuantiosa suma de dinero y no tener retorno de inversión. Hoy en día gracias a los avances tecnológicos y la apertura de información en la web, esta problemática podría minimizarse mediante el uso de herramientas que permitan a los emprendedores realizar un estudio previo del mercado, de tal forma que puedan aterrizar y orientar su propuesta de negocio de manera efectiva.

Objetivos

Objetivo General

Desarrollar un sistema de Bussines Intelligence basado en software libre y datos abiertos aplicado al perímetro de Quito, que ayude y facilite la toma de decisiones para el análisis e identificación de un segmento de mercado apto para la implementación de negocios orientados a la venta y mantenimiento de bicicletas.

Objetivos Específicos.

OE1: Realizar una revisión de la literatura en base a un SMS (Systematic Mapping Study) para determinar posibles soluciones y recomendaciones referentes a la identificación de segmentos de mercado mediante el uso de herramientas de BI.

OE2: Realizar una búsqueda en los diferentes portales públicos de datos abiertos, para recolectar información que ayude a identificar segmentos de mercado alineados a negocios de venta y mantenimiento de bicicletas en el perímetro urbano de Quito.

OE3: Construir un data mart mediante la metodología propuesta por Ralph Kimball para facilitar el procesamiento y análisis de los datos.

OE4: Construir un dashboard mediante el análisis de variables y la estructuración de preguntas alineadas al giro de negocio (venta y mantenimiento de bicicletas) para ayudar a la identificación un posible nicho de mercado.

OE5: Desarrollar un algoritmo de clusterización con software libre y técnicas de minería de datos para comparar los resultados del Bussines Intelligence en base a un modelo de segmentación de mercado.

Hipótesis de investigación

El desarrollo de un sistema de Business Intelligence brindará soporte a la toma de decisiones, a aquellos emprendedores que deciden dedicarse a la venta y mantenimiento de bicicletas en el perímetro urbano de Quito, mediante la identificación de un segmento de mercado apto para la implementación de este tipo de negocios.

Variable Independiente: Sistema de Business Intelligence basado en software libre y datos abiertos.

Variable Dependiente: Análisis e identificación de segmentos de mercado para negocios orientados a la venta y mantenimiento de bicicletas en el perímetro urbano de Quito.

Justificación, importancia y alcance del proyecto

Según (Lasio et al., 2018), en el Ecuador la poca visión global y el desconocimiento del mercado por parte del emprendedor, generan factores que limitan el crecimiento, sostenibilidad e incluso el éxito de negocios en el tiempo, la falta de acceso a información disponible, educación y entrenamiento en temas de marketing, planes de negocio o análisis de mercado son parámetros que aíslan la capacidad de emprender. El presente proyecto tiene la finalidad de otorgar a los emprendedores alineados a negocios de venta y mantenimiento de bicicletas, una herramienta que les permita realizar un análisis general del mercado, mediante la visualización de información comercial estratégica y que ayude a la toma de decisiones en cuanto a la definición de un posible segmento de mercado, apto para la ejecución de sus planes y estrategias de negocio. Estas herramientas son de gran importancia puesto que permiten optimizar tiempo, recursos humanos y económicos necesarios para la ejecución de estudios y análisis de segmentación de mercado. El proyecto comprende la construcción de un sistema de B.I tomando como punto de partida el análisis y evaluación de datos abiertos, la construcción de un data mart que incluye el dashboard para la visualización información estratégica y la construcción de un algoritmo de clusterización mediante técnicas de minería de datos para comprobar su efectividad.

Capítulo 2

Marco teórico y estado del arte

El presente capítulo tiene la finalidad de adoptar consideraciones teóricas que permiten sustentar el desarrollo del proyecto de una manera efectiva. Los temas citados están alineados a los objetivos generales y específicos pues son la base, en el análisis y la construcción de la solución al problema. Temas como; datos abiertos, inteligencia de negocios (B.I) y software libre son ejes fundamentales que se alinean de una u otra forma a la temática orientada al negocio como es: la segmentación de mercado, el uso e importancia de las bicicletas y su enfoque aplicado al distrito metropolitano de Quito. Adicionalmente en este capítulo, se ha incluido un estudio a la literatura, rescatando el criterio de otros autores que guardan relación con la propuesta del presente proyecto. El objetivo poder resolver el problema con bases sólidas y bien fundamentadas.

Datos Abiertos

(Dietrich et al., 2012) Los Datos Abiertos, en especial de fuentes gubernamentales son un gran recurso que todavía no ha sido explotado de forma adecuada. Muchas personas y organizaciones recogen una amplia gama de diferentes tipos de datos con el fin de utilizarlos en sus tareas. El gobierno es particularmente significativo en este sentido, no tanto por la cantidad y la centralidad de los datos que recoge, sino también porque la mayoría de los datos del gobierno por ley son públicos, y por lo tanto podrían ser abiertos y estar disponibles para que otros los puedan usar.

“Los datos abiertos pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen” (Dietrich et al., 2012).

Entre los atributos más importantes que destacan a los datos abiertos podemos destacar los siguientes:

Disponibilidad y acceso: La información debe estar disponible como un todo y de preferencia descargable de internet.

Reutilización y redistribución: Deben ser provistos bajo términos que permitan reutilizarlos y redistribuirlos.

Participación universal: Todos deben poder utilizar y redistribuir la información.

Hoy en día varias organizaciones públicas y privadas hacen uso de datos abiertos para analizar el comportamiento de su entorno, en ciertos casos la captación de este tipo de información se almacena y procesa en un data warehouse con el objetivo de generar indicadores estratégicos para la toma de decisiones.

Modelamiento Dimensional

El modelado de dimensiones se alinea a conceptos relacionados al diseño de almacenes de datos, posee las características de una base de datos relacional, debido a que se compone de claves primarias y foráneas que sirven para enlazar las tablas y generar una lógica de análisis. El modelado dimensional no está orientado a un esquema de trabajo transaccional, su objetivo se basa en la optimización entre la comunicación del usuario y la base de datos. La finalidad de este modelado es desarrollar mayor fluidez al momento de realizar consultas para ganar mayor comprensibilidad y rendimiento al momento de realizar tareas de análisis de datos.

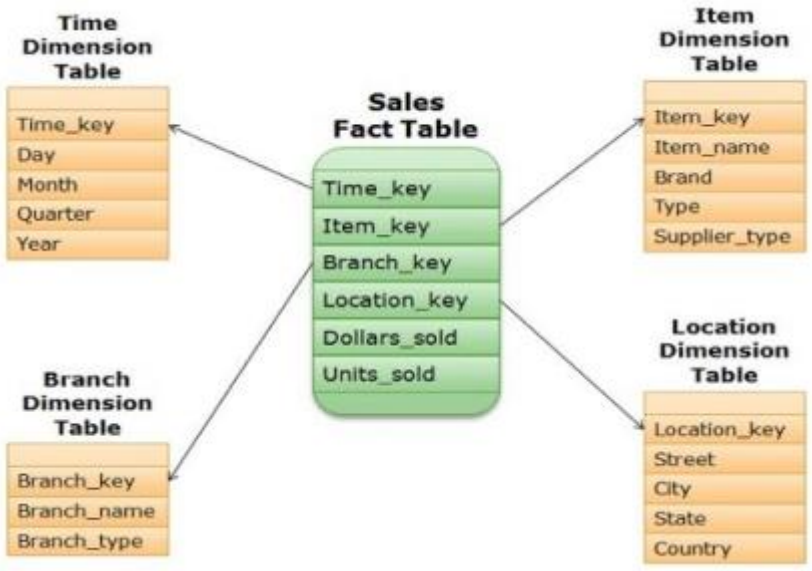
Esquema Estrella

En base al objeto al cual hace referencia por su nombre "estrella", su estructura se conforma por una o más tablas de dimensiones y una tabla de hechos central.

La tabla principal denominada de hechos contiene las claves primarias “PK” que se encuentran asociadas a las tablas de dimensiones con el nombre de las claves externas, estas claves son el canal de conectividad entre las tablas de dimensión y la tabla de hechos. Las tablas de dimensiones se caracterizan porque en ellas se almacenan datos cualitativos, mientras que la tabla de hechos contiene las medidas. El modelo estrella es des normalizado lo cual permite optimizar el rendimiento al momento de realizar consultas a la base de datos, ver Figura 2.

Figura 2

Esquema Estrella



Esquema Copo de Nieve

El esquema copo de nieve también adopta la particular forma del objeto al cual hace referencia, este esquema se caracteriza por estar conformado de tres tipos de tablas que son; a) tabla de hechos, b) tablas de dimensiones y c) tablas de sub dimensiones. Comúnmente el esquema copo de nieve, se ejecuta cuando las tablas de dimensiones son muy grandes o complejas de representar bajo un modelo estrella. Lo

distintivo de este modelo, es que las tablas de dimensiones representan relaciones normalizadas (3NF) y forman parte de un modelo relacional de base de datos, ver Figura 3.

Figura 3

Esquema copo de nieve

Data Warehouse

"Un Data Warehouse (DW) es una base de datos que almacena información para la toma de decisiones. Las características de los DWs hacen que los modelos de datos y estrategias de diseño sean diferentes a los utilizados para las bases de datos operacionales, requieren de nuevas técnicas y herramientas de diseño" (Peralta, 2015).

Según (Duque & Tamayo, 2001), el Data Warehouse es un producto resultante de la selección, clasificación y valoración de grandes cantidades de datos en base a criterios estratégicos del negocio. Estos datos en combinación con herramientas de procesamiento y visualización reflejan información que puede ayudar a la toma de

decisiones en base a la contestación de preguntas claves del giro empresarial. Las soluciones Data Warehousing han sido diseñadas para que evolucionen y se desarrollen en base a los requerimientos del negocio, donde la información varía de acuerdo a un período de tiempo. Existen diferentes metodologías que permiten la implementación de almacenes de datos, conservando una secuencia lógica alineada al soporte de los objetivos estratégicos de las organizaciones.

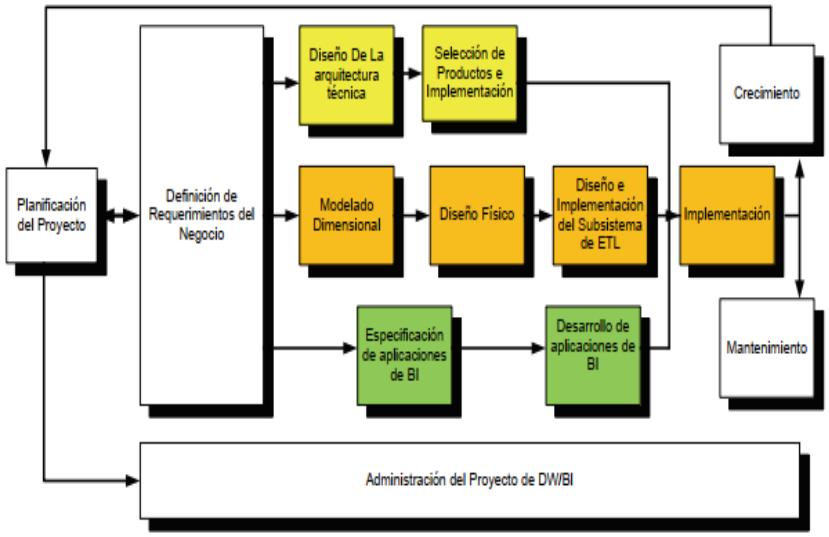
Metodología de Ralph Kimball

Según el estudio de (Rivadera, 2010), esta metodología se basa en cuatro principios básicos: 1) Centrarse en el negocio, 2) Construir una infraestructura de información adecuada, 3) Realizar entregas e incrementos significativos; y 4) Ofrecer la solución completa.

En la actualidad, la construcción de una solución de (Data warehouse/Business Intelligence) involucra un proceso complejo para lo cual, Kimball propone los siguientes pasos que ayudan a optimizar su desarrollo: La planificación, definición de requerimientos, modelo dimensional, diseño físico, desarrollo del proceso ETL y la implementación son tareas centrales que influyen de manera directa en el desarrollo del B.I. En paralelo a las fases principales se identifican actividades destinadas al análisis y evaluación de la arquitectura técnica, necesaria para la implementación del B.I. La administración del proyecto es otra fase que va de la mano con los procesos técnicos pues permite evidenciar el avance y ejecución de cada etapa. Una vez implementado el sistema de BI, el data warehouse puede crecer debido a nuevas necesidades y con él la ejecución de actividades basadas en el crecimiento y mantenimiento de la solución, ver Figura 4.

Figura 4

Metodología de Kimball



Nota:(Rivadera, 2010)

Planificación

Se caracteriza por determinar el propósito del proyecto de DW/BI, sus objetivos y el alcance. La ejecución de esta fase incluye el mapeo de actividades y tareas que permiten identificar la magnitud del proyecto, entre las características principales de la fase de planificación se encuentran las siguientes:

- Definir el alcance (entender los requerimientos del negocio).
- Identificar las tareas
- Programar las tareas
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos

Análisis de requerimientos

La definición de los requerimientos es un proceso de entendimiento y aprendizaje en el cual se realiza una interacción entre los analistas y el personal que tiene dominio en el negocio. Este proceso se puede ejecutar mediante entrevistas y tiene como objetivo captar información estratégica que ayude impulsar la creación y ejecución del proyecto. En base a las entrevistas se puede identificar temas analíticos y procesos de negocio que permitirán entender los datos y su transformación para la entrega de conocimiento. Los temas analíticos agrupan requerimientos comunes en un tema común, ver Tabla 1.

Tabla 1

Temas analíticos

Tema Analítico	Análisis o requerimiento inferido o pedido	Proceso de negocio de soporte	Comentarios
Planificación de ventas	Análisis Históricos de órdenes de revendedores	Ordenes de compras	Por cliente, por país, por región de ventas
	Proyección de ventas	Ordenes de compras	La proyección es un proceso de negocio que usa las ordenes como entradas

Nota:(Rivadera, 2010)

Diseño de la arquitectura técnica

Entorno Back Room: Hace referencia a la parte interna de la arquitectura técnica, aquí se especifican los procesos ETL que intervienen desde los orígenes de datos hasta la carga de los nuevos datos en el data warehouse.

Entorno Front Room: Es el entorno donde se presentan los datos que se obtienen del data warehouse al usuario a través de servicios de navegación, seguridad, monitoreo, generación de reportes y manejo de consultas.

Modelado dimensional

Se basa en el diseño del data warehouse bajo un modelo de alto nivel en base a la especificación de requerimientos realizada en el paso 2, este proceso consiste en la ejecución de los siguientes pasos:

Elegir el proceso de negocio: Decisión de la dirección y dependiente del análisis de requerimientos y de los temas analíticos anotados en la etapa anterior.

Establecer el nivel de granularidad: Enfocado en el nivel de detalle con el cual será desarrollado la DW. La granularidad depende de los requerimientos solicitados por el negocio. La sugerencia general es comenzar a diseñar el DW al mayor nivel de detalle posible, ya que se podrían realizar agrupamientos posteriores, al nivel deseado.

Elegir las dimensiones: Las dimensiones surgen a partir de las discusiones del equipo, las facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen un conjunto de atributos que brindan una perspectiva sobre una medida en una tabla hechos.

Identificar medidas y las tablas de hechos: Este paso, consiste en identificar las medidas que surgen de los procesos de negocios. Una medida es un atributo (campo) de una tabla que se desea analizar, sumando o agrupando sus datos y usando los criterios de corte conocidos como dimensiones. Las medidas habitualmente se vinculan con el nivel de granularidad de los datos, y se encuentran en tablas que denominamos hechos (fact en inglés). Cada tabla de hechos tiene como atributos una o más medidas de un proceso organizacional, de acuerdo a los requerimientos. Un registro contiene una medida expresada en números, como ser cantidad, tiempo, dinero, etc., sobre la cual se desea realizar una operación de agregación (promedio, conteo, suma, etc.) en función de una o más dimensiones. La granularidad, en este punto, es el nivel de detalle que posee cada registro de una tabla de hechos.

Diseño físico

Este paso hace referencia a las características técnicas de implementación del data warehouse tomando en cuenta hardware y software necesario. Los entornos de DW requieren la integración de numerosas tecnologías. Se debe tener en cuenta tres factores para poder establecer el diseño de su arquitectura técnica:

_Los requerimientos del negocio.

_ Los entornos técnicos disponibles.

_ Las directrices técnicas y estratégicas planificadas por la compañía.

La arquitectura técnica permite definir los servicios y elementos técnicos del proyecto. Se divide en dos entornos:

Diseño del sistema de Extracción, Transformación y Carga (ETL)

Este proceso es la base sobre el cual se alimenta el Data warehouse, su objetivo se centra en la extracción de los datos de los sistemas o fuentes de origen, la aplicación de ciertos parámetros para mejorar la calidad de los datos y la carga en la base de datos dimensional para luego alimentar a los sistemas de análisis

Especificación y desarrollo de aplicaciones de BI.

En esta etapa se prepara la data para su visualización a manera de información a los usuarios finales (Rivadera, 2010) define este proceso como la cara visible de la inteligencia de negocios puesto que se basa en la presentación de informes, reportes e indicadores que permiten identificar las respuestas a las peguntas del negocio.

Cubos OLAP

“Se entiende por OLAP, o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar y

manipular datos y combinaciones de los mismos a través de consultas o incluso informes" (Wrembel & Koncilia, 2007).

La tecnología OLAP permite optimizar la comunicación entre los sistemas DSS (Sistemas de soporte a las decisiones) y el data warehouse. La agilidad en la comunicación de estos dos entornos agilita el análisis de datos en línea, permitiendo la ejecución de respuestas rápidas a consultas analíticas complejas e iterativas. Los modelos de datos multidimensionales de OLAP sumado a las técnicas de agregación de datos, organizan y resumen grandes volúmenes de datos de tal manera, que puedan ser evaluados de manera instantánea. Entre las principales características de esta tecnología podemos citar las siguientes:

- Su arquitectura se basa en la creación de un modelo de datos intuitivo y multidimensional, capaz de facilitar la selección, recorrido y exploración de los datos. El modelado puede centrarse en la utilización de los esquemas conocidos en la actualidad como es el estrella y copo de nieve.
- Optimiza el rendimiento de hardware y software, gracias a su lenguaje analítico de consulta que proporciona la capacidad de explorar las complejas relaciones existentes entre los datos del negocio.
- Su estructura permite realizar un precálculo de los datos consultados con más frecuencia, de tal manera que agilita las respuestas a las consultas ad hoc.

Los cubos OLAP, vienen a ser una representación lógica del modelo dimensional físico (estrella o copo de nieve) rescatando sus principales componentes:

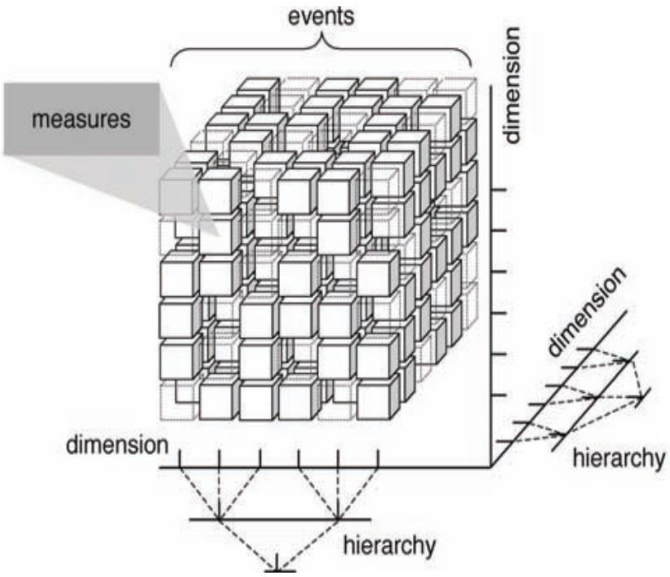
Dimensión. Perspectiva desde la que pueden ser vistos los datos del cubo. Una dimensión tiene al menos una jerarquía asociada.

Medida. Métrica del negocio que se encuentra en la intersección de las diferentes dimensiones del cubo.

Jerarquía. Estructura para navegar a través de los posibles valores de una dimensión. Se compone de diferentes niveles. La Figura 5, refleja la estructura de un cubo Olap.

Figura 5

Cubo OLAP



Nota: (Wrembel & Koncilia, 2007)

Software Libre

(Gonzales et al., 2007) , el término software libre (o programas libres), tal como fue concebido por Richard Stallman en su definición (The Free Software Foundation, 1996), hace referencia a las libertades que puede ejercer quien lo recibe, concretamente cuatro:

1) Libertad para ejecutar el programa en cualquier sitio, con cualquier propósito y

para siempre.

2) Libertad para estudiarlo y adaptarlo a nuestras necesidades. Esto exige el acceso al código fuente.

3) Libertad de redistribución, de modo que se nos permita colaborar con vecinos y amigos.

4) Libertad para mejorar el programa y publicar sus mejoras. Esto también exige el código fuente.

(Gonzales et al., 2007), en su trabajo resaltan la aparición del software libre, el cual en los últimos años ha revolucionado el mercado tecnológico debido a la oportunidad de uso e implementación de sistemas informáticos bajo un concepto gratuito. Los autores en su trabajo relacionado al análisis del software libre, realizan un estudio de las características, historia, financiamiento y otros atributos que han llevado a esta nueva tendencia de desarrollo a lograr grandes contribuciones al mundo del software.

ETL (Extracción, Extracción, Transformación y Carga de datos)

En la actualidad el procesamiento de datos se ha convertido en una actividad de gran importancia en el manejo de la información dentro de las organizaciones. ETL son las siglas en inglés de Extraer, Transformar y Cargar (Extract, Transform and Load). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos, limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para cumplir las exigencias del negocio, ver Figura 6.

Extraer: La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen, cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir bases

de datos no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

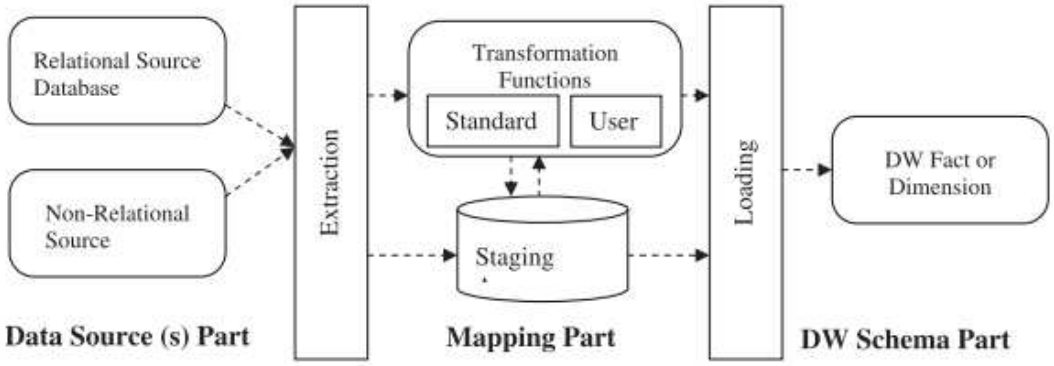
Transformar: La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos. No obstante en otros casos pueden ser necesarias aplicar algunas de las siguientes transformaciones:

- División. Que consiste en separar los datos de un campo en campos más sencillos y unitarios, es decir, en la base de datos de donde estos se obtuvieron, puede darse el caso de que, en el campo de dirección de un cliente, se incluya el teléfono de este. Durante este proceso, se separará la dirección y el teléfono en dos campos independientes.
- Corrección de errores sintácticos y comprobación de la veracidad de los datos.
- Estandarización (Standardizing) o transformación de los datos estableciendo formatos predefinidos, que faciliten el entendimiento de los datos y su procesado.
- Buscar la relación entre los datos para simplificarlos. Se suelen crear tablas nuevas con dicha información facilitando su representación.

Carga: La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos (Borja, 2018).

Figura 6

Modelado general de un proceso ETL



Nota:(El-Sappagh et al., 2011)

Software Libre Herramientas ETL

Los procesos ETL suelen ser muy complejos, sobre todo por la gran cantidad de información que existe en la actualidad, que proviene de diferentes fuentes de información con diferentes estructuras y que se intentan integrar en un entorno homogéneo. En el mercado actual existen todo tipo de herramientas especializadas, que se diferencian según el formato en el que se encuentran los datos, el objetivo perseguido, y tecnología utilizada (Borja, 2018). Algunas de las herramientas que se pueden encontrar fácilmente en la actualidad se presentan en la Tabla 2.

Tabla 2

Herramientas (ETL) de código abierto

ETL Software Libre			
Nombre	**Descripción**	**URL**	**Licencia/Modo**
Pentaho Data Integration	Integración de datos utilizando un enfoque basado en metadatos. Utiliza un entorno gráfico intuitivo. No hace falta escribir líneas de código para su utilización y dispone de plugins.	http://community.pentaho.com/projects/data-integration/	Con licencia y Versión Gratis
Talend Data Integration	Herramienta basada en Eclipse, para el proceso ETL que es uno de los procesos más importantes en la integración de datos.	https://es.talend.com/products/talend-open-studio/	Con licencia y Versión Gratis
OpenRefine	Es una poderosa herramienta para trabajar con datos desordenado, limpiándolos, y transformándolos a un formato deseado.	_http://openrefine.org/	Versión Gratis
Scriptella ETL Project	Herramienta de lanzamiento de script ETL. Utiliza sintaxis XML para sus scripts, los cuales pueden integrarse con scripts escritos en SQL, JavaScrot, JEXL, Velocity, etc. Algunas de las fuentes de entrada que acepta son LDAP, JDBC, XML, CSV, texto, entre otros.	http://scriptella.javaforge.com/	Versión Gratis
Together	Se compone de varias herramientas separadas con funcionalidades ETL. Están desarrolladas en código Java y soporta la conexión con diferentes tipos de bases de datos (MSSQL, Oracle, DB2, QED, JDBC, MySQL,…) y acepta como entrada varios tipos dearchivos (CSV, XML,…). Algunas herramientas son:TDC – Together Document Converter, TDT – Together Data Transformer, TXE – Together XML Extractor.	http://www.together.at/download	Versión Gratis
CloverETL Community Edition	Es una herramienta muy gráfica que permite varios tipos de transformaciones, así como diversos tipos de entrada y salida de datos, como son los procedentes de las BBDD MySQL, PostgreSQL, SQLite, MSSQL, Oracle, Sysbase y Derby, archivos CSV, XML, etc. Cuenta con versiones de pago que permiten muchas más opciones (clasificación, clusters).	*http://www.cloveretl.com/products/community-edition*	Versión Gratis

ETL Software Libre			
Nombre	**Descripción**	**URL**	**Licencia/Modo**
Apatar	Usa interfaz gráfica de trabajo mediante la cual se puede hacer el filtrado, la validación y la planificación de los datos. Los conectores incluyen MySQL, PostgreSQL, Oracle, MSSQL, Sybase, FTP, HTTP, SalesForce.com, SugarCRM, Compiere ERP, CRM Goldmine, XML, archivos planos, WebDAV, Buzzsaw, LDAP, Amazon y Flickr.	http://www.apatar.com/	Versión Gratis
Jaspersoft ETL	Herramienta basada en Eclipse, para el proceso ETL que es uno de los procesos más importantes en la integración de datos. Incluye flujos y procesa diferentes tipos de archivos. Fácil de desplegar.	*http://community.jaspersoft.com/project/jaspersoft-etl*	Versión Gratis
Data Pipeline	Transforma datos y los procesa. Puede leer y escribir archivos de tipo CSV, Excel, JDBC, JSON.	http://northconcepts.com/datapipeline/	Versión Gratis
KETL	Está basado en java. Incluye gestión de Jobs y alertas. Es capaz de gestionar varios hilos a la vez. Los Jobs están definidos en XML.	http://www.ketl.org/	Versión Gratis

Inteligencia de Negocios

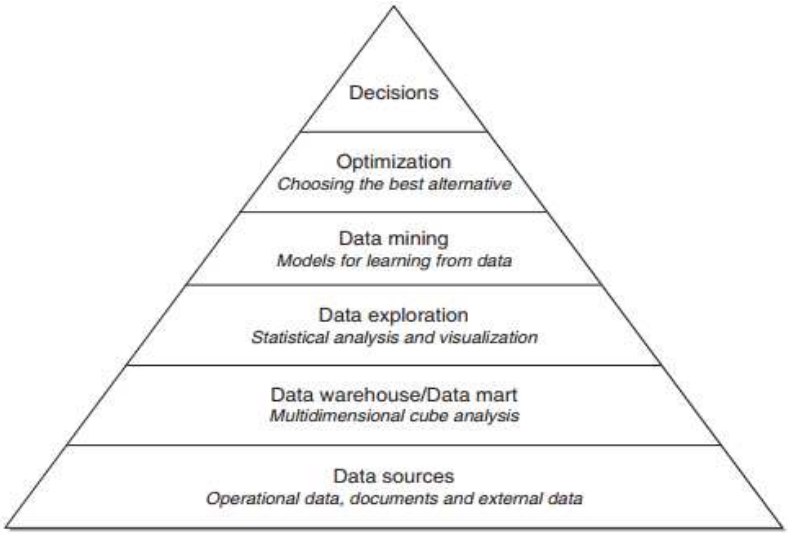
Conocimiento: se forma a partir de información que se utiliza para tomar decisiones y desarrollar las acciones correspondientes (Heang & Mohan, s. f.).

Aina et.al., enfocados en el uso de Bussines Intelligence como una revisión sistemática de la literatura, reflejan la importancia de usar este tipo de tecnología en los últimos diez años y la capacidad de integrar y analizar grandes cantidades de datos para comprender sus oportunidades, fortalezas y debilidades. Este tipo de sistemas se consideran estratégicos en las organizaciones. La distribución de los estudios seleccionados con respecto a los diferentes sectores / industrias mostró que la mayor parte de la investigación del sistema de BI se realizó en los sectores de servicios gubernamentales, transporte, seguros, comunicaciones, atención médica, banca, agricultura, construcción y servicios profesionales.(Ain et al., 2019)

La Figura 7, refleja los principales componentes que intervienen en un sistema de BI.

Figura 7

Principales componentes de un sistema de inteligencia de negocios



Nota:(Vercellis, 2009)

Bicicletas

"La Bicicleta es un vehículo que consta de dos ruedas alineadas fijas a un cuadro, se dirige mediante un manillar y es impulsada por una combinación de pedales y engranajes movidos por los pies" (Suero, 2010).

Suero destaca la importancia de incentivar el cuidado del medio ambiente y la adopción de hábitos saludables para el ser humando y propone un modelo de gestión para el transporte en bicicleta. Este modelo estudia factores como:

_Gestión política

_Organización privada

_Infraestructura y seguridad

_Promoción y formación.

La adecuada gestión de estos cuatro factores enfatiza la creación de una nueva propuesta de movilidad saludable, económica y ecológica.

Fuentes de datos abiertos y de acceso público en Ecuador

El Plan Nacional de Gobierno Electrónico 2014-2017 de Ecuador (http://www.gobiernoelectronico.gob.ec) articula 12 principios y un modelo compuesto de tres objetivos, 11 estrategias, cuatro pilares, personas, marco regulatorio, procesos y servicios, y TIC. Ecuador al igual que muchos países ha desarrollado un enorme trabajo en cuanto a la utilidad de la información y los datos. La iniciativa de apertura de datos surge desde el ámbito europeo para concienciar a los servidores públicos del valor de la información que se maneja dentro de la administración en función de la transparencia, la participación y la colaboración ciudadana.

Principios sobre datos abiertos y su liberación en Ecuador:

1. Dato Completo
2. Dato Primario
3. Dato Oportuno
4. Dato Accesible
5. Dato procesable por una máquina
6. Dato Inclusivo
7. Dato en formato abierto
8. Dato con Licencia Libre

Entre los usos de buenas prácticas se encuentran los cuatro pilares del Modelo de Gobierno Electrónico descrito en el Plan nacional Electrónica:

1. Personas

2. Marco Regulatorio

3. Servicios y Procesos

4. Tecnologías de la Información.

Cada uno cumple con objetivos específicos que ayuda a optimizar el proceso de tratamiento de datos abiertos en las organizaciones públicas (*GUÍA DE POLÍTICA PÚBLICA DE DATOS ABIERTOS*, 2014).

Datos de interés para el giro de negocio: (Uso, Venta y Mantenimiento de bicicletas) en el Distrito Metropolitano de Quito

Oleas y Albornoz desarrollaron un estudio a la histórica transformación del espacio público de la ciudad de Quito como una alternativa de inclusión para una nueva propuesta de movilidad saludable y amigable con el medio ambiente. Los autores destacan la gestión política, económica y social que ha trascendido en la implementación de organizaciones que actualmente fomentan el uso de la bicicleta, así como también las principales vías y espacios públicos creados para quienes prefieren el uso de este medio de transporte. El estudio refleja la creciente aceptación de una nueva propuesta de movilidad en dos ruedas y los constantes cambios a nivel de infraestructura que ha sufrido Quito para poder incluir de manera segura esta nueva faceta de los quiteños (Oleas & Albornoz, 2015).

(Diario La Hora, 2020), la capital tiene previsto ampliar sus ciclo vías a 120 kilómetros, de los 64 que tenía antes de la pandemia. La Alcaldía estima que hasta antes de la pandemia, casi el 70% de la población de Quito se movilizaba en transporte público, pero la suspensión de este servicio al inicio de la cuarentena alentó a que muchas personas buscaran la opción de las dos ruedas.

La ruta ciclística se iniciará desde Carapungo y Calderón, en el norte, hasta Quitumbe en el sur, y con ayuda del sector privado se construirán estacionamientos de corta distancia sin costo, en varios puntos de la ciudad.

Según Fernando De La Torre, Director Metropolitano Modos de Transporte Sostenible, el uso de bicicleta aumentó al menos un 600% en Quito. La falta de transporte público durante la cuarentena y el miedo a contraer covid-19 motivaron a los quiteños a movilizarse en este vehículo(Diario Primicias, 2020).

Segmentación de Mercado

Gichuru afirma en su estudio que la segmentación es muy importante para que los jugadores de la industria logren la satisfacción y retención del cliente. La segmentación del mercado implica ver un mercado heterogéneo como un número de mercados homogéneos más pequeños en respuesta a las diferentes preferencias, atribuibles a los deseos de los clientes para obtener satisfacciones más precisas de sus necesidades variables.(Gichuru & Limiri, s. f.).

(Tynan, 1987) afirma que el análisis de un adecuado segmento de mercado se basa en las siguientes bases:

1. Bases geográficas en las que los mercados se dividen en unidades geográficas.

2. Las bases demográficas incluyen estudios de segmentación basados en edad, sexo, grupo socioeconómico, tamaño de la familia, ciclo de vida, ingresos, ocupación, educación, etc.

3. Bases psicológicas en las que se utilizan factores de personalidad, actitudes, riesgos, motivación, etc. para dividir el mercado.

4. Las bases psicográficas incluyen estilo de vida, actividades, intereses, opiniones, necesidades, valores y similares como delineadores del mercado.

5. Las bases de comportamiento incluyen lealtad a la marca, tasa de uso, beneficios buscados.

Minería de datos

Metodología CRISP-DM

La metodología de minería de datos de IBM llamada CRISP-DM, que son las siglas de Cross-Industry Standard Process for Data Mining es un referente al momento de desarrollar proyectos orientados al análisis de datos. A continuación se describe de manera general las fases y actividades de la metodología (IBM, 1994), ver Figura 8.

Comprensión del negocio

Entre las características básicas de esta fase se encuentra la determinación de los objetivos del negocio, su evaluación, determinación del objetivo de la minería de datos y el desarrollo del plan.

Comprensión de Datos

Las actividades a resaltar en esta fase son; la obtención de los datos, su descripción, exploración y verificación de calidad.

Preparación de Datos

La preparación consiste en la selección de los datos con aporte estratégico para el proyecto, la ejecución de procesos de limpieza, su integración y finalmente la inclusión de formatos adecuados para su procesamiento.

Modelamiento

Dependiendo del modelo a utilizar las actividades de esta fase hacen referencia a la selección técnica de modelamiento, diseños de prueba, la construcción del modelo y su evaluación.

Evaluación

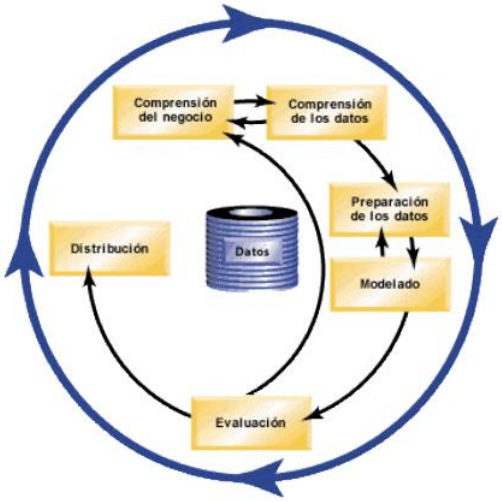
A diferencia de la evaluación ejecutada en la fase de modelamiento, en esta fase la evaluación se orienta a los resultados generados a partir de la ejecución del modelo.

Despliegue

Finalmente en esta fase las actividades a ejecutar se orientan al monitoreo y mantenimiento del proyecto de minería, generación de reportes y su respectiva documentación.

