



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**Desarrollo de un modelo de flujos de datos que permita identificar el comportamiento de
la morbimortalidad del adulto mayor ecuatoriano**

Malacatus Morocho, Juana Catalina

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

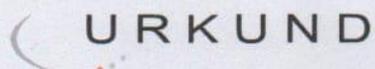
Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

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

Msc. Coral Coral, Henry Ramiro

25 de agosto de 2020



Analysed Document: Desarrollo de un modelo de flujos de datos que permita identificar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano.
 TRABAJO_TITULACIÓN_MALACATUS_CATALINA_POSGRADO.docx (D78162393)
 Submitted: 8/25/2020 5:19:00 PM
 Submitted By: hrcoral@espe.edu.ec
 Significance: 7%

Sources included in the report:

ALVAREZ- CAÑAREJO-29012018.docx (D35124987)
 proyecto de investigacion_LAFUENTE_LEMA.docx (D62358049)
 cap 1 2 y 3 subir.docx (D63878662)
 DERECHO A LA SALUD GRUPO 6 PDF.pdf (D77902301)
 Mortalidad_materna2006_2016.pdf (D47211727)
https://epidemiologiatalax.files.wordpress.com/2012/10/p_epi_del_adulto_mayor_en_mexico_2010.pdf
<https://www.medigraphic.com/cgi-bin/new/resumenl.cgi?IDARTICULO=37857Feng>,
https://apps.who.int/iris/bitstream/handle/10665/186466/9789240694873_spa.pdf?sequence=1Organizaci
https://www.paho.org/hq/index.php?option=com_content&view=article&id=3561:2010-clasificacion-internacional-enfermedades-cie&Itemid=2560&lang=esOrganizaci
<https://repositorio.umsa.bo/bitstream/handle/123456789/21076/TM-1406.pdf?sequence=1&isAllowed=y>
https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/Camas_Egresos_Hospitalarios/Cam_Egre_Hos_2017/Metodologia_ECEH_%202017.pdf
https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/Camas_Egresos_Hospitalarios/Cam_Egre_Hos_2018/Metodologia%20Camas%20y%20Egresos%20Hospitalarios%202018.pdf

Instances where selected sources appear:

51

Firma:

Msc. Coral Coral, Henry Ramiro

DIRECTOR



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, “Desarrollo de un modelo de flujos de datos que permita identificar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano” fue realizado por la señorita **Malacatus Morocho, Juana Catalina** el mismo que ha sido revisado en su totalidad, analizado 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, 25 de agosto de 2020

Firma:

Msc. Coral Coral, Henry Ramiro

Director

C.C.: 1714864830



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

RESPONSABILIDAD DE AUTORÍA

Yo, **Malacatus Morocho, Juana Catalina** con cédula de ciudadanía n° 1104858640, declaro que el contenido, ideas y criterios del trabajo de titulación: **Desarrollo de un modelo de flujos de datos que permita identificar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano** es de mi 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, 25 de agosto de 2020

Firma

Ing. Malacatus Morocho, Juana Catalina

C.C.: 1104858640



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **Malacatus Morocho, Juana Catalina** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Desarrollo de un modelo de flujos de datos que permita identificar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 25 de agosto de 2020

Firma

Ing. Malacatus Morocho, Juana Catalina

C.C.: 1104858640

Dedicatoria

A Dios, a mis abuelitos Agustín, Albertina y Juanita, a mis padres Manuelito y Amadita por su amor incondicional. A mi familia, a mis hermanos Rafael, Milton y hermanas Marie, Elizabeth, Fátima, Maryuri, Yalexí por su apoyo y motivación. A mi sobrina Alwenna y a mi bebé Matías.

A la memoria de mi tío-abuelo Nelson Eras quién fue la razón de esta investigación.

Agradecimiento

A la Universidad de las Fuerzas Armadas (ESPE) por permitirme ingresar a tan distinguida institución, a la coordinadora Ing. Tatiana Gualotuña, a cada uno de los docentes y a mi director de tesis por su guía y dedicación.

Índice de contenido

Carátula.....	1
Certificado del Director	3
Responsabilidad de autoría.....	4
Autorización de publicación.....	5
Dedicatoria.....	6
Agradecimiento.....	7
Índice de contenido	8
Índice de Tablas	11
Índice de Figuras	12
Resumen	13
Abstract.....	14
Capítulo I: Planteamiento del Problema de Investigación.....	15
Antecedentes	15
El problema de investigación	17
<i>Contexto del problema</i>	17
<i>Planteamiento del problema</i>	18
Objetivos	19
<i>Objetivo general</i>	19
<i>Objetivos específicos</i>	19
Justificación y alcance	20
Hipótesis de investigación	21
Categorización de las variables de investigación.....	21
<i>Variable dependiente</i>	21
<i>Variable independiente</i>	21
Revisión Inicial de Literatura	22
<i>Objetivo de la búsqueda</i>	22
<i>Preguntas de investigación</i>	22
<i>Criterios de inclusión y exclusión</i>	22
<i>Definición de la estrategia de búsqueda</i>	23
<i>Conclusión</i>	28
Capítulo II: Marco Teórico.....	30
Variable Independiente	30
Minería de datos	30

Técnicas de aprendizaje	32
<i>Aprendizaje supervisado</i>	32
<i>Aprendizaje no supervisado</i>	32
Técnicas de minería de datos	32
Datos de secuencias en minería de datos	35
Series Temporales	35
<i>Análisis de regresión y tendencias en series temporales</i>	36
<i>Pronósticos en series temporales</i>	37
Plataformas de análisis e inteligencia de negocios	39
Variable Dependiente	43
Envejecimiento	43
Adulto Mayor	45
Indicadores de salud	46
<i>Tasa de Prevalencia</i>	47
<i>Tasa de Morbilidad</i>	47
<i>Tasa de Mortalidad</i>	48
<i>Diez principales causas de morbilidad hospitalaria en adultos mayores</i>	48
<i>Diez principales causas de mortalidad en el adulto mayor</i>	48
Código Internacional de Enfermedades (CIE)	49
<i>Capítulos</i>	49
<i>Lista especial de tabulación para la morbilidad</i>	50
<i>Categorías de tres caracteres</i>	51
Capítulo III: Metodología	52
Métodos	52
Metodología	53
Capítulo IV: Aplicación de la Metodología CRISP-DM	56
Comprensión del negocio	56
<i>Evaluación de la situación actual del caso de estudio</i>	56
<i>Establecimiento de los objetivos de DM</i>	56
Comprensión de los datos	57
<i>Recopilación inicial de datos</i>	57
<i>Descripción de los datos</i>	58
<i>Exploración de los datos</i>	76
<i>Verificar la calidad de los datos</i>	78
Preparación de datos	85

	10
<i>Selección</i>	85
<i>Limpieza</i>	87
<i>Estructuración, integración y formateo</i>	87
<i>Análisis preliminar</i>	88
Modelado	100
<i>Selección de la técnica de modelado</i>	100
<i>Selección de atributos para el análisis del algoritmo y construcción del modelo</i>	100
<i>Evaluación del modelo</i>	105
Capítulo V: Análisis y Discusión de Resultados	106
Evaluación de resultados obtenidos	106
Despliegue	110
Discusión de los resultados	111
Capítulo VI: Conclusiones y Recomendaciones	114
Conclusiones	114
Recomendaciones	115
Bibliografía	117

Índice de Tablas

Tabla 1	<i>Grupo de control</i>	24
Tabla 2	<i>Artículos candidatos encontrados en los repositorios digitales</i>	25
Tabla 3	<i>Estudios seleccionados</i>	26
Tabla 4	<i>Interpretación de los valores típicos de MAPE</i>	39
Tabla 5	<i>Plataformas de análisis e inteligencia de negocios</i>	41
Tabla 6	<i>Capítulos CIE-10</i>	49
Tabla 7	<i>VARIABLES del archivo de egresos hospitalarios del año 2011</i>	58
Tabla 8	<i>VARIABLES del archivo de egresos hospitalarios del año 2012</i>	59
Tabla 9	<i>VARIABLES del archivo de egresos hospitalarios del año 2013</i>	60
Tabla 10	<i>VARIABLES del archivo de egresos hospitalarios del año 2014</i>	61
Tabla 11	<i>VARIABLES archivo de egresos hospitalarios, años 2015 -2018</i>	62
Tabla 12	<i>VARIABLES del archivo de defunciones generales del año 2011</i>	64
Tabla 13	<i>VARIABLES del archivo de defunciones generales del año 2012</i>	65
Tabla 14	<i>VARIABLES del archivo de defunciones generales del año 2013</i>	67
Tabla 15	<i>VARIABLES del archivo de defunciones generales del año 2014</i>	69
Tabla 16	<i>VARIABLES archivos de defunciones generales, años 2015 y 2016</i>	71
Tabla 17	<i>VARIABLES del archivo de defunciones generales del año 2017</i>	72
Tabla 18	<i>VARIABLES del archivo de defunciones generales del año 2018</i>	73
Tabla 19	<i>Principales variables de los archivos de egresos hospitalarios</i>	75
Tabla 20	<i>Principales variables de los archivos de defunciones generales</i>	76
Tabla 21	<i>Número de egresos y defunciones periodo 2011-2018</i>	85
Tabla 22	<i>Campos preliminares de los egresos hospitalarios</i>	86
Tabla 23	<i>Campos preliminares de las defunciones generales</i>	86
Tabla 24	<i>VARIABLES seleccionadas, pronóstico de causas de morbilidad</i>	100
Tabla 25	<i>Modelo usado para cada causa de morbilidad</i>	102
Tabla 26	<i>VARIABLES seleccionadas, pronóstico de causas de mortalidad</i>	102
Tabla 27	<i>Modelo usado para cada causa de mortalidad</i>	104
Tabla 28	<i>Evaluación modelos morbilidad</i>	105
Tabla 29	<i>Evaluación modelos mortalidad</i>	105
Tabla 30	<i>Métricas de calidad Morbilidad</i>	106
Tabla 31	<i>Métricas de calidad Mortalidad</i>	106
Tabla 32	<i>Resultados del pronóstico de morbilidad 2019-2020</i>	107
Tabla 33	<i>Modelo de líneas de tendencia para las causas de morbilidad</i>	108
Tabla 34	<i>Resultados del pronóstico de mortalidad 2019-2020</i>	109
Tabla 35	<i>Modelo de líneas de tendencia para las causas de mortalidad</i>	110
Tabla 36	<i>Pronóstico y evaluación, principales causas de morbilidad</i>	112
Tabla 37	<i>Pronóstico y evaluación de las principales causas de mortalidad</i>	113

Índice de Figuras

Figura 1 <i>Mapa estructura etaria de América Latina y el Caribe 2015-2060</i>	16
Figura 2 <i>Fundamentación de las variables de investigación</i>	30
Figura 3 <i>Técnicas de minería de datos</i>	34
Figura 4 <i>Cuadrante mágico de Gartner</i>	40
Figura 5 <i>Envejecimiento Saludable</i>	43
Figura 6 <i>Capacidad funcional, intrínseca y oportunidades para la salud pública</i>	44
Figura 7 <i>Características del Adulto Mayor Ecuatoriano</i>	46
Figura 8 <i>Recopilación inicial de datos</i>	57
Figura 9 <i>Exploración de datos de egresos hospitalarios</i>	76
Figura 10 <i>Exploración de datos de defunciones generales</i>	78
Figura 11 <i>Campo fecha_egr de los egresos hospitalarios</i>	79
Figura 12 <i>Campo cau221rx de los egresos hospitalarios</i>	80
Figura 13 <i>Campo edad de los egresos hospitalarios</i>	81
Figura 14 <i>Campo fecha_ingr de los egresos hospitalarios</i>	81
Figura 15 <i>Campo est_civil de las defunciones generales</i>	83
Figura 16 <i>Campo mor_viol de las defunciones generales</i>	83
Figura 17 <i>Campo anio_fall de las defunciones generales</i>	84
Figura 18 <i>Campo edad de las defunciones generales</i>	84
Figura 20 <i>Campo lc1 de las defunciones generales</i>	85
Figura 21 <i>ETL de los egresos hospitalarios con Tableau Prep</i>	87
Figura 22 <i>ETL de las defunciones generales con Tableau Prep</i>	88
Figura 23 <i>Número de egresos de adultos mayores según su sexo</i>	88
Figura 24 <i>Pirámide de edad del adulto mayor atendido</i>	89
Figura 25 <i>Mapa de la residencia de los adultos mayores egresados y porcentajes</i>	90
Figura 26 <i>Sectores de los establecimientos de salud</i>	90
Figura 27 <i>Mapa de la ubicación de los establecimientos de salud y porcentajes</i>	91
Figura 28 <i>Condición de egreso hospitalario</i>	91
Figura 29 <i>Grupos de enfermedades del adulto mayor</i>	92
Figura 30 <i>Porcentaje de las causas de morbilidad</i>	93
Figura 31 <i>Evolución de las causas de morbilidad</i>	93
Figura 32 <i>Diez principales causas de morbilidad</i>	94
Figura 33 <i>Adultos mayores fallecidos según su sexo</i>	95
Figura 34 <i>Pirámide de edad del adulto mayor fallecido</i>	96
Figura 35 <i>Mapa de la residencia de los adultos mayores fallecidos y porcentajes</i>	96
Figura 36 <i>Mapa según el lugar de fallecimiento y porcentaje</i>	97
Figura 37 <i>Porcentaje de causas de mortalidad</i>	98
Figura 38 <i>Evolución de las causas de mortalidad</i>	98
Figura 39 <i>Diez principales causas de mortalidad</i>	99
Figura 40 <i>Pronóstico principales causas de morbilidad obtenido en Tableau</i>	101
Figura 41 <i>Pronóstico principales causas de mortalidad obtenido en Tableau</i>	103

Resumen

El presente estudio abarcó el desarrollar un modelo de series temporales que permita identificar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano, a través del análisis de información histórica de los egresos hospitalarios y de las defunciones a nivel nacional, para determinar la prevalencia de enfermedades y las principales causas de fallecimiento. Para ello con el uso de la herramienta Tableau Public, se realizó un análisis preliminar de las variables principales de los archivos recopilados, así como el diseño y generación de los modelos de series temporales y su pronóstico a través de la técnica de homogenización exponencial de Holt y Winters que se ajusta a la evolución de las series tanto para la morbilidad como para la mortalidad obteniendo una predicción aceptable y un MAPE apropiado inferior al 19,30%, permitiendo cumplir con el objetivo propuesto. Se concluye que por cada 1000 adultos mayores atendidos en el sistema nacional de salud, la prevalencia de las tres principales causas de morbilidad en el año 2019 y 2020 será la Neumonía, la Colelitiasis y colecistitis y la Diabetes mellitus. A nivel nacional la prevalencia de las tres principales causas de mortalidad por cada 10000 adultos mayores, durante el año 2019 y 2020 se definen Enfermedades isquémicas del corazón, Diabetes Mellitus y Enfermedades hipertensivas.