Figura 8

Metodología de minería de datos CRISP-DM



Nota: Manual CRISP-DM de IBM SPSS Modeler

Algoritmo de minería de datos K-means

El algoritmo K-means, es un método de clustering conocido por su efectividad y dinamismo, tiene como objetivo asignar cada punto al clúster cuyo centro (también llamado centroide) es el más cercano. El centro es el promedio de todos los puntos en el grupo, es decir, sus coordenadas son la media aritmética de cada dimensión por separado sobre todos los puntos en el grupo. Su lógica se basa en la asignación estricta de información a un determinado conjunto de particiones. El valor de los datos se asigna a la partición más cercana en función de algunos parámetros de similitud, como

la distancia euclidiana de intensidad. Posteriormente, las particiones se recalculan en función de estas asignaciones. Con cada pasada sucesiva, un valor de datos puede cambiar particiones, alterando así los valores de las particiones en cada paso. Los algoritmos de K-Means suelen converger en una solución muy rápida a diferencia de otros clústeres algoritmos.

Los pasos del algoritmo son los siguientes

- Elegir el número de conglomerados, k.
- Generar aleatoriamente k conglomerados y determinar los centros de conglomerados, o generar directamente k puntos aleatorios como centros de clúster.
- Asignar cada punto al centro del grupo más cercano.
- Volver a calcular los nuevos centros de clústeres.
- Repetir los dos pasos anteriores hasta que alguno se cumpla (Nathiya & Punitha, 2010).

Trabajos relacionados (estado del arte)

El análisis sobre el estado del arte fue desarrollado mediante la inclusión de la estrategia de búsqueda en base a un SMS (Systematic Mapping Study) del cual se usaron como fuente los repositorios académicos: Springer Link, ScienceDirect y IEEE Xplore.

Definición del objetivo: El objetivo del estudio del estado del arte está enfocado en resolver las preguntas de los objetivos específicos 3.3.2.

Definición de los criterios de inclusión y exclusión: Las búsquedas en las bases digitales dependiendo del tema retornan una gran cantidad de artículos relacionados por lo cual es importante definir las características idóneas de los artículos

que serán tomados en cuenta para su análisis. Las características propuestas para el análisis de los artículos son las siguientes:

Criterios de Inclusión

Con el fin de analizar referencias teóricas con un enfoque actual, se tomarán en cuenta artículos publicados a partir del 2000.

- Se han valorado los artículos científicos publicados en el idioma inglés.
- Artículos que contengan información referente al cuidado del medio ambiente a través de un adecuado proceso de gestión de residuos sólidos alineados a la tecnología.
- Artículos que tengan temas relacionados al uso de sistemas de información geográfica en la optimización de la gestión de residuos sólidos.

Definición de la estrategia de búsqueda

Revisión Inicial: Se realiza una búsqueda inicial en los repositorios académicos elegidos para buscar estudios relacionados con las preguntas de investigación planteadas.

Validación cruzada de estudios: En esta fase se procede a verificar que los estudios cumplan con los criterios de inclusión y exclusión, el objetivo de este proceso es obtener el listado inicial de documentos académicos con los cuales se va a trabajar en las siguientes fases del estudio.

Integración del Grupo de Control: El grupo de control está conformado por los estudios que cumplen con los criterios de inclusión y exclusión. Es importante realizar un análisis inicial del título de los estudios, introducción, conclusiones y palabras claves. Los estudios seleccionados para el grupo de control se reflejan en la Tabla 3.

Tabla 3

Estudios por grupo de control

GRUPO DE CONTROL	TÍTULO	PALABRAS CLAVE
EC1	Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix	Business Intelligence ,Marketing
EC2	APPLICATION OF BUSINESS INTELLIGENCE TO SUPPORT MARKETING STRATEGIES: A CASE STUDY APPROACH	Business Intelligence, Marketing Strategies, Application
EC3	Market Segmentation Analysis and Visualization Using K-Mode Clustering Algorithm for E-Commerce Business	Market Segmentation, data visualization,
EC4	A DECISION SUPPORT SYSTEM FOR MARKET SEGMENT DRIVEN PRODUCT DESIGN	Market Segmentation, decision support, system
EC5	Business intelligence: The role of the internet in marketing research and business decision-making	Business Intelligence, marketing, business decision
EC6	USING DATA MINING FOR AN INTELLIGENT MARKETING CAMPIAN	Marketing campaign, intelligent, data mining

Construcción de la cadena de búsqueda

Para la construcción de la cadena de búsqueda se usan las palabras que más se repiten en cada contexto, definido a partir de los estudios del grupo de control como se indica en la Tabla 4.

Tabla 4

Cadena de Búsqueda

PALABRA CLAVE	EC1	EC2	EC3	EC4	EC5	EC6	NUMERO DE REPETICIONES
Business Intelligence	X	X	X		X	X	5
Market Segmentation			X	X	X	X	4
Marketing	X	X	X	X	X	X	6
Decision Support		X		X			2
Strategies	X	X	X	X		X	5
System	X	X	X	X	X	X	6

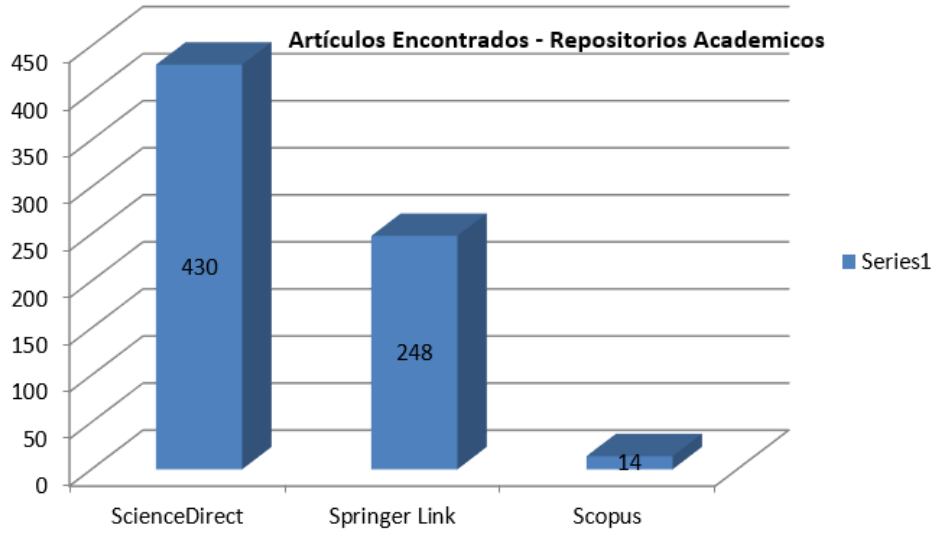
La cadena de búsqueda está formada por la unión de las palabras claves que más se repiten en cada contexto, los conectores usados son OR para las palabras que están dentro del mismo contexto y el conector AND para las palabras que están en contextos distintos, de esta manera se establece el siguiente resultado:

((Business Intelligence) or (Intelligent Decision Support) and (System)) and ((Market Segmentation) or (Marketing Segmentation)) and (Strategies)

Al aplicar la cadena de búsqueda en los dos repositorios académicos seleccionados se encontraron los siguientes resultados, ver Figura 9.

Figura 9

Resultados - cadena de búsqueda



Una vez obtenidos los resultados se realizó la revisión de los documentos citados a continuación:

Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix

Shaokun Fan Raymond Y.K. Lau, J. Leon Zhao Fan, Lau & Zhao realiza un análisis sobre el uso de la inteligencia empresarial combinada con big data y minería de datos como una nueva forma de crear estrategias de marketing. Exponen en su estudio la identificación de métodos y aplicaciones identificados en cinco importantes perspectivas de marketing: personas, producto, lugar, precio y promoción. Los datos recopilados en base a los cinco criterios combinados con el procesamiento de grandes cantidades de datos permiten la construcción de sistemas de inteligencia empresarial que apoyan la toma de decisiones (Fan et al., 2015).

APPLICATION OF BUSINESS INTELLIGENCE TO SUPPORT MARKETING STRATEGIES: A CASE STUDY APPROACH

Kurniawae et.al., proponen un estudio de caso con implementación de B.I. como una estrategia para tomar las mejores decisiones en el negocio, consideran que el análisis de productos, precios, promociones y variables de lugares son clave para crear planes de marketing. El estudio de caso se aplica en un restaurante en el cual desarrollan un marco de arquitectura para la inteligencia empresarial en sistemas de información que incluye herramientas etl, almacenamiento de datos, minería de datos y análisis e informes (Kurniawan et al., 2005).

Market Segmentation Analysis and Visualization Using K-Mode Clustering Algorithm for E-Commerce Business

Deepali Kamthania, Ashish Pahwa and Srijit S. Madhavan Kamthania et.al, hacen una segmentación del mercado utilizando técnicas de minería de datos y métodos de algoritmos de agrupamiento con datos de la web. Crean algunos grupos de personas basados en un comportamiento similar e información geográfica con el algoritmo de K-mode para identificar la posible segmentación del mercado. Aplican el mismo ejercicio con varios tipos de productos con el objetivo de definir el nivel de popularidad y conectarse con los clientes para la expansión del negocio.(Kamthania et al., 2018)

A DECISION SUPPORT SYSTEM FOR MARKET SEGMENT DRIVEN PRODUCT DESIGN

Ningrong LEI, Seung Ki MOON Lei y Moon desarrollaron un caso de estudio para construir un Sistema de Apoyo a la Decisión (DSS) con datos del mercado automotriz, aplican la minería de datos con métodos de algoritmos de agrupamiento para identificar un grupo de modelos y marcas de automóviles que permiten establecer

la segmentación del mercado. Descubrieron que, con un conjunto completo de parámetros del producto que el sistema propuesto puede lograr una precisión de clasificación del 92,40%. Otro resultado notable fue que la precisión de la clasificación aumentó a 93.83% cuando el precio se eliminó de la lista de parámetros de entrada, concluyendo que en este DSS el precio no es un buen indicador del segmento de mercado (Lei & Ki Mon, 2013).

Business intelligence: The role of the internet in marketing research and business decision-making

Ivana Kursan, Mirela Mihić Kursan y Mihic, hacen un estudio a partir de las ventajas sobre Internet y la gran cantidad de datos que podemos encontrar en esta gran fuente de información, en la actualidad internet representa un medio eficiente para la comunicación con los usuarios y una gran oportunidad para saber lo que quieren. Hacen referencia a técnicas de recopilación, procesamiento y análisis aplicados a sistemas de B.I como una manera óptima de mejorar las estrategias de marketing y la decisión comercial (Kursan & Mihi, 2010).

USING DATA MINING FOR AN INTELLIGENT MARKETING CAMPIAN

Shorouq Fathi Eletter Eletter desarrolla un estudio basado en métodos de minería de datos para identificar un modelo de segmentación de clientes utilizando un mapa auto organizado (SOM). El modelo desarrollado pudo clasificar el conjunto de datos en dos grupos. Los resultados mostraron que SOM es una técnica que permite a los usuarios visualizar con éxito datos multidimensionales en dos dimensiones y poder ver y comprender la relación entre las variables en el conjunto de datos y su efecto en los grupos fácilmente (Eletter, s. f.).

Oleas y Albornoz desarrollan un estudio a la histórica transformación del espacio público de la ciudad de Quito como una alternativa de inclusión para una nueva

propuesta de movilidad saludable y amigable con el medio ambiente. Los autores destacan la gestión política, económica y social que ha trascendido en la implementación de organizaciones que actualmente fomentan el uso de la bicicleta, así como también las principales vías y espacios públicos creados para quienes prefieren el uso de este medio de transporte. El estudio refleja la creciente aceptación de una nueva propuesta de movilidad en dos ruedas y los constantes cambios a nivel de infraestructura que ha sufrido Quito para poder incluir de manera segura esta nueva faceta de los quiteños (Oleas & Albornoz, 2015).

(Diario La Hora, 2020) La capital tiene previsto ampliar sus ciclovías a 120 kilómetros, de los 64 que tenía antes de la pandemia. La Alcaldía estima que hasta antes de la pandemia, casi el 70% de la población de Quito se movilizaba en transporte público, pero la suspensión de este servicio al inicio de la cuarentena, alentó a que muchas personas buscaran la opción de las dos ruedas.

La ruta ciclística se iniciará desde Carapungo y Calderón, en el norte, hasta Quitumbe en el sur, y con ayuda del sector privado se construirán estacionamientos de corta distancia sin costo, en varios puntos de la ciudad.

Según Fernando De La Torre, Director Metropolitano Modos de Transporte Sostenible, el uso de bicicleta aumentó al menos un 600% en Quito. La falta de transporte público durante la cuarentena y el miedo a contraer covid-19 motivaron a los quiteños a movilizarse en este vehículo(Diario Primicias, 2020)

Capítulo 3

Aplicación de metodología y desarrollo de la solución

En este capítulo se detalla paso a paso el proceso de desarrollo de la solución del Sistema de Business Intelligence en base a la metodología propuesta por Ralph Kimball.

Planificación del proyecto

Alcance, tareas, programación de las tareas, planificación del uso de recursos, asignación carga de trabajo.

Definición del proyecto

El desarrollo de este sistema nace de la necesidad de otorgar a los emprendedores con afinidad al mercado de bicicletas, una herramienta que les permita visualizar información general y estratégica, de un posible segmento de mercado al cual dirigirse en el perímetro urbano de la ciudad de Quito.

Justificación y Objetivos

En términos generales, la solución facilitará la toma de decisiones para el análisis e identificación de un segmento de mercado, apto para la implementación de negocios orientados a la venta y mantenimiento de bicicletas en el perímetro urbano de Quito. El capítulo 1 muestra información detallada en cuanto a objetivos generales, específicos su justificación e importancia del proyecto.

Alcance

La metodología de Kimball establece en la fase de planificación, la identificación de tareas a realizar, su programación y los recursos necesarios para ejecutarlas. El alcance del proyecto SEGDO_BICI_UIO (Segmentación de Mercado para bicicletas Quito) se orienta a aquellos emprendedores que inician sus negocios sin realizar un análisis y evaluación de su entorno, como proceso previo a la ubicación de sus

establecimientos. El producto resultante del sistema de BI es un dashboard que permitirá analizar la información de personas por edades, género, o nivel de estudio que usan bicicletas o realizan actividades físicas en un espacio geográfico desagregado a nivel parroquial. En la Tabla 5, se puede apreciar el desglose del trabajo basado en las actividades y tareas que contempla la ejecución del proyecto.

Tabla 5

Actividades del proyecto SEGDO_BICI_UIO

Actividad	Responsable	Rol	Tareas	Tiempo Meses					
				M1	M2	M3	M4	M5	M6
Análisis de Requerimientos	Darwin Aldas	Analista de Datos	_Realizar Entrevistas	X					
			_Desarrollar documento de requerimientos	X					
			_Identificar fuentes de datos abiertos	X					
			_Establecer fuente de datos	X					
			_Analizar y Evaluar Variables	X					
Generación del modelado dimensional	Darwin Aldas	Arquitecto de Datos	_Elegir proceso del negocio		X				
			_Establecer nivel de granularidad		X				
			_Elegir las dimensiones		X				
			_Identificar medidas y tablas de hecho		X				
			_Diagrama Lógico y Físico		X				

Actividad	Responsable	Rol	Tareas	Tiempo Meses					
				M1	M2	M3	M4	M5	M6
Diseño de la arquitectura física	Darwin Aldas	Arquitecto de Datos	_Diseñar arquitectura (Entorno Back Room)			X			
			_Diseñar arquitectura (Entorno Front Room)			X			
Desarrollo del proceso ETL	Darwin Aldas	Arquitecto de Datos	_Generar procesos (ETL), extracción, transformación y Carga.				X	X	
			_Realizar pruebas				X	X	
Desarrollo del Dashboard	Darwin Aldas	Analista de Datos	_Desarrollar panel de control (Dashboard)						X

Análisis de requerimientos

El desarrollo del sistema de Business Intelligence es una propuesta tecnológica orientada a satisfacer la necesidad de aquellos emprendedores cuyo giro de negocio es el mercado de bicicletas. El requerimiento general se basa en la posibilidad de poder identificar este segmento de mercado en la ciudad de Quito, mediante la captación, procesamiento y análisis de datos alojados en fuentes abiertas. La lógica del proyecto no está orientada a satisfacer la necesidad de una empresa o negocio (patrocinador) en específico, sin embargo se han realizado entrevistas a dueños de negocios orientados al mercado de las bicicletas para conocer su criterio y posibles preguntas estratégicas que permitan desarrollar indicadores mediante el uso de datos abiertos.

Entrevistas

Para alinear la propuesta a la metodología y establecer un adecuado proceso en la fase de requerimientos se realizó una entrevista a dos dueños de negocios

destinados a la venta, mantenimiento y reparación de bicicletas. El objetivo es saber si los actuales emprendedores realizaron algún estudio de mercado antes de establecer su negocio y si los datos que conformaran el sistema BI pueden considerarse importantes al momento de escoger un nicho de mercado al cual atacar. Los dos entrevistados coinciden en que la ubicación geográfica es importante, también resaltan el hecho de saber si existen personas que utilizan bicicletas o realizan actividades físicas. A pesar que ninguno de los entrevistados realizó un análisis de mercado para sus negocios, ambos consideran que el tener una herramienta informática que facilite la toma de decisiones al momento de escoger un nicho de mercado sería de gran ayuda.

El detalle de las entrevistas se encuentra en el Anexo 1.

Documentación de Requerimientos

En base a la estructura propuesta por (Rivadera, 2010) para identificar un documento de requerimientos y una vez realizadas las entrevistas se detectaron algunos temas analíticos claves para la identificación de segmentos de mercado, la Tabla 6 contiene los temas que pueden ser analizados en base a los datos abiertos provenientes de la encuesta.

Tabla 6

Requerimientos agrupados por temas analíticos

Ítem	Tema Analítico	Análisis o Requerimiento Inferido o Pedido	Proceso de Negocio de Soporte	Comentarios
1	Uso de la bicicleta	_Obtener el número de personas que usan bicicletas	Variables de la encuesta que hacen referencia a la actividad física y transporte.	Analizar las parroquias donde la gente utiliza bicicleta, frecuencia de uso y su combinación con las siguientes variables: _Edad _Genero _Sexo _Nivel de Instrucción
2	Transporte (La bicicleta como medio de desplazamiento)	_Obtener el número de personas que en su rutina diaria, utilizan la bicicleta para TRASLADARSE desde su hogar al trabajo, establecimiento educativo etc.	Variables de la encuesta que hacen referencia a la actividad física y transporte.	Analizar las parroquias donde la gente utiliza bicicleta para transportarse en su rutina diaria, saber cuáles son las razones por la cual hacen uso de este medio y su combinación con las siguientes variables: _Edad _Genero _Sexo _Nivel de Instrucción
3	Actividad Física (Ejercicio y Deporte)	Obtener el número de personas que realiza ejercicios o práctica un deporte	Variables de la encuesta que hacen referencia a la actividad física y transporte.	Analizar las parroquias donde la gente hace actividad física el tiempo que invierte en estas actividades y su combinación con las siguientes variables: _Edad _Genero _Sexo _Nivel de Instrucción

El objetivo general de un sistema de BI es transformar los datos en información y la información en conocimiento para optimizar los procesos de un negocio o empresa que cuenta con sus propios datos. El proyecto SEGDO_BICI_UIO (Segmentación de mercado para bicicletas Quito) no se basa en atender las especificaciones o requerimientos de una empresa específica, se fundamenta en la búsqueda, obtención y procesamiento de datos, captados en fuentes abiertas que permitan alcanzar su objetivo, dado que los datos equivalen a la materia prima del producto, se ha establecido las siguientes actividades para su adecuado tratamiento: 1) identificación de la fuente de datos (datos abiertos), 2) determinación de la fuente de datos y 3) análisis de los datos.

Identificación de la fuente de datos (Datos Abiertos)

La identificación de la fuente de datos es una actividad de gran importancia puesto que toda la materia prima (dato) será obtenida del repositorio identificado. Es importante saber que no todos los datos recolectados del mundo digital (internet) son considerados datos abiertos con calidad, existe en la actualidad. El Plan Nacional de Gobierno Electrónico en el Ecuador que a partir del año 2014 ha articulado principios y modelos para el uso de este tipo de información, estos principios permiten identificar una fuente de calidad y un adecuado manejo del proyecto

Los principios sobre datos abiertos y su liberación en Ecuador que se seguirán/aplicarán en este proyecto son: 1.Dato Completo, 2.Dato Primario, 3.Dato Oportuno, 4.Dato Accesible, 5.Dato procesable por una máquina, 6.Dato Inclusivo, 7.Dato en formato abierto, 8.Dato con Licencia Libre.

El Instituto Nacional de Estadísticas y Censos (INEC) como institución responsable de la estadística oficial es la entidad encargada de planificar, normar y

certificar la producción del Sistema Estadístico Nacional. Una de sus políticas es la difusión de la información estadística en forma oportuna a través de medios impresos y magnéticos a personas o entidades públicas y privadas a nivel nacional o internacional.

La institución en su afán de acatar sus políticas y los principios sobre datos abiertos ha desarrollado y publicado un repositorio digital en el cual se puede acceder a toda la información estadística generada en el INEC. En este repositorio se puede encontrar bases de datos, metodologías, sintaxis y tabulados de las operaciones estadísticas en formatos abiertos.

EL link de la fuente de datos abiertos es el siguiente: Instituto Nacional de Estadísticas y Censos. http://aplicaciones3.ecuadorencifras.gob.ec/BIINEC-war/

Determinación de la fuente de datos

El Instituto Nacional de Estadísticas y Censos (INEC) es una de las entidades que provee una gran fuente de datos aplicados a varios temas de interés, en sus portales de uso público se encuentra disponible información de ámbito económico, ambiental, social entre otros. Para satisfacer el objetivo del proyecto se ha investigado datos sociodemográficos y sociales que contengan variables afines al aprovechamiento del tiempo en actividades físicas y el uso de las bicicletas.

En el 2018 se ejecuta la Encuesta Nacional Multipropósito de Hogares como una respuesta a la necesidad de contar con una fuente de información para calcular los indicadores del Plan Nacional de Desarrollo (PND 2017-2021).

La encuesta multipropósito se alinea a tres ejes, cuatro objetivos y 21 indicadores del Plan Nacional de Desarrollo, en el análisis de la encuesta se valida y referencia al objetivo 1 como el que mejor aporta a la finalidad del proyecto, la Tabla 7, refleja sus características.

Tabla 7

Alineación de la encuesta multipropósito alineado al plan nacional de desarrollo

(Objetivo 1)

Contribución o alineación al Plan Nacional de Desarrollo			
Eje del Plan Nacional de Desarrollo	**Objetivo del eje**	**Meta**	**Indicador**
Derechos para todos durante toda una vida	Objetivo 1: Garantizar una vida digna con iguales oportunidades para todas las personas.	Aumentar la cobertura, calidad, y acceso a servicios de salud: Incrementar el porcentaje de percepción positiva de los hogares en relación con los servicios públicos de salud, de calidad al 2021.	Porcentaje de hogares con percepción positiva en relación a los servicios de salud pública
		Aumentar la cobertura, calidad, y acceso a servicios de salud: Incrementar el porcentaje de percepción positiva de los hogares en relación a servicios públicos de salud, de calidad al 2021	
		Incrementar de 12,2% a 14,4% la población mayor a 12 años que realiza más de 3.5 horas a la semana de actividad física al 2021	Porcentaje de la población de 12 y más años de edad, que realiza ejercicio o deporte en su tiempo libre más de 3.5 horas a la semana

Nota: Instituto Nacional de Estadística y Censos INEC

Análisis de los datos

El Instituto Nacional De Estadísticas y Censos (INEC) es la principal fuente de datos para el presente análisis y su temática está basada en la encuesta multipropósito de temas sociodemográficos de hogares. Las encuestas bajo una lógica de proyecto

son conocidas en el INEC como operaciones estadísticas y se desarrollan metodológicamente en base al Modelo de Producción Estadística el cual garantiza la calidad de sus productos, los objetivos de la encuesta multipropósito son los siguientes.

Objetivo General

"Obtener información que permita medir y hacer seguimiento a los indicadores de las metas planteadas en el Plan Nacional de Desarrollo 2017-2021, y demás agendas de desarrollo nacional e internacional"(Fernández et al., 2019)

Objetivos Específicos

"Producir información estadística cuantitativa que permita la medición y monitoreo de los indicadores del Plan Nacional de Desarrollo.", "Disponer de datos que posibiliten la actualización de indicadores para el seguimiento y evaluación de las políticas, planes y programas que se desarrollen.", "Visualizar el perfil social, demográfico y económico de la población total en el área urbana y en el área rural del país, a través de variables de carácter general como: sexo, edad, parentesco, nivel de instrucción, asistencia escolar, percepción, entre otras." En base a las metodologías de diseño muestral, el INEC delimita esta operación estadística como una encuesta basada en una muestra tipo probabilístico con un dominio de estudio demográfico y social. La población y migración, trabajo, educación, salud, ingresos y consumo, asentamientos humanos y vivienda, justicia y crimen son sus principales temas de estudio, mientras que la percepción sobre funcionamiento, uso, calidad de los servicios públicos, confianza en las Instituciones públicas y discriminación, características y servicios básicos de la vivienda, actividad física, el acceso a las tecnologías de la información y comunicación, y la tenencia de computadores y acceso a internet en los hogares son subtemas de interés político y social. El universo de estudio definido para la encuesta multipropósito son personas de 5 años y más residentes en el Ecuador excluyendo a

quienes habitan en viviendas colectivas, flotantes y población indigente (sin techo), su población objetivo se centra en los hogares y personas constituidas por la población de 5 años y más de quienes se ha de recabar información sociodemográfica y para investigar temas de percepción, son las personas de 16 años y más. La unidad de observación son todas las viviendas particulares ocupadas ubicadas en territorio nacional con sus dos unidades de análisis: el hogar particular y las personas miembros de hogar. La encuesta tiene una cobertura geográfica definida por las viviendas ocupadas a nivel nacional que incluye la región insular desagregados por área geográfica urbano y rural, esta encuesta se ha caracterizado por su periodicidad y continuidad anual, dado que las metas planteadas se basan en el Plan Nacional de Desarrollo 2017-2021 (Fernández et al., 2019).

Análisis y Evaluación de Variables

Para el proyecto se han analizado las variables investigadas mediante el formulario de la encuesta multipropósito, tomado en cuenta el objetivo 1: Vida digna con iguales oportunidades para todas las personas. Este objetivo refleja en uno de sus indicadores el porcentaje de la población de 12 y más años, que realiza ejercicio o deporte en su tiempo libre más de 3.5 horas a la semana.

En base al criterio de (Tynan, 1987) quién afirma que un adecuado segmento de mercado se basa en el análisis de datos geográficos, demográficos, psicológicos, psicográficos y de comportamiento se ha identificado al siguiente grupo de variables como base, debido a su contenido estratégico para alcanzar el objetivo del proyecto.

La Tabla 8, muestra las variables alineadas al proyecto SEGDO_BICI_UIO.

Tabla 8

Variables de la encuesta multipropósito alineadas al proyecto SEGDO_BICI_UIO

Código de Variable	Nombre Variable	Definición	Formato de la Variable	Categorías de la Variable	Tipo
área	Área	Especificación la ubicación de la vivienda: Urbano - Rural	Texto	No Aplica	Variable Geográfica
ciudad	Ciudad	Identificador del código de la ciudad.	Texto	No Aplica	Variable Geográfica
zona	Zona Censal	Es una división estadística que se define como carga de trabajo para la supervisión y control principalmente en los operativos censales.	Texto	No Aplica	Variable Geográfica
sector	Sector Censal	Es una división estadística que se define como carga de trabajo de los operativos de campo en investigaciones estadísticas.	Texto	No Aplica	Variable Geográfica
id_conglomerado	Conglomerado	Agrupación de manzanas integradas por un número de determinado de viviendas que comparten características en común (estrato) y que son próximas entre sí.	Texto	No Aplica	Variable Geográfica
viv	Vivienda	Identificador del código de la ciudad	Numérico	No Aplica	Variable Geográfica

Código de Variable	Nombre Variable	Definición	Formato de la Variable	Categorías de la Variable	Tipo
hog	Hogar	Identificador del código de la ciudad	Numérico	No Aplica	Variable Geográfica
persona	persona	Numero de elemento en el hogar	Numérico	No Aplica	Variable Demográfica
s1p2	Sexo	Es un atributo diferencial fundamental de análisis demográfico como también en el estudio de las características sociales y económicas de una población.	Categórico	_1. Hombre _2. Mujer	Variable Demográfica
s1p3	Edad	Tiempo transcurrido a partir del nacimiento de un individuo.	Numérico	Valores entre 0 a 98	Variable Demográfica
s1p4	Relación de Parentesco	Relación de parentesco del hogar respecto al jefe del hogar	Numérico	_1. Jefe _2. Cónyuge _3. Hijo o Hija _4. Yerno o nuera _5. Nieto o nieta _6. Padres o suegros _7. Otros parientes _8. Empleado/a domestica _9. Otros no parientes	Variable Demográfica
s1p6	Estado civil o conyugal actual	Estado civil de la persona entrevistada.	Categórico	_1. Casado _2. Separado _3. Divorciado _4. Viudo _5. Unión Libre _6. Soltero	Variable Demográfica

Código de Variable	Nombre Variable	Definición	Formato de la Variable	Categorías de la Variable	Tipo
s1p12a	Nivel de instrucción	El nivel de instrucción aprobado por el informante.	Categórico	_1. Ninguno _2. Centro de alfabetización _3. Jardín de Infantes _4. Primaria _5. Educación Básica _6. Secundaria _7. Bachillerato _8. Superior no Universitario _9. Superior Universitario _10. Postgrado	Variable Demográfica
s6p1a	Uso de la bicicleta	Especifica si usa o no la bicicleta como medio de transporte.	Categórico	_1. SI _2. NO	Variable Psicográfica
s6p1b	Frecuencia Uso de la Bicicleta	Identifica la frecuencia con la que usa la bicicleta.	Categórico	_1. Todos los días _2. Al menos una vez a la semana? _3. Al menos una vez al mes _4. Al menos una vez al año	Variable Psicográfica
s6p2	Medios de desplazamiento	Identifica el uso de la bicicleta como medio de transporte dentro de la rutina diaria	Categórico	_1. Vehículo Particular (solo) _2. Vehículo Particular (compartido) _3. Transporte Público _4. Bicicleta _5. Caminar _6. Otro ¿Cuál? (especifique) _99. No aplica	Variable Psicográfica

Código de Variable	Nombre Variable	Definición	Formato de la Variable	Categorías de la Variable	Tipo
s6p3	Especificación medio de transporte	Identifica la razón principal por la que usa un medio de transporte en especifico	Categórico	_1. Comodidad _2. Necesidad _3. Seguridad _4. Conciencia Ambiental _5. Salud/Deporte _6. Ahorro de dinero _7. Cercanía	Variable Psicográfica
s6p4	Uso del tiempo (deportes)	Especifica en un periodo de tiempo si realizo ejercicio o actividad física	Categórico	_1. Si _2. No	Variable Psicográfica
s6p4a	Tiempo Actividad Lunes a Viernes -Horas	Especifica el tiempo de actividad física de lunes a viernes en horas	Numérico	No Aplica	Variable Psicográfica

Código de Variable	Nombre Variable	Definición	Formato de la Variable	Categorías de la Variable	Tipo
s6p4b	Tiempo Actividad Lunes a Viernes - Minutos	Especifica el tiempo de actividad física de lunes a viernes en minutos	Numérico	No Aplica	Variable Psicográfica
s6p4c	Tiempo Actividad Sábado y domingo - Horas	Especifica el tiempo de actividad física de sábado y domingo en horas	Numérico	No Aplica	Variable Psicográfica
s6p4d	Tiempo Actividad Sábado y domingo - Minutos	Especifica el tiempo de actividad física de sábado y domingo en minutos	Numérico	No Aplica	Variable Psicográfica

Nota: Instituto Nacional de Estadística y Censos INEC

Diseño de la arquitectura técnica

La arquitectura técnica cubre los procesos y herramientas que se aplican a los datos en esta área existen dos conjuntos que tienen distintos requerimientos, brindan sus propios servicios y componentes de almacenaje de datos: El back room (habitación trasera) y el front room (habitación frontal). El back room es el responsable de la obtención y preparación de los datos, por lo que también se conoce como adquisición de datos y el front room es responsable de entregar los datos a la comunidad de usuario y también se le conoce como acceso de datos. En la Figura 10, se puede apreciar un esquema global de la arquitectura planteada para el desarrollo del sistema de BI. La arquitectura consta de tres capas, la capa inicial se compone de archivos .csv y .shape denominado fuente de datos, la segunda capa es un espacio de almacenamiento dimensional el cual se conforma de cubos OLAP (OnLine Analytical Processing) encargado de almacenar los datos estructurados, finalmente se tiene la capa aplicación BI donde se refleja la información de manera gráfica y sus principales indicadores en un dashboard o panel de control.

Figura 10

Arquitectura global de la aplicación

Ambiente Back Room

Para el desarrollo del ambiente back room de nuestro sistema de BI, se ha considerado un esquema estructurado en tres capas: repositorio de datos fuente, data staging area, y el modelo dimensional. Cada uno aporta a una lógica específica en el objetivo del proyecto.

En la Figura 11, se presenta el ambiente back room, que inicia su proceso en el repositorio de datos fuente y termina llenando las tablas del modelo dimensional, ubicado en el gestor para base de datos PostgreSQL.

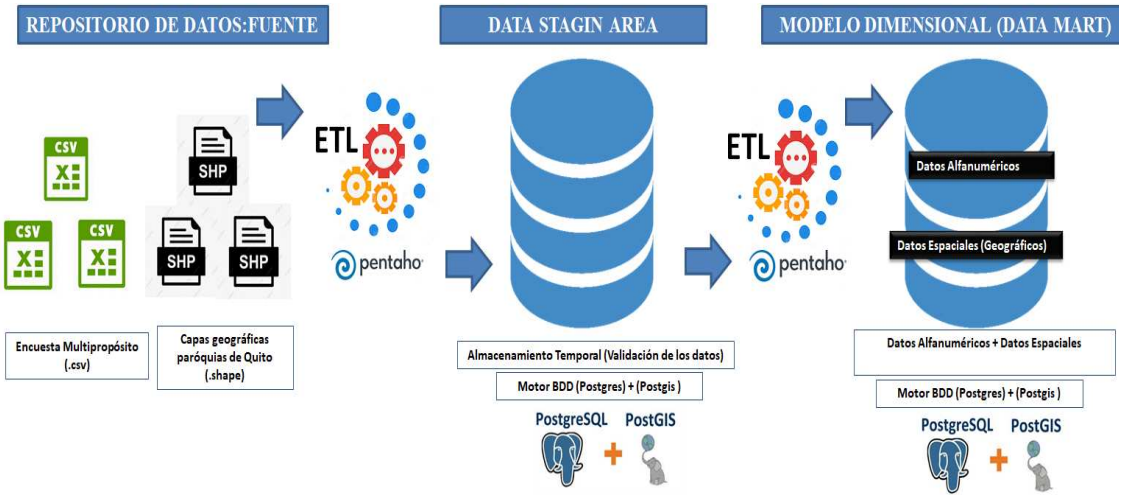
Repositorio de datos fuente: Se caracteriza por ser el disparador del proceso. La fuente de datos contiene la materia prima encargada de dar vida al proyecto en base al análisis de la fuente. Nuestro repositorio está conformado por archivos en formato csv y shape. Los archivos .csv contienen información alfanumérica de la encuesta multipropósito, dividida geográficamente a nivel de zonas censales. Mientras que los archivos .shape contienen la información geográfica municipal de Quito a nivel parroquial. Estos datos una vez que se han validado a nivel de estructura, formatos y calidad, están listos para ser migrados a una siguiente etapa.

Data Staging Area (DSA) o almacenamiento de datos temporal: Este componente es de gran ayuda al procesamiento de los datos, pues permite alojarlos en una estructura similar al repositorio destino. En este espacio virtual, se procede a cargar los datos provenientes de los archivos .csv y .shape en un gestor de base de datos PostgreSQL y PostGIS. El DSA permite manipular los datos mediante el uso de lenguaje de consulta estructurado (SQL) para poblar de forma adecuada las tablas del modelo dimensional del data mart.

Modelo Dimensional o Data Mart: El esquema data mart está alojado en la misma base que contiene el esquema DSA, sin embargo la diferencia radica en su diseño. El data mart tiene un concepto relacional y su forma es dimensional. Las tablas del esquema en mención son pobladas como etapa final al procesamiento de los datos posterior al llenado del espacio temporal y sobre ellas recae el análisis de la información obtenida de sus combinaciones. Se trata de modelos sencillos que aseguran buenos tiempos de respuesta, y su lógica permite interactuar con software de visualización y análisis de BI.

Figura 11

Esquema lógico ambiente back room



Ambiente Front Room

El ambiente destinado a la representación de los datos como información está compuesto por el modelo dimensional, el cuadro de mando y los usuarios.

La Figura 12 muestra los componentes que interactuaran con el usuario mediante la visualización de un dashboard o panel de control.

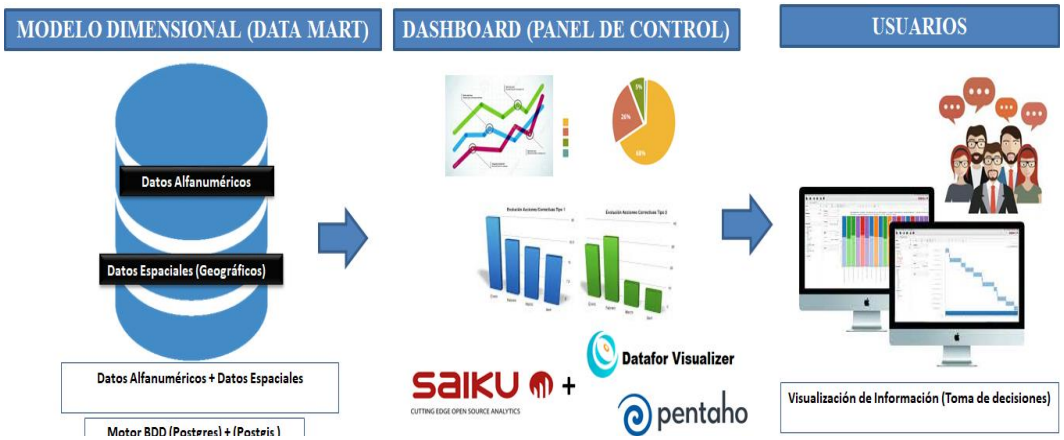
Modelo Dimensional (Data Mart): Para la lógica del ambiente front room, el modelo dimensional (data mart) desarrollado, proporciona los datos y la estructura que se requiere para ser analizado mediante un ambiente gráfico.

Dashboard (Cuadro de mando): Es una representación gráfica de los principales indicadores, que intervienen en el análisis de los objetivos principales del proyecto mediante el uso del visualizador analítico Sayku en su versión open source. Se procede a cargar los cubos de datos Online Analytical Processing (OLAP) generados a partir del data mart. La herramienta permite interactuar gráficamente con la estructura de los cubos, de tal manera que se puede dar forma a un indicador para su posterior inclusión en el panel de control (dashboard). Una vez realizado el análisis de los cubos OLAP, se procede a crear el dashboard con la herramienta Visualizer de Datafor que consiste en la instalación de un plugin compatible con Pentaho, esta herramienta se encargara de mostrar los indicadores de manera visual al usuario final.

Usuarios: Bajo nuestra lógica, el elemento (usuarios) hace referencia a un determinado número de personas que acceden al análisis de los datos mediante un panel de control que permite visualizar la información de manera gráfica. Gracias a la ayuda del software Sayku y Visualizer, los usuarios podrán tener acceso al panel de control mediante el uso de un navegador web.

Figura 12

Esquema lógico del ambiente front room



Selección de productos para el desarrollo del sistema BI

El desarrollo del sistema de BI requiere el uso de herramientas que cumplan con las necesidades y especificaciones del proyecto. En nuestro caso se ha planteado como estrategia el uso de software libre, en particular, debido al ahorro en la compra de licencias y las diferentes opciones que se pueden encontrar actualmente en el mundo digital.

Gestor de Base de Datos

Actualmente PostgreSQL también llamado Postgres se ha convertido en un poderoso sistema de gestión de bases de datos relacional orientado a objetos y de código abierto. Gracias a su librería espacial PostgGIS que permite manipular datos geográficos, es posible evidenciar su gran capacidad en cuanto al almacenamiento y las bondades que posee al momento de probar su rendimiento en actividades de procesamiento. La versión utilizada en el proyecto es la 11.9, para el tratamiento de datos alfanuméricos y la versión PostGIS es la 5.2 para el tipo de dato geométrico. La

Tabla 9, refleja las características del gestor de base de datos utilizado en el sistema de BI.

Tabla 9

Características del gestor de bases de datos

Gestor de Base de Datos: Características			
Software	**Versión**	**Tipo**	**Descripción**
PostgreSQL	11.9	Libre	Sistema de código abierto de administración de bases de datos de tipo relacional.
PostGIS	2.5	Libre	Es un software compatible con Open Geospatial Consortium (OGC) utilizado como una extensión para PostgreSQL, que es una forma de base de datos objeto-relacional.

Herramientas de Bussines Intelligence

Para el desarrollo del componente BI, se ha elegido Pentaho BI Community Edition, puesto que es una herramienta completa que permite trabajar con los datos enfocando su esfuerzo al procesamiento, análisis y visualización. Es una herramienta open source; está basada en Java y se define como una plataforma de BI “orientada a la solución” y “centrada en procesos”. Pentaho es una suite que incluye todas las herramientas para crear inteligencia de negocios como son: consultas, reportes, análisis dashboards, procesos ETL y minería de datos. La Tabla 10, muestra el kit de herramientas basadas en software libre utilizados en el desarrollo del sistema de BI.

Tabla 10

Características de las herramientas de la solución de bussines intelligence

Herramientas de Bussines Intelligence			
Software	**Versión**	**Tipo**	**Descripción**
Pentaho data Integration	8.3	Libre	Proporciona las capacidades de extracción, transformación y carga (ETL) que facilitan el proceso de captura, limpieza y almacenamiento de datos
Pentaho schema workbench - Mondiran	8.1	Libre	Interfaz de diseñador que permite crear y probar esquemas de cubos OLAP.
Pentaho analysis services	8.1	Libre	Servidor OLAP compatible con el MDX, y el lenguaje de conducta XML para el análisis y especificaciones de la interfaz.
Saiku analytics	3.0	Libre y comercial	Solución de análisis en su versión community que permite visualizar y analizar de forma amigable bases de datos dimensionales desarrollados en base a cubos OLAP creados con pentaho.
Datafor Visualizer for pentaho	3.3	Libre y comercial	Solución en versión free edition permite el análisis y creación de dashboards (panel de control) en base a cubos OLAP creados con pentaho)

Modelado Dimensional

En base a los requerimientos, las variables obtenidas de los datos abiertos y el objetivo principal del sistema de BI, se han identificado tres enfoques de la encuesta multipropósito que son de gran utilidad para la construcción del data mart. Estos tres enfoques se llaman modelos y ayudan a identificar posibles segmentos de mercado para aquellos negocios alineados a la venta y mantenimiento de bicicletas. La Tabla 11,

muestra las principales características de los modelos resultantes en base al análisis y evaluación de las variables.

Tabla 11

Resumen del proceso iterativo del modelo dimensional

Lógica del Negocio	Modelo	Granularidad	Dimensiones Compartidas	Dimensiones Compartidas	Hechos
Uso de la bicicleta en Quito	Modelo 1	Media	4	3	1
Actividad Física en los habitantes de Quito	Modelo 2	Media	3	3	1
Medios de transporte y la bicicleta como alternativa de desplazamiento en Quito	Modelo 3	Media	5	3	1

Para el análisis de los datos relacionados al desarrollo del BI se ha adoptado el modelo estrella, este modelo consta de una tabla central de "Hechos" y varias "dimensiones", incluida una dimensión de "Tiempo".

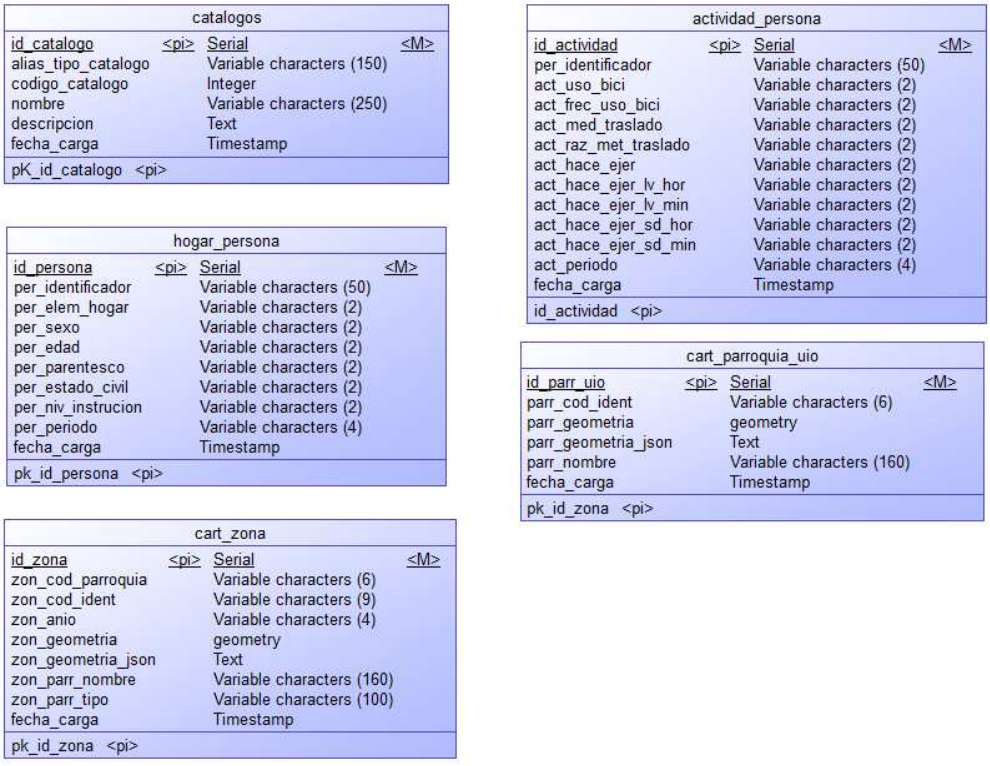
Para ejecutar un acertado diagrama de los modelos se han construido gráficos de burbujas que ayudan a entender de mejor manera las características del negocio.

Modelo 1: Esquema DSA (Data Staging Area)

El esquema está conformado por 5 entidades o tablas que carecen de relación y claves primarias, en el modelado la mayoría de campos que conforma cada tabla son de tipo texto puesto que los archivos origen .csv mantienen este formato en cada columna. Este entorno de almacenamiento representa un área intermedia entre la fuente y el data mart con el fin de afinar y pulir los datos. El modelado de datos del esquema DSA se lo ha desarrollado en base a un criterio técnico con el objetivo de hacer más fácil el procesamiento de la información. En base a las variables de la Tabla 8, se ha diseñado modelo lógico como se muestra en la Figura 13.

Figura 13

Modelo lógico del esquema DSA (Data Staging Area)



Es importante tomar en cuenta que los datos obtenidos de las fuentes abiertas están estructuradas en un solo archivo y su contenido a nivel de metadato en forma horizontal. Mediante el esquema de almacenamiento temporal, se tiene una estructura fácil de manipular y rápida de migrar hacia el data mart. A continuación, se describe mediante matrices los atributos de las tablas que forman parte del esquema de almacenamiento temporal. La Tabla 12, muestra los diferentes campos encargados de almacenar los datos referentes a personas y sus características como sexo, edad nivel de instrucción, parentesco y estado civil. Las actividades físicas tales como el andar en bicicleta o realizar ejercicio, son factores de análisis en nuestro proyecto. La Tabla 13, contiene atributos que involucran a las personas que realizan este tipo de actividades. La Tabla 14, se ha estructurado de una manera especial, puesto que en ella se están guardando datos referentes a catálogos y su detalle, por ejemplo, el catálogo llamado nivel de instrucción con su detalle: 1) Ninguno, 2) Centro de alfabetización, 3) Jardín de Infantes, 4) Primaria, 5) Educación Básica, 6) Secundaria, 7) Bachillerato, 8) Superior no Universitario, 9) Superior Universitario, 10) Post-grado. La Tabla 15, está compuesta por atributos que identifican las zonas geográficas censales utilizadas por el INEC para establecer una referencia geográfica. La Tabla 16, refleja atributos que permiten identificar una parroquia en el Distrito Metropolitano de Quito bajo la lógica de su administración municipal.

Tabla 12

Atributos de la tabla temporal incluida en el esquema DSA: Hogar_Persona

Tabla Dimensional			Archivo Origen			
Columna	**Descripción**	**Tipo de Dato - Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_persona	Secuencial tabla persona	Serial				Llenado secuencialmente al poblar los datos
per_identificador	Código Identificador de la persona	Varchar (50)	ident_per			Se llena directamente de la fuente (archivos.xls)
per_elem_hogar	Posición en el hogar	ineteger	persona	Numérico		Se llena directamente de la fuente (archivos.xls)
per_sexo	Código: sexo persona	integer	s1p2	Categórico	_1. Hombre _2. Mujer	Se llena directamente de la fuente (archivos.xls)

Tabla Dimensional					Archivo Origen	
Columna	**Descripción**	**Tipo de Dato - Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
per_edad	Edad de la persona	integer	s1p3	Numérico		Se llena directamente de la fuente (archivos.xls)
per_parentesco	Código de parentesco con respecto al jefe del hogar.	integer	s1p4	Numérico	_1. Jefe _2. Cónyuge _3. Hijo o Hija _4. Yerno o nuera _5. Nieto o nieta _6. Padres o suegros _7. Otros parientes _8. Empleado/a domestica _9. Otros no parientes	Se llena directamente de la fuente (archivos.xls)
per_estado_civil	Estado civil de la persona	integer	s1p6	Categórico	_1. Casado _2. Separado _3. Divorciado _4. Viudo _5. Unión Libre _6. Soltero	Se llena directamente de la fuente (archivos.xls)

Tabla Dimensional				Archivo Origen		
Columna	**Descripción**	**Tipo de Dato - Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
per_niv_instrucion	Código nivel de instrucción	integer	s1p12a	Categórico	_1. Ninguno _2. Centro de alfabetización _3. Jardín de Infantes _4. Primaria _5. Educación Básica _6. Secundaria _7. Bachillerato _8. Superior no Universitario _9. Superior Universitario _10. Post-grado	Se llena directamente de la fuente (archivos.xls)
per_periodo	Periodo de tiempo	Varchar(4)				Periodo de referencia de levantamiento de la encuesta
fecha_carga	Fecha de carga de los datos	Timestamp				Registrado mediante triggers

Tabla 13

Atributos de la tabla temporal incluida en el esquema DSA: Actividad_Persona

Tabla Dimensional			Archivo Origen			
Columna	**Descripción**	**Tipo de Dato - Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_actividad	Secuencial tabla dsa.actividad_persona	Serial				Llenado secuencialmente al poblar los datos
per_identificador	Código Identificador de la persona	Varchar(50)	ident_per			Se llena directamente de la fuente (archivos.xls)
act_uso_bici	Personas que usan bicicleta	integer	s6p1a	Categórico	_1. SI _2. NO	Se llena directamente de la fuente (archivos.xls)
act_frec_uso_bici	Identifica la frecuencia con que usan la bicicleta	integer	s6p1b	Categórico	_1. Todos los días _2. Al menos una vez a la semana? _3. Al menos una vez al mes _4. Al menos una vez al año	Se llena directamente de la fuente (archivos.xls)

Tabla Dimensional			Archivo Origen			
Columna	**Descripción**	**Tipo de Dato - Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
act_med_traslado	El medio que usa para desplazarse de un lugar a otro	integer	s6p2	Categórico	_1. Vehículo Particular (solo) _2. Vehículo Particular (compartido) _3. Transporte Público _4. Bicicleta _5. Caminar _6. Otro ¿Cuál? (especifique) _99. No aplica	Se llena directamente de la fuente (archivos.xls)
act_raz_met_traslado	Especifica el motivo por el cual usa un medio de desplazamiento	integer	s6p3	Categórico	_1. Comodidad _2. Necesidad _3. Seguridad _4. Conciencia Ambiental _5. Salud/Deporte _6. Ahorro de dinero _7. Cercanía	Se llena directamente de la fuente (archivos.xls)
act_hace_ejer	Tiempo utilizado para realizar alguna actividad física	integer	s6p4	Categórico	_1. Si _2. No	Se llena directamente de la fuente (archivos.xls)
act_hace_ejer_lv_hor	Tiempo de actividad física de lunes a viernes (horas)	integer	s6p4a	Numérico		Se llena directamente de la fuente (archivos.xls)

Tabla Dimensional			Archivo Origen			
Columna	**Descripción**	**Tipo de Dato - Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
act_hace_ejer_lv_min	Tiempo de actividad física de lunes a viernes (minutos)	integer	s6p4b	Numérico		Se llena directamente de la fuente (archivos.xls)
act_hace_ejer_sd_hor	Tiempo de actividad física de fin de semana (horas)	integer	s6p4c	Numérico		Se llena directamente de la fuente (archivos.xls)
act_hace_ejer_sd_min	Tiempo de actividad física fin de semana (minutos)	integer	s6p4d	Numérico		Se llena directamente de la fuente (archivos.xls)
act_periodo	periodo de levantamiento de datos	Varchar (4)				Periodo de referencia de levantamiento de la encuesta
fecha_carga	fecha de carga de datos	Timestamp				Registrado mediante triggers

Tabla 14

Atributos de la tabla temporal incluida en el esquema DSA: Catálogos

Tabla Dimensional	Archivo Origen

Columna	Descripción	Tipo de Dato-Longitud	Código de Variable	Formato de la Variable	Categorías de la Variable	Reglas de Conversión: Observaciones
id_catalogo	Secuencial tabla dsa.catalogos	Serial				Llenado secuencialmente al poblar los datos
alias_tipo_catalogo	Texto identificador del catálogo	Varchar(150)		Texto		Se llena directamente de la fuente (archivos.xls)
codigo_catalogo	Código identificador del catálogo	intger		Numérico		Se llena directamente de la fuente (archivos.xls)
nombre	Nombre del catálogo	Varchar(250)		Texto		Se llena directamente de la fuente (archivos.xls)
descripcion	Descripción del catálogo	text		Texto		Se llena directamente de la fuente (archivos.xls)
fecha_carga	fecha de carga de datos	Timestamp				Registrado mediante triggers

Tabla 15

Atributos de la tabla temporal incluida en el esquema DSA: Cart_Zona

Tabla Dimensional	Archivo Origen

Columna	Descripción	Tipo de Dato-Longitud	Código de Variable	Formato de la Variable	Categorías de la Variable	Reglas de Conversión: Observaciones
id_zona	Secuencial tabla dsa.cart_zona	Serial		Numérico		Llenado secuencialmente al poblar los datos
zon_cod_parroquia	Código identificador DPA del canton Quito	Varchar(6)	dpa_parroq	Texto		Se llena directamente de la fuente (archivos.shape)
zon_cod_ident	Código identificador DPA de la zona censal	Varchar(9)	dpa_zona	Texto		Se llena directamente de la fuente (archivos.shape)
zon_anio	Anio de actualización de los datos	Varchar(4)	dpa_anio	Texto		Se llena directamente de la fuente (archivos.shape)
zon_geometria	Campo geométrico de la capa	geometry	geometry	Geometria		Se llena directamente de la fuente (archivos.shape)
zon_geometria_json	Campo geométrico de la capa en formato JSON	text		Geometria		Se llena directamente de la fuente (archivos.shape)
zon_parr_nombre	Nombre de la parroquia	Varchar(160)	dpa_nombre	Texto		Se llena directamente de la fuente (archivos.shape)

Tabla Dimensional				Archivo Origen		
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
zon_parr_tipo	Tipo de la parroquia que contiene la zona	Varchar(6)	dpa_tipo	Texto		Se llena directamente de la fuente (archivos.shape)
fecha_carga	Fecha de carga de datos	Timestamp		Timestamp		Registrado mediante triggers

Tabla 16

Atributos de la tabla temporal incluida en el esquema DSA: Cart_Parr_Uio

Tabla Dimensional				Archivo Origen		
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_parr_uio	Secuencial tabla dsa.cart_parr_uio	Serial		Numérico		Llenado secuencialmente al poblar los datos
parr_cod_ident	Código identificador de la parroquia	Varchar(4)	parroq_id	Texto		Se llena directamente de la fuente (archivos.shape)

parr_geometria	Campo geométrico de la capa	geometry	geometry	Geometria	Se llena directamente de la fuente (archivos.shape)
parr_geometria_json	Campo geométrico de la capa en formato JSON	text		Texto	Se llena directamente de la fuente (archivos.shape)
parr_nombre	Nombre de la parroquia	Varchar(160)	nombre	Texto	Se llena directamente de la fuente (archivos.shape)
fecha_carga	fecha de carga de datos	Timestamp		Timestamp	Registrado mediante triggers

Modelo 2: Uso de la Bicicleta

En base a la meta del objetivo 1 del plan nacional de desarrollo que dice ”incrementar de 12,2% a 14,4% la población mayor a 12 años que realiza más de 3.5 horas a la semana de actividad física al 2021 (Fernández et al., 2019). Se ha diseñado preguntas para la encuesta multipropósito que permitan recolectar información de las personas que incluyen en sus actividades cotidianas, el uso de la bicicleta y la ejecución de ejercicio físico. La Figura 14, muestra la representación de un modelo estrella en el cual se tiene una tabla de hechos llamada “uso de la bicicleta” que combina las dimensiones tiempo, hogar persona, ubicación y frecuencia de uso.

Figura 14

Gráfico de burbujas del uso de la bicicleta

Las tablas que conforman el modelo dimensional que estructura el datamart, se caracterizan por su estructura des normalizada. A pesar de que su lógica es

redundante, esta permite optimizar su rendimiento al momento de trabajar en su análisis. Las tablas resultantes de este modelo se las detalla a continuación;

La Tabla 17, contiene los atributos de las personas encuestadas y sus características como sexo, edad, estado civil y su nivel de instrucción. Si bien es cierto su estructura se asemeja a la tabla del esquema temporal, su diferencia radica en la inclusión de un campo que describe el atributo, por ejemplo, el atributo sexo contiene el valor 1 y 2 para lo cual su equivalente es, 1=hombre, 2=mujer. La Tabla 18, muestra los atributos de la periodicidad de levantamiento de la encuesta, reflejada en meses y años. La Tabla 19, es de carácter especial, pues en esta se refleja los atributos resultantes de la intersección entre una zona censal y las parroquias ubicadas en el Distrito Metropolitano de Quito. La Tabla 20, refleja los atributos que identifican a la frecuencia de un usuario en cuanto al uso de la bicicleta. La Tabla 21, representa la lógica central del modelo, puesto que su estructura permitirá realizar un adecuado análisis de los datos. En ella se reflejan los atributos que enlazan al resto de combinaciones identificadas como dimensiones.

Tabla 17

Atributos de la tabla dimensión dim_hogar_persona, incluida en el esquema del data mart

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de conversión: Observaciones**
id_persona	Secuencial tabla persona	Serial	PK				Llenado secuencialmente al poblar los datos
per_identificador	Código Identificador de la persona	Varchar(50)		ident_per			
per_elem_hogar	Posición en el hogar	ineteger		persona	Numérico		
per_sexo	Código: sexo persona	integer					
per_sexo_desc	Descripción sexo	Varchar(12)		s1p2	Categórico	_1. Hombre _2. Mujer	
per_edad	Edad de la persona	integer		s1p3	Numérico		Valores entre 0 a 98

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de conversión: Observaciones**
per_parentesco	Código de parentesco con respecto al jefe del hogar.	integer		s1p4	Numérico	_1. Jefe _2. Cónyuge _3. Hijo o Hija _4. Yerno o nuera _5. Nieto o nieta _6. Padres o suegros _7. Otros parientes _8. Empleado/a domestica _9. Otros no parientes	
per_parentesco_desc	Descripción parentesco	Varchar(120)					
per_estado_civil	Estado civil de la persona	integer		s1p6	Categórico	_1. Casado _2. Separado _3. Divorciado _4. Viudo _5. Unión Libre _6. Soltero	
per_estado_civil_desc	Descripción estado civil	Varchar(120)					
per_niv_instrucion	Código nivel de instrucción	integer		s1p12a	Categórico	_1. Ninguno _2. Centro de alfabetización _3. Jardín de Infantes	

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de conversión: Observaciones**
per_niv_instruccion_desc	Descripción nivel de instrucción	Varchar(120)				_4. Primaria _5. Educación Básica _6. Secundaria _7. Bachillerato _8. Superior no Universitario _9. Superior Universitario _10. Post-grado	
per_periodo	Periodo de tiempo	Varchar(4)					Periodo de referencia de levantamiento de la encuesta
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

Tabla 18

Atributos de la tabla dimensión, dim_tiempo incluida en el esquema data mart

Tabla Dimensional	Archivo Origen

Columna	Descripción	Tipo de Dato-Longitud	Clave	Código de Variable	Formato de la Variable	Categorías de la Variable	Reglas de Conversión: Observaciones
id_tiempo	Secuencial, dato de referencia para la tabla tiempo	Serial	PK				Llenado secuencialmente al poblar los datos
anio	Referencia de tiempo en años	integer		periodo_anio	Numérico		Campo completado mediante transformación: año de levantamiento de datos
mes	Referencia de tiempo en meses	integer		periodo_mes	Numérico		Campo completado mediante transformación: mes de levantamiento de datos (diciembre)
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

Tabla 19

Atributos de la tabla dimensión, dim_uio_parroquia incluida en el esquema data mart

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_parroquia	Secuencial dato de las parroquias	Serial	PK				Llenado secuencialmente al poblar los datos
parr_codigo	Código de la parroquia Quito Urbano	Varchar(4)		parroquia_id			
parr_cod_zona_censal	Código de la zona censal (INEC)	Varchar(9)		dpa_zona			Fuente Actualización Cartográfica INEC(2010)
parr_nombre	Nombre de la parroquia	Varchar(120)		nombre			
parr_geometria	Dato geométrico de la parroquia	geometry			Dato geométrico		Se obtiene mediante transformaciones (Intersección entre zona censal y parroquias de Quito)
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

Tabla 20

Atributos de la tabla dimensión, dim_frecuencia_uso_bici, incluida en el esquema data mart

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_frecuencia_uso_bici	Secuencial datos frecuencia uso bicicleta	Serial	PK				Llenado secuencialmente al poblar los datos
cod_frecuencia	Código de registro frecuencia uso bicicleta	integer		s6p1b	Categórico	_1. Todos los días _2. Al menos una vez a la semana? _3. Al menos una vez al mes _4. Al menos una vez al año	Llenado desde la fuente mediante separación de variables bajo una lógica de catálogos
nom_frecuencia	Descripción de registro frecuencia uso bicicleta	Varchar(120)					
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

Tabla 21

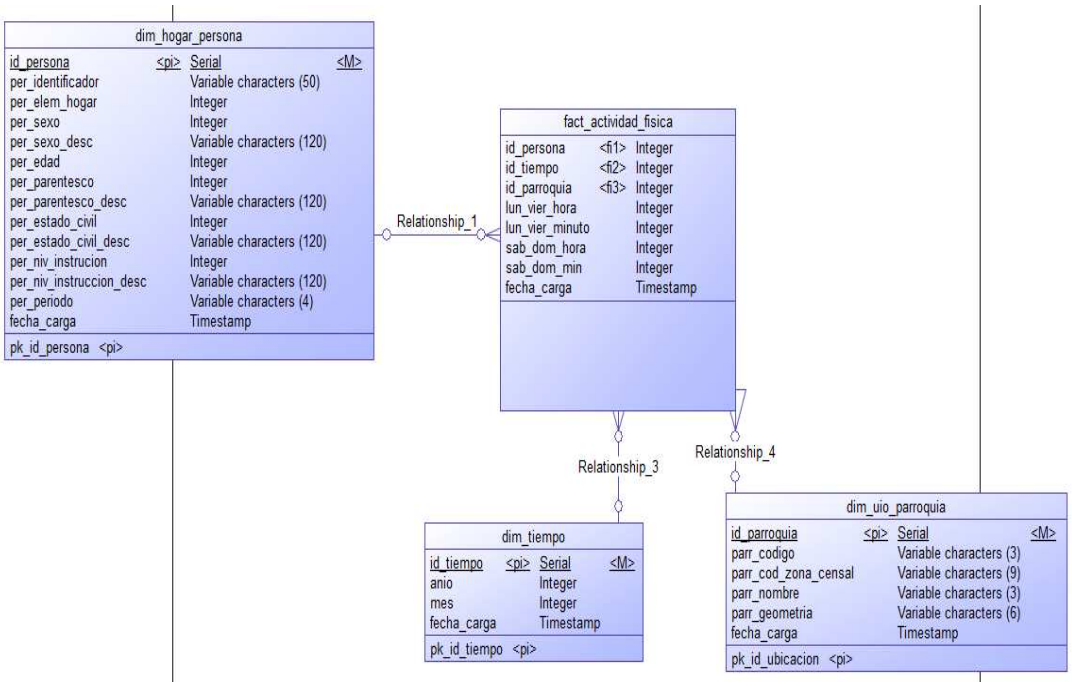
Atributos de la tabla de hechos, fact_uso_bicicletas incluida en el esquema data mart

Tabla Dimensional					**Archivo origen**		
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_persona	Clave foránea (asocia la dimensión dim_hogar_persona)	integer	FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y la dim_hogar_persona
id_freq_uso_bici	Clave foránea (asocia la dimensión dim_frecuancia_uso_bici)	integer	FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y la dim_frecuancia_uso_bici
id_tiempo	Clave foránea (asocia dimensión di _tiempo)		FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y la dim_tiempo
id_parroquia	Clave foránea (asocia dimensión dim_uio_parroquia)	integer	FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y su ubicación geográfica
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

La Figura 15, es el resultado gráfico del modelo dimensional desarrollarlo para realizar el análisis a las personas que hacen uso de la bicicleta en Quito.

Figura 15

Modelo dimensional: uso de las bicicletas (diagrama lógico)



Modelo 3: Actividad Física

Para poder determinar de mejor manera un segmento de mercado se ha incluido en el sistema de BI variables que tienen afinidad al cumplimiento del objetivo 1. Si bien es cierto que una bicicleta puede ser adquirida por cualquier persona, es importante reconocer que aquellas que realizan actividades físicas de manera constante, podrían convertirse en un conjunto de potenciales clientes.

El modelo 3 permite analizar información de aquellas personas que incluyen en su rutina tiempo invertido en actividades relacionadas al ejercicio físico, que podrían

impactar en la práctica de algún deporte y por ende la necesidad de adquirir una bicicleta, ver Figura 16.

Figura 16

Gráfico de burbujas, actividad física y su periodicidad

El modelo representativo para el análisis de la actividad física y su periodicidad comparte las dimensiones de las tablas: Tabla 17, Tabla 18, Tabla 19 explicadas anteriormente y la inclusión de la Tabla 22, la cual guarda el enfoque central “llamada de hechos”. Esta tabla contiene los atributos que permiten medir el número de personas que realizan ejercicios físicos en sus rutinas diarias.

Tabla 22

Atributos de la tabla de hechos, fact_actividad_física Incluida en el esquema data mart

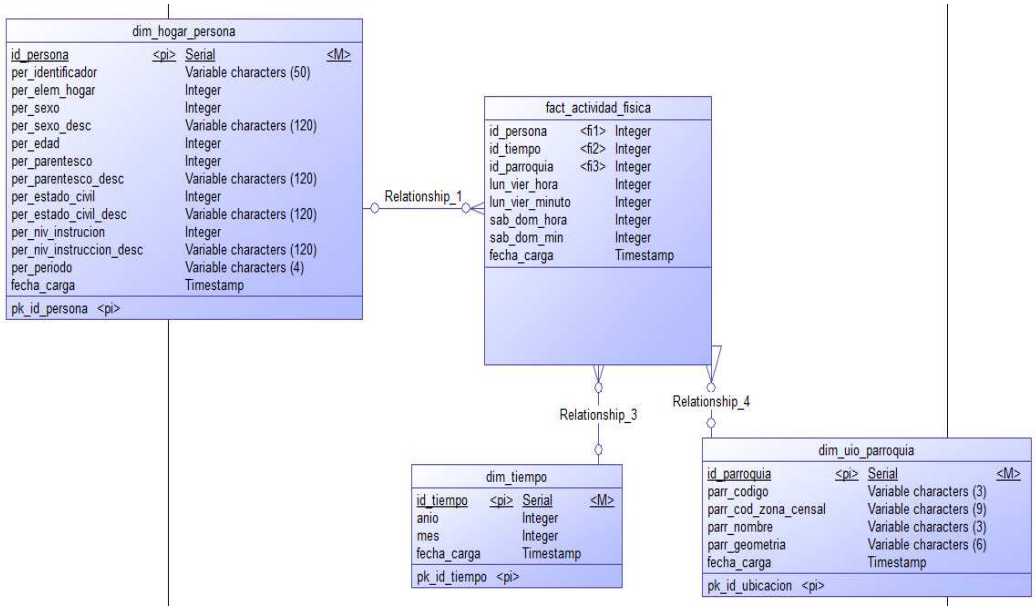
Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_persona	Clave foránea (asocia la dimensión dim_hogar_persona)	integer	FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y la dim_hogar_persona
id_tiempo	Clave foránea (asocia dimensión di _tiempo)		FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y la dim_tiempo
id_parroquia	Clave foránea (asocia dimensión dim_uio_parroquia)	integer	FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y su ubicación geográfica
lun_vier_hora	Tiempo en horas de actividad física (lunes a viernes)	integer			Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona
lun_vier_minuto	Tiempo en minutos actividad física (lunes a viernes)	integer			Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
sab_dom_hora	Tiempo en horas de actividad física (sábados y domingos)	integer			Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona
sab_dom_minuto	Tiempo en minutos de actividad física (sábados y domingos)	integer			Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

La Figura 17, establece el modelo dimensional bajo un esquema lógico, que permite identificar la tabla de hechos y sus dimensiones combinadas. Dimensiones necesarias para el análisis de las personas que se ejercitan a nivel de parroquia en el Distrito Metropolitano de Quito.

Figura 17

Modelo dimensional: actividad física y periodicidad (diagrama lógico)



Modelo 4: Medios de transporte y la bicicleta como alternativa

La bicicleta es considerada como un medio de transporte sostenible y amigable con el medio ambiente. Hoy en día, muchos países a nivel mundial han acogido el uso de la misma como una opción de movilidad en sus ciudades, construyendo ciclo vías y en varios casos desarrollando estrategias que promuevan el uso de este medio de transporte. Acoger esta forma alternativa de trasladarse, no solo es amigable con

nuestro planeta, sino también ayuda a mejorar la calidad de vida de las personas. Está comprobado que el uso continuo de la bicicleta mejora el rendimiento físico y emocional de quien lo usa (Suero, 2010). En consecuencia, el uso de la bicicleta como alternativa de transporte es otro factor de análisis para el proyecto SEGDO_BICI_UIO, por lo tanto se lo ha incluido como parte de nuestro data mart.

La Figura 18, refleja su representación gráfica bajo un enfoque dimensional.

Figura 18

Gráfico de burbujas, medios de transporte y la bicicleta como alternativa

El análisis dimensional para el modelo relacionado a los medios de transporte y el uso de la bicicleta comparte las dimensiones de las tablas: Tabla 17, Tabla 18, Tabla 19 descritas anteriormente. La Tabla 23, contiene los atributos que identifican las razones, por las cuales los usuarios eligen usar una bicicleta para trasladarse a sus

trabajos o lugares de estudio. La Tabla 24, incluye atributos relacionados a los medios de transporte usados comúnmente para trasladarse de un lugar a otro y entre ellos la bicicleta.

La Tabla 25, es nuestra tabla central o de hechos y contiene los atributos relacionados a las dimensiones que sirven para generar los indicadores de análisis para las personas que usan la bicicleta para transportarse.

Tabla 23

Atributos de la tabla dimensión, dim_trans_traslada incluida en el esquema data mart

Tabla Dimensional				**Archivo Origen**			
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_raz_medio_traslado	Secuencial datos (razón para usar un medio de traslado)	Serial	PK				Llenado secuencialmente al poblar los datos
raz_codigo	Código de registro razón para usar un medio de traslado	integer		s6p3	Categórico	_1. Comodidad _2. Necesidad _3. Seguridad _4. Conciencia Ambiental _5. Salud/Deporte _6. Ahorro de dinero _7. Cercanía	Llenado desde la fuente mediante separación de variables bajo una lógica de catálogos
raz_nombre	Opción de catálogo (razón para usar un medio de traslado)	Varchar(120)					
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

Tabla 24

Atributos de la tabla dimensión, dim_medios_traslado incluida en el esquema data mart

Tabla Dimensional				Archivo Origen			
Columna	**Descripción**	**Tipo de Dato - Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de Conversión: Observaciones**
id_medio_traslado	Secuencial datos (medios de traslado)	Serial	PK				Llenado secuencialmente al poblar los datos
med_codigo	Código de registro medios de traslado	integer		s6p2	Categórico	1. Vehículo Particular (solo) _2. Vehículo Particular (compartido) _3. Transporte Público _4. Bicicleta _5. Caminar _6. Otro ¿Cuál? (especifique) _99. No aplica	Llenado desde la fuente mediante separación de variables bajo una lógica de catálogos
med_nombre	Opción de catálogo (medios de traslado)	Varchar(120)					
fecha_carga	Fecha de carga de los datos	Timestamp					Registrado mediante triggers

Tabla 25

Atributos de la tabla de hecho, fact_uso_bicicleta incluida en el esquema data mart

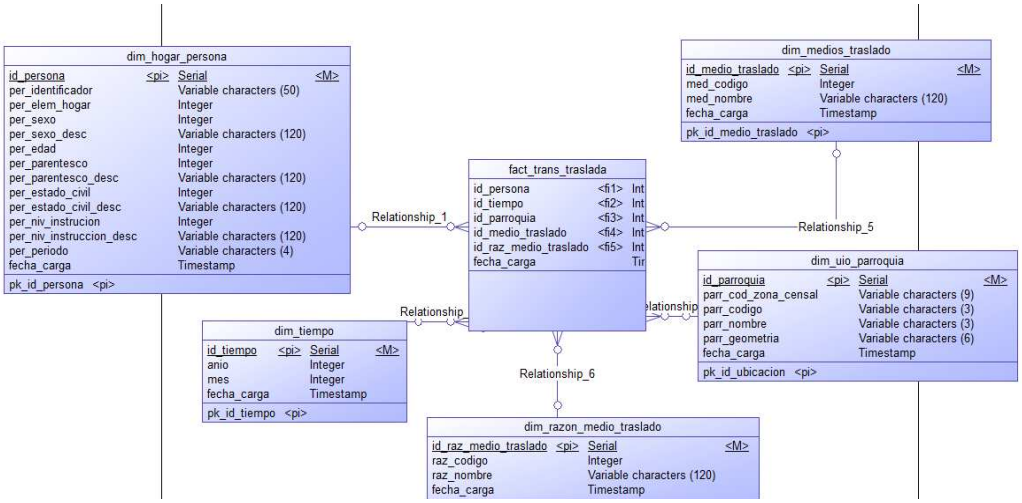
Tabla Dimensional			**Archivo Origen**				
Columna	**Descripción**	**Tipo de Dato-Longitud**	**Clave**	**Código de Variable**	**Formato de la Variable**	**Categorías de la Variable**	**Reglas de conversión: Observaciones**
id_persona	Clave foránea (asocia la dimensión dim_hogar_persona)	integer	FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y la dim_hogar_persona
id_tiempo	Clave foránea (asocia dimensión di _tiempo)		FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y la dim_tiempo
id_parroquia	Clave foránea (asocia dimensión dim_uio_parroquia)	integer	FK		Numérico		Se llena a partir de la fuente: tabla dsa. actividad_persona y su ubicación geográfica

id_medio_traslado	Clave foránea (asocia dimensión dim_medios_traslado)	integer	FK	Numérico	Se llena a partir de la fuente: tabla dsa.actividad_persona y dim_medios_traslado
id_raz_medio_traslado	Clave foránea (asocia dimensión dim_razon_medio_traslado)	integer	FK	Numérico	Se llena a partir de la fuente: tabla dsa.actividad_persona y dim_razon_medios_traslado
fecha_carga	Fecha de carga de los datos	Timestamp			Registrado mediante triggers

La Figura 19, representa el diagrama lógico del modelo dimensional, que servirá de análisis, para los indicadores basados en el uso de la bicicleta como un medio de transporte y la razón que impulsa el uso de este medio.