PALABRAS CLAVE:

- **MORBILIDAD**
- **MORTALIDAD**
- **ADULTO MAYOR**
- **SERIES TEMPORALES**

Abstract

The purpose of this research was to develop a time series model that allows identifying the morbidity and mortality behavior of the Ecuadorian elderly, through the analysis of historical information on patient discharges and deaths recorded throughout the country, to determine the prevalence of diseases and the main causes of death. For this, with the use of the Tableau Public tool, a preliminary analysis of the main variables of the files found was carried out, also the design and generation of the time series models, and the forecast was developed through the exponential homogenization technique of Holt - Winters that adjusts to the evolution of both morbidity series and for mortality, obtaining an acceptable prediction and a convenient MAPE of less than 19,30%, achieving the proposed objective. It is concluded that for every 1000 older adults treated in the national health system, the prevalence of the three main causes of morbidity in 2019 and 2020 will be Pneumonia, Cholelithiasis and cholecystitis and Diabetes mellitus. In the country, the prevalence of the three main causes of mortality per 10,000 older adults, during the year 2019 and 2020 will be Ischemic Heart diseases, Mellitus diabetes and Hypertensive diseases.

KEY WORDS:

- **MORBIDITY**
- **MORTALITY**
- **ELDERLY**
- **TIME SERIES**

Capítulo I: Planteamiento del Problema de Investigación

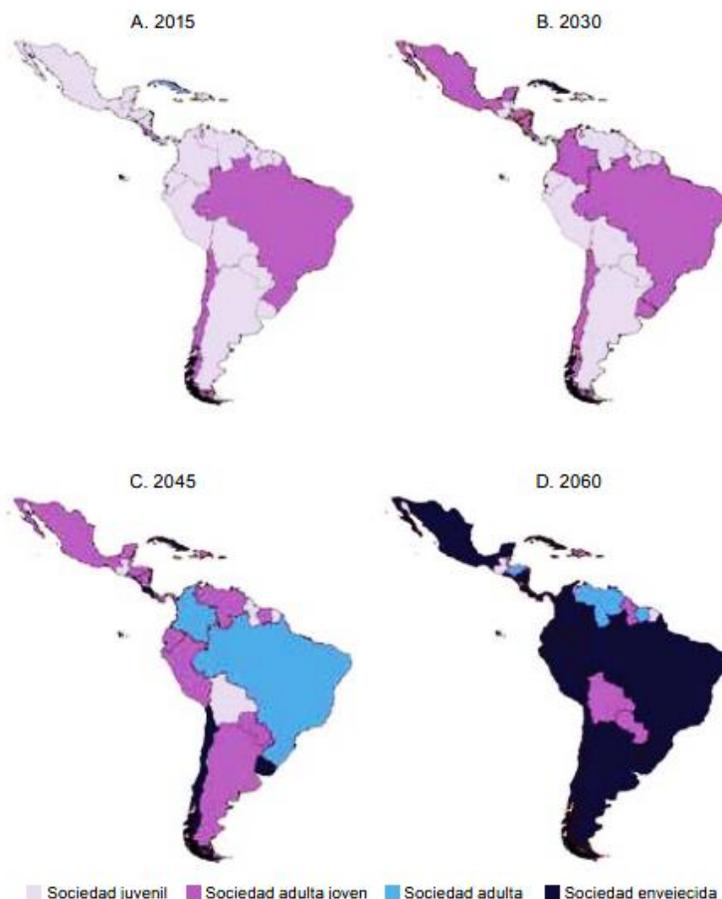
Antecedentes

La Organización Mundial de la Salud (OMS) y la Organización de Naciones Unidas (ONU) consideran que un inmenso desafío que afrontará la sociedad en el siglo XXI, es el envejecimiento poblacional. De modo que para la OMS (2015) entre 2015 y 2050, el porcentaje de personas mayores de 60 años en el planeta pasará del 12% al 22%, en el 2020 este grupo será superior al de niños menores de cinco años, además en 2050, el 80% de adultos mayores vivirá en países con ingresos medianos y bajos. Se presume que dentro de treinta años en China, el país más poblado del mundo, las personas mayores de 65 años corresponderán al 30% de la población total, en comparación con la proporción del 11,9% de 2018 (Wu, 2018). Por consiguiente, el subsistir más años no garantiza una calidad de vida mejor o más digna.

Según estudios del desarrollo social y sostenible de la Comisión Económica para América Latina y el Caribe (CEPAL) en el quinquenio 2030-2035 aumentará el envejecimiento demográfico, los países transitarán a etapas más avanzadas del envejecimiento, Ecuador, por ejemplo, se ubicará en la fase moderadamente avanzada (Huenchuan, 2018). Como se puede observar en la figura 1 para el año 2060, solo 2 países seguirán presentando características de una sociedad juvenil (Guyana Francesa y Guatemala), 6 países se ubicarán en la fase de sociedad adulto joven, en tanto que 3 países (Honduras, República Bolivariana de Venezuela y Surinam) presentarán características de una sociedad adulta, mientras que en el resto de los países de la región presentará una sociedad donde las personas de edad son mayoría y por lo tanto conforman el grupo más relevante en términos numéricos.

Figura 1

Mapa estructura etaria de América Latina y el Caribe 2015-2060



Nota: Tomado de (Envejecimiento, personas mayores y Agenda 2030 para el Desarrollo Sostenible: perspectiva regional y de derechos humanos., 2018, pág. 36).

De acuerdo a las proyecciones del Instituto Nacional de Estadística y Censos (INEC) sobre población y demografía, en el 2030, Ecuador empezará a tener una sociedad que envejece, pues las mujeres tendrán menos de dos hijos. También el INEC (2015), indica que en el 2020 de un total de 17'510.643 de habitantes habrá 1'310.297 personas catalogados como adultos mayores correspondiente al 7,48% de la población.

El problema de investigación

Contexto del problema

Los datos estadísticos de INEC (2018), revelan que a pesar de existir la ley del adulto mayor, políticas públicas y estrategias como Médico del Barrio y Mis Mejores Años, en el país aún se vulnera los derechos de los adultos mayores. Por ejemplo, la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) da a conocer que en diciembre del 2018 solo 112.330 adultos mayores es decir únicamente el 8,88% goza de un Empleo Adecuado o Pleno. La Ley Orgánica de Servicio Público (LOSEP) impone la edad de retiro forzoso del campo laboral, siendo los adultos mayores obligados a vender sus renuncias. La carencia de ingresos económicos repercute en diferentes áreas como: salud, alimentación, seguridad social y vivienda (Romero, 2015).

En el año 2018, el 65,9% de establecimientos de salud que prestan internación hospitalaria se encuentran en el sector privado con fines de lucro, y el 27,8% se encuentran en el público. También el INEC (2018) registró 1'143.765 de egresos hospitalarios en el sector público y privado de los cuales 163.955 es decir un 14,33% correspondían a pacientes de 65 años y más. Además, al cierre del año 2017 se registraron 114.248 adultos mayores con discapacidad de un total de 441.988. En Ecuador, la atención gratuita y especializada de salud, así como el acceso gratuito a medicinas es un derecho (Consejo de Igualdad Intergeneracional - CNII, 2019). Sin embargo, hay costos indirectos asociados con el acceso de estos servicios, como el costo de transporte (escasez de medios de transporte en los sectores rurales) y los medicamentos recetados (no disponibles en los hospitales, centros o subcentros públicos) en estos casos los pacientes tienen que pagar los costos de su bolsillo (Waters & Gallegos, 2012).

Además el (Ministerio de Inclusión Económica y Social - MIES, 2013) determina las siguientes características en los adultos mayores ecuatorianos:

- 45% se encuentra en condiciones de pobreza y extrema pobreza.
- 42% vive en el sector rural.
- 14,6% de hogares pobres se compone de un adulto mayor viviendo solo.
- 14,9% son víctimas de negligencia y abandono.
- 74,3% no posee acceso a la seguridad social contributiva.

Planteamiento del problema

El envejecimiento puede estar influenciado por factores genéticos (individuales) y puede ser acelerado por diversas variables externas que repercuten en el adulto mayor, entre las causas más relevantes se encuentra (Betancourt, 2014):

Con respecto a la asistencia sanitaria, los adultos mayores se encuentran expuestos a insuficientes campañas o estrategias de prevención de enfermedades; limitada capacitación del cuidado integral, paliativo y geriátrico al personal de salud; incremento en la demanda hospitalaria – sector público; costos elevados de atención, medicamentos y rehabilitación; detección tardía de enfermedades; uso inadecuado de recursos; iatrogenia.

Con respecto al entorno social el adulto mayor se enfrenta a condicionantes sociales y asistenciales como: deficiente apoyo social, negligencia, abandono, pobreza, desempleo, indigencia, fraude, inequidad, discriminación, hábitos tóxicos, hábitos sedentarios, escasa educación alimentaria, transformación demográfica, envejecimiento poblacional.

Estas causas ocasionan graves efectos en nuestros adultos mayores tales como: enfermedades crónicas y agudas, deterioro cognitivo, discapacidad – deterioro

físico, incapacidad, discriminación laboral, pérdida de autonomía, depresión, aterosclerosis, malnutrición y finalmente la muerte.

Una vez definidas las causas y efectos con base al análisis inicial del presente estudio, se identifica que la problemática se enfoca en tratar de prevenir **el deterioro de la salud y calidad de vida del adulto mayor ecuatoriano**.

Objetivos

Objetivo general

Desarrollar un modelo de series temporales que permita identificar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano, a través del análisis de información histórica, referente a egresos de pacientes de establecimientos de salud que prestan internación hospitalaria, para determinar la prevalencia de enfermedades, así como las principales causas de defunción de personas de la tercera edad.

Objetivos específicos

OE1. Realizar una revisión inicial de literatura, sobre la aplicación de series temporales en el estudio de la morbilidad y mortalidad del adulto mayor.

OE2. Recolectar, procesar y analizar las bases de datos disponibles sobre la morbimortalidad del adulto mayor ecuatoriano, a través de las herramientas más propicias.

OE3. Identificar y seleccionar la plataforma más conveniente, para diseñar y generar el modelo.

OE4. Evaluar el modelo propuesto utilizando técnicas estadísticas de validación de minería de datos.

Justificación y alcance

El presente trabajo pretende promover el desarrollo de investigaciones que permitan conocer la evolución de uno de los sectores ecuatorianos que requieren atención prioritaria como lo son los adultos mayores. Dentro de pocos años en el mundo se presentará una transición demográfica avanzada, ya que habrá más adultos mayores que niños, y nuestro país no está preparado para afrontar esta realidad inminente. El sistema sanitario sobre todo el público podría colapsar al no poder albergar a esta población que requiere de atención inmediata.

Para estudiar más a fondo la realidad de nuestros adultos mayores se trata de aprovechar fuentes primarias como la información histórica que ofrece el INEC (datos abiertos), eligiendo las principales variables de los datasets que se enfocan en estadísticas anuales de egresos hospitalarios de establecimientos de salud generales, cantonales, especializados y clínicas particulares a nivel nacional durante el período 2011-2018, en los cuales se tratará de observar el comportamiento de las diferentes enfermedades y problemas de salud o morbilidad de la población adulta mayor ecuatoriana, así como la información encontrada en las defunciones generales sobre la mortalidad de dicha población, aplicando minería de datos. Posteriormente se seleccionará la herramienta más propicia para realizar la limpieza de los datos a través del proceso de ETL. También se seleccionará la herramienta de minería de datos, que permita identificar los algoritmos más acordes con el modelo que se desea diseñar. Finalmente se evaluará el modelo propuesto utilizando técnicas de validación de minería de datos.

Para cumplir con los objetivos de la investigación se hará uso de métodos como el cuantitativo, deductivo, inductivo, hipotético-deductivo e histórico (tendencial); así como la metodología CRISP-DM y técnicas de minería de datos.

Ya que el enfoque referente a la salud del adulto mayor, exige la adopción de acciones que tomen en cuenta las necesidades especiales y características particulares del individuo, el presente proyecto propone aplicar series temporales para identificar los patrones de comportamiento de la salud del adulto mayor ecuatoriano en base a los datos registrados en los egresos hospitalarios y defunciones generales, y que permitan determinar un modelo de la morbilidad que se ajuste a esta población.

Esta investigación, brindaría datos que permitan a organizaciones, instituciones públicas, privadas o startups enfocarse en este sector vulnerable.

Hipótesis de investigación

La creación del modelo de series temporales permitirá determinar el comportamiento de la morbilidad del adulto mayor ecuatoriano.

Categorización de las variables de investigación

Variable dependiente

Morbilidad del adulto mayor

Variable independiente

Modelo de series temporales

La demostración de la hipótesis planteada será mediante la comprobación estadística evaluando los resultados del modelo usando técnicas de validación de minería de datos.

Revisión Inicial de Literatura

Objetivo de la búsqueda

Realizar una revisión de literatura para determinar el estado del arte de la aplicación de series temporales en el estudio de enfermedades y de la mortalidad humana.

Preguntas de investigación

Al realizar la revisión de literatura se plantea contestar las siguientes interrogantes:

¿Hay estudios sobre enfermedades o mortalidad humana, en los cuales se aplique series temporales?

¿Cuáles son los modelos de series temporales que se acoplan mejor?

¿Cómo se evalúan estos modelos?

Criterios de inclusión y exclusión

Criterios de Inclusión. Las características o criterios de inclusión que se tomarán en cuenta para la presente SMS fueron los siguientes:

- Artículos en los cuales se reporten estudios sobre la aplicación de minería de datos en la salud humana.
- Artículos en los cuales se reporten estudios de revisión de literatura que sean actuales (a partir del año 2012).
- Artículos científicos y documentos de conferencias publicados en el idioma inglés y español.

Criterios de Exclusión. Los criterios de exclusión que se tomarán en cuenta para la presente SMS serán los siguientes:

- Artículos en los cuales se reporten estudios sobre la evaluación de los signos vitales.
- Artículos en los cuales se reporten estudios sobre hogares inteligentes y datos recolectados por sensores.
- Artículos en los cuales se reporten estudios aplicados a plantas y animales.
- Artículos en los cuales se reporten estudios sobre adquisiciones y ventas.

Definición de la estrategia de búsqueda

Revisión Inicial. Se realizó una revisión bibliográfica como primer paso hacia la realización de un Estudio de Mapeo Sistemático (SMS), este método proporcionó una estructura básica para la búsqueda, selección y revisión de documentos, así como para identificar investigaciones recientes con relevancia para explorar las preguntas de investigación. Las bases digitales que se incluyeron en el proceso de búsqueda fueron: ACM, Google Scholar, IEEE, Scielo, Scopus y Springer.

Validación cruzada de estudios. Se verifica que los estudios cumplan con los criterios de inclusión y exclusión, obteniendo un listado inicial de documentos que servirán de base en las siguientes fases.

Integración del Grupo de Control. En la recopilación de información para determinar el grupo de control, se verifica que los artículos seleccionados cumplan con los criterios de inclusión - exclusión, y analizando inicialmente el título, resumen, y palabras claves, seleccionando los descritos en la tabla 1:

Tabla 1

Grupo de control

Grupo de control	Título de la publicación	Año	Palabras Claves
EC1	(Data mining techniques applied to diagnosis of clinical entities. Revista Cubana de Informática Médica)	2012	CRISP-DM, data mining, arterial hypertension, KDD, clinical diagnosis, WEKA.
EC2	(Ambient temperature and coronary heart disease mortality in Beijing, China: A time series study)	2012	Ambient temperature, Coronary heart disease, Mortality, Cold effect, Hot effect, Gender, Age
EC3	(Text Mining and Medicine: Usefulness in Respiratory Diseases)	2013	Text mining, Respiratory diseases, Chronic obstructive pulmonary disease, Lung cancer, Critically ill respiratory patients
EC4	(Minería de datos de salud: estudio de los factores personales, familiares y vivienda que influyen en las enfermedades de diabetes e hipertensión a partir de la encuesta de atención primaria en salud del Área Metropolitana del Valle de Aburrá)	2014	Atención Primaria en Salud (APS), Minería de datos, Análisis de encuestas
EC5	(An Adaptive Expert System for Automated Advices Generation-Based Semi-continuous M-Health Monitoring)	2014	Expert System, continuous monitoring, analytics, diabetes, blood pressure, healthy advice generation.
EC6	The Use of Data Mining Techniques to Predict Mortality and Length of Stay in an ICU (The use of data mining techniques to predict mortality and length of stay in an ICU)	2016	Data mining, MIMIC 2, eHealth, Mortality Prediction, ICU
EC7	(Descriptive and Predictive Time Series Analysis of Premature Mortality from Noncommunicable Disease among Filipinos)	2016	Noncommunicable disease, cardiovascular disease, cancer, diabetes mellitus, chronic respiratory disease, ARIMA modeling
EC8	(A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals)	2017	Clinical time series, data mining, decision tree, double exponential smoothing, Fuzzy, tolerance rough set
EC9	(Forecasting mortality rates: multivariate or univariate models?)	2018	Mortality rates, multivariate model, univariate model, exponential smoothing, Lee-Carter model
EC10	(Modelling the Frequency of Depression using Holt-Winters Exponential Smoothing Method)	2018	Depression, incidence, time series, stochastic process
EC11	(Merging heterogeneous clinical data to enable knowledge discovery)	2018	Data fusion, interoperability, multi-modal data, big data, phenotyping

Construcción de la cadena de búsqueda. La cadena de búsqueda se formó por la unión de las palabras que más se repiten en el grupo de control, así como los conectores OR para las palabras con contexto semejante, el conector AND para las palabras de diferentes contextos, y el conector NOT que excluye las palabras predecesoras:

(diseases OR morbidity OR health) AND (mortality) AND ("time series" OR "time series analysis") AND (forecast* OR predict*) AND NOT ("ambient assisted living") AND NOT ("vital sign recognition") AND NOT ("smart homes") AND NOT ("sensor data") AND NOT (purchase) AND NOT (sales) AND NOT (plants) AND NOT (animals) AND NOT (weather)

Al aplicar la cadena de búsqueda en los repositorios académicos seleccionados se procede a filtrar solo los documentos que estén en el idioma inglés y español, cuya fecha de publicación sea mayor al 2012 y que sean artículos científicos o documentos de conferencias. Una vez obtenidos los resultados se establecieron los estudios primarios detallados en la Tabla 2.

Tabla 2

Artículos candidatos encontrados en los repositorios digitales

Biblioteca digital	Número de Artículos
Google scholar	8710
Springer	2817
Scopus	285
ACM Digital Library	92
IEEEExplore	47
Scielo	8

Una vez obtenidos los resultados se realizó la revisión de los documentos más relevantes los cuales se listan a continuación en la Tabla 3.

Tabla 3*Estudios seleccionados*

Autor(es)	Año	Título de la publicación	Resumen
Sarimah Abdullah, Napisah Sapii, Sharina Dir, Tg Mardhiah Tg Jalal	2012	(Application of univariate forecasting models of tuberculosis cases in Kelantan)	Este estudio fue diseñado para determinar el patrón y el mejor modelo de series temporales de casos de Tuberculosis (TB) 2003 a 2010 en Kelantan para pronosticar el número de casos de TB en los próximos 2 años. A través de una encuesta con casos registrados que han sido notificados al Departamento de Salud de Kelantan, los datos de la serie son mensuales. Se ha aplicado el modelo univariado (Método Naive, Pronóstico promedio, Técnicas de suavizado exponencial y Método Box-Jenkins) a la serie de datos para determinar el mejor modelo de pronóstico para la serie de datos.: Basado en 72 series de datos mensuales de casos de TB, se observó un aumento en el patrón de tendencia. Se descubrió que la técnica de suavizado exponencial doble es el mejor modelo de serie temporal en comparación con el suavizado exponencial simple, el método de Holt, ARRES y la tendencia y la estacionalidad de Holt Winter para efectos multiplicativos y aditivos. Se estima que los valores futuros aumentarán gradualmente de 115 casos en enero de 2011 a 121 casos en diciembre de 2012. Conclusión: La tendencia creciente de los casos de TB debe tomarse en serio en términos de controlar y prevenir los casos para alcanzar los Objetivos de Desarrollo del Milenio (ODM) para 2015 (Abdullah, Sapii, Dir, & Jalal, 2012).
Ali Al-Sakkaf and Geoff Jones	2013	(Comparison of Time Series Models for Predicting Campylobacteriosis Risk in New Zealand)	La finalidad del estudio es encontrar el modelo de serie temporales más apropiado para los datos de campylobacteriosis de Nueva Zelanda, se tiene consideraciones prácticas adicionales dado un posible cambio estructural, es decir, un cambio específico y repentino en respuesta a las intervenciones implementada. Se utilizó un enfoque metodológico univariado para predecir los casos mensuales de enfermedad utilizando los datos de vigilancia de Nueva Zelanda de la incidencia de campylobacteriosis de 1998 a 2009, dejando el año 2009 fuera del conjunto de datos para

Autor(es)	Año	Título de la publicación	Resumen
Anna L. Buczak, Benjamin Baugher, Erhan Guven, Liane C. Ramac-Thomas, Yevgeniy Elbert, Steven M. Babin and Sheri H. Lewis	2015	(Fuzzy association rule mining and classification for the prediction of malaria in South Korea)	<p>Modelo de validación. Los dos mejores modelos se ajustaron a los datos completos de 1998-2009 y se usaron para predecir cada mes de 2010. Los modelos de intervención Holt-Winters (multiplicativo) y ARIMA (aditivo) se consideraron los mejores modelos para predecir la campylobacteriosis en Nueva Zelanda, proporcionando un buen rendimiento de predicción (Al-Sakkaf & Jones, 2013).</p> <p>Se describe la aplicación de un método para crear modelos de predicción utilizando la Minería de reglas de la Asociación Fuzzy para extraer relaciones entre datos epidemiológicos, meteorológicos, climáticos y socioeconómicos de Corea. Construyendo dos clasificadores y sus resultados se fusionaron para convertirse en un modelo de predicción de la malaria. Los futuros casos de malaria de 7 a 8 semanas se predicen como BAJO, MEDIO o ALTO, según las recomendaciones de los usuarios, ALTO se considera un brote en el futuro. Este documento demuestra que el enfoque basado en datos de esta investigación puede usarse para la predicción de diferentes enfermedades (Buczak, y otros, 2015).</p>
Matthew G. Crowson, Kristine Schulz, and Debara L. Tucci	2016	(National Utilization and Forecasting of Otopical Antibiotics: Medicaid Data Versus “Dr. Google”)	<p>Los investigadores obtuvieron datos trimestrales nacionales de utilización de medicamentos resumidos de Medicaid y datos de motores de búsqueda Google Trends (GT) semanales para ciprofloxacina-dexametasona (CD), ofloxacina (OF) y cortisporina (CS) ototópicos desde enero de 2008 hasta julio de 2014. Se utilizó un análisis de series de tiempo para evaluar estacionalidad de las recetas, el método de Holt-Winter para pronosticar los volúmenes de recetas trimestrales y las correlaciones de Pearson para comparar los datos de GT y Medicaid. Encontrando que los ototópicos OF, CD y CS tienen una variación estacional sinusoidal con picos de volumen de prescripción de Medicaid que ocurren en el verano. Después de 2012, OF fue el ototópico más comúnmente recetado, y se pronostica que esta tendencia continuará. Se pronostica que el uso de CS disminuirá. El interés de búsqueda de los usuarios de Google en estos agentes ototópicos demostró una variación estacional análoga.</p>

Autor(es)	Año	Título de la publicación	Resumen
Eunjoo Yang, Hyun Woo Park, Yeon Hwa Choi, Jusim Kim, Lkhagvadorj Munkhdalai, Ibrahim Musa and Keun Ho Ryu	2018	(A Simulation-Based Study on the Comparison of Statistical and Time Series Forecasting Methods for Early Detection of Infectious Disease Outbreaks)	<p>Los análisis de GT sobre el interés en los antibióticos ototópicos pueden ser útiles para los proveedores y administradores de atención médica como un método complementario para evaluar las tendencias de utilización de la atención médica (Crowson, Schulz, & Tucci, 2016).</p> <p>Se compara cuatro algoritmos de detección de brotes temporales diferentes: el CUmulative SUM (CUSUM), el Early Aberration Reporting System (EARS), el promedio móvil integrado autorregresivo (ARIMA) y el algoritmo Holt-Winters. La comparación se realizó basándose no solo en 42 series temporales diferentes generadas teniendo en cuenta las tendencias, la estacionalidad y los brotes que se producen al azar, sino también los datos diarios y semanales del mundo real relacionados con la infección por diarrea. Los algoritmos fueron evaluados utilizando diferentes métricas. Estos fueron a saber, sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), puntaje F1, error porcentual absoluto medio simétrico (sMAPE), error cuadrático medio (RMSE) y desviación absoluta media (MAD). Aunque los resultados de la comparación mostraron un mejor rendimiento para el método EARS C3 con respecto a los otros algoritmos, a pesar de las características de los datos subyacentes de la serie temporal, Holt-Winters mostró un mejor rendimiento cuando la frecuencia de referencia y los valores de los parámetros de dispersión fueron inferiores a 1,5 y 2, respectivamente (Yang, y otros, 2018).</p>

Conclusión

Al efectuar la revisión inicial de literatura no se encuentra estudios que describan específicamente la aplicación de series temporales en la identificación de patrones de comportamiento de la morbimortalidad del adulto mayor ecuatoriano, sin embargo, en forma general con respecto a la morbilidad y mortalidad como se puede evidenciar en la Tabla 3 se obtienen diferentes estudios, en los cuales se hallan

detallados aspectos relevantes como: series temporales y detección de enfermedades, predicción de la mortalidad, pronosticar de volúmenes de recetas, diagnóstico y tratamientos clínicos, que pueden servir como referencia en el desarrollo del proyecto. El estudio de los datos ha sido de gran apoyo en el área de la salud, ya que ha permitido múltiples estudios sobre atención médica, atención crítica, diagnóstico y tratamiento de enfermedades respiratorias y crónicas como: diabetes, cáncer, hipertensión, insuficiencia cardiaca, lesiones renales, fibrilación auricular, accidentes cerebrovasculares, tuberculosis, campylobacteriosis, etc.

Las series temporales han cobrado gran relevancia en el transcurso de los años, el analizar los datos médicos y de ser posible hacer predicciones precisas han cobrado una mayor notoriedad. Para el análisis y pronósticos de estas series se ha empleado algunas técnicas entre las cuales se encuentran minería de reglas de la Asociación Fuzzy, método Naive, suavizado exponencial como el método de Holt, y Holt-Winters, método CUMulative SUM (CUSUM), el Early Aberration Reporting System (EARS), Box-Jenkins y el promedio móvil integrado autorregresivo (ARIMA). En algunos estudios se consideraron los mejores modelos para predecir a los de suavizado exponencial en especial al de Holt Winter para efectos multiplicativos y aditivos.

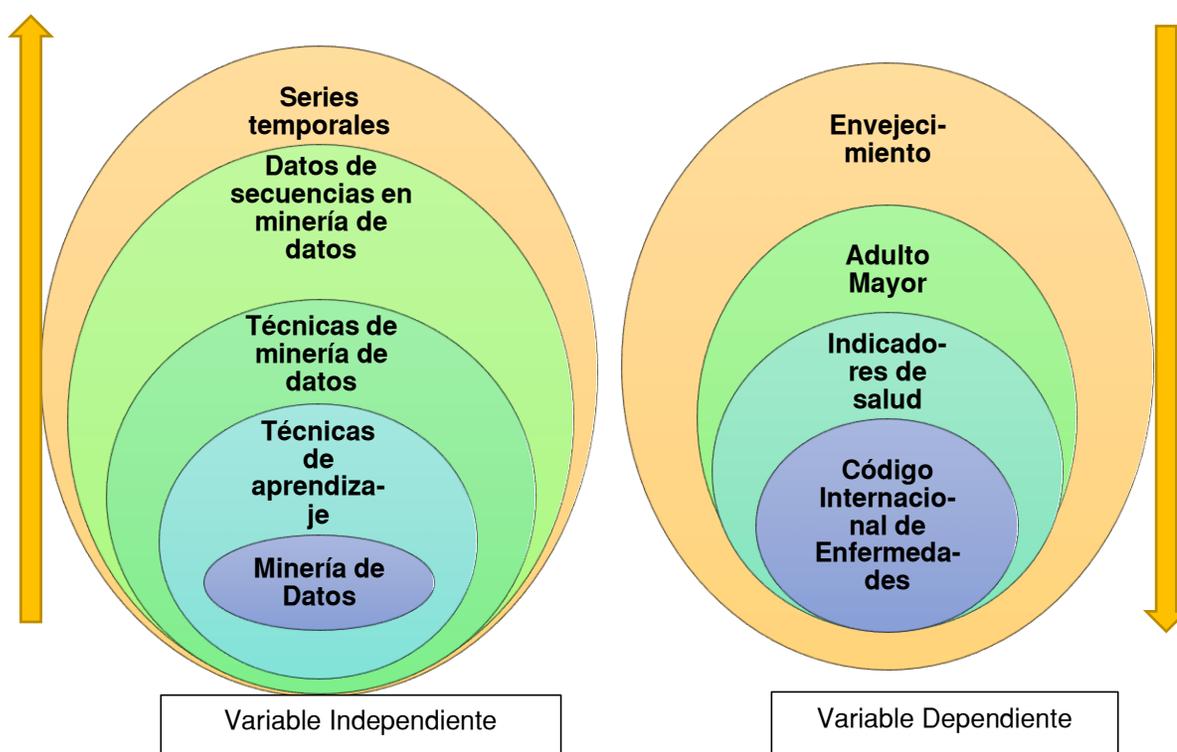
Los algoritmos fueron evaluados utilizando diferentes métricas, entre estas se encuentran sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), error porcentual absoluto medio simétrico (sMAPE), error cuadrático medio (RMSE) y desviación absoluta media (MAD), sin embargo, el Error Porcentual Absoluto Medio (MAPE), es más comúnmente usado en la evaluación del desempeño predictivo.