Figura 19

Modelo dimensional: medios de transporte y la bicicleta como alternativa (diagrama lógico)



Al finalizar la generación de los modelos dimensionales que conforman el data mart, es necesario realizar una matriz de procesos / dimensiones (Bus Matrix), en la cual se lista a detalle las dimensiones y tablas de hechos que forman parte de cada proceso de negocio resultante del análisis de requerimientos. Dentro de esta actividad se requiere diagramar el modelo dimensional completo del data mart, en donde se pueden observar todas las tablas de hechos y las dimensiones mediante un esquema gráfico ver Tabla 26.

Tabla 26

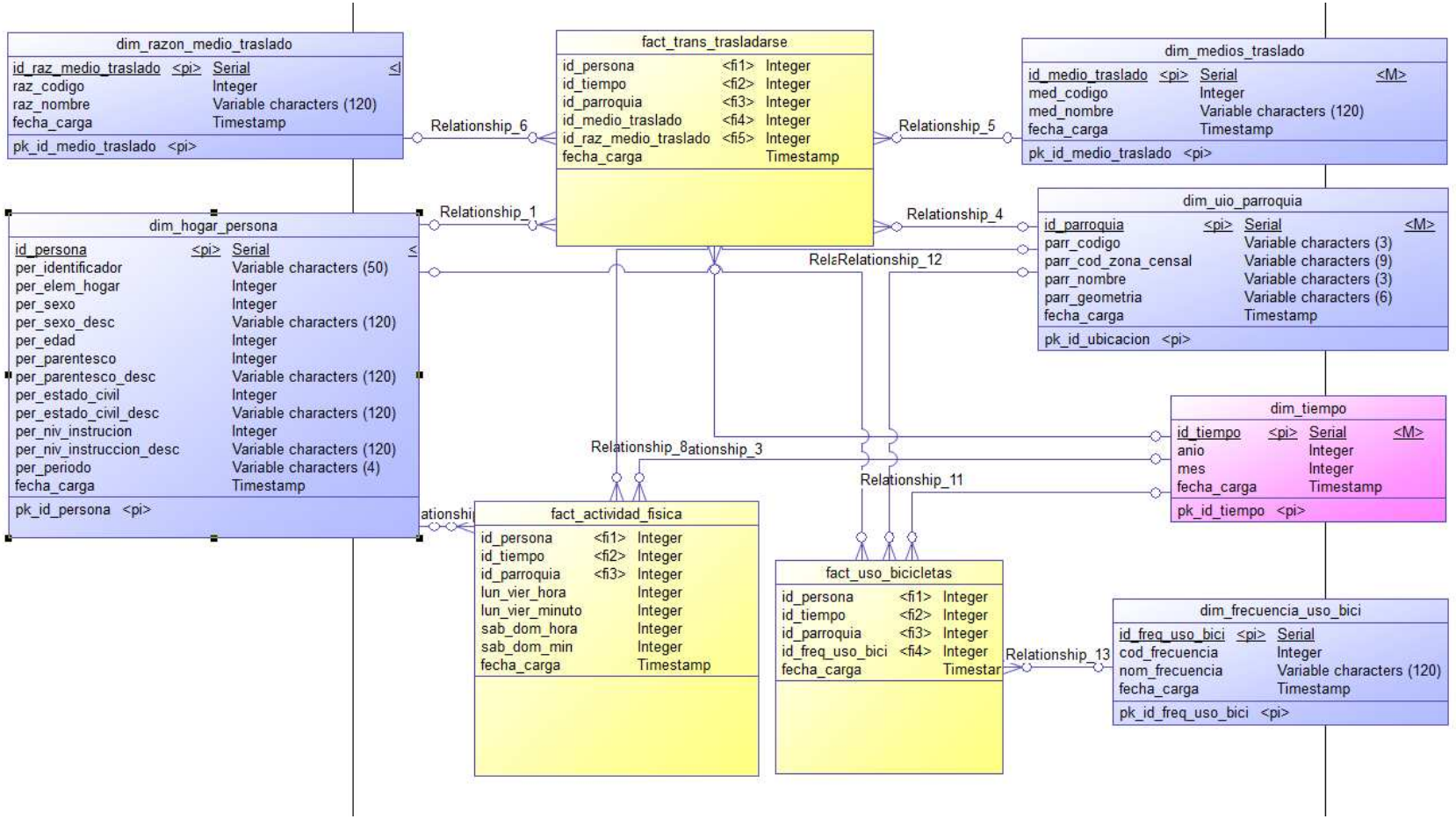
Matriz de procesos: dimensiones detalladas para el data mart segmentación de mercado

Tabla de Hechos	Granularidad	Dimensiones					
		dim_hogar_persona	dim_frecuencia_uso_bici	dim_medios_traslado	dim_uio_parroquia	dim_razon_medio_trasladado	dim_tiempo
fact_uso_bicicletas	Media	X	X		X		X
fact_actividad_fisica	Media	X			X		X
fact_trans_trasladarse	Media	X		X	X	X	X

El resultado del modelamiento de los datos, en base al objetivo principal del sistema del proyecto SEGDO_BICI_UIO se refleja en la Figura 20. Esta figura contempla el modelo lógico del data mart y su estructura conformada por dimensiones y hechos. En base a nuestro modelado de tipo estrella, su estructura se basa en tres tablas centrales que actúan como pilares del análisis del data mart y a sus alrededores 6 tablas de dimensión, que permite establecer una mejor apreciación de la información al momento de interactuar con el software de visualización.

Figura 20

Modelo lógico dimensional (data mart segmentación de mercado bicis UIO)



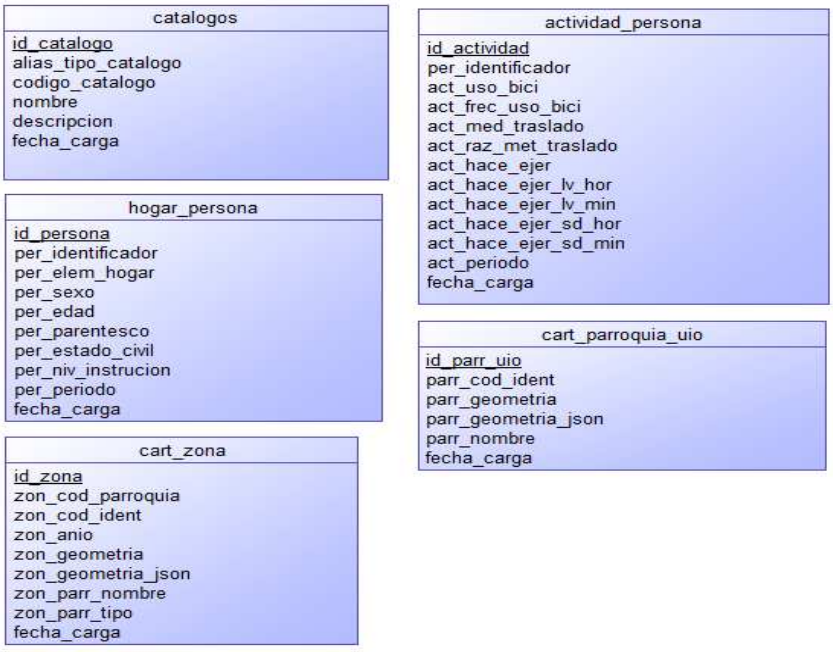
Diseño físico

Modelado del DSA (Data Staging Area)

El modelado de los datos, resultado de las diferentes fuentes de recolección, se ha diagramado en su representación final mediante el modelado físico. La Figura 21, muestra el diseño del DSA, considerado como nuestro espacio de almacenamiento temporal acorde al esquema de la sección 0.

Figura 21

Modelo físico dimensional data staging area (DSA)



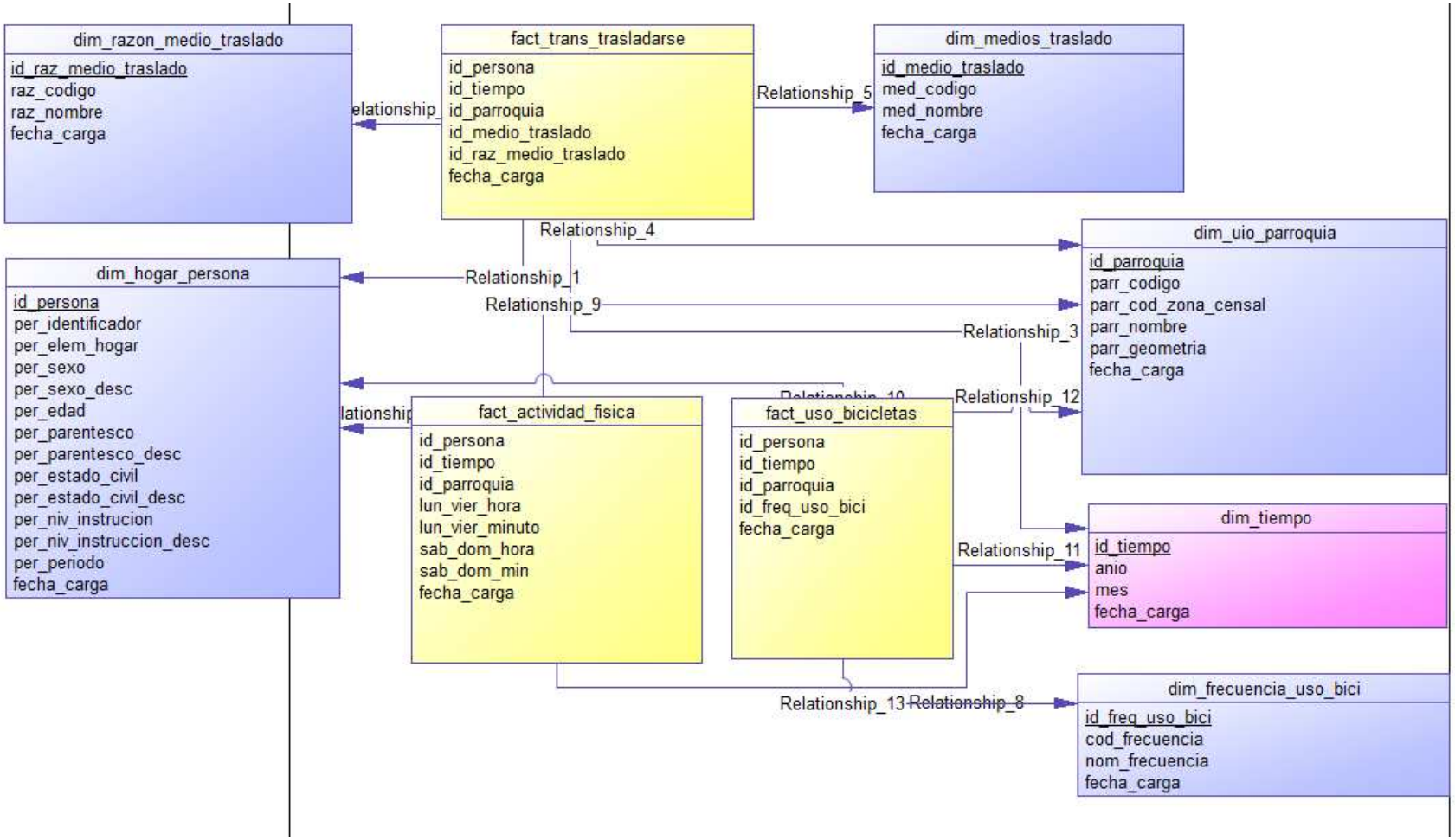
Modelado Total del Data Mart

A partir del modelo lógico dimensional del data mart se genera el modelo físico el cual contiene una especificación de todas las tablas y columnas, se incluye también las claves externas usadas para identificar las relaciones entre cada entidad. Este modelo representa la forma en cómo se construirá la base de datos en su forma final dependiendo del software usado para la administración de la base de datos. Para el

proyecto SEGDO_BICI_UIO a partir del modelo físico se ha generado un script con formato legible para el gestor de base de datos PostgreSQL. La estructura del data mart caracterizada por tres tablas de hechos y seis tablas de dimensiones permite almacenar los datos bajo una concepto de lectura optimo, una velocidad adecuada de acceso a los datos permite un buen análisis y la garantía de transformación de datos en información. El modelo físico del data mart en su versión final se puede apreciar en la Figura 22.

Figura 22

Modelo físico dimensional (Data Mart Segmentación de Mercado Bicis UIO)

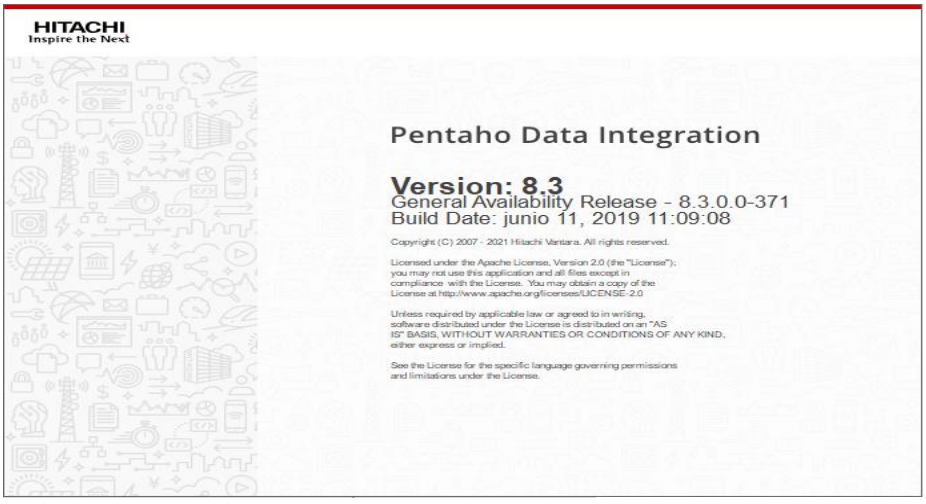


ETL (extracción, transformación y carga) diseño y desarrollo

El modelado del repositorio dimensional contenedor de datos permite tener una visión más clara de la forma en que se puede visualizar la información. El proceso de extracción, transformación y carga es la actividad encargada de manipular los datos desde su fuente y su manipulación para su almacenarlo en el data mart. La herramienta seleccionada para el trabajo (ETL) es Pentaho Data Integration debido a su licencia gratuita y la variedad de herramientas que permite realizar un óptimo trabajo de procesamiento. La Figura 23, muestra la versión de la herramienta utilizada.

Figura 23

Herramienta (ETL) Pentaho Data Integration

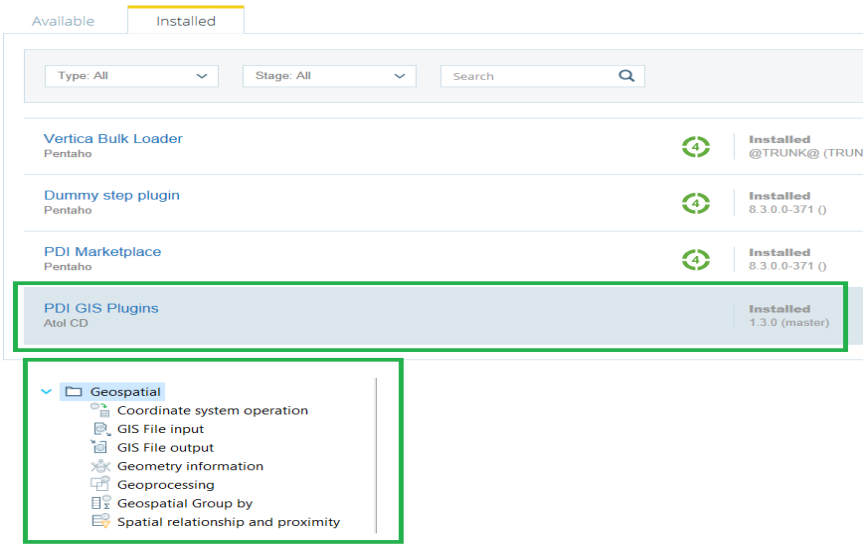


La versión 8.3 soporta la utilización de un componente llamado PDI GIS Plugin, el cual permite realizar la iteración con datos geométricos. Este atributo es de gran importancia para el proyecto, puesto que nos ha ayudado a identificar las parroquias de

análisis en base a la intersección de las capas geométricas de la base INEC con la base municipal de la ciudad de Quito. De esta forma, identificamos a las parroquias urbanas que tienen más afinidad en el uso de la bicicleta. La figura Figura 24, muestra el detalle del plugin utilizado.

Figura 24

PDI GIS plugin (Manejo de datos geométricos)

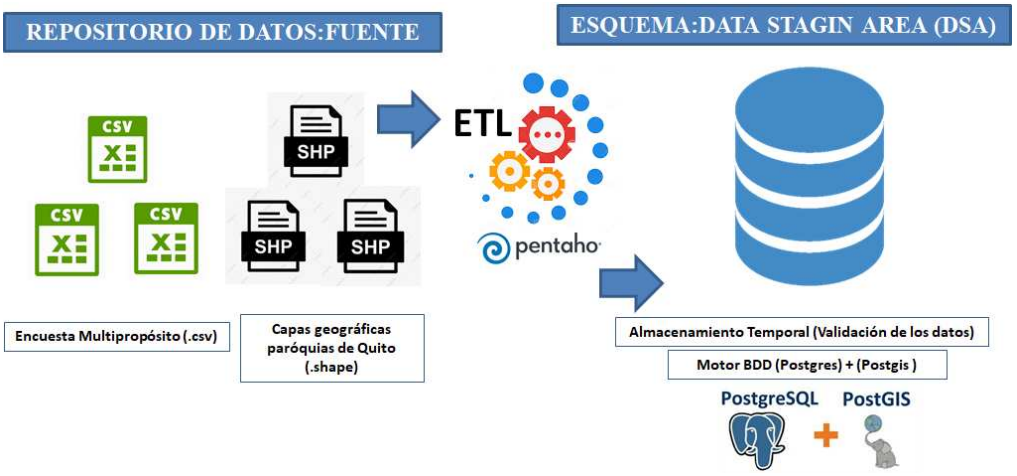


Extracción transformación y carga para el esquema de procesamiento temporal DSA (Data Staging Area)

El objetivo del DSA es generar un espacio o esquema temporal de almacenamiento en el cual se pueda realizar actividades de limpieza y depuración para poder trasladar los datos desde la fuente a su destino final (data mart), ver Figura 25.

Figura 25

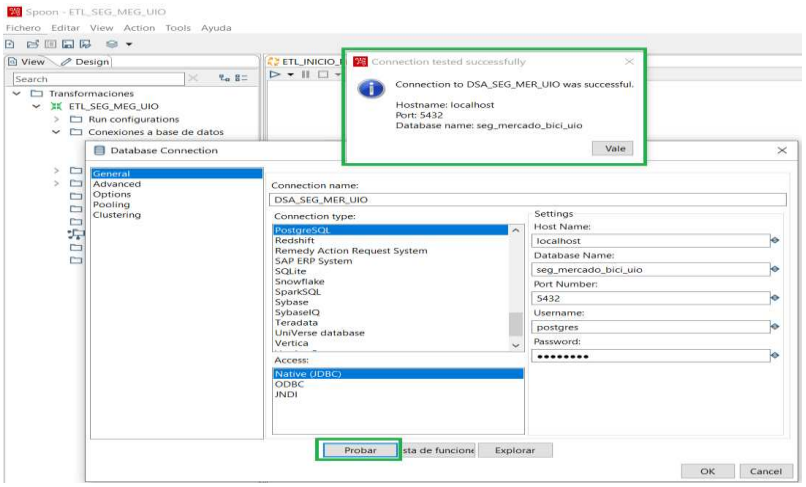
Lógica de extracción, transformación y carga para esquema DSA



La propuesta lógica para la migración, se basa en la utilización del esquema temporal de almacenamiento (data staging área), el cual se encuentra en PostgreSQL y su librería espacial PostGIS. Para poder interactuar con el gestor, es necesario crear una conexión con la base de datos. Esta conexión se la realiza mediante una librería JDBC (Java Database Connectivity) y servirá como un puente de comunicación entre PostgreSQL y la fuente. La Figura 26, muestra la creación de la conexión para poder acceder a los datos desde Pentaho Data Integration a la base de datos relacional.

Figura 26

Conexión con la base de datos (Gestor BDD PostgreSQL + PostGIS)



En la herramienta data Integration de Pentaho, se tiene la opción “conexiones a base de datos”, la cual se activa ingresando los parámetros de conexión. Los parámetros que se usan son el usuario y la contraseña de PostgreSQL, el nombre de la base de datos y el puerto de comunicación que por lo general es el 5432. Una vez ingresados los parámetros necesarios para la comunicación con la base de datos, se realiza una comprobación presionando el botón probar, el cual reflejará una ventana auxiliar con el resultado.

A continuación, se detallan los procesos realizados para el llenado del esquema temporal DSA.

A) Tabla temporal dsa.actividad_persona

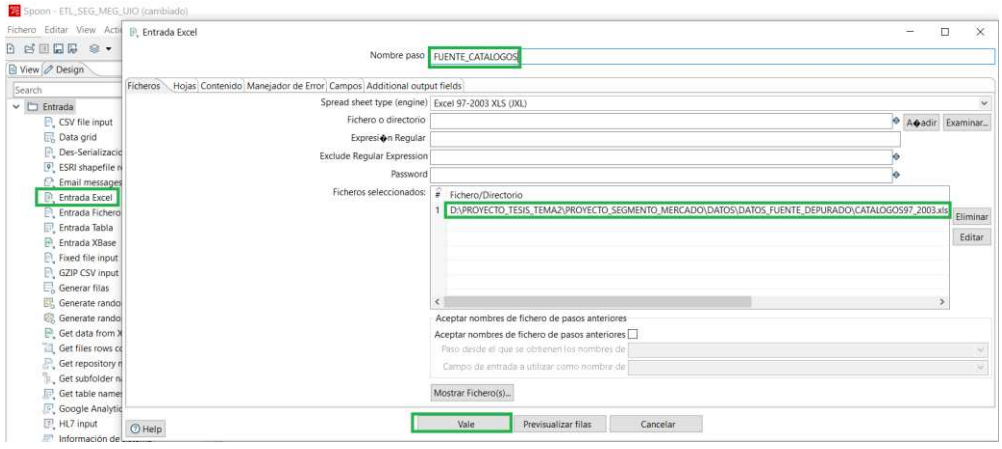
El adecuado proceso de transformación de los datos depende de la fuente, el tipo de dato y la calidad de su contenido. La información que se ha obtenido desde las

diferentes fuentes abiertas ha permitido la utilización de formatos conocidos como .csv, .xlsx y .shape. A continuación, se detalla el proceso de extracción, transformación y carga realizada para poblar de datos la tabla que contiene variables psicográficas referentes al uso de bicicletas y la ejecución de actividades físicas.

Mediante la creación de una transformación, se usa dos componentes de Pentaho que permiten la entrada y salida de datos. El componente de entrada nos permite interactuar con archivos independientes en varios formatos. Para el caso de los catálogos se ha estructurado un archivo Excel con todas las variables categóricas de la fuente. La lógica de la transformación consiste en cargar el archivo Excel en la herramienta y seleccionar todos los campos que se requiere migrar. Ver Figura 27.

Figura 27

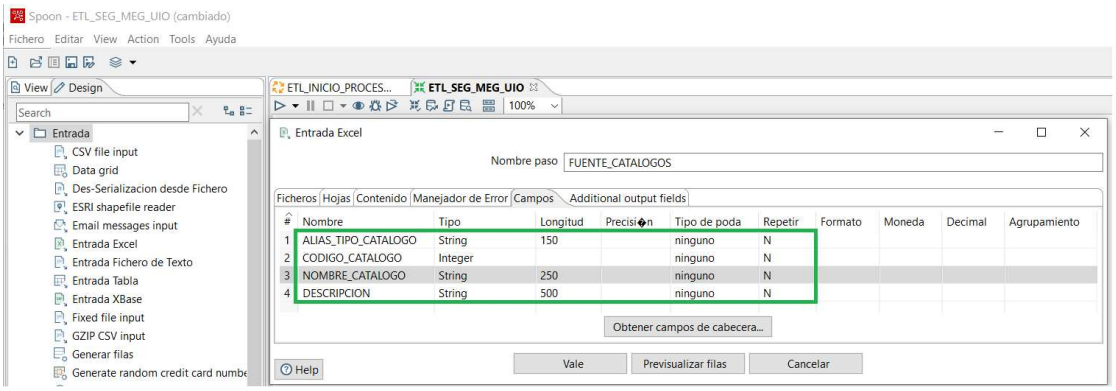
Carga de archive excel que contiene datos categóricos



Al cargar en la herramienta de extracción el archivo contenedor de los datos, es necesario escoger los campos de la estructura Excel que son migrados a su tabla equivalente del esquema destino, tal como se muestra en la Figura 28.

Figura 28

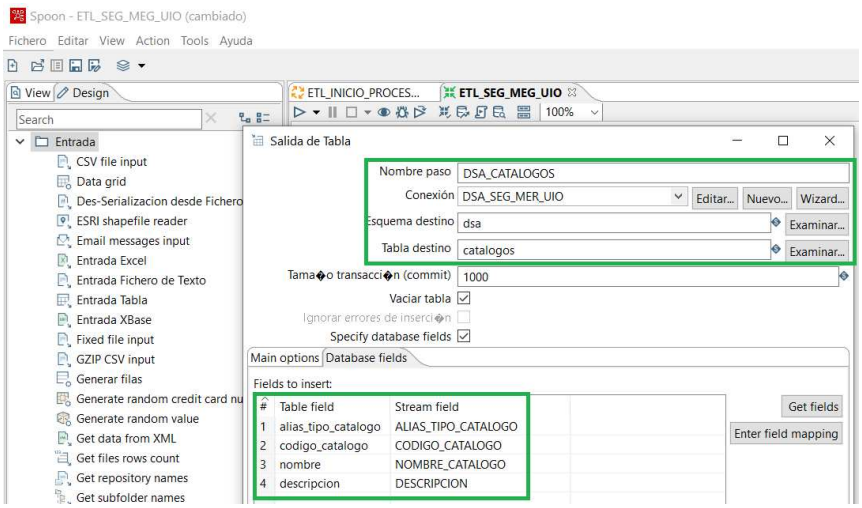
Selección de campos para carga en dsa.catalogos



Una vez realizado el proceso de carga y selección, es necesario concatenar el componente con el elemento de salida de datos. Este proceso se encarga de llenar la tabla de catálogos del esquema dsa.catalogos. En la Figura 29, se puede referenciar los parámetros de conexión a la base, el esquema, la tabla destino y todos los campos que se han seleccionado del archivo Excel que equivalen a las columnas de la nueva tabla.

Figura 29

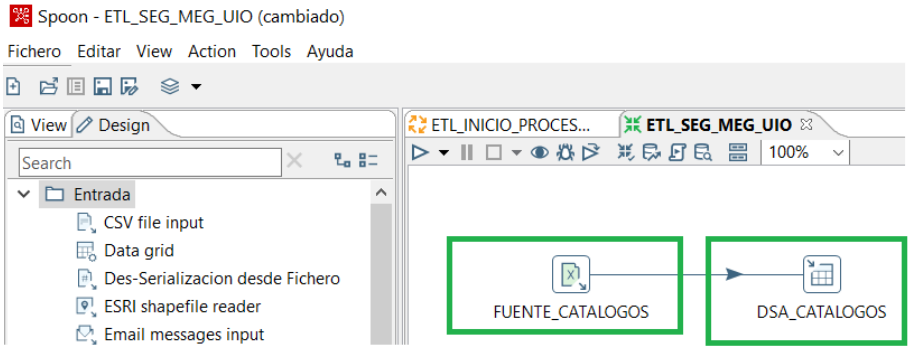
Almacenamiento de datos en esquema destino dsa.catalogos



El resultado final de esta transformación se resume en dos pasos compuestos de un componente de entrada y salida, tal como se puede apreciar en la Figura 30.

Figura 30

Transformación para la tabla temporal dsa.catalogos



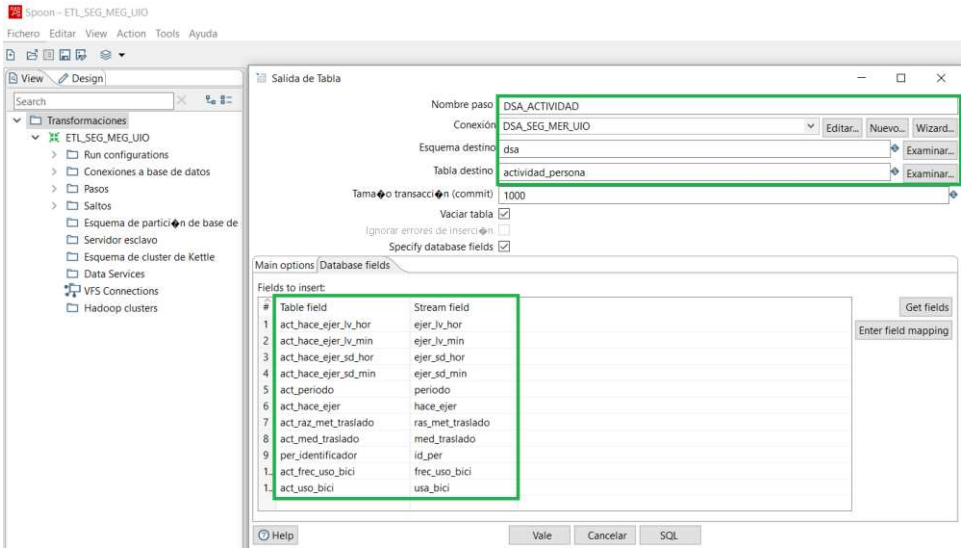
B) Tabla temporal dsa.actividad_persona

El proceso para el llenado de la tabla dsa.actividad_persona conlleva la misma lógica y secuencialidad de la tabla dsa.catalogos. La diferencia radica en el conjunto de variables a transformar, puesto que en estas entidades se puede encontrar los datos de aquellas personas que han realizado actividades referentes al uso de bicicletas o en su defecto la ejecución de actividades físicas.

La Figura 31, muestra los campos a ser migrados y la nueva base de datos destino con sus respectivos parámetros de conexión.

Figura 31

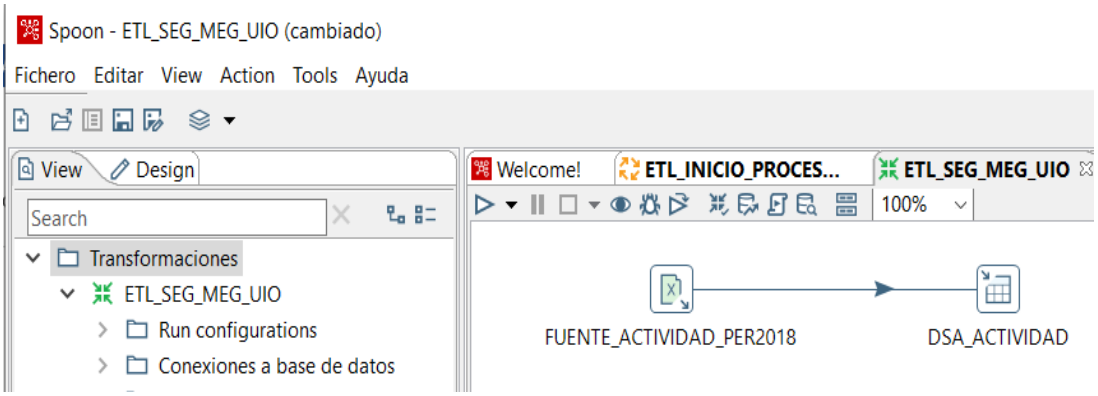
Almacenamiento de datos en esquema destino dsa.actividad_persona



El resultado final de esta transformación se refleja en el uso de dos componentes de entrada y salida de datos, tal como se muestra en la Figura 32.

Figura 32

Transformación para la tabla temporal dsa.actividad_persona



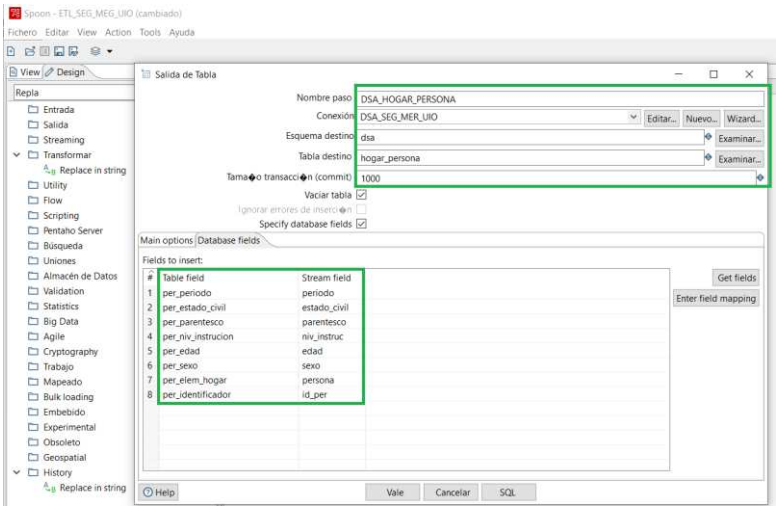
C) Tabla temporal dsa.hogar_persona

La transformación de los datos para la tabla temporal dsa.hogar_persona, incluye en su lógica un componente de entrada, en el cual se cargan los datos desde la fuente .xlsx. Una herramienta de procesamiento llamada “replace in string”, permite buscar los campos del archivo excel con datos nulos y transformarlos en ceros.

El componente de salida que contiene la conexión a la base de datos conectado al componente replace in string se encarga de llevar datos depurados al gestor de BDD PostgreSQL, ver Figura 33.

Figura 33

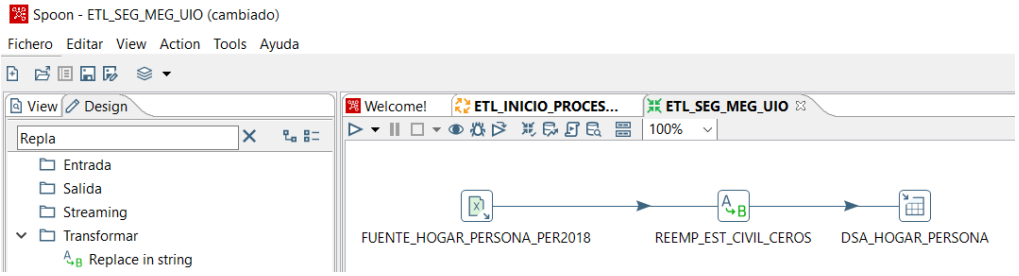
Almacenamiento de datos en esquema destino dsa.actividad_persona



La entidad hogar_persona está compuesta de variables que hacen referencia a los atributos de la persona como: sexo, edad, estado civil, nivel de instrucción y el parentesco dentro del hogar. La Figura 34 muestra la transformación realizada para el llenado de la tabla temporal dsa.hogar_persona.