Capítulo II: Marco Teórico

En el marco teórico se intenta describir los fundamentos técnicos de las variables dependiente e independiente del tema de estudio, por ello se propone la siguiente jerarquía descrita en la figura 2.

Figura 2

Fundamentación de las variables de investigación



Variable Independiente

Minería de datos

La minería de datos o data mining en inglés, es encontrar patrones útiles en los datos, también se conoce como descubrimiento de conocimiento, aprendizaje automático y análisis predictivo. Sin embargo, cada término tiene una connotación ligeramente diferente dependiendo del contexto (Kotu & Deshpande, 2015).

La minería de datos que puede ir desde una simple matriz de pocas observaciones numéricas, hasta una matriz compleja de millones de observaciones y con miles de variables. El acto de minería de datos utiliza algunos métodos computacionales especializados para descubrir estructuras significativas y útiles en los datos. Estos métodos computacionales se han derivado de los campos de las estadísticas, el aprendizaje automático y la inteligencia artificial. La disciplina de la minería de datos coexiste y está asociada con varias áreas relacionadas, como los sistemas de bases de datos, la visualización de datos, el análisis exploratorio de datos y evaluación del desempeño (Kotu & Deshpande, 2015).

El propósito de la minería de datos es normalmente crear un modelo descriptivo o un modelo predictivo. De acuerdo con (Jain & Srivastava, 2013), un modelo descriptivo presenta en forma concisa las principales características del conjunto de datos. Es esencialmente un resumen de los puntos de datos, lo que permite estudiar aspectos importantes del conjunto de datos. Típicamente, un **modelo descriptivo** se encuentra a través de la minería de datos no supervisada; es decir, un enfoque ascendente donde los datos hablan por sí mismos. La minería de datos no dirigida encuentra patrones en el conjunto de datos, pero deja la interpretación de los patrones al minero de datos.

El propósito de un **modelo predictivo** es permitir que el minero de datos prediga un valor desconocido (a menudo futuro) de una variable específica; la variable objetivo. Si el valor objetivo es uno de un número predefinido de etiquetas discretas (de clase), la tarea de minería de datos se llama clasificación. Si la variable objetivo es un número real, la tarea es la regresión (Jain & Srivastava, 2013).

Técnicas de aprendizaje

Otro concepto crucial en minería de datos es el modelo de aprendizaje. Hay dos enfoques o técnicas de aprendizaje, supervisadas y no supervisadas (Trivedi, Dey, Kumar, & Panda, 2017).

Aprendizaje supervisado

Según (Trivedi, Dey, Kumar, & Panda, 2017) la técnica de aprendizaje supervisado utiliza un conjunto de datos para hacer predicciones. Los datos de capacitación incluyen la etiqueta de la clase. El modelo supervisado requiere la disponibilidad de clase para cada muestra de entrenamiento que no siempre se da en un problema del mundo real. Agrupa la técnica supervisada en clasificación y regresión, las redes neuronales y las máquinas de vectores de soporte (SVM) son las aplicaciones más populares de técnicas supervisadas.

Aprendizaje no supervisado

Este método no requiere etiquetas de clase para la construcción del modelo. Se construye utilizando un comportamiento normal sin usar etiquetas de clase, y luego se usa para "detectar casos atípicos que se desvían del comportamiento normal". Los modelos particularmente no supervisados pueden detectar los nuevos tipos de fraude, ya que no se limitan al patrón que en el conjunto de datos de entrenamiento. A menudo tienen más exploración en comparación con los modelos supervisados (Trivedi, Dey, Kumar, & Panda, 2017).

Técnicas de minería de datos

Conforme con (Trivedi, Dey, Kumar, & Panda, 2017) la predicción y descripción en minería de datos se logra mediante el uso de las siguientes técnicas:

- Clasificación, es uno de los modelos de aprendizaje más comunes en la aplicación de minería de datos que puede clasificar el registro de datos en una de las clases predefinidas. Las técnicas de clasificación más populares son los árboles de decisión, los clasificadores bayesianos, la máquina de vectores de soporte y la red neuronal.

- Clustering o agrupamiento, es el enfoque de minería de datos más común que se basa en el aprendizaje no supervisado. La tarea de la técnica de agrupamiento es similar a la tarea de la técnica de clasificación. La técnica de clasificación intenta separar los registros de datos en un conjunto de clases predefinidas, mientras en el método de agrupamiento, las etiquetas de clase son desconocidas (no está disponible en el conjunto de datos) y depende del algoritmo de agrupamiento encontrar una clase aceptable. El resultado de la agrupación definida como los objetos en un grupo tienen una gran similitud, mientras que son muy diferentes a los objetos en los otros grupos. es un método útil que puede usarse como una herramienta de detección atípica para detectar atributos que se desvían de los patrones normales. Algunos métodos de agrupación, mezclan datos numéricos y nominales en grandes bases de datos y las técnicas de agrupamiento más comunes son K-means.

- Predicción, es muy similar al método de clasificación; estima que los próximos valores se basan en los patrones de un conjunto de datos. La única diferencia es que, en la predicción, las clases no son un atributo discreto sino un continuo. Hay dos tipos principales de predicciones; predecir valores de datos no disponibles o predecir una etiqueta de clase para el conjunto de datos. Este método utiliza una gran cantidad de valores pasados como patrones predefinidos para estimar probables valores futuros. Los modelos de redes lógicas y neuronales son el ejemplo más común de este método.

- Detección de valores atípicos, ocurre cuando cualquier elemento de datos es muy diferente o incompatible con el conjunto de datos restante. Los valores atípicos son elementos de datos que no se pueden agrupar en una clase o clúster dado.

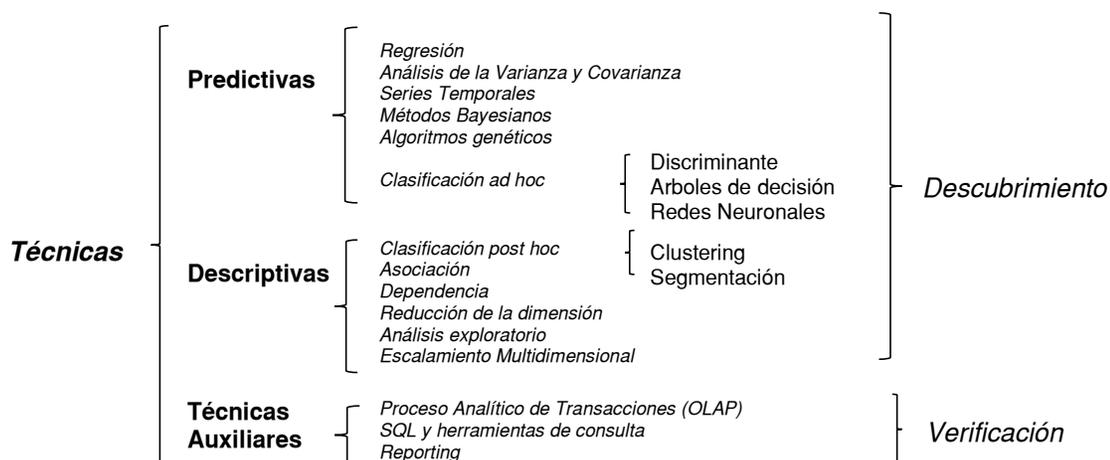
- Regresión, este método es un modelo estadístico que expone la relación entre una o más variables. Muchas experiencias prácticas utilizaron la regresión logística como estándar.

- Visualización, es la mejor técnica para proporcionar patrones complejos a través de la presentación clara de los datos.

Según (Pérez, 2007) se distingue entre técnicas orientadas al descubrimiento de conocimiento, como las predictivas (en las que las variables pueden clasificarse inicialmente en dependientes e independientes) y las descriptivas (en las que todas las variables tienen inicialmente el mismo estatus), así como las auxiliares basadas en estadísticas descriptivas, consultas e informes enfocados a la verificación (ver figura 3).

Figura 3

Técnicas de minería de datos



Nota: Tomado de (Pérez, 2007)

Datos de secuencias en minería de datos

De acuerdo con (Han, Kamber, & Pei, 2011) una secuencia es una lista ordenada de eventos, los datos se clasifican en función de las características de los eventos que se describen en tres grupos:

- Los datos de series temporales, consisten en secuencias largas de datos numéricos, registrados a intervalos de tiempo iguales (por ejemplo, por hora, por día, - por mes o por año). Los datos de series temporales se pueden generar mediante muchos procesos naturales y económicos, como los mercados de valores, y observaciones científicas, médicas o naturales.
- Los datos de secuencia simbólica, son secuencias largas de eventos o datos nominales, que generalmente no se observan a intervalos de tiempo iguales. Para muchas de esas secuencias, las brechas (es decir, lapsos entre eventos grabados) no importan mucho. Los ejemplos incluyen secuencias de compras de clientes y transmisiones de clics web, así como secuencias de eventos en ciencia e ingeniería y en desarrollos naturales y sociales.
- Las secuencias biológicas, incluyen secuencias de ADN y proteínas. Dichas secuencias suelen ser muy largas y tienen un significado semántico importante, complicado pero oculto. Aquí, las brechas suelen ser importantes.

Series Temporales

Una serie temporal se puede definir como un conjunto ordenado de mediciones en un intervalo de tiempo; formalmente, se considera una secuencia TS de datos ordenados en el tiempo $TS = \{TSt, t=1, \dots, N\}$, donde t representa el tiempo, N es el número de observaciones realizadas durante ese período de tiempo y TSt es el valor medido en el instante t . En algunos dominios, el valor de no una sino varias

observaciones puede medirse en cada punto de tiempo, lo que lleva a series de tiempo multidimensionales (Anguera, Lara, Lizcano, Martínez, & Pazos, 2012). Las tareas principales de la minería de datos en series de datos incluyen agrupamiento, clasificación, detección de fallas, reconocimiento de patrones y predicción. (Li & Wang, 2018).

Análisis de regresión y tendencias en series temporales

El análisis de regresión de datos de series de tiempo se ha estudiado sustancialmente en los campos de estadísticas y análisis de señales, sin embargo, a menudo es necesario ir más allá del análisis de regresión puro y realizar análisis de tendencias para muchas aplicaciones prácticas. El análisis de tendencias crea un modelo integrado que utiliza los siguientes cuatro componentes o movimientos principales para caracterizar datos de series temporales (Han, Kamber, & Pei, 2011):

Tendencia o movimientos a largo plazo. Indican la dirección general en la que se mueve un gráfico de series de tiempo a lo largo del tiempo, por ejemplo, utilizando el promedio móvil ponderado y los métodos de mínimos cuadrados para encontrar curvas de tendencia.

Movimientos cíclicos. Son las oscilaciones a largo plazo sobre una línea o curva de tendencia.

Variaciones estacionales. Son patrones casi idénticos que una serie temporal parece seguir durante las temporadas correspondientes de años sucesivos, por ejemplo, las temporadas de compras navideñas.

Movimientos aleatorios. Estos caracterizan cambios esporádicos debido a eventos fortuitos tales como disputas laborales o cambios de personal anunciados dentro de las empresas.

Pronósticos en series temporales

Indican (Hyndman & Athanasopoulos, 2018) que los métodos de homogenización o suavizamiento exponencial se propusieron a fines de la década de 1950, y han motivado algunos de los métodos de pronóstico más exitosos. Los pronósticos producidos usando métodos de suavizado exponencial son promedios ponderados de observaciones pasadas, y los pesos decaen exponencialmente a medida que las observaciones envejecen. En otras palabras, cuanto más reciente sea la observación, mayor será el peso asociado.

Método estacional de Holt-Winters. Holt y Winters extendieron el método de Holt para capturar la estacionalidad.

El método estacional Holt-Winters comprende la ecuación de pronóstico y tres ecuaciones de suavizado: una para el nivel $\hat{\mu}_t$, otra para la tendencia b_t y otra para el componente estacional s_t , con los parámetros de suavizado correspondientes α , β * y γ (Hyndman & Athanasopoulos, 2018).

Se usa m para denotar la frecuencia de la estacionalidad, es decir, el número de estaciones en un año. Por ejemplo, para datos trimestrales $m = 4$, y para datos mensuales $m = 12$.

Hay dos variaciones de este método que difieren en la naturaleza del componente estacional. Se prefiere el método aditivo cuando las variaciones estacionales son aproximadamente constantes a través de la serie, mientras que el método multiplicativo se prefiere cuando las variaciones estacionales cambian proporcionalmente al nivel de la serie.

Con el método aditivo, el componente estacional se expresa en términos absolutos en la escala de la serie observada, y en la ecuación de nivel, la serie se ajusta estacionalmente restando el componente estacional. Dentro de cada año, el componente estacional sumará aproximadamente cero. Con el método multiplicativo, el componente estacional se expresa en términos relativos (porcentajes), y la serie se ajusta estacionalmente dividiendo entre el componente estacional. Dentro de cada año, el componente estacional sumará aproximadamente m (Hyndman & Athanasopoulos, 2018).