Figura 34

Transformación para la tabla temporal dsa.hogar_persona



D) Tabla temporal dsa.cart_zona

La información cartográfica que maneja el Instituto Nacional de Estadísticas y Censos (INEC), se basa en la generación de zonas y sectores censales. Estos espacios geográficos están compuestos por la unión de manzanas, formadas en base al criterio metodológico de la institución. El INEC ha dispuesto en su repositorio la información cartográfica del último censo 2010, y se encuentra cargada en formato.gdb (geo data base). Para obtener la capa geográfica necesaria para el sistema de BI, se ha realizado un filtrado de las zonas a nivel nacional, y se ha exportado en formato. shape únicamente las zonas que conforman la ciudad de Quito.

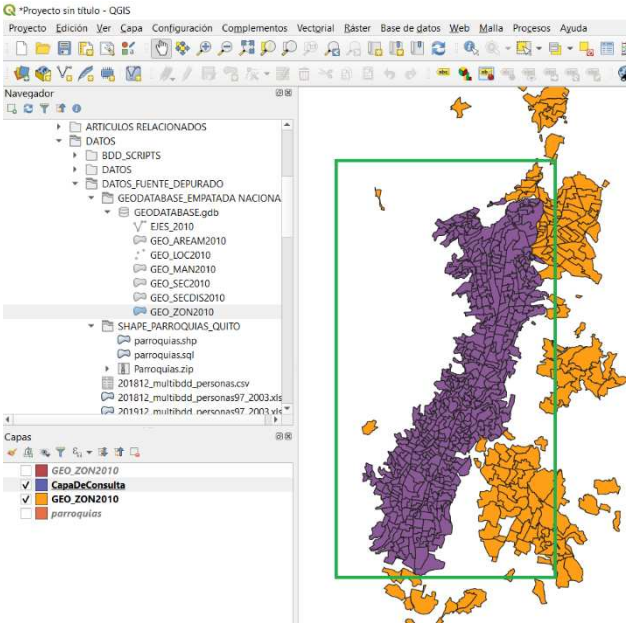
Se puede descargar las capas del último censo 2010, usadas para identificar las zonas censales, del siguiente link:

Instituto Nacional de Estadísticas y Censos. Información cartográfica V2. https://www.ecuadorencifras.gob.ec/category/cartografia-2/

La Figura 35, muestra el filtro ejecutado para la obtención de la capa geográfica. Una vez obtenida la capa zonal de Quito en formato. shape, se puede proceder a migrar a nuestro esquema de trabajo bajo el gestor de base de datos PostgreSQL y PostGIS.

Figura 35

Capa geográfica de las diferentes zonas de Quito

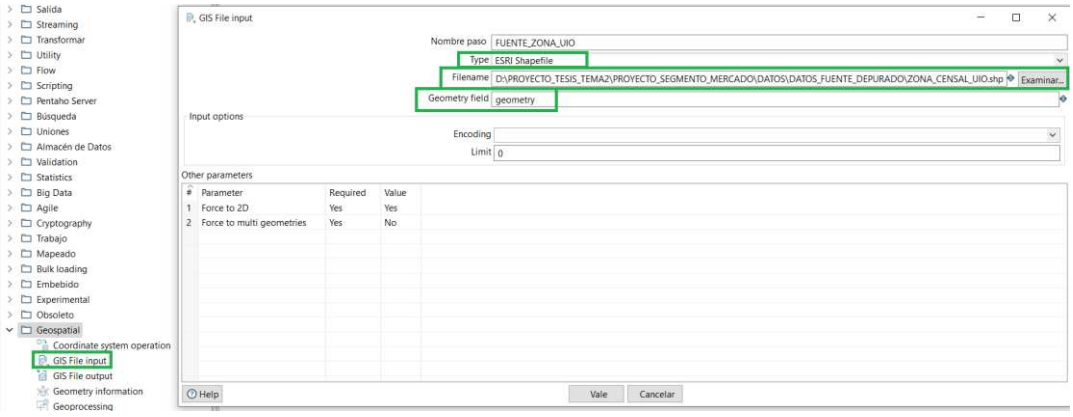


La entidad dsa.cart_zona se diferencia de las anteriores en su contenido, debido a que una de sus variables posee un dato geométrico que requiere de un componente especial para su transformación. La lógica de carga y migración a la nueva base de datos se mantiene, lo que difiere es el componente de entrada puesto que para este proceso se necesita el componente llamado “Gis file input”.

El componente de entrada compatible con datos geométricos permite especificar la fuente y el tipo de archivo, ver Figura 36.

Figura 36

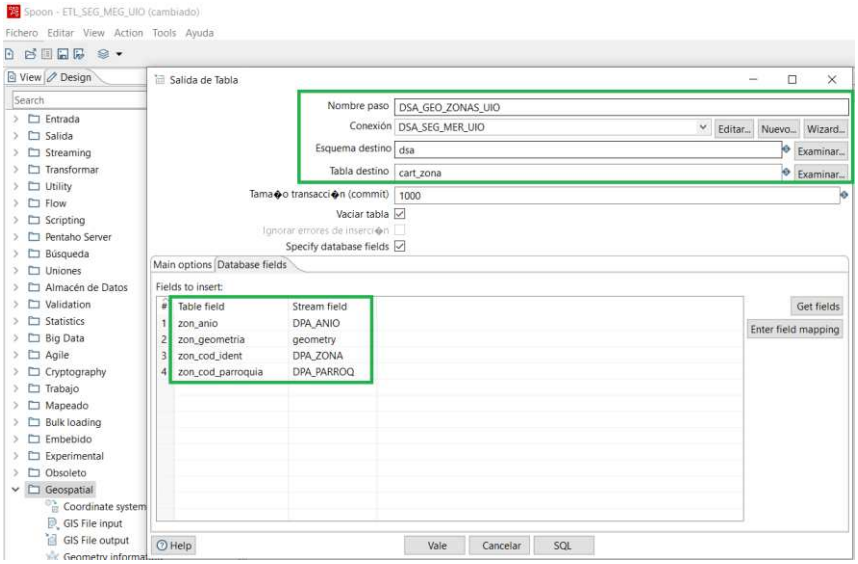
Carga archivo zonas Quito (shape)



El componente de salida contiene la conexión de la base de datos y la especificación de las columnas con las cuales se debe hacer match. Cada variable de la nueva tabla tiene su propio equivalente a nivel de archivo. shape. De esta manera se puede distinguir cuales son los datos destino y el espacio que le pertenece en el repositorio fuente. Ver Figura 37.

Figura 37

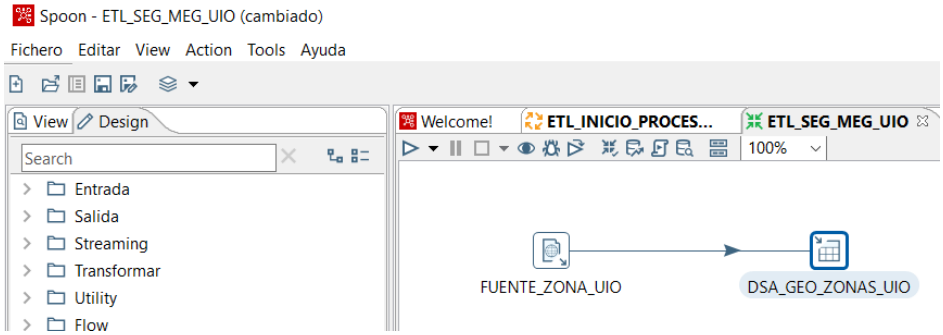
Almacenamiento de datos en esquema destino dsa.cart_zona



El resultado final de la transformación para la entidad dsa.cart_zona se muestra en la Figura 38. En esta, se puede apreciar el componente de entrada caracterizado por su lógica geográfica y otro componente extra destinado para la salida de los datos.

Figura 38

Transformación para la tabla temporal dsa.cart_zona



E) Tabla temporal dsa.cart_parr_uio

La última entidad del data staging area (DSA), también se caracteriza por su contenido a nivel de datos geográficos, similar a la entidad dsa.cart_zona. El proceso ejecutado para poblar de datos esta tabla, sigue la lógica de creación de una transformación con un elemento de entrada y uno de salida con características geográficas. La capa geométrica en formato .shape, se ha obtenido de la página municipal del distrito metropolitano de Quito, este portal tiene a su disposición la descarga de un archivo que contiene todas las parroquias de la ciudad.

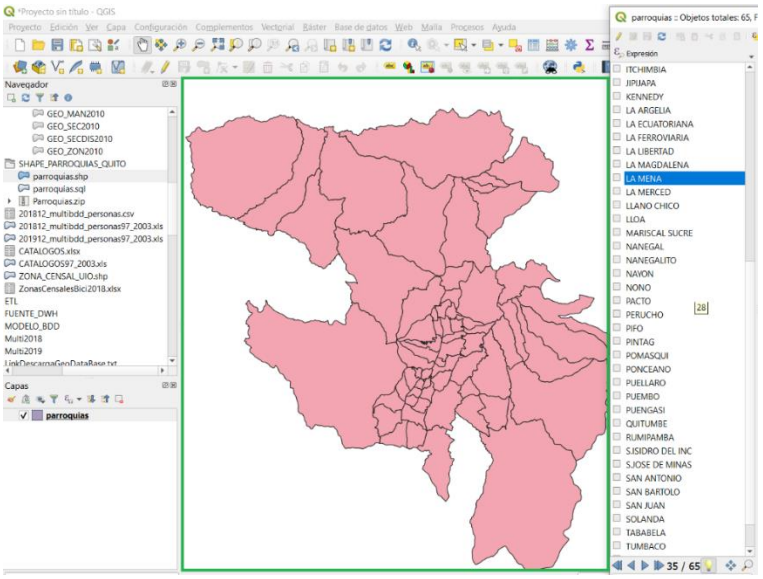
Las capas en formato .shape de las parroquias urbanas y rurales de Quito administradas a nivel municipal, se las puede descargar del siguiente link:

Municipio de Quito. Información Geográfica de Descarga. http://gobiernoabierto.quito.gob.ec/?page_id=1122

La capa geográfica parroquial de Quito permite al sistema de BI identificar a las parroquias con una buena afluencia de personas que practican bicicleta y un posible segmento de mercado al cual atacar. Debido a que en la base de datos se puede encontrar variables como edad, sexo y la especialidad de una persona; se puede establecer un posible nicho de mercado. La Figura 39 hace referencia a la capa parroquial obtenida del link municipal del distrito metropolitano de Quito y sus diferentes atributos.

Figura 39

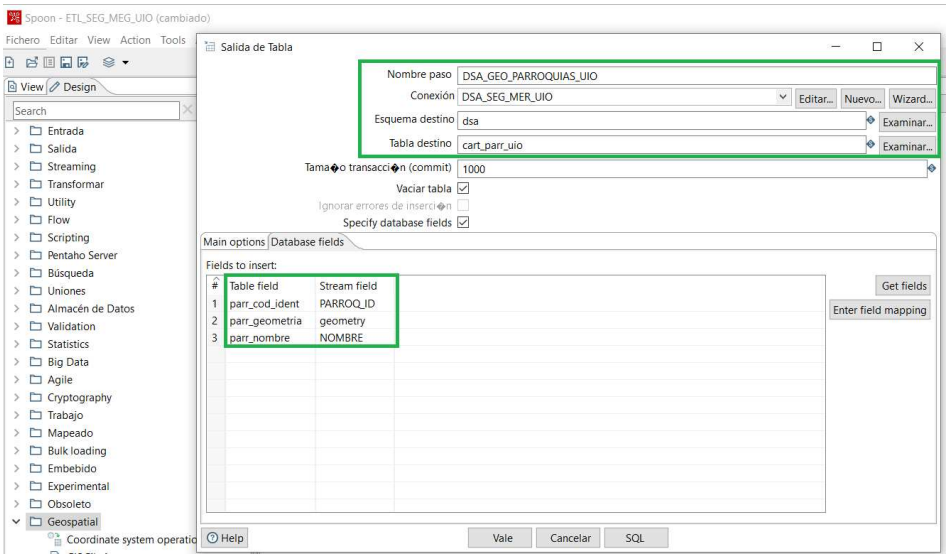
Capa geográfica parroquias de Quito



El archivo .shape que contiene las características geográficas de las zonas censales y combinadas con los espacios parroquiales, son muy importantes. El resultado de esta intersección se convierte en el insumo necesario para la migración de los datos a la nueva tabla. Una vez cargado el archivo en el componente de entrada, se lo debe concatenar al componente de salida, especificando las columnas que deben hacer match entre el archivo y la nueva tabla. La Figura 40, hace referencia a la estructura de la fuente a nivel de variables y su correspondencia en la nueva base de datos relacional.

Figura 40

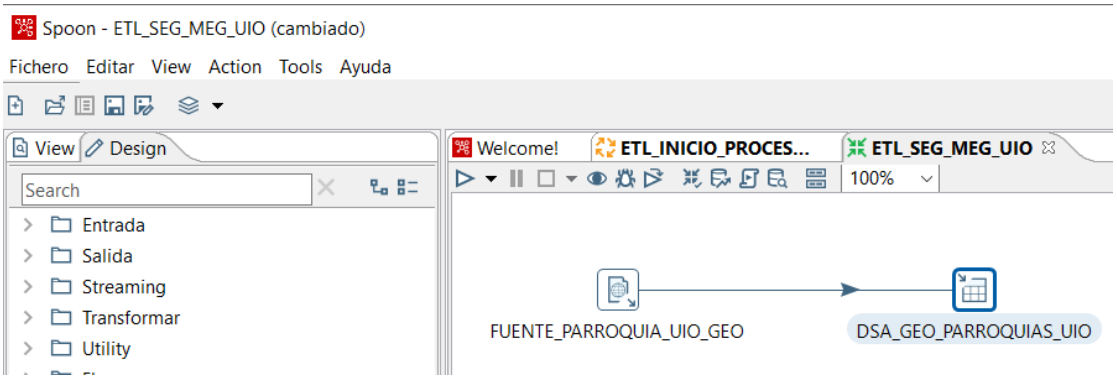
Almacenamiento de datos en esquema destino dsa.cart_parr_uio



La transformación de la tabla final del esquema de almacenamiento temporal se refleja en la Figura 41, la cual muestra los dos componentes necesarios para el desplazamiento de los datos.

Figura 41

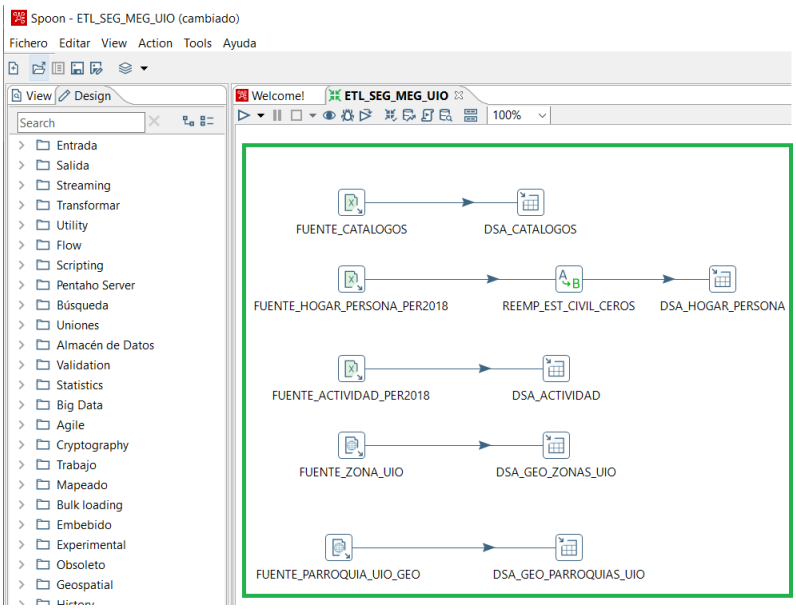
Transformación para la tabla temporal dsa.cart_parr_uio



Una de las características del esquema (DSA) es que todas las tablas que reciben información de las diferentes fuentes vacían su información para volverlas a llenar. Este proceso es de gran relevancia puesto que, al ser un espacio de almacenamiento temporal, los datos transitan de un lugar a otro con el objetivo de ser depurados y procesados de la mejor manera. La Figura 42, contempla la transformación de todas y las entidades que forman parte del esquema temporal de almacenamiento.

Figura 42

Extracción, transformación y carga para el esquema DSA



El esquema DSA, ha permitido tener un panorama claro y estructurado de los datos provenientes de la fuente. Gracias a su lógica de temporalidad se pudo ejecutar actividades de limpieza y depuración de los mismos, con el fin de poblar el modelo dimensional. En base a nuestra lógica de procesamiento, una vez poblado el esquema DSA se procedió a poblar las tablas del esquema dimensional para su posterior análisis.

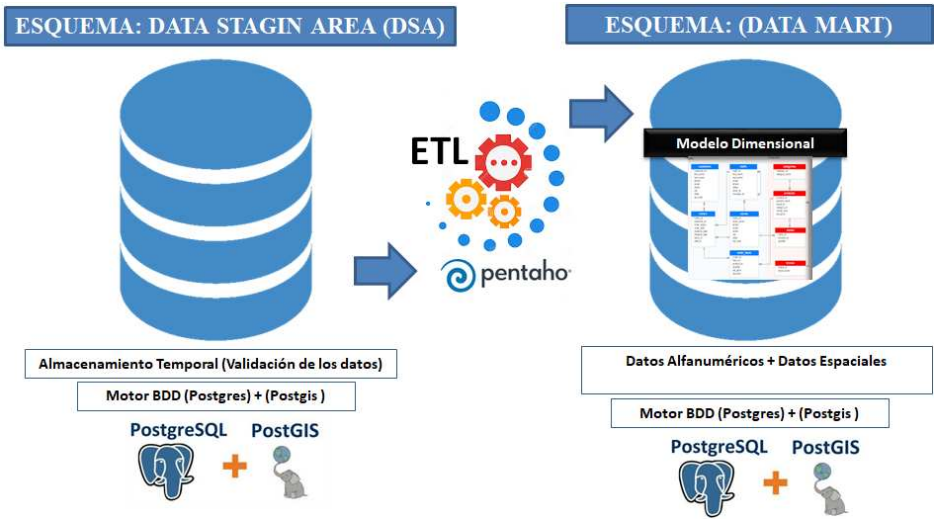
Extracción transformación y carga para el esquema de análisis Data Mart

El esquema data mart es el contenedor de toda la lógica dimensional que permite visualizar los datos e interpretar la información resultante. En base a la lógica de desarrollo (ETL), el esquema DSA y DATA MART pueden alojarse en bases de datos diferentes; sin embargo, debido a que la cantidad de datos del proyecto no tiene un crecimiento substancial, se ha incorporado en una sola base de datos diferenciada en esquemas.

La Figura 43, hace referencia al tratamiento realizado a los datos para poblar nuestro esquema de análisis llamado dtmrt (data mart).

Figura 43

Lógica de extracción, transformación y carga para esquema Data Mart



A continuación se detalla el proceso realizado para el llenado del esquema dimensional del data mart.

El proceso de extracción, transformación y carga se lo ha dividido en tres partes: 1) procesamiento en tablas dimensionales estáticas, 2) procesamienot den las tablas dimensionales dinámicas y 3) procesamiento en las tablas de hechos. Se ha ejecutado

esta lógica debido a que el llenado de las tablas dimensionales es más complejo y requieren un tratamiento especial. En base a nuestra propuesta se ha creado tres Jobs en el cual cada uno, ejecuta las transformaciones necesarias.

A) Procesamiento de datos en tablas de dimensiones estáticas

En base al modelo desarrollado y la estructura lógica de los datos, se ha realizado un proceso de carga para aquellas tablas que contienen datos categóricos fijos; es decir, su contenido no cambia y sus valores a nivel de variables se mantienen.

Las tablas consideradas como dimensiones estáticas son: dim_frecuencia_uso_bici, dim_medios_traslado y dim_razon_medio_traslado del esquema data mart, ver Figura 44.

Figura 44

JOB; extracción carga y transformación para las tablas de dimensión estáticas

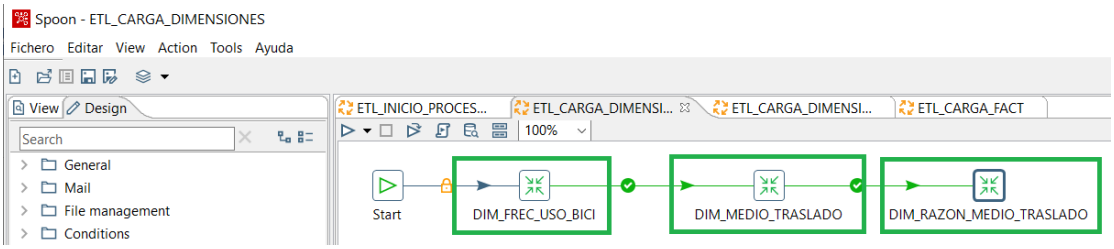


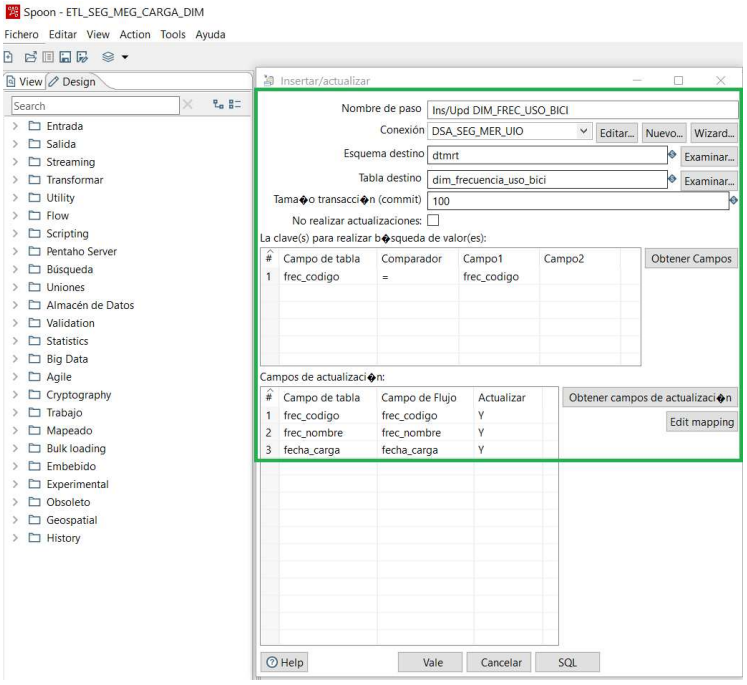
Tabla data mart: dtmrt.dim_frecuencia_uso_bici

Las tablas de dimensiones conllevan un tratamiento singular al momento de almacenar sus datos. Es necesario tomar en cuenta que a diferencia del esquema de almacenamiento temporal en el cual los datos se limpian cada vez que se realiza una inserción, el llenado de las tablas de dimensión requiere la ejecución de un merge.

La instrucción merge básicamente une datos de un resultado de origen establecido en una tabla destino. Se envían los datos al merge, este evento los compara (por la llave primaria). Si existe el registro, lo actualiza y si no existe, lo ingresa como nuevo registro. En la Figura 45 se puede verificar la transformación realizada para ejecutar un merge a la tabla dtmrt.dim_frecuencia_uso_bici. En este proceso, el identificador único a comparar es el campo frec_codigo, mediante este campo se verificó el contenido de la tabla y se realizó la comparación de registros.

Figura 45

Merge para la tabla dtmrt.dim_frecuencia_uso_bici



La transformación que requiere la ejecución de un merge varía en el número de componentes. Para la ejecución de este tipo de transacción recurrimos al uso de dos entradas de datos. La primera entrada debe comparar los registros de la segunda entrada, se ha incluido un ordenador (order by) de registros, para mantener el orden de

comparación y finalmente se ha agregado el componente de actualización e inserción para poblar las tablas destino. La Figura 46 muestra los componentes necesarios para la ejecución de un merge.

Figura 46

Transformación merge para dimensión dtmrt.dim_frecuencia_uso_bici

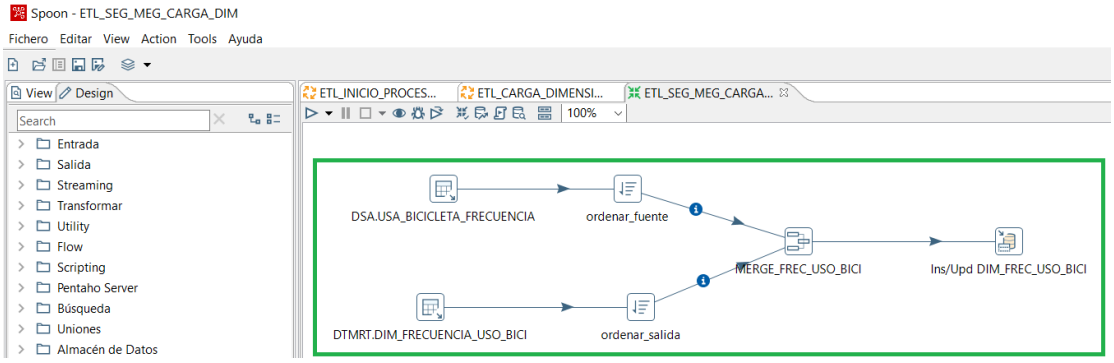
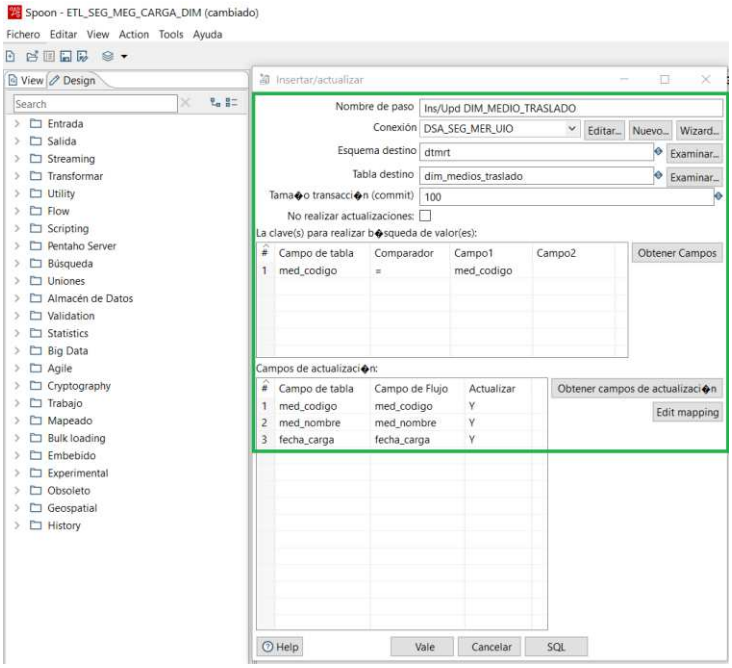


Tabla data mart dtmrt.dim_medios_traslado

Esta tabla almacena los datos de los diferentes medios de transporte, entre ellos la bicicleta. Cabe recalcar que entre las personas encuestadas hay quienes destacan las diferentes formas de trasladarse desde un lugar a otro. La lógica de transformación para esta tabla sigue el mismo proceso ejecutado en la dimensión del data mart dim_frecuencia_uso_bici. El campo identificador clave para la ejecución del merge es el med_codigo, este campo permite realizar la comparación a nivel de registros para distinguir el tipo de evento en la base de datos (insert o update), tal como se muestra en la Figura 48.

Figura 48

Merge para la tabla dtmrt.dim_medio_traslado



La transformación de la dimensión para especificar el medio de transporte que comúnmente usan las personas para poder movilizarse de un lugar a otro se puede apreciar en la Figura 47.

Figura 47

Transformación merge para dimensión dtmrt.dim_medio_traslado

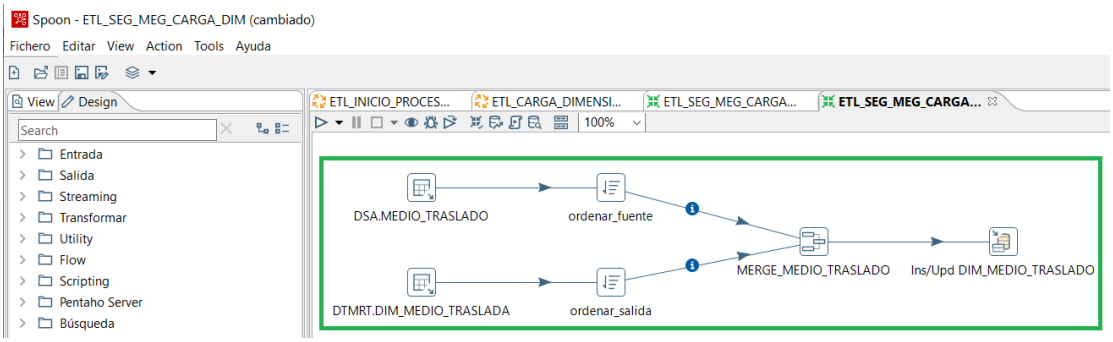


Tabla data mart dtmrt.dim_razon_medio_traslado

La dimensión dim_raz_medio_traslado es el complemento a la dimensión del tipo de transporte, esta tabla almacena los datos que ayudan a identificar la razón por la cual se da preferencia a un medio de desplazamiento en específico. La transformación trabaja en la misma lógica que el resto de las dimensiones estáticas en lo cual predomina el uso del merge para su actualización.

B) Procesamiento de datos en tablas de dimensiones dinámicas

A diferencia de las tablas de dimensión estáticas las cuales almacenan valores de variables fijas como nombres de catálogos, se tiene también dimensiones dinámicas las cuales se conforman de datos cambiantes y prácticamente forman parte del proceso lógico del negocio. Las tablas consideradas como dimensiones dinámicas son: dim_hogar_persona y dim_uio_parroquia.

Se ha desarrollado un job que contiene dos transformaciones, la primera se encarga de llenar los datos de las dimensiones dinámicas, que hacen referencia a los atributos de las personas encestadas. La segunda tiene como objetivo poblar de datos la tabla que contiene las características de las parroquias de Quito, la Figura 49 muestra la estructura del Job desarrollado.

Figura 49

Job; extracción carga y transformación para las tablas de dimensiones dinámicas

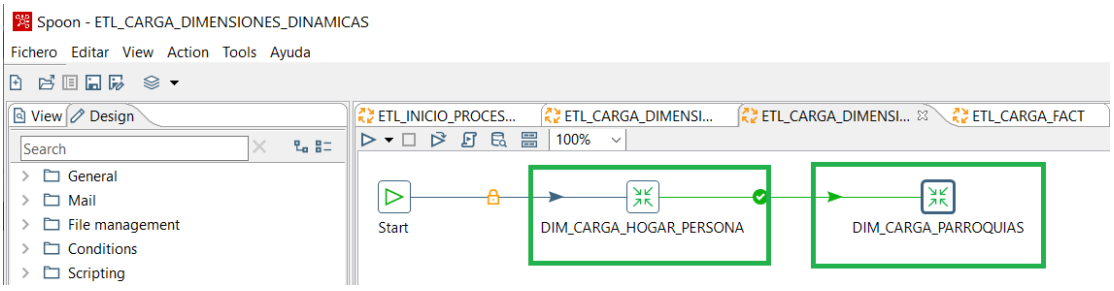
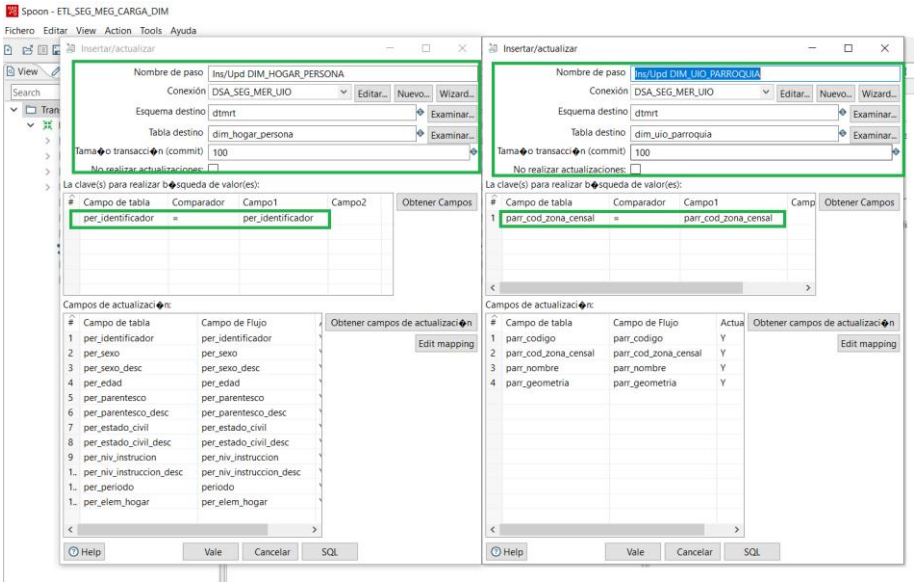


Tabla data mart dtmrt.dim_hogar_persona y dtmrt.dim_uio_parroquia

Las dos dimensiones dinámicas, caracterizadas por su concurrencia de datos se llenan en base a la lógica de transaccionalidad merge, estas entidades conforman la base del BI puesto que involucran la lógica del negocio y alojan el principal sustento de análisis. La principal característica de estas tablas es la cantidad de datos de almacenamiento puesto que su nivel de concurrencia es mayor a las entidades estáticas, en la Figura 50 se puede apreciar los merge para las dos entidades de contenido dinámico.

Figura 50

Merge para las tablas dtmrt.dim_hogar_persona y dtmrt.dim_uio_parroquia



C) Procesamiento de datos en tablas de hechos

Las tablas de hechos son la parte medular del modelo dimensional, estas tablas contienen campos claves que guardan relación con las tablas de dimensiones y reflejan

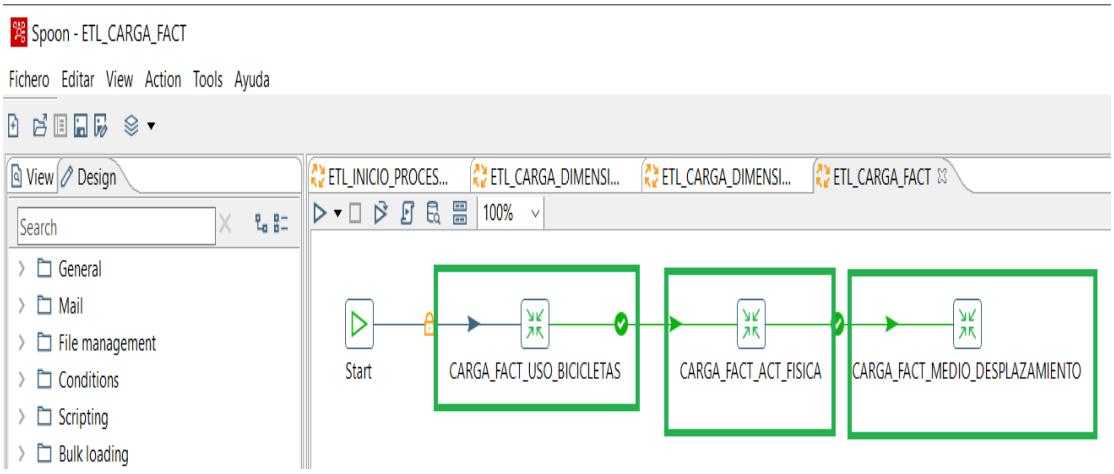
aquello que deseamos medir o analizar. Es importante tomar en cuenta que los hechos equivalen a la representación de un proceso del negocio.

Las tablas consideradas como hechos son: fact_uso_bicicletas, fact_actividad_fisica y fact_trans_trasladarse. Estas tablas estan ubicadas en la base de datos bajo el esquema data mart (dtmrt).

Los ETL de las tablas de hechos incluye el desarrollo de un Job dividido en tres transformaciones por cada tabla, tal como se muestra en la Figura 51.

Figura 51

JOB; extracción carga y transformación para las tablas de hechos



Tablas de hechos dtmrt.fact_uso_bicicletas: Permite analizar el número de personas y sus características, tomando en cuenta su ubicación geográfica a nivel de parroquia en el perímetro urbano de Quito. Esta información refleja estratégicamente los nombres de las parroquias en las cuales habitan personas que hacen uso de las bicicletas para entretenerse y realizar ejercicio físico.

La transformación realizada para llenar de datos la tabla de hechos se puede verificar en la Figura 52.

Figura 52

Extracción carga y transformación para la tabla de hechos dtmrt.fact_uso_bicicletas

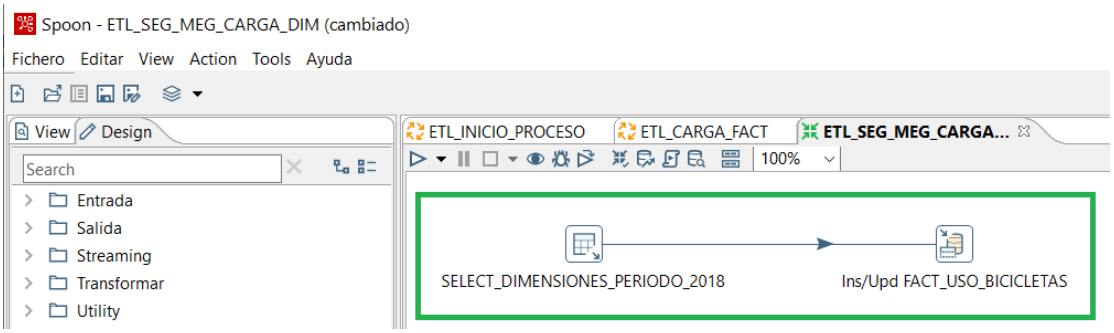


Tabla de hechos dtmrt.fact_actividad_fisica: Permite analizar el número de personas que realizan actividad física y su ubicación geográfica a nivel de parroquia. Estos datos podrían ser de gran utilidad puesto que podrían existir parroquias en los cuales existen grandes grupos de personas que practican deportes y potenciales usuarios de bicicletas. La Figura 53, muestra los componentes utilizados para llenar la tabla de hechos.