La forma del componente para el método aditivo es:

$$\hat{Y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta*(\ell_t - \ell_{t-1}) + (1-\beta*)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

Mientras que la forma del componente para el método multiplicativo es:

$$\hat{Y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta*(\ell_t - \ell_{t-1}) + (1-\beta*)b_{t-1}$$

$$s_t = \gamma(y_t / \ell_{t-1} + b_{t-1}) + (1-\gamma)s_{t-m}$$

donde k es la parte entera de $(h-1)/m$, lo que garantiza que las estimaciones de los índices estacionales utilizados para pronosticar procedan del último año de la muestra. La ecuación de nivel muestra un promedio ponderado entre la

observación ajustada estacionalmente ($y_{t-s_{t-m}}$) y el pronóstico no estacional ($\hat{y}_{t-1} + b_{t-1}$) para el tiempo t (Hyndman & Athanasopoulos, 2018).

Evaluación del pronóstico. La principal guía para la estimación del desempeño predictivo, que pondera que tan cerca está la variable pronosticada de la serie de datos representada, es el Error Porcentual Absoluto Medio (MAPE) y obedece a la siguiente expresión matemática:

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \frac{|\hat{Y}_t^s - Y_t^a|}{Y_t^a} * 100$$

donde T es el número de periodos, Y_t^a es el valor predicho por el modelo, y \hat{Y}_t^s es el valor observado.

El MAPE mide el valor medio del error absoluto en términos porcentuales al valor real de la variable, cuanto más bajo sea, mejor será el rendimiento del método de minería de datos (Ríos, 2008), la tabla 4 contiene valores típicos de MAPE y su interpretación.

Tabla 4

Interpretación de los valores típicos de MAPE

MAPE	Interpretación
<10	Pronósticos altamente precisos
10-20	Buen pronóstico
20-50	Previsión razonable
> 50	Pronósticos inexactos

Nota: Tomado de (Montaño Moreno, Palmer Pol, Sesé Abad, & Cajal Blasco, 2013)

Plataformas de análisis e inteligencia de negocios

En el Cuadrante Mágico para plataformas de análisis e inteligencia de negocios (Gartner, Inc., 2020) se evalúa las fortalezas y debilidades de 20 proveedores que

considera más importantes en el mercado, y ofrece a los lectores un gráfico que representa a los proveedores según su capacidad de ejecución y su visión completa. El gráfico está dividido en cuatro cuadrantes: jugadores de nicho, retadores, visionarios y líderes (ver figura 4).

Microsoft y Tableau dominan el mercado, ofreciendo carteras inigualables que cubren los "dos modos" de los que habla Gartner en el informe. Power BI de Microsoft presenta un amplio conjunto de métodos de implementación y funciones para prácticamente cada caso de uso. Tableau ofrece el mejor conjunto de capacidades de visualización de datos en el mercado. Qlik redondea a los líderes en esta iteración, con una fuerte búsqueda de la supremacía del mercado para el caso de uso integrado.

Figura 4

Cuadrante mágico de Gartner



Nota: Tomado de (Gartner, Inc., 2020)

Las principales características de las plataformas líderes se describen en la tabla

5.

Tabla 5

Plataformas de análisis e inteligencia de negocios

ANÁLISIS	POWER BI (MICROSOFT)	TABLEAU	QLIK
ANTECEDENTES	<ul style="list-style-type: none"> • Por doce años Microsoft se posiciona en la categoría de Líderes. • Es la herramienta más simple de usar entre los proveedores principales de BI. • Integra y visualiza datos de fuentes dispares, ya sean basadas en nube, locales o Hadoop. • Soporta visualizaciones basadas en lenguaje R, incluyendo previsión agrupamiento y árboles de decisión. • Disponible a través de la plataforma Azure de Microsoft. Opción de escritorio disponible. Las cuentas de nube son necesarias para compartir visualizaciones. • Precio Power BI Pro: \$9.99 por usuario y por mes. Power BI Premium: incluye almacenamiento y actividades en la nube tiene un precio mensual de \$4,995. Power BI Desktop: versión de prueba gratuita. 	<ul style="list-style-type: none"> • Lleva siete años cómo uno de los líderes. • Es una herramienta de visualización más atractiva e intuitiva, propicia la exploración ilimitada de datos gracias a la tecnología VizQL. • Permite el procesamiento de datos en memoria para rápidas visualizaciones al explorar datos. Conjuntos de datos más grandes requieren llamadas directas a la fuente. • Soporte totalmente integrado para lenguajes R y Python. Además, ofrece herramientas nativas para clustering y pronósticos. • Puede desplegarse en su propia nube o en terceras plataformas incluyendo Amazon y Microsoft. • Precio Tableau Desktop Personal: \$35 por usuario al mes Tableau Desktop Professional: \$70 por usuario al mes. Tableau Server: \$35 por usuario, al mes: Tableau Online: \$42 por usuario al mes. Tableau Public: software gratuito • Límite de suscripción de Tableau Online: 100GB de almacenamiento de datos en la nube. 	<ul style="list-style-type: none"> • Por nueve años se encuentra entre los líderes. • Incluye un motor de datos en memoria, guardar datos en memoria permite actualizar gráficos rápidamente mientras los usuarios exploran visualmente los datos. • Motor de datos robusto permite combinar y visualizar múltiples fuentes de datos. • No hay soporte para R o Python, pero ya vienen incluidos. • El análisis predictivo, agrupamiento y regresiones, son posibles a través de las conexiones API del software de terceros. • Ofrece producto de nube SaaS completamente gestionado. La mayoría de los clientes escoge ejecutar la versión de servidor. • Precio Enterprise: ofrece para un usuario un token de uso ilimitado, o 10 pases temporales de inicio de sesión, por un valor de \$1.500. Nube: por un usuario \$20 al mes, para múltiples inicios de sesión por usuario \$25 al mes.

ANÁLISIS	POWER BI (MICROSOFT)	TABLEAU	QLIK
CARACTERÍSTICAS	<ul style="list-style-type: none"> • Límite de suscripción: Pro 10 GB almacenamiento por usuario. Premium hasta 100 TB. Puede requerir costos adicionales para escalar la capacidad de datos. • Incremento de la eficiencia en las compañías para extraer informes de manera autónoma, sin requerir conocimientos avanzados. • Ofrece una herramienta denominada Quick Insights que determina las correlaciones y patrones de los datos, produciendo gráficos personalizados o automáticos. • Integración del análisis avanzado a través de scripts y objetos visuales de R, Microsoft Azure Machine Learning y Azure Stream Analytics, asegurando la escalabilidad de las aplicaciones para proyectos de mayor envergadura. • Amplio soporte de administración, seguridad e infraestructura de Microsoft. 	<ul style="list-style-type: none"> • Los usuarios pueden profundizar y explorar datos sin ninguna experiencia de programación, es independiente y autónomo lo cual permite utilizarlo en cualquier punto del proceso de tratamiento de datos. • Tiene la capacidad de conectarse a una multitud de fuentes de datos como Excel, texto, Json, Access, pdf, estadístico, etc. e incluso a conexiones directas de sistemas incluso servicios Cloud • Puede utilizar Tableau Prep o la API de esta herramienta para la extracción sistemática de datos. • Integra análisis matemático y estadísticas sencillas de aplicar. • Si el usuario posee conocimientos de programación, puede usar a Tableau front-end de estadísticas y análisis avanzados. • Tableau ha sido capaz de capitalizar el impulso de su mercado y obtener más socios de integración y soporte. 	<p>QlikView Personal Edition: versión de prueba gratis Qlik Sense Cloud Basic: software gratuito.</p> <ul style="list-style-type: none"> • Límite de suscripción de Qlik Sense Cloud Business 500 GB de almacenamiento en la nube por grupo de trabajo. • Permite uniones en tiempo real y destaca las relaciones entre las entidades a través de su motor asociativo subyacente. • Admite el ETL autónomo y el almacenamiento de datos. • Está diseñada para escalar en las empresas de manera más eficiente al proporcionar a TI el soporte necesario, las herramientas y, si es necesario, el soporte de los socios, para ponerse en marcha rápidamente. • Los conocimientos de datos se pueden generar rápidamente a partir de un usuario competente, ya que las capacidades de análisis de datos están limitadas solo por sus capacidades de creación de scripts. • Debe mejora su estrategia de plataforma y fortalecer sus conjuntos de características SmartData Discovery y Embedded Advanced Analytics.

Variable Dependiente

Envejecimiento

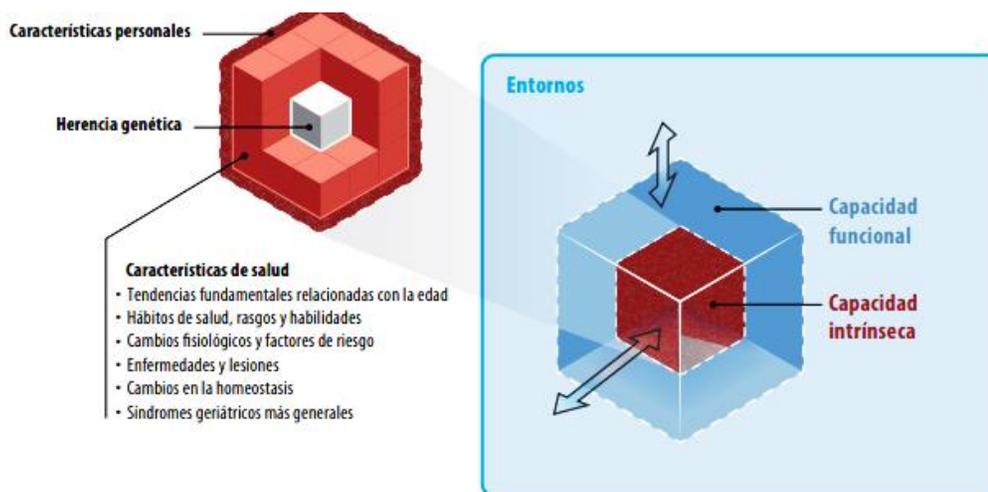
Para (Herrera, 2008) el envejecimiento es un proceso universal, alcanza a todos los individuos sin excepción; es intrínseco, genéticamente programado; es progresivo, comienza prácticamente desde el nacimiento; y es deletéreo, por lo tanto ineludiblemente termina con la muerte. El envejecimiento puede ser fisiológico si los parámetros biológicos son normales y se conserva la posibilidad de relación con el medio en que se mueve, o bien patológico en presencia de enfermedades que alteren el curso normal del envejecimiento.

En el plano biológico, conforme la OMS (2015) el envejecimiento está asociado con la acumulación de una gran variedad de daños moleculares y celulares. Con el tiempo, estos daños reducen gradualmente las reservas fisiológicas, aumentan el riesgo de muchas enfermedades y disminuyen en general la capacidad del individuo; es posible controlar ciertas enfermedades con medicina, mientras que otras requieren múltiples tratamientos con grandes efectos secundarios, provocando en ciertos casos incapacidades, impidiéndoles realizar simples actividades como leer o preparar alimentos.

Además según la OMS (2015), el envejecimiento saludable como se puede visualizar en la figura 5, es el proceso de desarrollar y sostener la capacidad funcional que permite el bienestar en la vejez. La figura 6 indica que cuándo se tiene en cuenta a la población en conjunto, las capacidades intrínseca y funcional pueden variar en la segunda mitad de la vida.

Figura 5

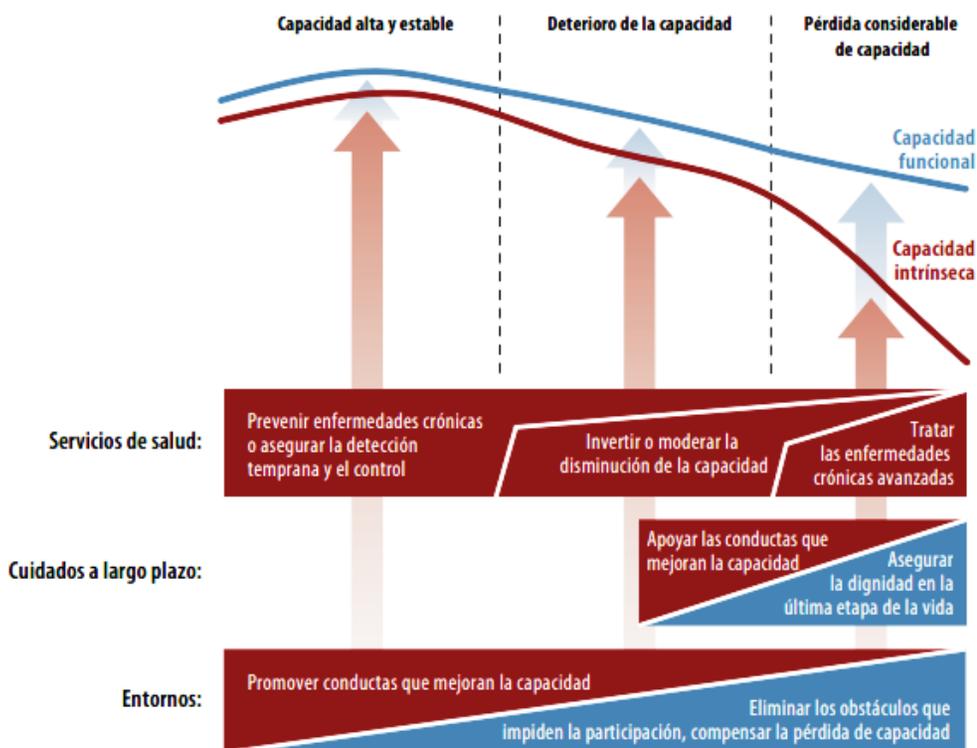
Envejecimiento Saludable



Nota: Tomado de (Organización Mundial de la Salud - OMS, 2015).

Figura 6

Capacidad funcional, intrínseca y oportunidades para la salud pública



Nota: Tomado de (Organización Mundial de la Salud - OMS, 2015)

Adulto Mayor

De acuerdo con la OMS (2015), aunque todos los individuos se agrupan igualmente en el ciclo del adulto mayor, las personas de 60 a 74 años son consideradas de edad avanzada; de 75 a 90 viejas o ancianas y los mayores de 90 suelen designar como grandes viejos o grandes longevos. La Naciones Unidas considera ancianos a toda persona mayor de 65 años para los países desarrollados y de 60 para los países en desarrollo (Huenchuan, 2018).

La Dra. Herrera (2008), indica que al momento de abordar estudios sobre el adulto mayor se debe tener presente las siguientes categorías:

- Adulto mayor enfermo, aquel que padece alguna enfermedad, aguda o crónica, pero que no cumple criterios de anciano frágil.
- Adulto mayor inmovilizado, pasa la mayor parte de su tiempo en cama, la cual solo puede abandonar con la ayuda de una tercera persona, hay una importante dificultad para desplazarse.
- Adulto mayor frágil o de alto riesgo, es aquel que cumple con tres o más de las siguientes pautas: psicogeriatría, edad >60 años preferiblemente >75 años, pluripatología relevante, polifarmacia, enfermedad principal incapacitante, algún grado de deterioro mental, trastornos afectivos (ansiedad o depresión), hospitalización en el último año, problemática social en relación con su estado de salud (vive solo, viudez reciente, cambio de domicilio reciente, bajo nivel socioeconómico, institucionalizado, etc.).
- Paciente geriátrico, es el adulto mayor frágil al cual se le añade algún proceso agudo que lo obliga a consultar médico y/o a hospitalizarse.

La constitución política del estado ecuatoriano (Constitución de la República del Ecuador, 2008) considera personas adultas mayores a aquellas que hayan cumplido los sesenta y cinco años de edad. En el Art. 36. de dicha constitución se señala:

Las personas adultas mayores recibirán atención prioritaria y especializada en los ámbitos público y privado, en especial en los campos de inclusión social y económica, y protección contra la violencia...

En la figura 7 se puede observar las características más relevantes del adulto mayor ecuatoriano.

Figura 7

Características del Adulto Mayor Ecuatoriano.



Indicadores de salud

Debido a alteraciones existentes en el ambiente o por la falta de interés del cuidado de la salud, el ser humano se encuentra propenso a un proceso que modifica su estado normal de salud conocido como enfermedad (o morbosidad). Existen varios

tipos de enfermedades los cuales pueden ser infecciosos o virales pueden producirse internamente o externamente causando daño en el cuerpo del ser humano (Ceballos & Gutierrez, 2014). La epidemiología se puede medir estudiando a las personas que enferman es decir analizando su morbilidad, o a su vez a las personas que fallecen esto es la mortalidad de las enfermedades (Organización Panamericana de la Salud - OPS, 2018).

Entre los indicadores más comunes que usa la OPS (2018) se encuentran:

Conteo, indica el impacto de una enfermedad en términos numéricos exactos, dentro de un período específico y en un lugar determinado.

Tasa, muestra la relación entre dos números, el numerador es el número absoluto de veces que ocurre el evento de interés en un período específico. El denominador es la población de referencia o estudiada en el mismo tiempo.

Tasa de Prevalencia

Es la probabilidad de que un individuo perteneciente a la población base sea afectado por el evento (enfermedad) de interés en un momento dado, es esencial para planificar y organizar los servicios y recursos existentes y obtener apoyo adicional, cuando sea necesario. Se calcula de la siguiente manera:

$$\text{Tasa de prevalencia} = \frac{\text{Número de casos existentes en un lugar y momento en el tiempo}}{\text{Número total de personas de la población en el mismo lugar y tiempo}} * 10^n$$

Tasa de Morbilidad

Los indicadores de morbilidad tienen la finalidad de medir la ocurrencia de enfermedades, lesiones y discapacidades en las poblaciones, en un lugar y tiempo determinado, entre ellos tenemos la incidencia y la prevalencia.

Tasa de Mortalidad

Indica el número de defunciones por lugar, intervalo de tiempo y causa, los datos de mortalidad se usan para cuantificar los problemas de salud, así como para determinar o monitorear prioridades o metas en salud. Conforme la OMS (2013) a diferencia de la morbilidad, la muerte es un evento único, claramente identificable, que refleja la ocurrencia y la gravedad de una enfermedad, sin embargo también se puede hacer uso de la prevalencia y de la incidencia para su cálculo.

En nuestro país el (Sistema de Indicadores Sociales del Ecuador - SIISE, 2015) impulsa la creación de herramientas efectivas y actuales para apoyar la toma de decisiones de política social, sus indicadores de resultado miden las consecuencias de los procesos sociales y reflejan los niveles de satisfacción de necesidades básicas alcanzados, dentro de los orientados al adulto mayor se registran dos indicadores.

Diez principales causas de morbilidad hospitalaria en adultos mayores

Para (SIISE, 2015) el indicador recoge las diez principales causas de morbilidad de los adultos mayores durante cada año, expresadas como porcentaje del total de casos de morbilidad en ese año. Las estadísticas hospitalarias del INEC, registran las causas de enfermedad y atención según el manual de clasificación estadística internacional de enfermedades de la OMS (CIE - 10), lista de 298 causas.

% de morbilidades hospitalarias

$$\text{por causa(x) en adultos mayores} = \frac{\text{nro de egresos de mayores de 64 años por la causa (x) en el año t}}{\text{total de mayores de 64 años que presentaron morbilidad en el año t}} \times 100$$

Diez principales causas de mortalidad en el adulto mayor.

Según (SIISE, 2015) el indicador recoge las diez principales causas de mortalidad en el adulto mayor durante cada año, expresadas como porcentaje del total de muertes en ese año. Las estadísticas vitales, nacimientos y defunciones registran las

causas de muerte según el CIE-10, para lo cual tomamos la codificación de las causas de muerte a tres dígitos o caracteres.

$$\% \text{ de muertes por la causa (x)} = \frac{\text{nro. de muertes de adultos mayores por la causa (x) en el año t}}{\text{total de adultos mayores en el año t}} \times 100$$

Código Internacional de Enfermedades (CIE)

La Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud, Décima Revisión (CIE-10) fue respaldada por la Cuadragésima Tercera Asamblea Mundial de la Salud en mayo de 1990 y se empezó a usar en los Estados Miembros de la Organización Mundial de la Salud a partir de 1994 (Organización Mundial de la Salud - OMS, 2017).

Tiene como propósito permitir el registro sistemático, análisis, interpretación, y comparación de los datos de mortalidad y morbilidad recolectados en diferentes países o áreas, y en diferentes épocas. Se utiliza para convertir los términos diagnósticos y otros problemas de salud, de palabras a códigos alfanuméricos que permiten su fácil almacenamiento y posterior recuperación para el análisis de la información (Organización Mundial de la Salud - OMS, 2017).

Capítulos

Conforme a la OPS (2003), la clasificación está dividida en 22 capítulos (ver tabla 6). Cada capítulo contiene suficientes categorías de tres caracteres para cubrir sus necesidades; no se han usado todos los códigos posibles, de manera de dar espacio para una futura revisión y expansión.

Tabla 6

Capítulos CIE-10

Valor	Categoría
1	I Ciertas enfermedades infecciosas y parasitarias (A00 -B99)
2	II Tumores [Neoplasias] (C00-D48)
3	III Enfermedades de la sangre y de los órganos hematopoyéticos y ciertos trastornos que afectan el mecanismo de la inmunidad (D50-D89)
4	IV Enfermedades endocrinas, nutricionales y metabólicas (E00-E90)
5	V Trastornos mentales y del comportamiento (F00 - F99)
6	VI Enfermedades del sistema nervioso (G00-G99)
7	VII Enfermedades del ojo y sus anexos (H00-H59)
8	VIII Enfermedades del oído y de la apófisis mastoides (H60-H95)
9	IX Enfermedades del sistema circulatorio (I00-I99)
10	X Enfermedades del sistema respiratorio (J00-J99)
11	XI Enfermedades del aparato digestivo (K00-K93)
12	XII Enfermedades de la piel y el tejido subcutáneo (L00-L99)
13	XIII Enfermedades del sistema osteomuscular y del tejido conectivo (M00-M99)
14	XIV Enfermedades del aparato genitourinario (N00-N99)
15	XV Embarazo, parto y puerperio (O00-O99)
16	XVI Ciertas afecciones originadas en el período perinatal (P00-P96)
17	XVII Malformaciones congénitas, deformidades y anomalías cromosómicas (Q00-Q99)
18	XVIII Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte (R00-R99)
19	XIX Traumatismos, envenenamientos y algunas otras consecuencias de causas externas (S00-T98)
20	XX Causas externas de morbilidad y de mortalidad (V01-Y98)
21	XXI Factores que influyen en el estado de salud y contacto con los servicios de salud (Z00-Z99)
22	XXII Códigos para propósitos especiales (U00-U99)

Nota: Tomado de (Organización Panamericana de la Salud, 2008)

Lista especial de tabulación para la morbilidad

Según la OPS (2003), la lista de tabulación para la morbilidad contiene 298 rúbricas detalladas. El propósito de la lista de morbilidad es que sirva de base para la elaboración de listas nacionales y también para comparaciones entre países. La lista permite presentar datos sobre pacientes hospitalizados y, con adaptaciones adecuadas en las que se agreguen algunas rúbricas y se expandan otras, especialmente del Capítulo XVIII (Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte) y del Capítulo XXI (Factores que influyen en el estado de

salud y contacto con los servicios de salud), sirve también para presentar información de otras fuentes, tales como atención ambulatoria y encuestas e investigaciones sobre salud. De acuerdo con la OPS (2008) en el volumen I SE recopila a mayor detalle la lista de 298 causas.

Categorías de tres caracteres

Dentro de cada grupo, conforme la OPS (2003) se señala que algunas de las categorías de tres caracteres corresponden a afecciones únicas seleccionadas debido a su frecuencia, gravedad o vulnerabilidad a las acciones de salud pública, en cambio otras corresponden a enfermedades con alguna característica común. Se provee generalmente espacio para “otras” afecciones o condiciones que deben ser clasificadas y contienen enfermedades diferentes, pero poco frecuentes, así como afecciones “no especificadas”. Según la OPS (2008) en el volumen I, además de describir la lista de 298 causas de morbilidad, también se refiere a la clasificación de estas categorías de mortalidad.

Capítulo III: Metodología

Se considera a la presente investigación con un enfoque cuantitativo y se categoriza como analítica, por ello se realizó un estudio de caso para obtener los resultados correspondientes, siguiendo los pasos descritos a continuación:

- Definición del caso de estudio
- Elaboración de las preguntas de investigación
- Localización de las fuentes y recopilación de datos
- Análisis e interpretación de resultados
- Elaboración del informe

Métodos

La información se obtuvo de fuentes primarias -datos abiertos- provenientes del INEC, y se hizo uso de los siguientes métodos:

Método cuantitativo, este método es uno de los instrumentos principales de la estadística. Se basa en los números para investigar, analizar y comprobar datos e información; este método permitió especificar y delimitar la asociación, generalizar y objetivar los resultados obtenidos, así como analizar toda la información numérica.

Método deductivo, este método va de lo general a lo particular, proporcionando la lógica de las conclusiones finales a partir de premisas determinadas, el mismo permitió la recolección de la información relacionada a las actividades y problemas suscitados en el tema planteado.

Método inductivo, este método va de lo particular a lo general, mediante la generalización del comportamiento del estudio observado. El empleo de este método en el proyecto permitió buscar y encontrar soluciones adecuadas para el desarrollo del mismo.

Método hipotético-deductivo, Permitió la verificación de la hipótesis, para determinar las respectivas conclusiones.

Método histórico (tendencial), está relacionado al conocimiento de las distintas etapas del estudio en sucesión cronológica con el fin de conocer la evolución del objeto de investigación, en este caso se aplicó en el estudio de prevalencia de la morbimortalidad y mortalidad.

Metodología

Para dar respuesta al **OE1** del CAPÍTULO I, en la elaboración de la **revisión** inicial de la **literatura** se siguieron los siguientes pasos:

- a) Planteamiento
- b) Criterios de inclusión y exclusión
- c) Grupo de control
- d) Cadena de búsqueda
- e) Selección de estudios
- f) Conclusiones del estado del arte

Para cumplir con los objetivos **OE2**, **OE3** y **OE4** planteados, con respecto al desarrollo de la propuesta del modelo se decidió hacer uso de **CRISP-DM**.

Cross Industry Standard Process for Data Mining en abreviatura CRISP-DM, es una metodología de minería de datos que proporciona una guía del ciclo de vida de un proyecto de análisis de datos y comprende 6 fases principales. Las fases I, II y III sirvieron de apoyo para el cumplimiento del OE2; la fase IV permitió la realización del OE3; y las fases V y VI propiciaron la ejecución del OE4; seguidamente se describen cada una de ellas en base a (Villena, 2016).

Fase I. Comprensión del negocio (Business Understanding) es la fase inicial, permite comprender los objetivos de proyecto y así convertirlos en conocimiento, aquí se plantean las siguientes actividades:

- Evaluación de la situación actual del caso de estudio
- Establecimiento de los objetivos

Fase II. Comprensión de los datos (Data Understanding) en esta etapa se analizan los datos para comprenderlos, reconocer problemas de calidad, descubrir conocimiento preliminar, y/o encontrar datos interesantes que permitan formar una hipótesis de acuerdo a la información oculta de acuerdo a la colección de datos iniciales, esta fase incluye las actividades siguientes:

- Recopilación inicial de datos
- Descripción de los datos
- Exploración de los datos
- Verificación de calidad de datos

Fase III. Preparación de datos (Data Preparation) en esta fase se realiza el análisis y selección de los datos de la recolección inicial para adaptarlos a las herramientas de modelado. Además, se eligió los registros, atributos y las tablas, también se realizó la transformación y limpieza de datos para luego ser usados en la herramienta de modelamiento. Esta fase comprende las siguientes actividades:

- Selección de datos
- Formateo de datos
- Estructuración, integración y formateo

Fase IV. Modelado (Modeling) en esta etapa se escoge y utiliza las técnicas de modelado más apropiadas para el proyecto, y se calibran sus parámetros a valores óptimos. En esta fase se consideraron las siguientes actividades:

- Selección de la técnica de modelado
- Selección de atributos para el análisis del algoritmo y construcción del modelo
- Evaluación del modelo

Fase V. Evaluación de resultados obtenidos (Evaluation) la fase de evaluación está orientada a la evaluación de los resultados obtenidos mediante la construcción del modelo que permitan alcanzar la calidad suficiente desde un enfoque de análisis de datos. Además, en esta fase se determina si los resultados del proceso de análisis de datos deben ser aplicados o no.

Fase VI. Despliegue (Deployment) con el modelo construido y validado se procede a documentar y presentar los resultados de manera comprensible.