Figura 53

Extracción carga y transformación para la tabla de hechos dtmrt.fact_actividad_fisica

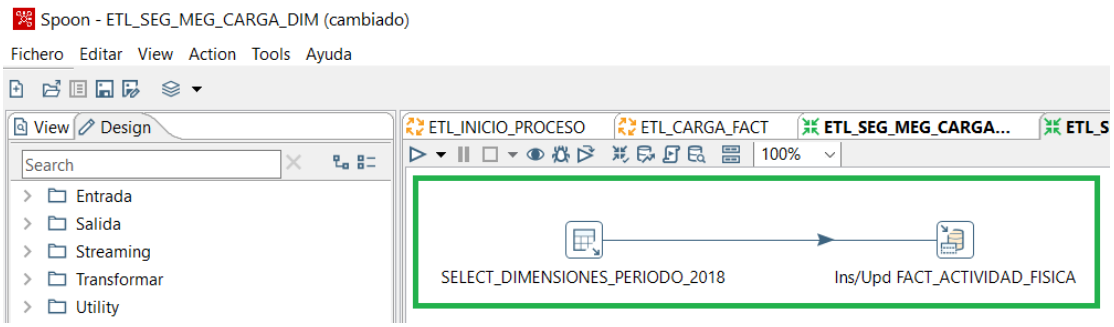
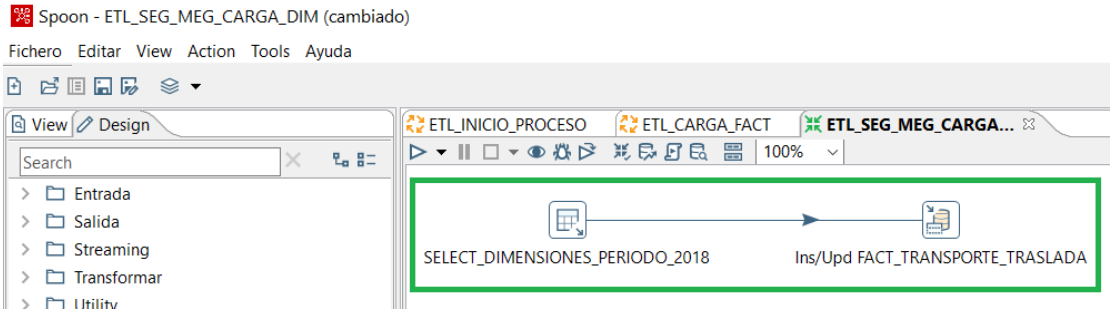


Tabla de hechos dtmrt.fact_trans_trasladarse: Permite analizar el número de personas que usan la bicicleta como un medio de desplazamiento. Al hablar de desplazamiento se destaca la utilización de la bicicleta para llegar a su respectivo lugar de trabajo o a un establecimiento educativo en caso de estudiantes. Esta información podría permitir identificar un segmento en el cual se tiene un amplio número de usuarios que utilizan bicicletas y potenciales clientes para la compra de accesorios y repuestos. La Figura 54 muestra los componentes utilizados para llenar la tabla de hechos en mención.

Figura 54

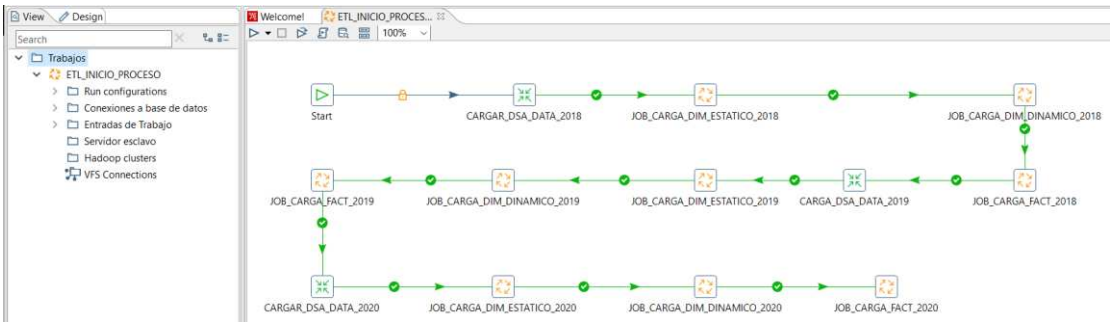
Extracción carga y transformación para la tabla de hechos dtmrt.fact_transporte_traslada



El levantamiento de la encuesta multipropósito de hogares se ha efectuado desde el 2018, por tanto, sus datos están publicados en el banco de datos del Instituto Nacional de Estadísticas y Censos en tres periodos 2018, 2019 y 2020, el proceso total de extracción, transformación y carga para los tres periodos se puede visualizar en la Figura 55.

Figura 55

Extracción, carga y transformación para el proyecto SEGDO_BICI_UIO



El producto resultante de todo el proceso ETL, es el poblado de datos en las tablas que conforman el esquema dimensional del datamart. Este esquema estructurado en base a la lógica de negocio es el responsable de alojar las características de los requerimientos para su representación en base a una visualización gráfica.

Desarrollo y visualización del BI (dashbord)

Parte del objetivo del proyecto SEGDO_BICI_UIO es la construcción del dashboard o panel de control. Esta herramienta es la encargada de mostrar los datos transformados en información de una manera amigable y legible mediante el uso de gráficos y tablas. A continuación, se detalla el proceso de construcción y el software necesario utilizado en su desarrollo.

Construcción y análisis de cubos OLAP (On-Line Analytical Processing)

Los cubos OLAP permiten un uso eficaz de las bodegas de datos para el análisis en línea, y proporcionan respuestas rápidas a consultas analíticas complejas e iterativas. En la construcción de los cubos OLAP se utilizó la herramienta de Pentaho (Schema Workbench), esta herramienta permite estructurar un cubo a partir de un modelo dimensional.

En Schema Workbench es posible construir el cubo OLAP en base a una lógica aplicada al modelo dimensional. La herramienta permite diferenciar entre tablas de hecho y dimensiones tomando en cuenta la propuesta de análisis definida en los requerimientos.

Cubo OLAP Uso de bicicletas

El cubo que hace referencia al uso de las bicicletas en el perímetro urbano de Quito, se encuentra relacionado al modelo dimensional número 2 y entre las principales características de análisis se tiene la obtención del número de usuarios que usan bicicleta tomando en cuenta las siguientes combinaciones:

_Conteo de usuarios que hacen uso de la bicicleta desagregado a nivel de parroquia, combinado con el sexo de las personas.

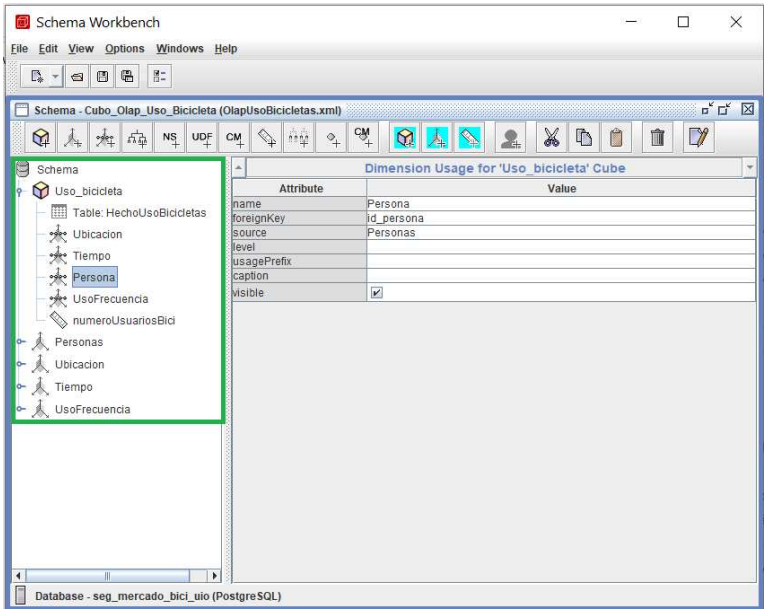
_Conteo de usuarios que hacen uso de la bicicleta desagregado a nivel de parroquia y combinado con su frecuencia de uso.

_Conteo de usuarios que usan bicicleta y su desagregación por nivel de instrucción.

La Figura 56 muestra la estructura del cubo OLAP en la interfaz de Schema Workbench. Aquí se puede referenciar gráficamente las medidas, dimensiones y hechos que permitirán el análisis de los datos referentes al número de personas que usan la bicicleta.

Figura 56

Cubo OLAP uso de la bicicleta



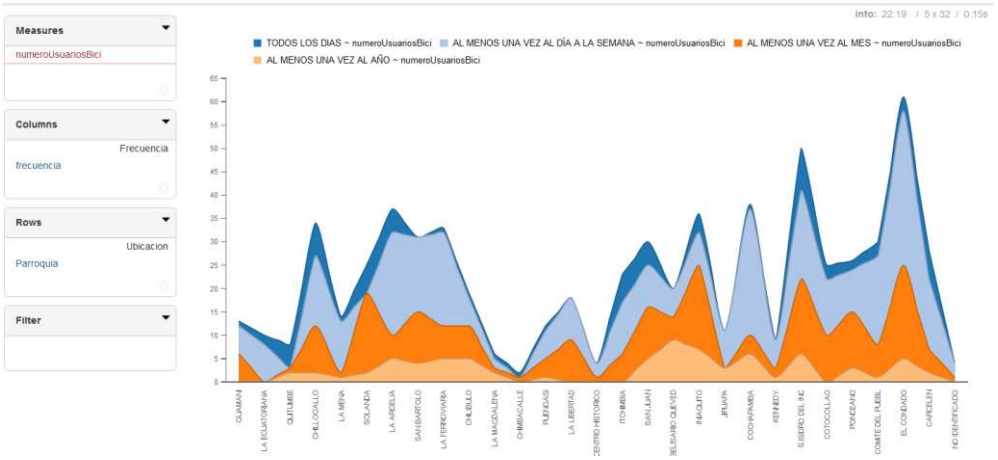
Análisis: Cubo OLAP uso de la bicicleta

La construcción de los cubos OLAP son de gran ayuda al momento de analizar los datos. En base a su estructura se puede mostrar la información requerida y su comportamiento. Para realizar el análisis de los diferentes modelos dimensionales desarrollados, se utilizó el software Saiku Analytics en su versión open source. Este sistema permite sondear los datos de manera gráfica y definir los indicadores que serán incluidos en el dashboard.

La Figura 57 representa gráficamente a nivel de parroquia el número de personas que utilizan la bicicleta y una definición de su frecuencia de uso, esta característica de tiempo se mide en base a 4 parámetros: a) todos los días, b) al menos una vez al día a la semana, c) al menos una vez al mes, d) al menos una vez al año. Este análisis nos podría ayudar a identificar las parroquias con ciclistas que hacen uso de la bicicleta de manera diaria y desarrollar una estrategia para brindar un servicio a este tipo de usuarios.

Figura 57

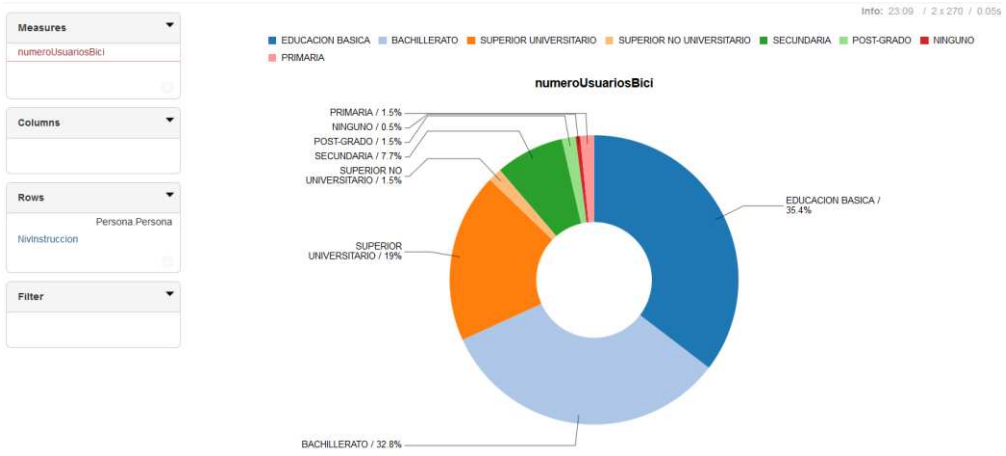
Frecuencia de uso de la bicicleta por parroquia



El nivel de instrucción de la persona es un dato presente en la encuesta multipropósito, este dato puede ser de interés puesto que nos podría dar una referencia de las personas que usan bicicleta y su perfil en cuanto a su nivel académico. La Figura 58 refleja la cantidad de personas que practican bicicleta combinado con el nivel de instrucción que poseen: a) superior universitario, b) secundaria, c) educación básica, d) bachillerato, e) superior no universitario, f) primaria, g) post grado.

Figura 58

Personas que usan bicicleta por nivel de instrucción



Cubo OLAP actividad física

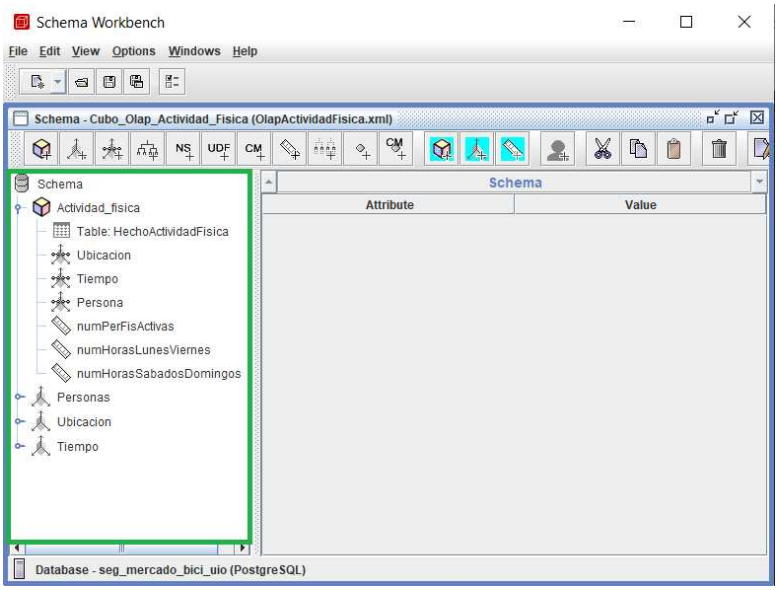
La actividad física es un factor determinante en la salud de una persona, una de las variables estudiadas en la encuesta multipropósito es el tiempo que una persona dedica a la práctica de ejercicio físico en el transcurso de la semana, si bien es cierto al hablar de ejercicio físico implicaría correr, trotar, hacer bicicleta etc., sin embargo consideramos interesante desarrollar un cubo con estos datos. La Figura 59, representa el modelo dimensional número 3 y el objetivo es mostrar un indicador con las siguientes características:

__Conteo de personas que hacen ejercicio físico desagregado a nivel de parroquia

combinado con su género.

Figura 59

Cubo OLAP actividad física

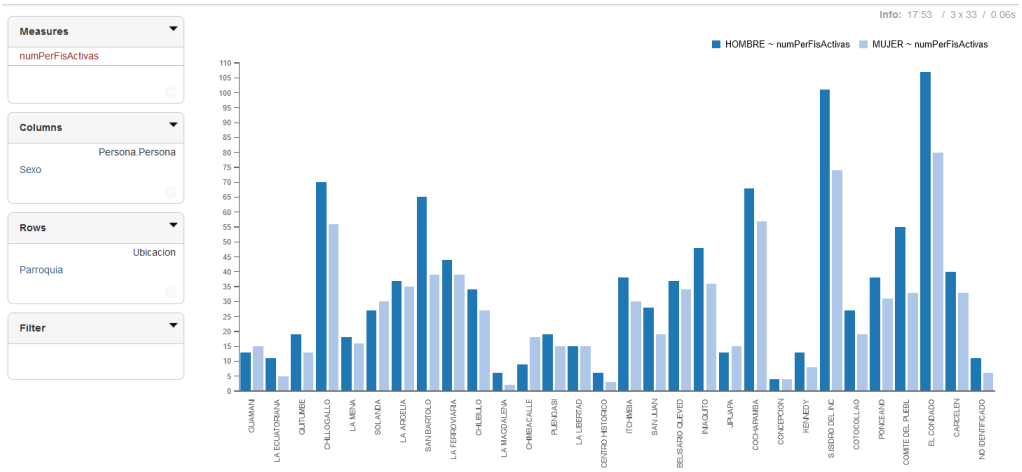


Análisis: Cubo OLAP actividad física

La representación gráfica del indicador que analiza el número de personas que realizan alguna actividad física desagregada a nivel de parroquia y género se muestra en la Figura 60.

Figura 60

Personas que hacen ejercicio por parroquia y sexo



Cubo OLAP medios de desplazamiento y la bicicleta como alternativa

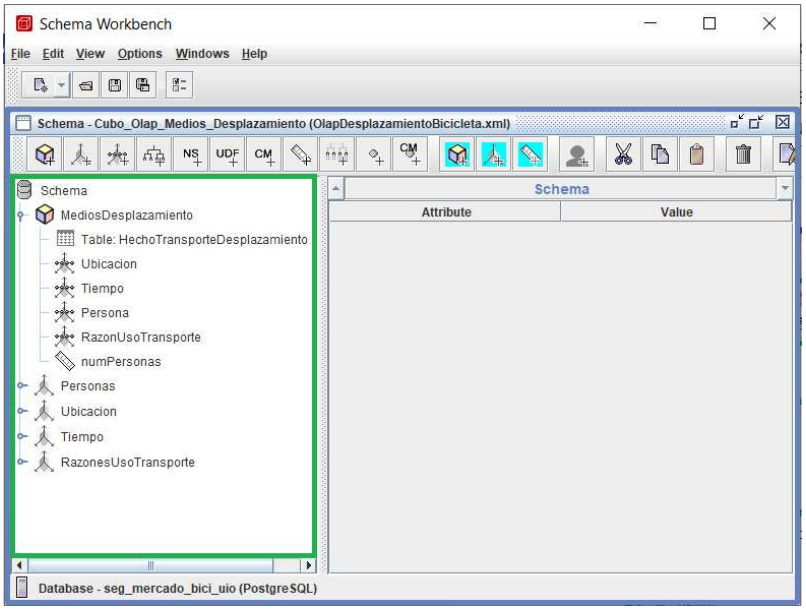
La encuesta multipropósito incluye en su estudio al uso de la bicicleta como una herramienta destinada a la ejecución de actividades físicas que combina salud con entretenimiento, y también el análisis de la bicicleta como un medio de transporte para trasladarse de un lugar a otro. Existe también entre las variables de estudio la razón por la cual se da preferencia al uso de bicicleta como un medio de transporte, mediante estos datos se ha construido un cubo que permite desarrollar un análisis con las siguientes características.

_El número de personas que usan la bicicleta como un medio de transporte para trasladarse de un lugar a otro a nivel de parroquia y las razonas por la cual prefieren este medio.

La Figura 61, hace referencia a la visualización del cubo en cuanto a estructura mediante la herramienta Schema Workbench.

Figura 61

Cubo OLAP medios de desplazamiento y la bicicleta como alternativa



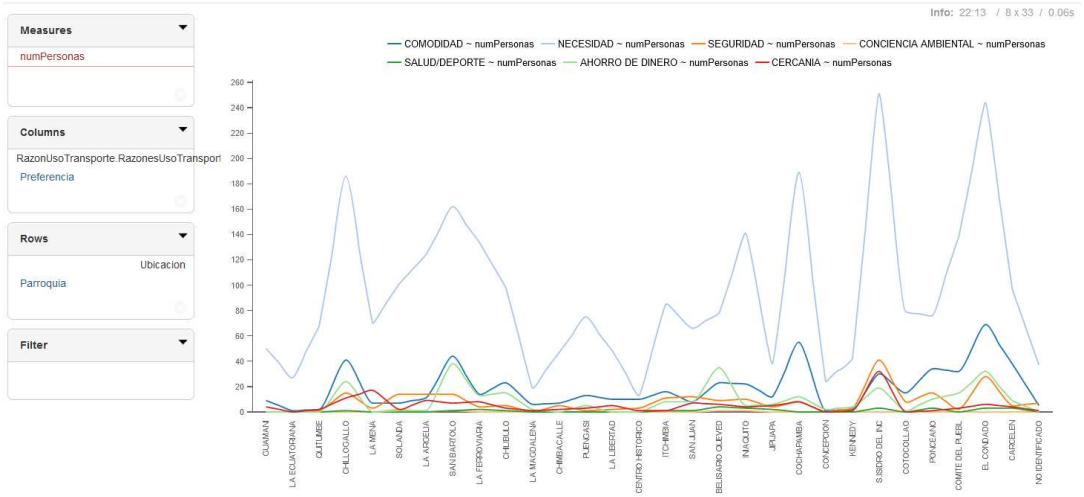
Análisis: Cubo OLAP medios de desplazamiento y la bicicleta como alternativa

En varias parroquias de la ciudad existen personas que hacen uso de la bicicleta como una forma de transportarse, a esta actividad se le suma la razón por la cual prefieren esta modalidad; la necesidad, seguridad, o la concientización ambiental son las preferencias de los usuarios.

La Figura 62 refleja el análisis del cubo construido en base al modelo dimensional 3 que hace referencia al número de usuarios por parroquia que usan la bicicleta como un medio de transporte y la razón por la que prefiere este medio.

Figura 62

Razones para el uso de la bicicleta como medio de transporte por parroquia



Gracias al uso de Schema Workbench, se pudo realizar el modelado de varios cubos en base al esquema dimensional desarrollado, sin embargo, la posibilidad de analizar estas nuevas formas de representación de información fue a base del software Sayku Analitics, con el cual se pudo evidenciar gráficamente los indicadores formados en base a dimensiones, medidas y hechos. Una vez analizados todos los cubos y con la idea clara de su representación a nivel gráfico, se procedió a desarrollar el dashboard o panel de control, incluyendo en su estructura todos los indicadores alineados a los requerimientos

Construcción del dashboard (panel de control)

El análisis de todos los cubos OLAP mediante la herramienta Visualizer permite tener un panorama claro de los indicadores a incluir en el dashboard. Gracias a esta herramienta que tiene la capacidad de interpretar modelos dimensionales de una manera gráfica, se puede establecer la estructura de un indicador en base a sus dimensiones y medidas. Una vez analizados los cubos y con una idea clara de los indicadores a mostrar, se procede a construir el dashboard o panel de control. La herramienta Visualizer de Datafor permite diseñar y personalizar paneles de control en base a los cubos desarrollados con Pentaho. Esta herramienta se obtiene como un plugin de Pentaho server y puede ser descargada e instalada desde el marketplace o desde su página oficial: http://www.datafor.com.cn/

La ventaja de Visualizer es que su interfaz es muy amigable e intuitiva. La manera de diseñar un panel de control se basa en el arrastre de componentes y su personalización mediante parametrización de forma, tamaño, color etc.

La estructura de nuestro panel de control se basa en la inclusión de los diagramas analizados con la herramienta Visualizer, cada indicador representa información alineada al requerimiento. Tomando en cuenta que a partir de nuestro data mart, se desarrollaron tres cubos de análisis, cada uno fue evaluado adoptando las diferentes temáticas de interés para un criterio de segmentación de mercado. El resultado de esta evaluación fue representado en un gráfico que incluye variables de tiempo y del negocio, los indicadores incluidos en el panel de control son los siguientes:

_Personas que usan bicicleta por parroquia y sexo: Permite visualizar información de las parroquias con afluencia de ciclistas y reconocer si existen más hombres o mujeres.

_Frecuencia de uso de la bicicleta por parroquia: Permite conocer en que parroquias de la ciudad de Quito, las personas usan bicicleta de forma diaria, al menos un día a la semana, al menos una vez al mes o al menos una vez al año.

_Personas que usan bicicleta por edad: Permite distinguir las edades en las cuales las personas aprovechan el uso de la bicicleta.

_Personas que hacen ejercicio por parroquia y sexo: Este indicador no tiene una estrecha relación con el uso de la bicicleta, sin embargo, me permite conocer las parroquias físicamente activas y potenciales usuarios de bicicletas.

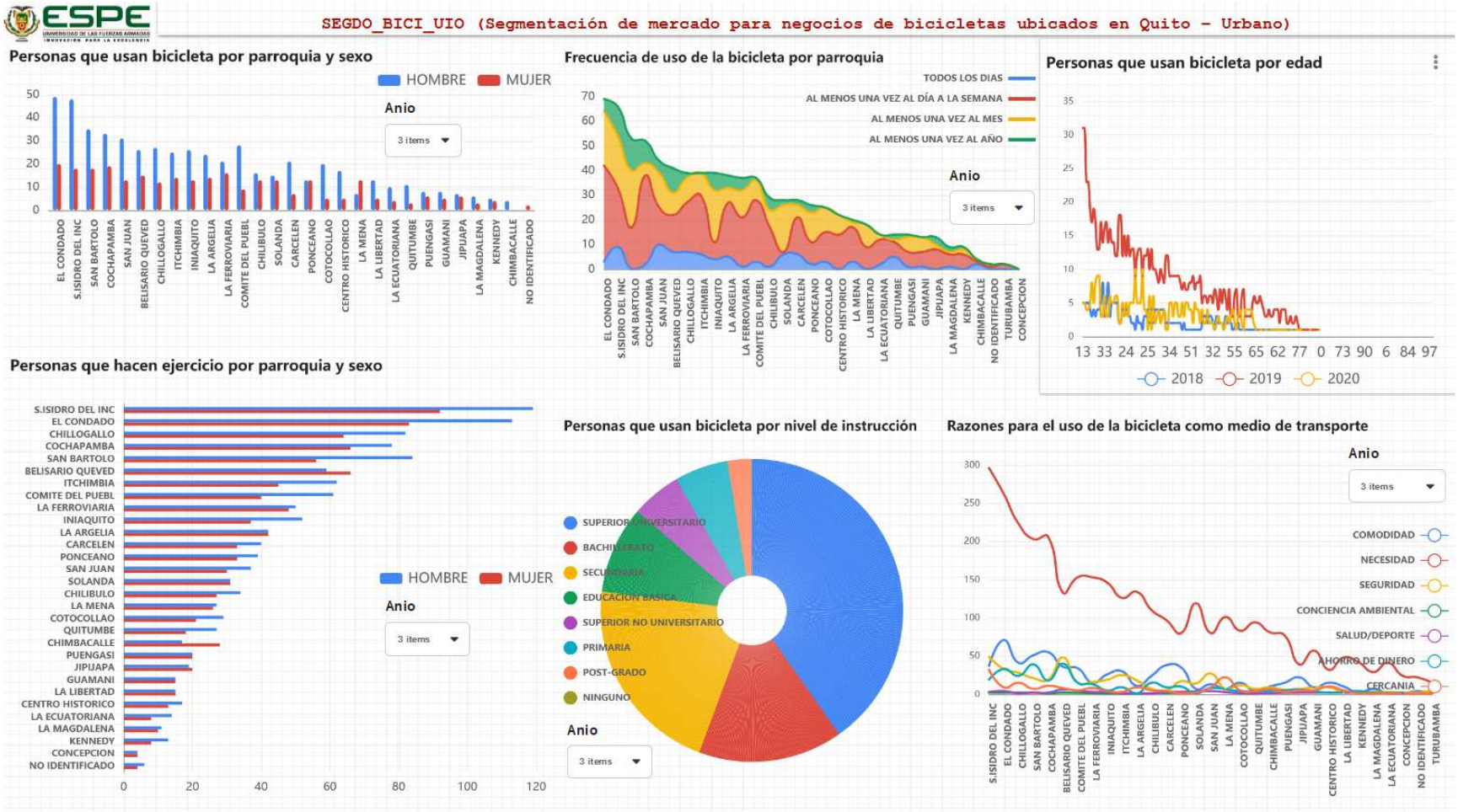
_Personas que usan bicicleta por nivel de instrucción: Permite conocer el tipo de usuarios que comúnmente utiliza una bicicleta, esta diferenciación se la realiza en base al siguiente catálogo; a) superior universitario, b) segundaria, c) educación básica, d) bachillerato, e) primaria, f) post grado, g) ninguno.

_Razones para el uso de la bicicleta como medio de transporte: A diferencia del indicador que resalta la medida de usuarios que practican bicicleta por entretenimiento, este indicador permite analizar a las personas que usan la bicicleta, como un medio de desplazamiento para trasladarse a un lugar de trabajo o estudio y conocer las razones por la cual prefiere este medio de transporte.

La Figura 63, refleja la estructura del dashboard construido para nuestro análisis de segmentación de mercado dirigido a negocios dedicados a las bicicletas.

Figura 63

Dashboard o panel de control SEGDO_BICI_UIO



Planteamiento del algoritmo

En este capítulo se desarrolla un algoritmo de clusterización (ver Anexo 2), utilizando la metodología CRISP-DM para minería de datos con el objetivo de realizar un análisis de segmentación de mercado utilizando el BI del proyecto SEGDO_BICI_UIO.

El principal objetivo del proyecto SEGDO_BICI_UIO es desarrollar un visualizador de datos o panel de control, que permita visualizar información referente al uso de la bicicleta en el perímetro urbano de Quito. Esta información permite brindar soporte, para aquellos emprendedores dedicados a este giro de negocio, que requieren definir un nicho de mercado al cual atacar.

Si bien es cierto los indicadores presentados en el dashboard pueden ser un gran referente al momento de analizar y establecer un segmento de mercado, en la práctica comprobar su efectividad puede ser subjetiva. Demostrar que la instalación de un negocio, orientado a las bicicletas, en una parroquia determinada de la ciudad, en base a la información del visualizador puede tardar mucho tiempo. Para realizar esta comprobación tendríamos que monitorear y dar seguimiento al emprendedor evaluando su rentabilidad y comparándola con otro emprendedor que haya iniciado su negocio sin la ayuda de nuestro proyecto.

En base a uno de los objetivos del proyecto se ha decidido realizar una comprobación teórica mediante el uso de minería de datos. La finalidad de la minería es cotejar la perspectiva que brinda el visualizador con un algoritmo de clusterización que permita identificar posibles clientes.

La metodología usada para el desarrollo de estas actividades se denomina CRISP-DM y se detalla a continuación.

Comprensión del negocio

En base al proceso realizado en el desarrollo del BI, la minería de datos ha permitido complementar el criterio de análisis, para el uso de la bicicleta, en las parroquias urbanas del distrito metropolitano de Quito. Dado que el sistema de BI nos ha brindado la facilidad de analizar los datos, de una manera visual mediante indicadores basados en requerimientos. El objetivo de la minería es otorgar un valor agregado al análisis de la información, mediante algoritmos que complementan la toma de decisiones, al momento de escoger un nicho de mercado apto, para el giro de negocios orientado a las bicicletas.

Comprensión de los datos

En base al modelo dimensional del proyecto SEGDO_BICI_UIO representado en la Figura 22 y que hace referencia a los data mart de análisis. Se adoptó el desarrollo de un modelo de minería de datos no supervisado denominado clustering. Este modelo permitió identificar y establecer grupos en base a su valor medio más cercano. Los datos utilizados para el desarrollo del cluster fueron el número de personas entre hombres y mujeres que usan la bicicleta como entretenimiento combinado con la frecuencia de uso, desagregado a nivel de parroquias.

Preparación de los datos

Una de las ventajas de realizar un proceso de minería de datos sobre un esquema de datos dimensional, es la agilidad presente al momento de manipular la información mediante la creación de vistas y consultas. Otro factor importante es la calidad, puesto que al data mart los datos llegan después de haber pasado por un proceso de limpieza y depuración.

Para la creación del modelo se utilizó R - Studio como herramienta de minería y PostgreSql como gestor de almacenamiento. Mediante una conexión ODBC (Open Data

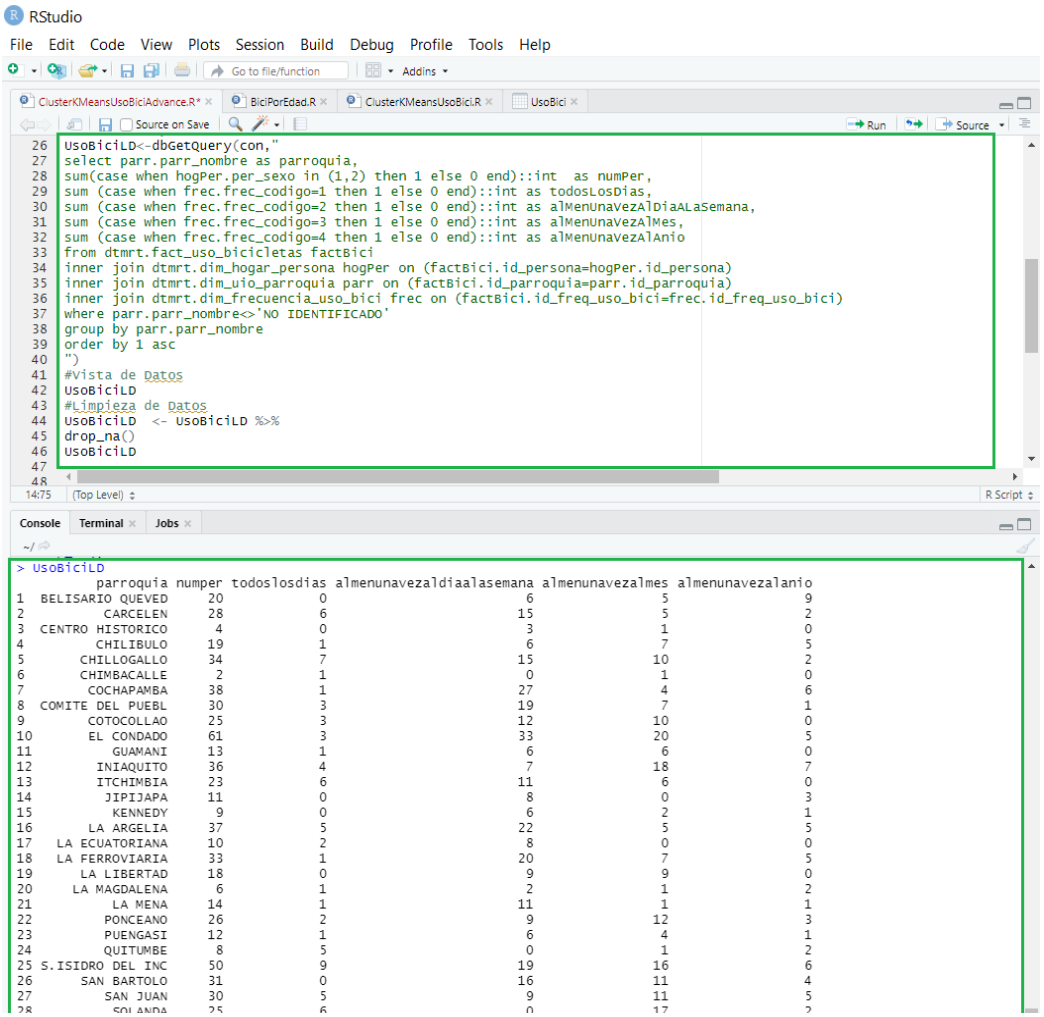
Base Connectivity) y la importación de ciertas librerías de R, se estableció la comunicación entre las dos herramientas. Los datos de entrada fueron obtenidos del data mart mediante una consulta y sirvió para dar origen al modelo de minería. La Figura 64 refleja la interfaz de la herramienta R-Studio y la visualización de los datos mediante el uso de las bibliotecas: 1) library (DBI), 2) library (RODBC), 3) library (odbc). La lógica de conexión a la base de datos se realiza mediante la ejecución de comandos,

los cuales activan ciertas funcionalidades para mostrar en pantalla el resultado deseado.

Los datos del conjunto, se reflejan en la parte inferior de la Figura 64, la tabla muestra el conteo de personas que hacen uso de la bicicleta entre hombres y mujeres,

Figura 64

Visualización de Datos (Conexión entre PostgreSql y R-Studio)



con la frecuencia de uso desagregado entre: a) Todos los días, b) Al menos una vez al día a la semana c) Al menos una vez al mes y d) Al menos una vez al año.