Capítulo IV: Aplicación de la Metodología CRISP-DM

Comprensión del negocio

Evaluación de la situación actual del caso de estudio

El proceso de envejecimiento difiere de acuerdo a las condiciones sociales, educativas, culturales y económicas. Los adultos mayores en el estado ecuatoriano son aquellos habitantes que poseen sesenta y cinco años de edad o más, a los cuales el gobierno otorga beneficios como la exención o devolución del pago del impuesto a la renta, el pago de tarifas reducidas en espectáculos públicos, transportes y el acceso preferencial a varios servicios de salud. “Algunos cuentan con seguridad social, un pequeño porcentaje cuenta con seguridad privada mientras que otros acceden a la pensión jubilar no contributiva. No obstante, en su mayoría las personas adultas mayores no tienen ninguno de estos servicios y prestaciones” (Consejo de Igualdad Intergeneracional - CNII, 2019).

En el estudio se procura aprovechar la información generada por el INEC, en base a los egresos hospitalarios y a las defunciones generales, se trata de promover el análisis de datos, ya que se encuentra poca información relacionada con este grupo vulnerable, además, se pretende dar una visión general de lo que se ha suscitado con la población del adulto mayor ecuatoriano durante el período 2011-2018, así como dar a conocer las principales causas de morbilidad y mortalidad para los años 2019 y 2020.

Establecimiento de los objetivos de DM

- Pronosticar la prevalencia de las tres principales causas de morbilidad del adulto mayor durante el 2019 y 2020 por cada 1000 pacientes egresados.
- Pronosticar la prevalencia de tres principales causas de mortalidad del adulto mayor durante el 2019 y 2020 por cada 10000 habitantes.

Comprensión de los datos

Recopilación inicial de datos

Para estudiar más a fondo la realidad de nuestros adultos mayores se trata de aprovechar fuentes primarias y fidedignas como la información histórica que ofrece el INEC (datos abiertos), eligiendo datasets correspondientes a los años 2011 - 2018 y que se enfocan en estadísticas anuales de egresos hospitalarios del sistema nacional de salud y en las defunciones generales, los cuales incluyen información del sector público, privado con fines de lucro y sin fines de lucro, permitiendo observar el comportamiento de los diferentes problemas de salud o morbilidad de la población adulta mayor ecuatoriana así como la mortalidad de la misma.

Los archivos descargados fueron descomprimidos y se encontraban en formato .sav. (ver figura 8), y a pesar de sobrepasar el millón de registros (en el caso de los egresos) no superaban los 250 MB en tamaño. Posteriormente se proceden a guardar en el mismo directorio para su posterior recuperación y tratamiento desde Tableau Prep Builder.

Figura 8

Recopilación inicial de datos

 defunciones_2011.sav	Archivo SAV	26.967 KB
 defunciones_2012.sav	Archivo SAV	15.699 KB
 defunciones_2013.sav	Archivo SAV	19.030 KB
 defunciones_2014.sav	Archivo SAV	15.546 KB
 defunciones_2015.sav	Archivo SAV	15.708 KB
 defunciones_2016.sav	Archivo SAV	14.920 KB
 defunciones_2017.sav	Archivo SAV	15.445 KB
 defunciones_2018.sav	Archivo SAV	16.464 KB
 egresos_hospitalarios_2011.sav	Archivo SAV	238.697 KB
 egresos_hospitalarios_2012.sav	Archivo SAV	144.320 KB
 egresos_hospitalarios_2013.sav	Archivo SAV	188.358 KB
 egresos_hospitalarios_2014.sav	Archivo SAV	171.800 KB
 egresos_hospitalarios_2015.sav	Archivo SAV	210.342 KB
 egresos_hospitalarios_2016.sav	Archivo SAV	204.327 KB
 egresos_hospitalarios_2017.sav	Archivo SAV	199.931 KB
 egresos_hospitalarios_2018.sav	Archivo SAV	222.789 KB

Descripción de los datos

En cuanto a la explicación de la información se inicia con el archivo de egresos hospitalarios del año 2011 el que posee 1'133.556 de registros y 29 variables o campos (ver tabla 7).

Tabla 7

Variables del archivo de egresos hospitalarios del año 2011

CAMPO	SIGNIFICADO	TIPO
cod_est	Código del establecimiento de salud	Categórica
ani_inv	Año de la estadística	Numérico
prov_ubie	Provincia ubicación del establecimiento	Categórica
cant_ubie	Cantón ubicación del establecimiento	Categórica
par_ubie	Parroquia ubicación del establecimiento	Categórica
Clase	Clase de establecimiento	Categórica
Tipo	Tipo de establecimiento	Categórica
entidad	Entidad a la que pertenece el establecimiento	Categórica
sector	Sector al que pertenece el establecimiento	Categórica
mes_inv	Mes de registro/investigación	Numérico
negreso	Número de orden de egreso	Numérico
cond_edad	Condición de la edad del paciente	Categórica
edad_pac	Edad del paciente	Numérico
sexo_pac	Sexo del paciente	Categórica
prov_pac	Provincia de residencia habitual del paciente	Categórica
cant_pac	Cantón de residencia habitual del paciente	Categórica
parr_pac	Parroquia de residencia habitual del paciente	Categórica
dia_ingr	Día de ingreso	Numérico
mes_ingr	Mes de ingreso	Numérico
anio_ingr	Año de ingreso	Numérico
dia_egr	Día egreso	Numérico
mes_egr	Mes egreso	Numérico
dia_estad	Días de estada	Numérico
con_egrpa	Condición del egreso	Categórica
cau_cie10	Causa de lista internacional detallada a 4 o más dígitos de la CIE-10	Categórica
cau221rx	Lista especial de 221 grupos	Categórica

CAMPO	SIGNIFICADO	TIPO
cap221rx	Capítulo lista 221	Categórica
cau298rx	Lista de causas 298	Categórica
edad_repac	Edad real del paciente	Numérico

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2011)

El archivo de egresos del año 2012 incluye 1'156.237 de registros y los 29 campos descritos en la tabla 8.

Tabla 8

Variables del archivo de egresos hospitalarios del año 2012

CAMPO	SIGNIFICADO	TIPO
Cod_est	Código del establecimiento de salud	Categórica
Ani_inv	Año de la estadística	Numérica
Prov_ubie	Provincia ubicación del establecimiento	Categórica
Cant_ubie	Cantón ubicación del establecimiento	Categórica
Parr_ubie	Parroquia ubicación del establecimiento	Categórica
Clase	Clase de establecimiento	Categórica
Tipo	Tipo de establecimiento	Categórica
Entidad	Entidad a la que pertenece el establecimiento	Categórica
Sector	Sector al que pertenece el establecimiento	Categórica
Mes_inv	Mes de registro/investigación	Numérico
Cond_edad	Condición de la edad del paciente	Categórica
Edad_pac	Edad del paciente	Numérico
Sexo_pac	Sexo del paciente	Categórica
Prov_pac	Provincia de residencia habitual del paciente	Categórica
Cant_pac	Cantón de residencia habitual del paciente	Categórica
Parr_pac	Parroquia de residencia habitual del paciente	Categórica
Dia_ingr	Día de ingreso	Numérico
Mes_ingr	Mes de ingreso	Numérico
Anio_ingr	Año de ingreso	Numérico
Dia_egr	Día egreso	Numérico
Mes_egr	Mes egreso	Numérico
Dia_estad	Días de estada	Numérico

CAMPO	SIGNIFICADO	TIPO
Con_egrpa	Condición del egreso	Categoría
Cau_cie10	Causa de lista internacional detallada a 4 o más dígitos de la CIE-10	Categoría
Cau221rx	Lista especial de 221 grupos	Categoría
Cap221rx	Capítulo lista 221	Categoría
Cau298rx	Lista de causas 298	Categoría
Causa3	Causa de lista internacional detallada a 3 dígitos de la CIE-10	Categoría

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2012)

El archivo de egresos del año 2013 contiene 1'178.989 de registros y 30 campos (ver tabla 9).

Tabla 9

Variables del archivo de egresos hospitalarios del año 2013

CAMPO	SIGNIFICADO	TIPO
cod_est	Código del establecimiento de salud	Categoría
prov_ubi	Provincia ubicación del establecimiento	Categoría
cant_ubi	Cantón ubicación del establecimiento	Categoría
parr_ubi	Parroquia ubicación del establecimiento	Categoría
clase	Clase de establecimiento	Categoría
tipo	Tipo de establecimiento	Categoría
entidad	Entidad a la que pertenece el establecimiento	Categoría
sector	Sector al que pertenece el establecimiento	Categoría
mes_inv	Mes de registro/investigación	Número
sexo	Sexo del paciente	Categoría
edad	Edad del paciente	Número
cod_edad	Condición de la edad del paciente	Categoría
prov_res	Provincia de residencia habitual del paciente	Categoría
cant_res	Cantón de residencia habitual del paciente	Categoría
parr_res	Parroquia de residencia habitual del paciente	Categoría
día_ingr	Día de ingreso	Número
mes_ingr	Mes de ingreso	Número
anio_ingr	Año de ingreso	Número

CAMPO	SIGNIFICADO	TIPO
fecha_ingr	Fecha de ingreso	Cronológica
dia_egr	Día egreso	Numérico
mes_egr	Mes egreso	Numérico
anio_egr	Año del egreso	Numérico
fecha_egr	Fecha de egreso	Cronológica
dia_estad	Días de estada	Numérico
con_egrpa	Condición del egreso	Categorica
cau_cie10	Causa de lista internacional detallada a 4 o más dígitos de la CIE-10	Categorica
cap221rx	Capítulo lista 221	Categorica
cau221rx	Lista especial de 221 grupos	Categorica
cau298rx	Lista de causas 298	Categorica
causa3	Causa de lista internacional detallada a 3 dígitos de la CIE-10	Categorica

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2013)

El archivo de egresos del año 2014, incorpora 1'192.749 registros y 32 campos (ver tabla 10).

Tabla 10

Variables del archivo de egresos hospitalarios del año 2014

CAMPO	SIGNIFICADO	TIPO
cod_est	Código del establecimiento de salud	Categorica
prov_ubi	Provincia ubicación del establecimiento	Categorica
cant_ubi	Cantón ubicación del establecimiento	Categorica
parr_ubi	Parroquia ubicación del establecimiento	Categorica
clase	Clase de establecimiento	Categorica
tipo	Tipo de establecimiento	Categorica
entidad	Entidad a la que pertenece el establecimiento	Categorica
sector	Sector al que pertenece el establecimiento	Categorica
mes_inv	Mes de registro/investigación	Numérico
sexo	Sexo del paciente	Categorica
edad	Edad del paciente	Numérico
cod_edad	Condición de la edad del paciente	Categorica

CAMPO	SIGNIFICADO	TIPO
etnia	Definición étnica del paciente	Categoría
dis_pac	Discapacidad del paciente	Categoría
prov_res	Provincia de residencia habitual del paciente	Categoría
cant_res	Cantón de residencia habitual del paciente	Categoría
parr_res	Parroquia de residencia habitual del paciente	Categoría
dia_ingr	Día de ingreso	Número
mes_ingr	Mes de ingreso	Número
anio_ingr	Año de ingreso	Número
fecha_ingr	Fecha de ingreso	Cronológica
dia_egr	Día egreso	Número
mes_egr	Mes egreso	Número
anio_egr	Año del egreso	Número
fecha_egr	Fecha de egreso	Cronológica
dia_estad	Días de estada	Número
con_egrpa	Condición del egreso	Categoría
cau_cie10	Causa de lista internacional detallada a 4 o más dígitos de la CIE-10	Categoría
cap221rx	Capítulo lista 221	Categoría
cau221rx	Lista especial de 221 grupos	Categoría
cau298rx	Lista de causas 298	Categoría
causa3	Causa de lista internacional detallada a 3 dígitos de la CIE-10	Categoría

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2014)

El archivo de egresos del año 2015 consta de 1'161.044 de registros, el del año 2016 posee 1'128.004 registros, el del 2017 registra 1'143.765, el archivo de egresos hospitalarios más reciente corresponde al del año 2018 consta de 1'164.659 registros, todos estos archivos tienen 36 campos que se describen a continuación en la tabla 11:

Tabla 11

Variables archivo de egresos hospitalarios, años 2015 -2018

CAMPO	SIGNIFICADO	TIPO
prov_ubi	Provincia ubicación del establecimiento	Categoría
cant_ubi	Cantón ubicación del establecimiento	Categoría

CAMPO	SIGNIFICADO	TIPO
parr_ubi	Parroquia ubicación del establecimiento	Categórica
area_ubi	Área ubicación del establecimiento	Categórica
clase	Clase de establecimiento	Categórica
tipo	Tipo de establecimiento	Categórica
entidad	Entidad a la que pertenece el establecimiento	Categórica
sector	Sector al que pertenece el establecimiento	Categórica
mes_inv	Mes de registro/investigación	Numérico
nac_pac	Nacionalidad del paciente	Categórica
nom_pais	Nombre del País	Categórica
cod_pais	Código del país	Numérico
sexo	Sexo del paciente	Categórica
cod_edad	Condición de la edad del paciente	Categórica
edad	Edad del paciente	Numérico
etnia	Definición étnica del paciente	Categórica
prov_res	Provincia de residencia habitual del paciente	Categórica
cant_res	Cantón de residencia habitual del paciente	Categórica
parr_res	Parroquia de residencia habitual del paciente	Categórica
area_res	Área de residencia habitual del paciente	Categórica
anio_ingr	Año de ingreso	Numérico
mes_ingr	Mes de ingreso	Numérico
dia_ingr	Día de ingreso	Numérico
fecha_ingr	Fecha de ingreso	Cronológica
anio_egr	Año del egreso	Numérico
mes_egr	Mes egreso	Numérico
dia_egr	Día egreso	Numérico
fecha_egr	Fecha de egreso	Cronológica
dia_estad	Días de estada	Numérico
con_egrpa	Condición del egreso	Categórica
esp_egrpa	Especialidad del egreso	Categórica
cau_cie10	Causa de lista internacional detallada a 4 o más dígitos de la CIE-10	Categórica
causa3	Causa de lista internacional detallada a 3 dígitos de la CIE-10	Categórica
cap221rx	Capítulo lista 221	Categórica
cau221rx	Lista especial de 221 grupos	Categórica
cau298rx	Lista de causas 298	Categórica

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2017)

Continuando con la descripción de los datos en los archivo de defunciones generales se visualiza diferentes variables para cada año, en el 2011 se encontraron 62.304 registros que incluye 48 campos (ver Tabla 12), en el 2012 se enumeran 63.511 registros con 66 campos (ver Tabla 13), en el año 2013 se contabiliza 63.104 registros con 52 campos (ver Tabla 14), el año 2014 contiene 64.770 registros y 48 campos (ver Tabla 15), el año 2015 consta de 66.598 de registros y el 2016 posee 68.848 registros con los mismos 45 campos (ver Tabla 16), el del 2017 registra 70.841 y 44 campos (ver Tabla 17), finalmente el archivo más reciente corresponde al del año 2018 incorporando 72.789 registros y 45 campos (ver Tabla 18).