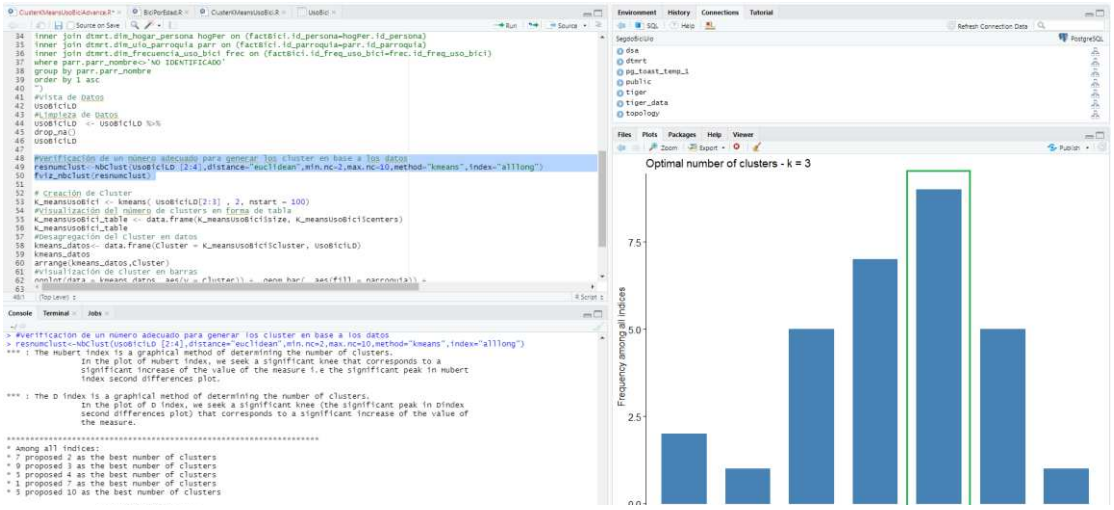
Modelamiento

La verificación y comprobación del sistema de BI se lo realizó con el algoritmo de minería K-means. Este método de clustering, que busca generar grupos en base a características similares, es uno de los más conocidos y usados, debido a su dinamismo y popularidad. Basados en una consulta generada al data mart, en la cual se tiene una matriz con datos de parroquias, personas y frecuencia del uso de la bicicleta, se procedió a realizar el análisis del número adecuado de clustres, mediante la función "fviz_nbclust" de R.

La Figura 65 refleja la ejecución de esta función, la cual evalúo los datos y determino que el número adecuado de grupos que pueden formarse a partir de las variables seleccionadas y la cantidad de registros es 3.

Figura 65

Número de clústers del modelo K-means



Una vez determinado el número de clusters ideal para nuestro análisis, se procedió a ejecutar la función "Kmeans" la cual realizó el agrupamiento de las parroquias, en base a las características de nuestro conjunto de datos. La matriz

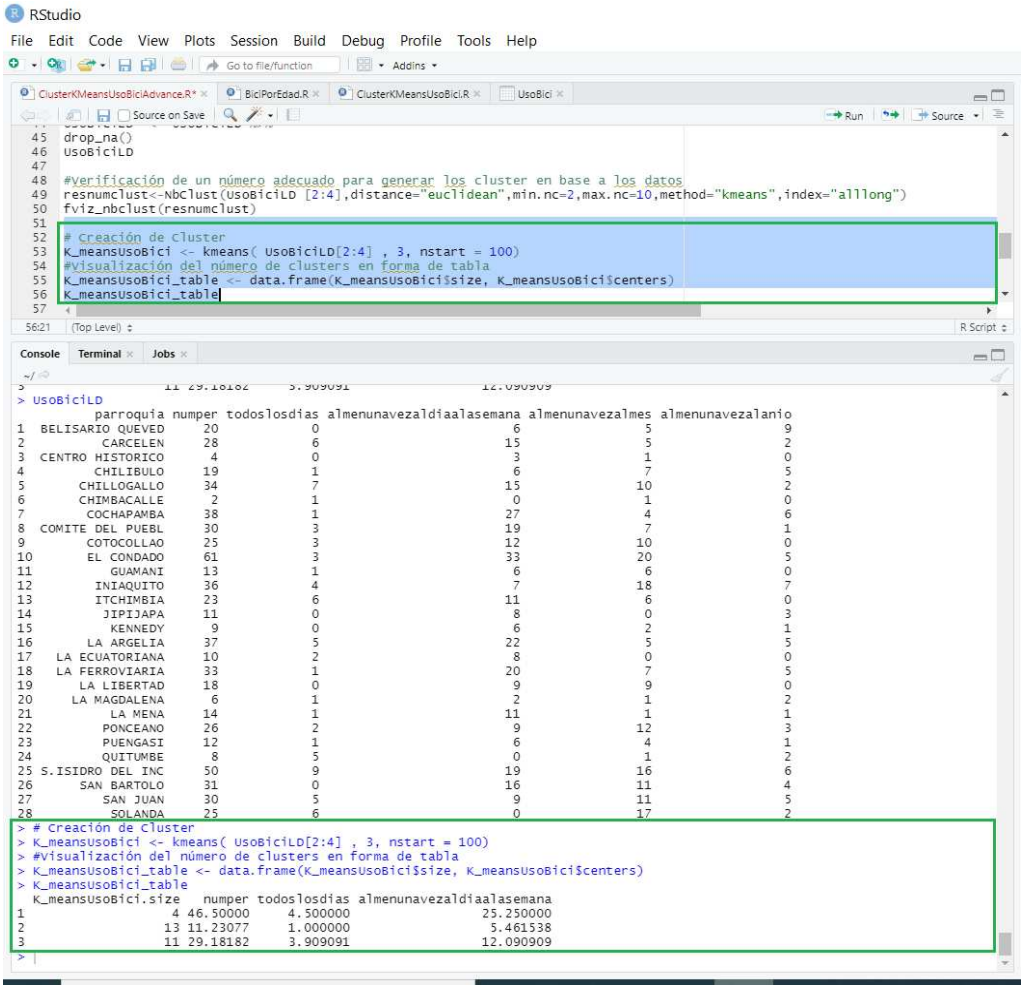
seleccionada para nuestro análisis se conformó por cuatro columnas; a) parroquias de Quito donde se usa la bicicleta, b) número de personas (hombres y mujeres), c) frecuencia de uso (todos los días), d) frecuencia de uso (al menos una vez al día a la semana).

Al aplicar la función Kmeans incluyendo como parámetro el número de clústeres=3 y las tres columnas numéricas de nuestro conjunto (2,3,4), se tiene una tabla resultante que incluye los clusters generados.

La Figura 66 refleja el resultado de aplicar Kmeans a nuestro conjunto de datos en una matriz conformada de tres filas y cuatro columnas con los valores generados por el algoritmo.

Figura 66

Resultado aplicación de modelo Kmeans

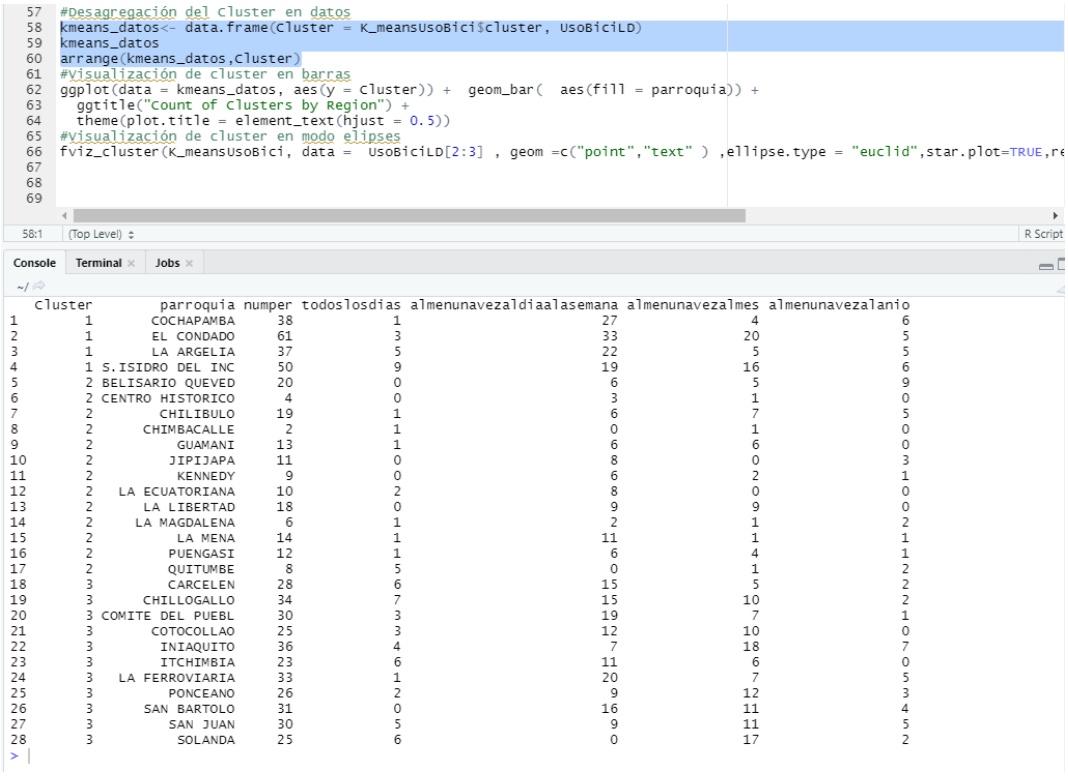


Dado que el resultado del cluster no es amigable al momento de interpretar su contenido, se procedió a generar una matriz con los clusters resultantes y los datos de nuestro conjunto. La función “data.frame” de R, permite construir una tabla en la cual se pudo diferenciar la agrupación de parroquias. La Figura 67 refleja el resultado de los 3 clusters que incluyen las variables parroquia, el conteo de personas que practican bicicleta, y su frecuencia de uso.

Los resultados arrojan los 3 grupos de parroquias que comparten similitud en cuanto a las cuatro variables analizadas en el conjunto de datos.

Figura 67

Clusters generados en base a conjunto de datos



```
#Desagregación del Cluster en datos
kmeans_datos<- data.frame(Cluster = K_meansUsoBici$cluster, UsoBiciLD)
kmeans_datos
arrange(kmeans_datos,Cluster)
#Visualización de cluster en barras
ggplot(data = kmeans_datos, aes(y = Cluster)) +  geom_bar(  aes(fill = parroquia)) +
  ggtitle("Count of Clusters by Region") +
  theme(plot.title = element_text(hjust = 0.5))
#Visualización de cluster en modo elipses
fviz_cluster(K_meansUsoBici, data =  UsoBiciLD[2:3] , geom =c("point","text" ) ,ellipse.type = "euclid",star.plot=TRUE,re
```

58:1 (Top Level) R Script

Console Terminal Jobs

```
   Cluster        parroquia numper todoslosdias almenunavezaldiaalasemana almenunavezalmes almenunavezalanio
1        1       COCHAPAMBA     38           1                        27                4                6
2        1       EL CONDADO     61           3                        33               20                5
3        1       LA ARGELIA     37           5                        22                5                5
4        1 S.ISIDRO DEL INC     50           9                        19               16                6
5        2 BELISARIO QUEVED     20           0                         6                5                9
6        2 CENTRO HISTORICO      4           0                         3                1                0
7        2        CHILIBULO     19           1                         6                7                5
8        2      CHIMBACALLE      2           1                         0                1                0
9        2          GUAMANI     13           1                         6                6                0
10       2         JIPIJAPA     11           0                         8                0                3
11       2          KENNEDY      9           0                         6                2                1
12       2   LA ECUATORIANA     10           2                         8                0                0
13       2      LA LIBERTAD     18           0                         9                9                0
14       2     LA MAGDALENA      6           1                         2                1                2
15       2          LA MENA     14           1                        11                1                1
16       2         PUENGASI     12           1                         6                4                1
17       2         QUITUMBE      8           5                         0                1                2
18       3         CARCELEN     28           6                        15                5                2
19       3      CHILLOGALLO     34           7                        15               10                2
20       3 COMITE DEL PUEBL     30           3                        19                7                1
21       3       COTOCOLLAO     25           3                        12               10                0
22       3        INIAQUITO     36           4                         7               18                7
23       3        ITCHIMBIA     23           6                        11                6                0
24       3   LA FERROVIARIA     33           1                        20                7                5
25       3         PONCEANO     26           2                         9               12                3
26       3      SAN BARTOLO     31           0                        16               11                4
27       3         SAN JUAN     30           5                         9               11                5
28       3          SOLANDA     25           6                         0               17                2
>
```

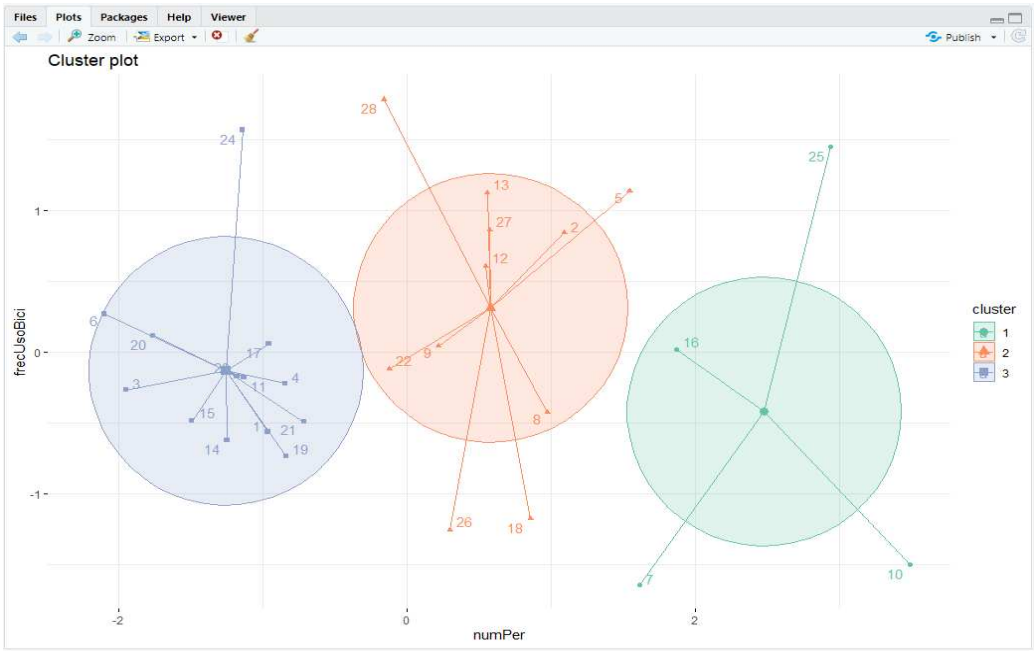
La representación gráfica de resultados es una de las ventajas que presenta R al momento de trabajar en procesamiento de datos, la función “ggplot” y “fviz_cluster” ayudan a realizar diagramas de los modelos aplicados.

La función “fviz_cluster” que permite construir diagramas con un enfoque alineado al modelo Kmeans, generó las agrupaciones a manera de conjuntos, tomando en cuenta sus centros. La Figura 68 representa tres conjuntos agrupados en base al conteo de personas practicantes del ciclismo y su frecuencia de uso. El eje X marca el número de personas y el eje Y la frecuencia de uso combinado entre a) Todos los días y

b) Al menos una vez al día a la semana. En el eje (X) a partir del centro cero hacia la derecha se expresa mayor número de personas que usan bicicleta, mientras que en el eje (Y) las parroquias que se acercan a cero son aquellas que utilizan la bicicleta con mayor frecuencia.

Figura 68

Gráfico de cluster (euclideo), agrupados por parroquia



En base al comportamiento de los datos según la gráfica; las parroquias San Isidro del Inca, El Condado, Cocha Pamba y La Argelia se caracterizan por tener un buen número de personas que practican bicicleta, sin embargo, no todas la usan de manera periódica. Parroquias como: San Juan, Iñaquito y Carcelén son parroquias con un buen número de personas que practican bicicleta y que además la usan frecuentemente. El resultado reflejo éxito en la ejecución del algoritmo, la selección de las variables y las funciones ejecutadas demuestran que el proceso realizado es apto para la ejecución de análisis mediante modelos de minería de datos.

En la etapa de evaluación se realizó el análisis de clusters, tomando en cuenta las variables de frecuencia de uso separando su periodicidad, esto permitió desarrollar un mejor análisis de segmentación puesto que al disminuir las variables de estudio la generación de grupos podría ser más específica.

Evaluación

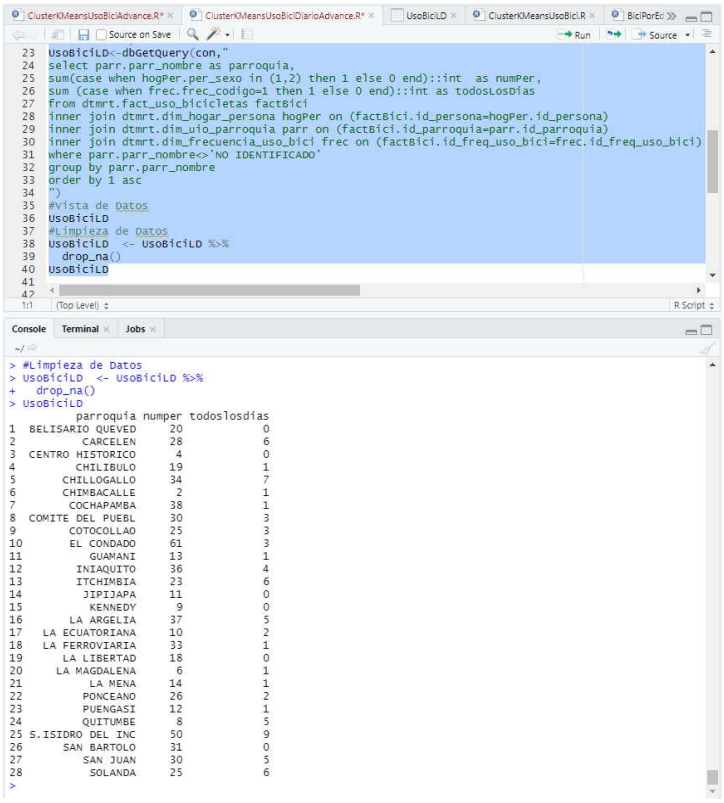
El proceso desarrollado para obtener los clusters ha permitido identificar las variables que pueden aportar mayor análisis y evaluación a nuestro módelo. Dado que la vista obtenida de nuestro data mart enfocada al análisis de personas que utilizan bicicleta y su frecuencia se ha desagregado en; a) frecuencia de uso (todos los días), b) frecuencia de uso (al menos una vez al día a la semana). Hemos decidido hacer una evaluación de los datos aplicando nuestro modelo en base a dos criterios de forma independiente.

Caso 1: Evaluación del modelo para el número de personas que usan bicicleta y su frecuencia de uso diaria.

Conocer las parroquias en las cuales existen usuarios que usan bicicleta de manera diaria puede ser de gran ayuda al momento de situar un emprendimiento orientado al mundo de las bicicletas. Para solventar esta duda trabajamos sobre el algortimo de nuestro modelo disminuyendo el número de variables en el conjunto de datos. La Figura 69 refleja las variables seleccionadas para la generación de nuestros nuevos clusters. A diferencia del conjunto de datos utilizados en el modelo de la Figura 68, este nuevo conjunto debe mostrar una nueva lógica debido a que el análisis se centra en una sola periodicidad (Personas que utilizan la bicicleta diariamente).

Figura 69

Conjunto de datos, usuarios que usan la bicicleta diariamente a nivel de parroquia



```
UsoBiciLD<-dbGetQuery(con,"
select parr.parr_nombre as parroquia,
sum(case when hogPer.per_sexo in (1,2) then 1 else 0 end)::int  as numPer,
sum (case when frec.frec_codigo=1 then 1 else 0 end)::int as todosLosDias
from dtmrt.fact_uso_bicicletas factBici
inner join dtmrt.dim_hogar_persona hogPer on (factBici.id_persona=hogPer.id_persona)
inner join dtmrt.dim_uio_parroquia parr on (factBici.id_parroquia=parr.id_parroquia)
inner join dtmrt.dim_frecuencia_uso_bici frec on (factBici.id_freq_uso_bici=frec.id_freq_uso_bici)
where parr.parr_nombre<>'NO IDENTIFICADO'
group by parr.parr_nombre
order by 1 asc
")
#Vista de Datos
UsoBiciLD
#Limpieza de Datos
UsoBiciLD  <- UsoBiciLD %>%
  drop_na()
UsoBiciLD
```

```
> #Limpieza de Datos
> UsoBiciLD  <- UsoBiciLD %>%
+   drop_na()
> UsoBiciLD
          parroquia numper todoslosdias
1  BELISARIO QUEVED     20            0
2           CARCELEN     28            6
3  CENTRO HISTORICO      4            0
4          CHILIBULO     19            1
5        CHILLOGALLO     34            7
6        CHIMBACALLE      2            1
7         COCHAPAMBA     38            1
8  COMITE DEL PUEBL     30            3
9         COTOCOLLAO     25            3
10       EL CONDADO     61            3
11          GUAMANI     13            1
12        INIAQUITO     36            4
13        ITCHIMBIA     23            6
14         JIPIJAPA     11            0
15          KENNEDY      9            0
16       LA ARGELIA     37            5
17   LA ECUATORIANA     10            2
18   LA FERROVIARIA     33            1
19      LA LIBERTAD     18            0
20     LA MAGDALENA      6            1
21          LA MENA     14            1
22         PONCEANO     26            2
23         PUENGASI     12            1
24         QUITUMBE      8            5
25 S.ISIDRO DEL INC     50            9
26      SAN BARTOLO     31            0
27         SAN JUAN     30            5
28          SOLANDA     25            6
>
```

Al ejecutar los pasos de nuestro algoritmo mediante la ejecución de la función “fviz_nbclust”, la cual proporciona el número de clusters ideal, se pudo apreciar que el número se mantiene en 3. La ejecución del modelo Kmeans sobre el nuevo conjunto de datos nos muestra la agrupación de parroquias en base a dos grandes conjuntos expresados con la siguiente lógica;

Conjunto 1) Buen número de personas que usan la bicicleta, pero con pocos referentes que lo hacen diariamente.

Conjunto 2) Bajo número de personas que usan bicicleta y por ende una baja referencia de su uso diario.

Conjunto 3) Buen número de personas que usan la bicicleta y buena referencia de uso diario.

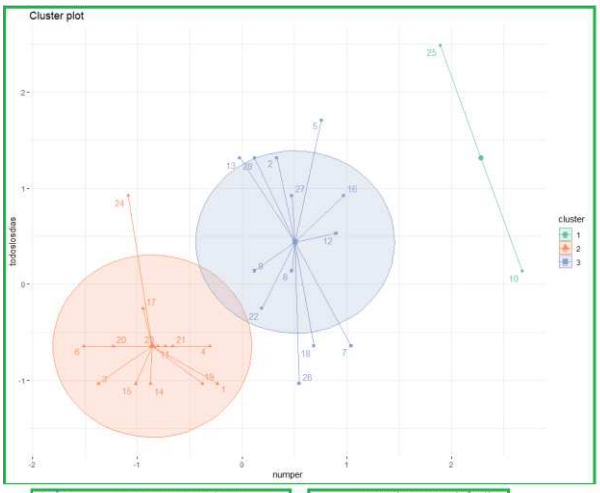
De los conjuntos formados, el conjunto número 3 es el conjunto que conviene analizar y evaluar pues de sus resultados saldrán alternativas para la toma de decisiones al momento de analizar la segmentación de mercado.

Las parroquias con un número adecuado de personas que practican bicicleta y que lo hacen de manera diaria son: Iñaquito, San Juan, Chillogallo, La Argelia.

Las parroquias San Isidro del Inca y El Condado se destacan por abarcar una gran cantidad de personas que usan bicicleta, pero muy pocas lo hacen de manera diaria. Ver Figura 70.

Figura 70

Gráfico de cluster (euclideo), caso de estudio 1



Cluster	parroquia	numper	todoslosdias
1	EL CONDADO	61	3
1	S.ISIDRO DEL INC	50	9
2	BELISARIO QUEVED	20	0
2	CENTRO HISTORICO	4	0
2	CHILIBULO	19	1
2	CHIMBACALLE	2	1
2	GUAMANI	13	1
2	JIPIJAPA	11	0
2	KENNEDY	9	0
2	LA ECUATORIANA	10	2
2	LA LIBERTAD	18	0
2	LA MAGDALENA	6	1
2	LA MENA	14	1
3	PUENGASI	12	1
2	QUITUMBE	8	5
3	CARCELEN	28	6
3	CHILLOGALLO	34	7
3	COCHAPAMBA	38	1
3	COMITE DEL PUEBL	30	3
3	COTOCOLLAO	25	3
3	INAQUITO	36	4
3	ITCHIMBIA	23	6
3	LA ARGELIA	37	5
3	LA FERROVIARIA	33	1
3	PONCEANO	26	2
3	SAN BARTOLO	31	0
3	SAN JUAN	30	5
3	SOLANDA	25	6

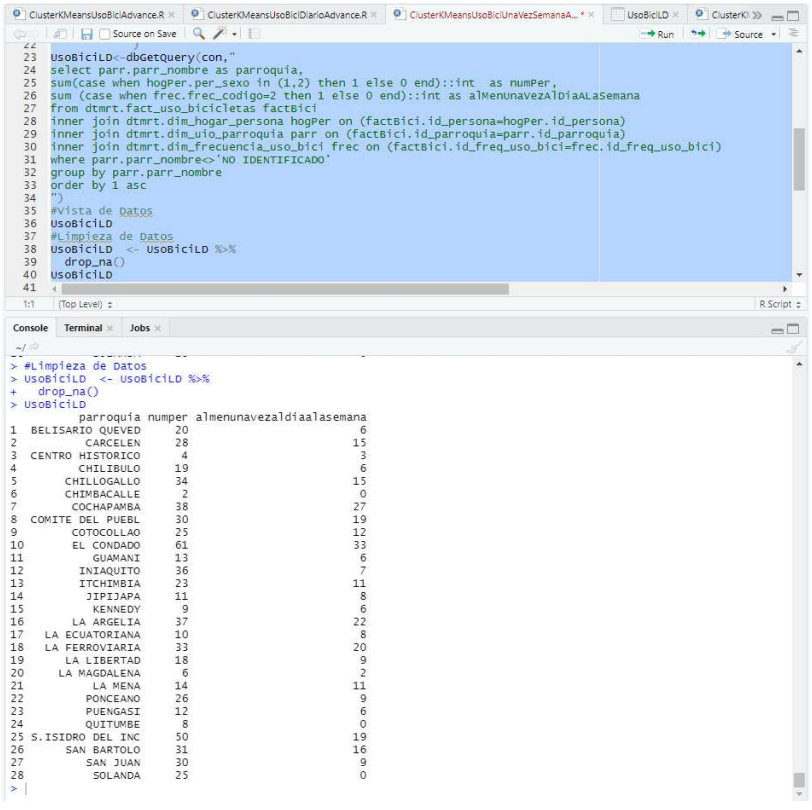
	parroquia	numper	todoslosdias
1	BELISARIO QUEVED	20	0
2	CARCELEN	28	6
3	CENTRO HISTORICO	4	0
4	CHILIBULO	19	1
5	CHILLOGALLO	34	7
6	CHIMBACALLE	2	1
7	COCHAPAMBA	38	1
8	COMITE DEL PUEBL	30	3
9	COTOCOLLAO	25	3
10	EL CONDADO	61	3
11	GUAMANI	13	1
12	INAQUITO	36	4
13	ITCHIMBIA	23	6
14	JIPIJAPA	11	0
15	KENNEDY	9	0
16	LA ARGELIA	37	5
17	LA ECUATORIANA	10	2
18	LA FERROVIARIA	33	1
19	LA LIBERTAD	18	0
20	LA MAGDALENA	6	1
21	LA MENA	14	1
22	PONCEANO	26	2
23	PUENGASI	12	1
24	QUITUMBE	8	5
25	S.ISIDRO DEL INC	50	9
26	SAN BARTOLO	31	0
27	SAN JUAN	30	5
28	SOLANDA	25	6

Caso 2: Evaluación del modelo para el número de personas que usan bicicleta y su frecuencia es de al menos una vez al día a la semana.

Identificar las parroquias que practican bicicleta diariamente es dato muy importante, pero otra alternativa de análisis es también saber, quiénes usan esta herramienta al menos un día a la semana. Para obtener esta información adoptamos de nuestro conjunto de datos, las dos variables que pueden ser útiles para el análisis. La Figura 71 muestra la matriz de evaluación.

Figura 71

Conjunto de datos, usuarios que usan la bicicleta al menos una vez al día a la semana



A diferencia del caso de evaluación número 1, el cual se analizó con el parámetro de clusters igual a 3, al aplicar “fviz_nbclust” a las nuevas variables, el número resultante de clusters fue igual a 2. Al ejecutar Kmeans para este nuevo caso nos dio como resultado dos grandes agrupaciones de parroquias. Estos dos grandes grupos formados en base a las características similares nos muestra el siguiente resultado:

Conjunto 1) Parroquias en las cuales el número de personas que usan bicicleta es bajo y el comportamiento reflejado a nivel de frecuencia de uso (al menos una vez al día a la semana) no es favorable.

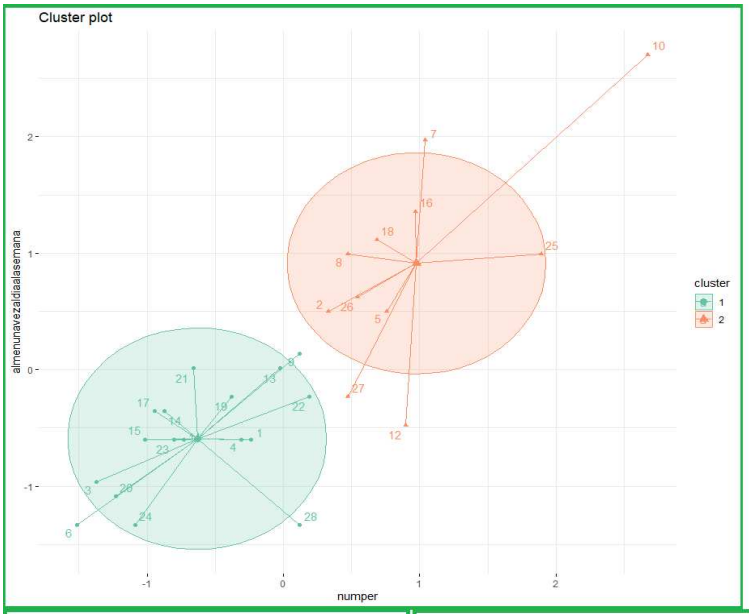
Conjunto 2) Parroquias en las cuales el número de personas que usan bicicleta es bueno y la referencia en cuanto a la frecuencia de uso es favorable a nuestro estudio.

Es así como: Chillogallo, La ferroviaria y La Argelia se caracterizan por la cantidad de ciclistas en el área y también porque la frecuencia de uso de la misma es favorable.

Parroquias del conjunto 1 del cual se puede destacar a: Chilibulo, Jipi Japa , La Libertad y Puengasi que se caracterizan por tener pocos usuarios de las bicicletas y a su vez su indicador de frecuencia de uso de la misma es débil, ver Figura 72.

Figura 72

Gráfico de cluster (euclideo), caso de estudio 2



	Cluster	parroquia	numper	almenunavezaldiaalasemana
1	1	BELISARIO QUEVED	20	6
2	1	CENTRO HISTORICO	4	3
3	1	CHILIBULO	19	6
4	1	CHIMBACALLE	2	0
5	1	COTOCOLLAO	25	12
6	1	GUAMANI	13	6
7	1	ITCHIMBIA	23	11
8	1	JIPIJAPA	11	8
9	1	KENNEDY	9	6
10	1	LA ECUATORIANA	10	8
11	1	LA LIBERTAD	18	9
12	1	LA MAGDALENA	6	2
13	1	LA MENA	14	11
14	1	PONCEANO	26	9
15	1	PUENGASI	12	6
16	1	QUITUMBE	8	0
17	1	SOLANDA	25	0
18	2	CARCELEN	28	15
19	2	CHILLOGALLO	34	15
20	2	COCHAPAMBA	38	27
21	2	COMITE DEL PUEBL	30	19
22	2	EL CONDADO	61	33
23	2	INIAQUITO	36	7
24	2	LA ARGELIA	37	22
25	2	LA FERROVIARIA	33	20
26	2	S.ISIDRO DEL INC	50	19
27	2	SAN BARTOLO	31	16
28	2	SAN JUAN	30	9

	parroquia	numper	almenunavezaldiaalasemana
1	BELISARIO QUEVED	20	6
2	CARCELEN	28	15
3	CENTRO HISTORICO	4	3
4	CHILIBULO	19	6
5	CHILLOGALLO	34	15
6	CHIMBACALLE	2	0
7	COCHAPAMBA	38	27
8	COMITE DEL PUEBL	30	19
9	COTOCOLLAO	25	12
10	EL CONDADO	61	33
11	GUAMANI	13	6
12	INIAQUITO	36	7
13	ITCHIMBIA	23	11
14	JIPIJAPA	11	8
15	KENNEDY	9	6
16	LA ARGELIA	37	22
17	LA ECUATORIANA	10	8
18	LA FERROVIARIA	33	20
19	LA LIBERTAD	18	9
20	LA MAGDALENA	6	2
21	LA MENA	14	11
22	PONCEANO	26	9
23	PUENGASI	12	6
24	QUITUMBE	8	0
25	S.ISIDRO DEL INC	50	19
26	SAN BARTOLO	31	16
27	SAN JUAN	30	9
28	SOLANDA	25	0

La clusterización de K-means aplicado a la comprobación de nuestro BI fue un modelo exitoso por cuanto nos ha permitido resaltar la información que refleja el dashboard a nivel de indicadores. A continuación, en el despliegue se destaca el uso del modelo y los dos casos de evaluación.

Despliegue

Las técnicas de minería de datos permitieron generar un valor agregado al desarrollo del sistema de BI. Los resultados del modelo de clusterización contrastaron con la información reflejada en el dashboard. El objetivo de poder empatar el resultado de los indicadores del panel de control, con un proceso de minería fue posible, gracias al uso de R y K-means. En el modelo ejecutado mediante K-means se obtuvieron dos casos de estudio ligados directamente al uso de las bicicletas. Cada caso forma parte del despliegue de resultados, visto bajo un enfoque de minería.

Caso 1: Evaluación del modelo para el número de personas que usan bicicleta y su frecuencia de uso diaria.

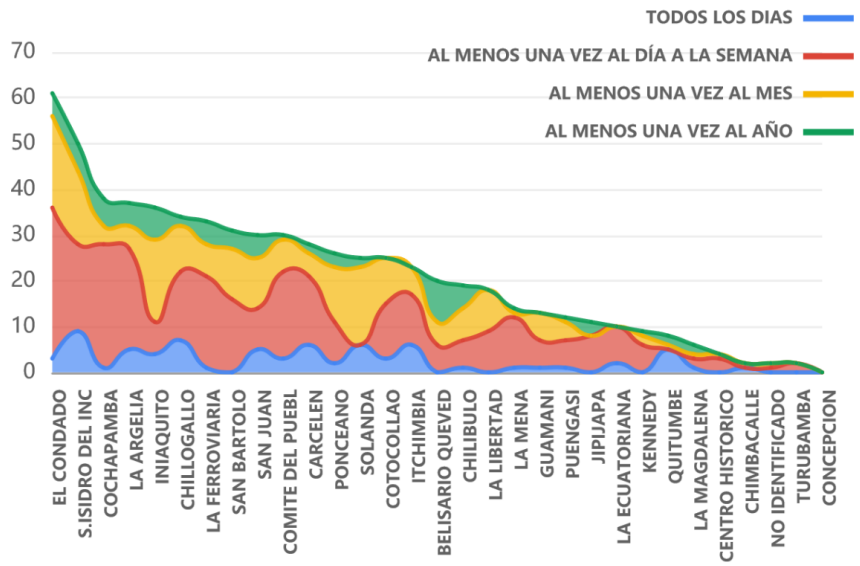
Caso 2: Evaluación del modelo para el número de personas que usan bicicleta y su frecuencia es de al menos una vez al día a la semana.

Los dos casos evaluados con el algoritmo de clusterización, permitieron identificar a las parroquias en base al número de personas que usan la bicicleta combinada con su frecuencia de uso. A diferencia de los resultados del visualizador el cual refleja el indicador mediante gráficos, ver Figura 73.

La clusterización permitió identificar grupos a partir de las mismas variables en base a sus similitudes, tal como se muestra en la Figura 71 y Figura 72.

Figura 73

Indicador de frecuencia (uso de la bicicleta por parroquia)



El proceso de minería ayudo a observar y analizar los datos desde un ángulo diferente, cabe recalcar que el panel de control refleja la información gráficamente mediante barras, líneas, donas etc. dependiendo del criterio del analista. El resultado de los modelos de minería, fueron más precisos puesto que no solo nos muestra un gráfico a interpretar, sino que además ha generado grupos de parroquias en base al comportamiento de sus variables.

Análisis de resultados

El principal objetivo del proyecto SEGDO_BICI_UIO, es permitir el análisis de la información, que se ha generado a partir del procesamiento de datos abiertos provenientes de encuestas. Esta información reflejada en un dashboard o panel de control debe ser una guía para aquellos emprendedores que tienen pensado iniciar su

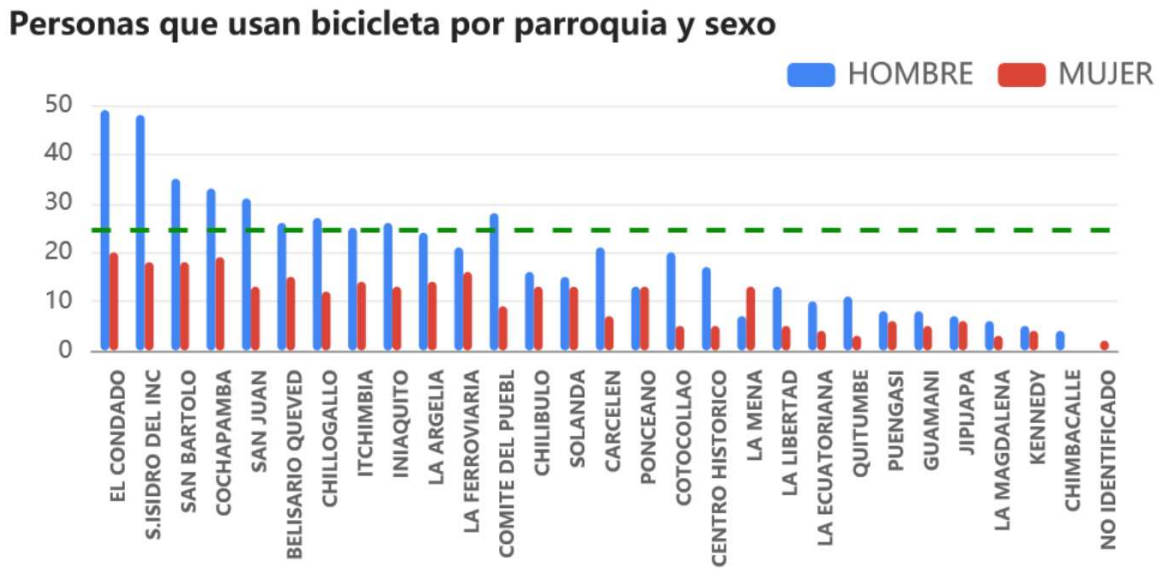
actividad económica en negocios orientados al mercado de las bicicletas. En base al análisis de requerimientos expresado en la Tabla 6 y tomando en cuenta el criterio de dos emprendedores, se han establecido las siguientes interrogantes y sus respuestas:

1. **En cuanto al uso de la bicicleta como una forma de actividad física; ¿Cuáles son las parroquias en las cuales se tiene mayor número de personas que usan bicicleta y su tendencia de uso, entre hombres y mujeres?**

   Para responder la pregunta hemos centrado el análisis en el dashboard, tomando en cuenta el indicador que refleja un conteo de personas practicantes del ciclismo por parroquia entre hombres y mujeres. En base al análisis de la Figura 74 , podemos darnos cuenta de que las parroquias: El Condado, San Isidro del Inca, San Bartolo, CochaPamba, San Juan, Iñaquito, Chillogallo, y Comité del Pueblo tienen valores por encima de la media y que representan una referencia de espacios geográficos donde existe una buena acogida al uso de bicicletas. Por otro lado, se puede evidenciar que en todas las parroquias los hombres generan más demanda de este medio de movilización que las mujeres.

Figura 74

Personas que usan bicicleta por parroquia y sexo



El análisis efectuado recae en los tres periodos de levantamiento de datos que van del 2018 al 2020, sin embargo, el 2020 se volvió característico debido a la pandemia. En este año los resultados reflejaron que las parroquias: San Bartolo, Belisario Quevedo y Centro Histórico aumentaron el número de usuarios y sobresalieron del resto de parroquias.

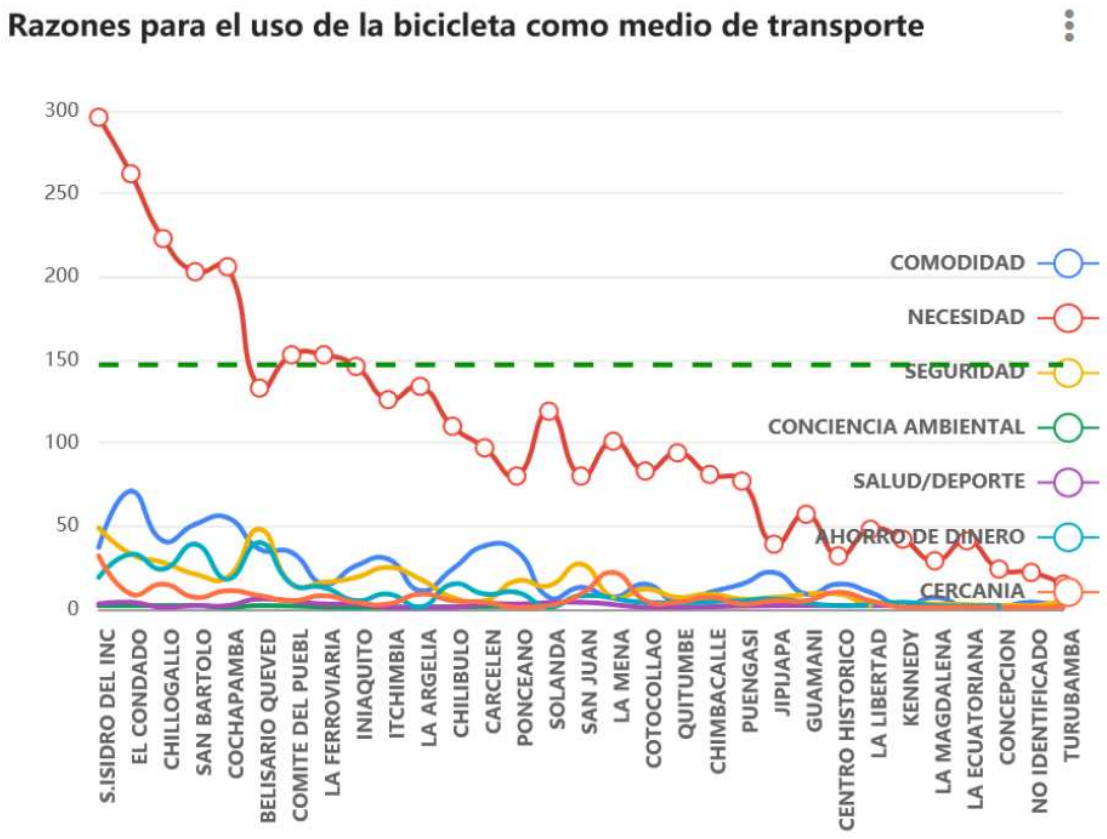
2. **En cuanto al uso de la bicicleta como medio de transporte para desplazarse a sus trabajos o centros educativos; ¿Cuáles son las parroquias con mayor número de personas que usa la bicicleta para movilizarse y sus razones?**

   Para responder a la pregunta se analizó el indicador que refleja las razones para el uso de la bicicleta bajo una perspectiva de medio de transporte. A diferencia del caso anterior en el cual el uso de la bicicleta estaba enfocado al entretenimiento. La Figura 75 nos muestra que las parroquias; El Condado, San Isidro del Inca, San Bartolo, Cochapamba y Chillogallo siguen

siendo referentes para el uso de la bicicleta, sin embargo, es importante tomar en cuenta que la mayoría de las personas adoptan la movilización en bici por necesidad, ubicando esta opción en primer lugar. Otro grupo refleja su uso por comodidad, ubicándolos en un segundo lugar, un tercer grupo muestra que el ahorro de dinero también es una razón importante, un cuarto grupo se inclina por la seguridad, mientras que un disminuido grupo se basa en su conciencia ambiental y finalmente en último lugar se encentra un grupo que usa la bicicleta por la cercanía.

Figura 75

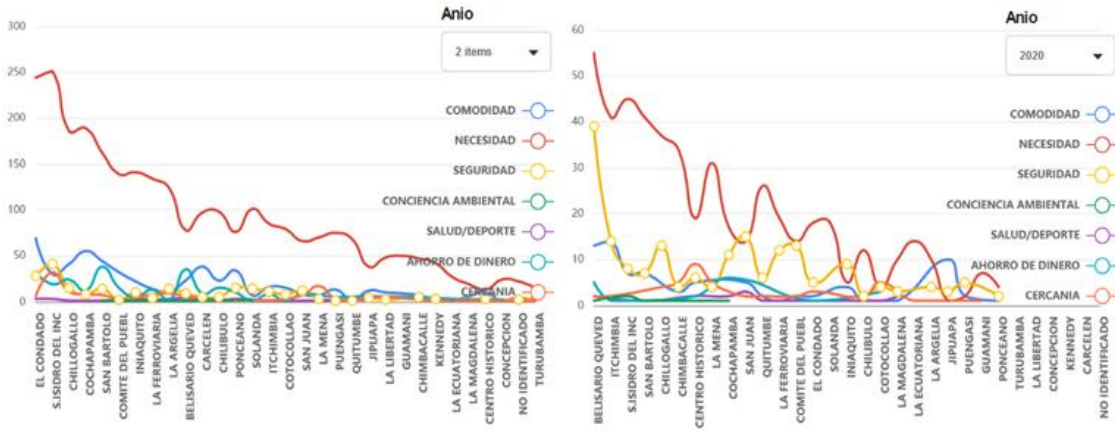
Razones para el uso de la bicicleta como medio de transporte



Cabe recalcar que la información analizada en la Figura 75 enfoca datos históricos desde el 2018 al 2020, en los cuales el factor seguridad para los periodos 2018 y 2019 fueron parte de un segundo plano. En el 2020 debido a la pandemia, la bicicleta como medio de transporte, se convirtió en una alternativa viable para movilizarse de un lugar a otro evitando posibles contagios. El uso de la bicicleta podría ayudar a evitar el contacto entre personas manteniendo el distanciamiento social, por tal motivo en ese año el factor seguridad pasó a ser una de las principales razones para el uso de la bicicleta. La Figura 76 hace mención al análisis realizado.

Figura 76

La seguridad como factor principal en el uso de la bicicleta para el 2020



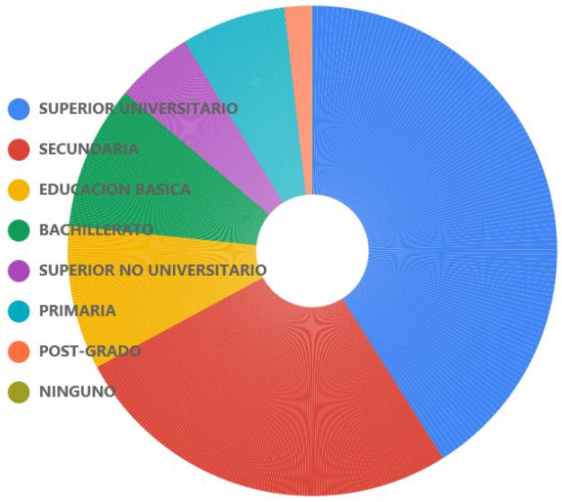
3. **¿Qué nivel de instrucción tienen las personas que usan la bicicleta como entretenimiento?**

   Los resultados analizados en base a los indicadores de la Figura 74 y Figura 75 respectivamente, han permitido identificar a las parroquias que pueden convertirse en lugares potenciales, para la instalación de un negocio orientado a bicicletas. La fuente de datos tiene como atributos el nivel de

instrucción de las personas encuestadas lo cual ha servido para generar un indicador y dar respuesta a la pregunta. La Figura 77 permite identificar la existencia de cuatro grandes grupos diferenciadores; El primer grupo está conformado por personas que tienen un nivel de instrucción superior, el segundo grupo está conformado por personas que se encuentran estudiando en la segundaria, el tercer grupo se compone de personas que tienen educación básica y el cuarto grupo con personas estudiantes de bachillerato. La identificación de estos grupos podría ser útil al momento de analizar estrategias orientadas a productos y posibles compradores.

Figura 77

Personas que usan bicicleta y su nivel de instrucción

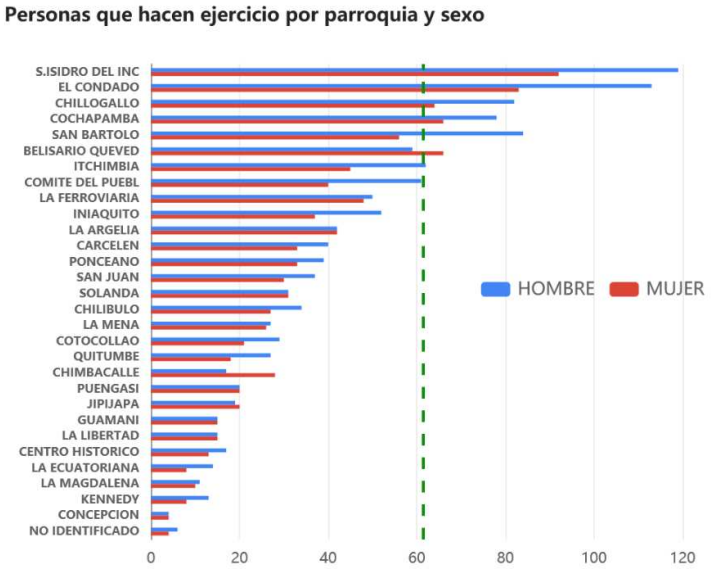


4. **En cuanto a la ejecución de ejercicio físico; ¿Cuáles son las parroquias con personas físicamente activas y su tendencia, entre hombres y mujeres?**

La encuesta multipropósito incluye en su encuesta variables que recaban información de las personas que realizan ejercicio físico, si bien es cierto este indicador no refleja información directa acerca de personas que usan bicicletas, podría convertirse en un potencial segmento, si se ejecuta una campaña que fomente la adquisición de una bicicleta para ejercitarse. La Figura 78 destaca a las parroquias; San Isidro del Inca, El Condado Chillogallo, Cochapamba, San Bartolo y Belisario Quevedo. Estas parroquias concentran sus datos por encima de la media y se convierten en candidatas al momento de analizar un segmento de mercado apto para un emprendimiento relacionado al mundo de las bicicletas. Si analizamos el sexo de las personas físicamente activas podemos darnos cuenta que los hombres lideran en la mayoría de parroquias, sin embargo en la parroquia Belisario Quevedo el número de mujeres es mayor.

Figura 78

Personas que hacen ejercicio por parroquia y sexo



5. **En cuanto a la frecuencia de uso de la bicicleta; ¿Cuáles son las parroquias con un considerable número de personas que usan la bicicleta de forma diaria?**

   La información reflejada en el panel de control también incluye un indicador que muestra las parroquias con mayor número de personas que usan bicicleta y su frecuencia desagregadas en los siguientes parámetros; a) todos los días, b) al menos una vez al día a la semana, c) al menos una vez al mes y d) al menos una vez al año. Este indicador fue sometido a un algoritmo de minería de datos llamado K means, con el objetivo de realizar grupos entre parroquias evaluando la cantidad de personas y frecuencia de uso diario. El resultado arrojo tres grupos de análisis;

   Grupo 1: Conformado por parroquias con un buen número de personas que usan bicicleta, pero pocas de ellas la usan diariamente.

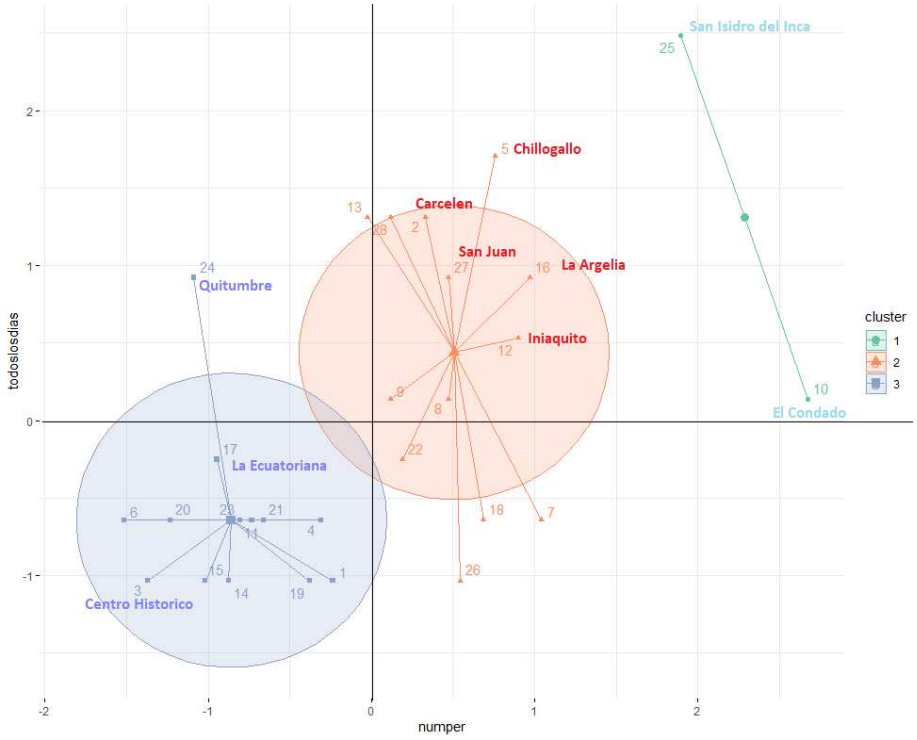
   Grupo 2: Conformado por parroquias con un número considerable de personas que usan bicicleta y una positiva respuesta de uso diario.

   Grupo 3: Conformado por parroquias con pocas personas que usan bicicleta y baja frecuencia en cuanto a su uso diario.

   En base al análisis realizado por los emprendedores, el grupo 2 es considerado el mejor candidato y un segmento ideal al cual atacar, es decir; Chillogallo, Carcelén, San Juan, La Argelia e Iñaquito son lugares en los cuales se podría ubicar un taller con accesorios destinado a la reparación y mantenimiento de bicicletas.

Figura 79

Agrupación de parroquias que usan la bicicleta diariamente



En la actualidad el mercado de las bicicletas se ha expandido considerablemente, el emprendedor tiene la oportunidad de concentrarse en la instalación de negocios orientados a la venta, mantenimientos, repuestos y accesorios de bicicletas. El análisis de resultados ha permitido identificar a varias parroquias con características asociadas al uso de la bicicleta. Según el visualizador y tomando en cuenta el criterio de nuestros emprendedores, El Condado, San Isidro del Inca, San Bartolo, Cochapamba, San Juan, Iñaquito, Chillogallo, y Comité del Pueblo, podrían considerarse como áreas potencialmente aptas para la instalación de almacenes dedicados a la venta de bicicletas y accesorios. Chillogallo, Carcelén, San Juan, La Argelia e Iñaquito se caracterizaron por el uso frecuente de la bicicleta, estas parroquias

podrían convertirse en áreas potenciales para la instalación de talleres debido al uso periódico de este medio de desplazamiento.

Según los datos del 2020 la bicicleta se convirtió en un medio de transporte seguro debido a la pandemia. Las personas adoptaron este medio de transporte gracias a la posibilidad de mantener el distanciamiento social y el aumento de las ciclo vías en los últimos años.

El proyecto SEGDO_BICI_UIO fue presentado a un experto en Marketing, para conocer su criterio y la utilidad que puede reflejar el sistema de BI en los emprendedores. Según el experto, el visualizador y la información que este refleja pueden contribuir de forma parcial en la identificación de un posible segmento de mercado. Afirmó que la información obtenida del panel de control es muy valiosa, y que puede servir como una referencia ante la incertidumbre de querer emprender en la ciudad de Quito, pero no saber dónde. El visualizador se convertiría en el punto de partida para un estudio de segmentación más profundo y preciso.

El análisis de los resultados ejecutado en conjunto con los emprendedores y la evaluación de la herramienta en base al criterio de un experto en marketing, genero resultados positivos al proyecto y su objetivo principal. Para verificar parte de la información se procedió a ejecutar una encuesta a 10 personas de cuatro parroquias de Quito destacadas por el número de usuarios que usan bicicleta reflejadas en el sistema de BI.

Se seleccionó a las parroquias El Condado y San Isidro del Inca como buenos referentes en cuanto al considerable número de personas que usan bicicletas mientras que las parroquias Chimbacalle y La Magdalena como lo opuesto a las dos parroquias mencionadas.

Las variables consideradas en la encuesta fueron las siguientes:

Tabla 27

Variables encuesta de verificación

Código de Variable	Nombre Variable	Definición	Formato de la Variable	Categorías de la Variable	Tipo
1	Parroquia	Especificación la ubicación de la vivienda: Urbano - Rural	Texto	No Aplica	Variable Geográfica
2	Sexo	Es un atributo diferencial fundamental de análisis demográfico como también en el estudio de las características sociales y económicas de una población.	Categórico	_1. Hombre _2. Mujer	Variable Demográfica
3	Edad	Tiempo transcurrido a partir del nacimiento de un individuo.	Numérico	Valores entre 0 a 98	Variable Demográfica
4	Uso de la bicicleta	Especifica si usa o no la bicicleta como medio de transporte.	Categórico	_1. SI _2. NO	Variable Psicográfica
4.1	Frecuencia Uso de la Bicicleta	Identifica la frecuencia con la que usa la bicicleta.	Categórico	_1. Todos los días _2. Al menos una vez a la semana _3. Al menos una vez al mes _4. Al menos una vez al año	Variable Psicográfica
4.2	Especificación medio de transporte	Identifica la razón principal por la que usa un medio de transporte en especifico	Categórico	_1. Comodidad _2. Necesidad _3. Seguridad _4. Conciencia Ambiental _5. Salud/Deporte _6. Ahorro de dinero _7. Cercanía	Variable Psicográfica

Código de Variable	Nombre Variable	Definición	Formato de la Variable	Categorías de la Variable	Tipo
5	Medios de desplazamiento	Identifica el uso de la bicicleta como medio de transporte dentro de la rutina diaria	Categórico	_1. Vehículo Particular (solo) _2. Vehículo Particular (compartido) _3. Transporte Público _4. Bicicleta _5. Caminar _6. Otro ¿Cuál? (especifique) _99. No aplica	Variable Psicográfica

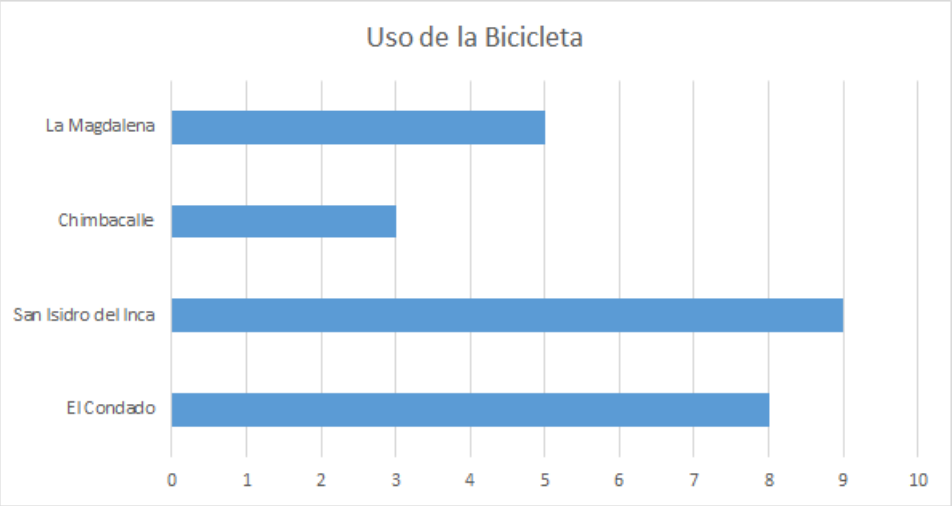
Las variables utilizadas en la encuesta de verificación fueron las mismas que incluye la encuesta multipropósito de modo resumido. Los datos fueron tabulados mediante la herramienta Excel y se desarrollaron 4 indicadores que reflejan la información resultante.

El primer indicador hace referencia al uso de la bicicleta en las diferentes parroquias de la ciudad seleccionadas para la comprobación. Como se puede observar en la Figura 80 la información recolectada concuerda con el visualizador, es decir; las parroquias San Isidro del Inca y El Condado se destacan por tener mayor número de usuarios que usan bicicleta que las parroquias Chimbacalle y La Magdalena.

Cabe recalcar que la encuesta se realizó a personas mayores de 12 años tal como lo realiza la encuesta multipropósito en su sección relacionada a la actividad física y transporte.

Figura 80

Resultados encuesta de verificación (uso de la bicicleta)

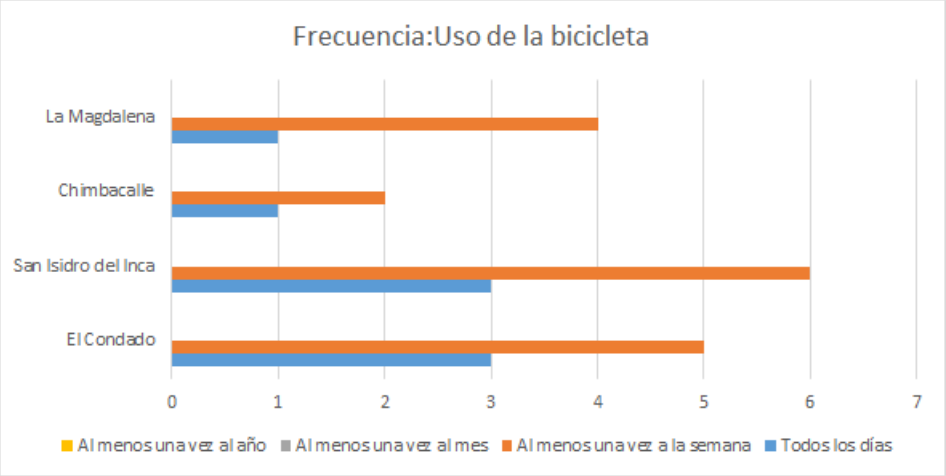


La periodicidad con la que usan la bicicleta es otro indicador contemplado en el proyecto SEGDO_BICI_UIO y una variable importante en la encuesta de verificación. En las cuatro parroquias prevalece el uso de la bicicleta de al menos una vez a la semana, si comparamos con los datos del visualizador podemos corroborar que las dos parroquias San Isidro del Inca y El condado se destacan por el considerable número de personas que usan bicicleta y mas no por su uso diario ver Figura 81.

AL recabar las conclusiones de los emprendedores estas parroquias no serían candidatas para alojar talleres que se dediquen al mantenimiento y reparación de este medio de transporte.

Figura 81

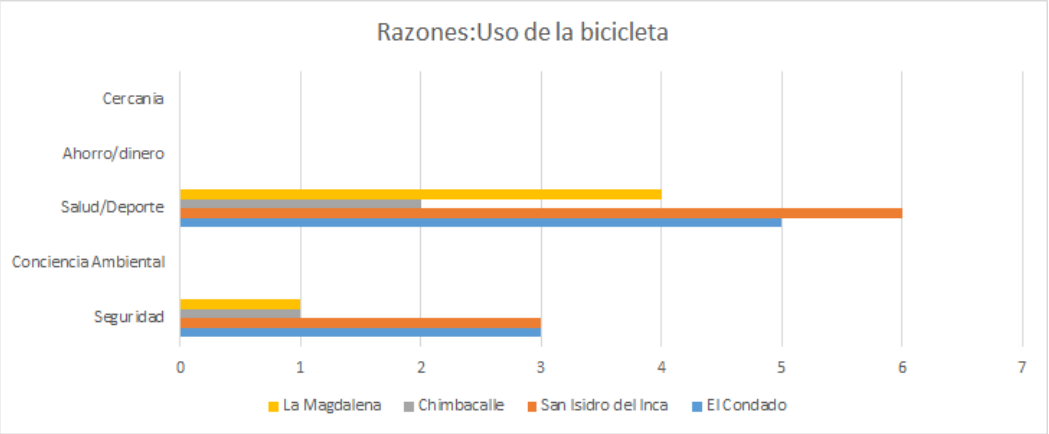
Resultados encuesta de verificación (frecuencia de uso de la bicicleta)



El último indicador analizado nos indica las razones por lo cual las personas encuestadas hacen uso de la bicicleta. La Figura 82 demuestra que la tendencia es salud, seguridad y deporte. En varias de las parroquias encuestadas supieron mencionar que el desplazamiento en bicicleta ayuda a respetar las normas de distanciamiento que exige la pandemia por tanto lo categorizaron como un medio seguro. Otro grupo de personas afirma que es una de las mejores vías para mantenerse en forma y saludable.

Figura 82

Resultados encuesta de verificación (razones de uso de la bicicleta)



Los resultados de las encuestas concuerdan con la información reflejada en el dashboard.

Capítulo 4

Conclusiones y recomendaciones

Conclusiones

- El estado del arte permitió conocer sobre la existencia de varios trabajos orientados a la segmentación de mercado, que incluyen el uso de tecnologías basadas en big data, minería de datos e inteligencia de negocios, sin embargo no todos apuntan al uso de datos abiertos (generados por entidades públicas) y la mayoría trabaja con datos generados por el negocio.
- El uso de datos abiertos ha tomado fuerza en la última década. En nuestro país, entidades gubernamentales como el Instituto Nacional de Estadísticas y Censos (INEC), publican en sus espacios digitales datos de interés social, que pueden ser explotados mediante técnicas de BI y minería de datos.
- El INEC se convirtió en la principal fuente de datos del proyecto. Adicionalmente fue necesario el uso de datos provenientes de fuentes municipales para la organización de los datos. El INEC destaca en su metodología de estudio geográfico a las zonas censales, por otro lado, el municipio tiene disponible una división geográfica a nivel de parroquias. La intersección geográfica, entre zonas censales y parroquias, hizo posible el análisis de los indicadores desagregado a nivel parroquial urbano.
- La agilidad y claridad en el proceso de modelamiento y creación de los data marts, fue un factor evidente gracias a la adopción del enfoque metodológico propuesto por Ralph Kimbal.
- La creación de un espacio temporal de almacenamiento de datos (staging área) fue determinante para el éxito en el proceso de extracción

transformación y carga (ETL). Es importante destacar, que el procesamiento de datos basado en diferentes fuentes, puede dificultarse debido a diferencias en formatos, presentaciones y estructuras.

- El kit de herramientas de Pentaho utilizadas en el proyecto, permitieron desarrollar con éxito las fases de; procesamiento, análisis y visualización de los inidicadores.
- El análisis de los datos mediante el software Saiku Analytics, permitió analizar y evaluar gráficamente los indicadores de mayor relevancia. La visualización del dashboard permitió conocer que parroquias como; El Condado, San Isidro del Inca, San Bartolo, CochaPamba, San Juan, Iñaquito, Chillogallo, y Comité del Pueblo pueden ser considerados como candidatos para la ubicación de un negocio orientado al mundo de las bicicletas.
- El algoritmo de minería de datos “K means” permitió comprobar y corroborar los resultados de los datos que alimentan el BI reflejados en el dashboard. La lógica del algoritmo permitió identificar por grupos las parroquias con características que favorecen a los emprendedores dedicados al mercado de las bicicletas.

Recomendaciones

- La Encuesta Nacional Multipropósito de Hogares, principal fuente de datos del proyecto SEGDO_BICI_UIO, tiene una proyección de 4 años, inicio en el 2018 y culminará en el 2021. Se recomienda alimentar el sistema de BI con los datos de los cuatro períodos para mantener una buena referencia a nivel de datos históricos.

- Nuestro sistema de BI está basado en el uso de datos abiertos y en la actualidad Internet aloja una gran cantidad de ellos, pero es necesario tomar en cuenta la seriedad que implica el desarrollo de proyectos académicos y es recomendable hacer uso de datos oficiales debidamente justificados por entidades que se dedican a la publicación de información.
- El proyecto SEGDO_BICI_UIO refleja su lógica de presentación de información en gráficos y tablas, sin embargo; sería importante incorporar el concepto de GEO BI, puesto que muchos emprendedores toman en cuenta las ciclo vías al momento de instalar su negocio. Mediante el uso de GEO BI se podría incluir las capas de las rutas e inclusive la geo referenciación de aquellos locales comerciales con afinidad al giro de negocio, que podrían identificarse como competidores.
- Se recomienda la suite de Pentaho en la ejecución de actividades relacionadas al procesamiento y análisis de los datos, la versión libre de estas herramientas son de gran utilidad y aportan considerablemente en la construcción del BI.
- El algoritmo de minería K-means utilizado en el proyecto, fue de gran utilidad en el proceso de análisis de los datos, su lógica permitió identificar grupos basados en la semejanza de su comportamiento y discriminarlos. Se recomienda su uso en proyectos que requieran identificar grupos específicos sobre un conjunto de datos.

Bibliografía

Ain, N., Vaia, G., DeLone, W. H., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review. *Decision Support Systems*, *125*, 113113. https://doi.org/10.1016/j.dss.2019.113113

Azma, F., & Mostafapour, M. A. (2012). Business intelligence as a key strategy for development organizations. *Procedia Technology*, *1*, 102-106. https://doi.org/10.1016/j.protcy.2012.02.020

Balcazar, J. (2015). *ANALISIS DE CASOS DE LOS FACTORES POTENCIALES QUE ESTAN INCIDIENDO EN EL FRACASO DE LAS PYMES.* 17.

Borja, J. E. R. (2018). *Comparación de herramientas ETL de código abierto*. 94.

Camilleri, M. A. (2018). *Market Segmentation, Targeting and Positioning*. 17.

Diario La Hora. (2020). *Quito apuesta por la bicicleta para evadir al Covid-19*. Quito apuesta por la bicicleta para evadir al Covid-19. https://lahora.com.ec/noticia/1102319910/quito-apuesta-por-la-bicicleta-para-evadir-al-covid-19

Diario Primicias. (2020). *El uso de bicicleta en Quito aumentó un 600% y se planifican nuevas ciclovías*. https://www.primicias.ec/primicias-tv/sociedad/uso-bicicleta-quito-aumenta-planifican-ciclovias/

Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, R., Tait, J., & Zijlstra, T. (2012). *Manual de los Datos Abiertos*.

Duque, N., & Tamayo, A. (2001). *DATA WAREHOUSE (BODEGA DE DATOS)HERRAMIENTA PARA LA TOMA DE DECISIONES (PARTE1)*.

Eletter, S. F. (s. f.). *USING DATA MINING FOR AN INTELLIGENT MARKETING CAMPIAN*. 8.

Fan, S., Lau, R. Y. K., & Zhao, J. L. (2015). Demystifying Big Data Analytics for

Business Intelligence Through the Lens of Marketing Mix. *Big Data Research*, *2*(1), 28-32. https://doi.org/10.1016/j.bdr.2015.02.006

Fernández, C., Morales, S., Urcuango, A., Albán, A., & Nabernegg, M. (2019). *Encuesta Nacional Multipropósito de Hogares (Seguimiento al Plan Nacional de Desarrollo 2018)*. 66.

Gichuru, M. J., & Limiri, E. K. (s. f.). *MARKET SEGMENTATION AS A STRATEGY FOR CUSTOMER SATISFACTION AND RETENTION*. 10.

Gómez, A. A. R., & Bautista, D. W. R. (2010). Inteligencia de negocios: Estado del arte. *Scientia et technica*, *1*(44), 321-326. https://doi.org/10.22517/23447214.1803

Gonzales, J., Seoane, J., & Robles, G. (2007). *Software Libre*.

GUÍA DE POLÍTICA PÚBLICA DE DATOS ABIERTOS. (2014).

Heang, R., & Mohan, R. (s. f.). *LITERATURE REVIEW OF BUSINESS INTELLIGENCE*. 10.

IBM. (1994). *Manual CRISP-DM de IBM SPSS Modeler*. 56.

Kamthania, D., Pawa, A., & Madhavan, S. (2018). Market Segmentation Analysis and Visualization using K-Mode Clustering Algorithm for E-Commerce Business. *Journal of Computing and Information Technology*, *26*(1), 57-68. https://doi.org/10.20532/cit.2018.1003863

Kurniawan, Y., Gunawan, A., & Kurnia, S. G. (2005). APPLICATION OF BUSINESS INTELLIGENCE TO SUPPORT MARKETING STRATEGIES: A CASE STUDY APPROACH. . . *Vol.*, *64*, 9.

Lasio, V., Ordeñana, X., Caicedo, G., Samaniego, A., & Izquierdo, E. (2018). *Global Enterpreneurship Monitor Ecuador 2017*.

Lei, N., & Ki Mon, S. (2013). *A DECISION SUPPORT SYSTEM FOR MARKET SEGMENT DRIVEN PRODUCT DESIGN RNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED07*. 10.

Oleas, D., & Albornoz, M. B. (2015). La bicicleta y la transformación del espacio público en Quito (2003-2014). *Letras Verdes. Revista Latinoamericana de Estudios Socioambientales*.

Peralta, V. (2015). *Un caso de estudio sobre diseño lógico de Data Warehouses*. 20.

RACINES, A. (2016). *PROPUESTA DE UN PLAN DE MARKETING PARA IMPULSAR EL CRECIMIENTO DE LOS CLIENTES DEL PALACIO DEL JUGO UBICADO EN LAS PARROQUIAS QUITUMBE Y LA MAGDALENA EN EL SECTOR SUR DE LA CIUDAD DE QUITO PARA EL AÑO 2017*.

Rivadera, G. R. (2010). *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*. *5*, 16.

Suero, D. (2010). *Factibilidad del uso de la bicicleta como medio de transporte en la ciudad de Bogotá*. 9.

The Free Software Foundation. (1996). *¿What is Free software?* https://www.gnu.org/philosophy/free-sw.html

Tynan, C. (1987). *Market Segmentation*.

Wrembel, R., & Koncilia, C. (Eds.). (2007). *Data Warehouses and OLAP: Concepts, Architectures and Solutions*. IGI Global. https://doi.org/10.4018/978-1-59904-364-7

Glosario

- OLAP: On Line Analytical Processing
- BI: Business Intelligence (Inteligencia de negocio)
- DASHBOARD: Tablero de Control
- ETL: Extract, Transform and Load (Extraer, transformar y cargar) Refiere a la
- transformación de los datos.
- DSA: Data storage area (Espacio temporal para crear el Data Warehouse)
- DW: Referencia a Data Warehouse