Tabla 12

Variables del archivo de defunciones generales del año 2011

CAMPO	SIGNIFICADO	TIPO
prov_insc	Provincia de inscripción	Categórica
cant_insc	Cantón de inscripción	Categórica
parr_insc	Parroquia de inscripción	Categórica
anio_insc	Año de inscripción	Numérico
mes_insc	Mes de inscripción	Numérico
acta_insc	Acta de inscripción	Numérico
ofi_insc	Oficina de inscripción	Categórica
sexo	Sexo	Categórica
anio_nac	Año de nacimiento	Numérico
mes_nac	Mes de nacimiento	Numérico
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
edad	Edad al fallecer	Numérico
cod_edad	Código de edad al fallecer	Categórica
prov_fall	Provincia de fallecimiento	Categórica
cant_fall	Cantón de fallecimiento	Categórica
parr_fall	Parroquia de fallecimiento	Categórica
area_fall	Área de fallecimiento	Categórica
reg_fall	Región de fallecimiento	Categórica
lugar_fall	Lugar de ocurrencia del fallecimiento	Categórica

CAMPO	SIGNIFICADO	TIPO
cert_por	Certificación de defunción	Categórica
est_civ	Estado civil y/o conyugal	Categórica
prov_res	Provincia de residencia habitual del fallecido	Categórica
cant_res	Cantón de residencia habitual del fallecido	Categórica
parr_res	Parroquia de residencia habitual del fallecido	Categórica
area_res	Área de residencia habitual del fallecido	Categórica
reg_res	Región Natural de residencia habitual	Categórica
resident	Tipo de residencia del fallecido	Categórica
sabe_leer	Sabe leer y escribir	Categórica
niv_inst	Nivel de instrucción alcanzado	Categórica
p_etnica	Autoidentificación étnica	Categórica
embara	Defunción de Mujeres en edad Fértil	Categórica
fallecioaluz	Falleció dando a luz	Categórica
semana6	Últimas 6 semanas dio a luz	Categórica
tipo_cert	Tipo de certificación	Categórica
causa4original	Causa de fallecimiento CIE-10 a 4 dígitos	Categórica
causa4	Causa básica de fallecimiento CIE-10 a 4 dígitos	Categórica
causa	Causa básica de fallecimiento CIE-10 a 3 dígitos	Categórica
mu_violen	Muerte violentas	Categórica
pro_med	Se realizó necropsia/autopsia	Categórica
paren	Parentesco del fallecido con la persona que inscribe	Categórica
cauletra	Código del cuarto caracteres de la causa básica	Categórica
mu_matern	Casos de muertes maternas	Categórica
causa80	Lista de tabulación 2 para la mortalidad subcategorías (80 causas)	Categórica
causa103	Lista condensada (103 causas)	Categórica
causa67A	Lista condensada (67 causas) A	Categórica
causa67B	Lista condensada desagregada (67 causas) B	Categórica
lc	Lista corta de causas de defunción	Categórica

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2011)

Tabla 13

Variables del archivo de defunciones generales del año 2012

CAMPO	SIGNIFICADO	TIPO
ofi_insc	Oficina del Registro Civil	Categórica

CAMPO	SIGNIFICADO	TIPO
prov_insc	Provincia de inscripción	Categórica
cant_insc	Cantón de inscripción	Categórica
parr_insc	Parroquia de inscripción	Categórica
anio_insc	Año de inscripción	Numérico
mes_insc	Mes de inscripción	Numérico
acta_insc	Acta de inscripción	Numérico
Sexo	Sexo	Categórica
anio_nac	Año de nacimiento	Numérico
mes_nac	Mes de nacimiento	Numérico
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
edad	Edad al fallecer	Numérico
cod_edad	Código de edad al fallecer	Categórica
prov_res	Provincia de residencia habitual del fallecido	Categórica
cant_res	Cantón de residencia habitual del fallecido	Categórica
parr_res	Parroquia de residencia habitual del fallecido	Categórica
area_res	Área de residencia habitual del fallecido	Categórica
est_civil	Estado civil y/o conyugal	Categórica
sabe_leer	Sabe leer y escribir	Categórica
niv_inst	Nivel de instrucción alcanzado	Categórica
p_etnica	Autoidentificación étnica	Categórica
lugar_ocur	Lugar de ocurrencia del fallecimiento	Categórica
cod_esta	Código del Establecimiento de Salud	Categórica
prov_fall	Provincia de fallecimiento	Categórica
cant_fall	Cantón de fallecimiento	Categórica
parr_fall	Parroquia de fallecimiento	Categórica
area_fall	Área de fallecimiento	Categórica
causa4	Causa básica de fallecimiento CIE-10 a 4 dígitos	Categórica
causa	Causa básica de fallecimiento CIE-10 a 3 dígitos	Categórica
mu_matern	Mortalidad Materna	Categórica
mor_mat	Mortalidad Materna (Si muerte materna - No muerte materna)	Categórica
mu_violen	Muertes accidentales y/o violentas	Categórica
lugar_muviolen	Lugar donde ocurrió el hecho muerte accidental o violenta	Categórica
pro_med	Se realizó necropsia/autopsia	Categórica
cer_por	Certificación de defunción	Categórica
paren_fall	Parentesco del fallecido con la persona que inscribe	Categórica
tipo_cer	Certificación médica	Categórica

CAMPO	SIGNIFICADO	TIPO
causa103	Lista condensada (103 causas)	Categórica
cauletra	Especificación de causa	Categórica
causa67A	Lista condensada (67 causas) A	Categórica
causa67B	Lista condensada desagregada (67 causas) B	Categórica
causa80	Lista de tabulación 2 para la mortalidad subcategorías (80 causas)	Categórica
t1	Total de registros	Numérico
certimed	Tipo de certificación	Categórica
reg_fall	Región de fallecimiento	Categórica
residente	Residentes y no residentes	Categórica
gedad	Grupo de edad (Grandes grupos)	Categórica
gedad1	Grupo de edad (12 años y más)	Categórica
gedad2	Grupo de edad 2 (Grandes grupos)	Categórica
grupoedad	Grupos de edad	Categórica
dised14	Grupos de edad	Categórica
gedad3	Grupos de edad 3	Categórica
lc1	Lista corta de causas de defunción	Categórica
lceti	Códigos de Agrupamiento	Categórica
lcnun	Código de la lista corta	Categórica
redad1	Rangos de edad (1 año y más)	Categórica
reg_res	Región de residencia	Categórica
gcausa	Grupo de causas (Grandes grupos)	Categórica
gcausa103	Grupo de causa 103	Categórica
codigo	Código de grupo de causa 103	Categórica
grucau11	Grandes grupos de causas	Categórica
causa667	Causa667	Categórica
meses	Grupo de edad (meses)	Categórica
días	Grupo de edad (días)	Categórica
grupoedad1	Grupos de edad del fallecido en días, meses y años	Categórica

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2012) y (Archivo Nacional de Datos y Metadatos Estadísticos - ANDA, 2012)

Tabla 14

Variables del archivo de defunciones generales del año 2013

CAMPO	SIGNIFICADO	TIPO
Folio	Número del informe estadístico	Numérico
ofi_insc	Oficina del Registro Civil	Categórica
prov_insc	Provincia de inscripción	Categórica
cant_insc	Cantón de inscripción	Categórica
parr_insc	Parroquia de inscripción	Categórica
anio_insc	Año de inscripción	Numérico
mes_insc	Mes de inscripción	Numérico
dia_insc	Día de inscripción	Numérico
fecha_insc	Fecha de inscripción	Cronológica
acta_insc	Acta de inscripción	Categórica
Sexo	Sexo	Numérico
anio_nac	Año de nacimiento	Numérico
mes_nac	Mes de nacimiento	Numérico
dia_nac	Día de nacimiento	Numérico
fecha_nac	Fecha de nacimiento	Cronológica
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
dia_fall	Día de fallecimiento	Numérico
fecha_fall	Fecha de fallecimiento	Cronológica
nac_fall	Nacionalidad	Categórica
cod_pais	Código del país	Categórica
Edad	Edad al fallecer	Numérico
cod_edad	Código de edad al fallecer	Categórica
prov_res	Provincia de residencia habitual del fallecido	Categórica
cant_res	Cantón de residencia habitual del fallecido	Categórica
parr_res	Parroquia de residencia habitual del fallecido	Categórica
loc_res	Código de localidad de residencia habitual del fallecido	Categórica
area_res	Área de residencia habitual del fallecido	Categórica
est_civil	Estado civil y/o conyugal	Categórica
sabe_leer	Sabe leer y escribir	Categórica
niv_inst	Nivel de instrucción alcanzado	Categórica
Etnia	Autoidentificación étnica	Categórica
lugar_ocur	Lugar de ocurrencia del fallecimiento	Categórica
cod_esta	Código del establecimiento de Salud	Categórica
prov_fall	Provincia de fallecimiento	Categórica
cant_fall	Cantón de fallecimiento	Categórica
parr_fall	Parroquia de fallecimiento	Categórica
loc_fall	Código de localidad de fallecimiento	Categórica

CAMPO	SIGNIFICADO	TIPO
area_fall	Área de fallecimiento del fallecido	Categórica
causa4	Causa básica de defunción (categorías - 4 caracteres)	Categórica
mor_mat	Mortalidad materna	Categórica
mor_viol	Tipo presuntivo de la muerte accidental o violenta	Categórica
lug_viol	Lugar donde ocurrió el hecho muerte accidental o violenta	Categórica
autopsia	Se realizó autopsia/necropsia	Categórica
cer_por	Certificación de defunción	Categórica
parent_fall	Parentesco del fallecido con la persona que inscribe	Categórica
causa	Causa básica de defunción (categorías - 3 caracteres)	Categórica
causa103	Lista condensada (103 causas)	Categórica
causa80	Lista de tabulación 2 para la mortalidad subcategorías (80 causas)	Categórica
causa67A	Lista condensada (67 causas) A	Categórica
causa67B	Lista condensada desagregada (67 causas) B	Categórica
lc1	Lista corta de causas de defunción	Categórica

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2013)

Tabla 15

Variables del archivo de defunciones generales del año 2014

CAMPO	SIGNIFICADO	TIPO
ofi_insc	Oficina del Registro Civil	Numérico
prov_insc	Provincia de inscripción	Categórica
cant_insc	Cantón de inscripción	Categórica
parr_insc	Parroquia de inscripción	Categórica
anio_insc	Año de inscripción	Numérico
mes_insc	Mes de inscripción	Numérico
día_insc	Día de inscripción	Numérico
fecha_insc	Fecha de inscripción	Cronológica
nac_fall	Nacionalidad	Categórica
cod_pais	Código del país	Categórica
Sexo	Sexo	Categórica
anio_nac	Año de nacimiento	Numérico
mes_nac	Mes de nacimiento	Numérico
dia_nac	Día de nacimiento	Numérico
fecha_nac	Fecha de nacimiento	Cronológica

CAMPO	SIGNIFICADO	TIPO
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
dia_fall	Día de fallecimiento	Numérico
fecha_fall	Fecha de fallecimiento	Cronológica
cod_edad	Código de edad al fallecer	Catagórica
Edad	Edad al fallecer	Numérico
prov_res	Provincia de residencia habitual del fallecido	Catagórica
cant_res	Cantón de residencia habitual del fallecido	Catagórica
parr_res	Parroquia de residencia habitual del fallecido	Catagórica
area_res	Área de residencia habitual del fallecido	Catagórica
est_civil	Estado civil y/o conyugal	Catagórica
sabe_leer	Sabe leer y escribir	Catagórica
niv_inst	Nivel de instrucción alcanzado	Catagórica
Etnia	Autoidentificación étnica	Catagórica
lugar_ocur	Lugar de ocurrencia del fallecimiento	Catagórica
prov_fall	Provincia de fallecimiento	Catagórica
cant_fall	Cantón de fallecimiento	Catagórica
parr_fall	Parroquia de fallecimiento	Catagórica
area_fall	Área de fallecimiento del fallecido	Catagórica
causa4	Causa básica de defunción (categorías - 4 caracteres)	Catagórica
causa	Causa básica de defunción (categorías - 3 caracteres)	Catagórica
mor_viol	Tipo presuntivo de la muerte accidental o violenta	Catagórica
lug_viol	Lugar donde ocurrió el hecho muerte accidental o violenta	Catagórica
mor_mat	Mortalidad materna	Catagórica
mor_relac	Mujer en edad fértil relacionada con el embarazo, parto y posparto, pero no es muerte materna	Catagórica
autopsia	Se realizó autopsia/necropsia	Catagórica
cer_por	Certificación de defunción	Catagórica
parent_fall	Parentesco del fallecido con la persona que inscribe	Catagórica
lc1	Lista corta de causas de defunción	Catagórica
causa103	Lista condensada (103 causas)	Catagórica
causa80	Lista de tabulación 2 para la mortalidad subcategorías (80 causas)	Catagórica
causa67A	Lista condensada (67 causas) A	Catagórica
causa67B	Lista condensada desagregada (67 causas) B	Catagórica

Nota: Tomado de (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2014)

Tabla 16*Variables archivos de defunciones generales, años 2015 y 2016*

CAMPO	SIGNIFICADO	TIPO
prov_insc	Provincia de inscripción	Categoría
cant_insc	Cantón de inscripción	Categoría
parr_insc	Parroquia de inscripción	Categoría
anio_insc	Año de inscripción	Numérico
mes_insc	Mes de inscripción	Numérico
día_insc	Día de inscripción	Numérico
fecha_insc	Fecha de inscripción	Cronológica
nac_fall	Nacionalidad	Categoría
cod_pais	Código del país	Categoría
Sexo	Sexo	Categoría
anio_nac	Año de nacimiento	Numérico
mes_nac	Mes de nacimiento	Numérico
dia_nac	Día de nacimiento	Numérico
fecha_nac	Fecha de nacimiento	Cronológica
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
día_fall	Día de fallecimiento	Numérico
fecha_fall	Fecha de fallecimiento	Cronológica
cod_edad	Código de edad al fallecer	Categoría
Edad	Edad al fallecer	Numérico
prov_res	Provincia de residencia habitual del fallecido	Categoría
area_res	Área de residencia habitual del fallecido	Categoría
est_civil	Estado civil y/o conyugal	Categoría
sabe_leer	Sabe leer y escribir	Categoría
niv_inst	Nivel de instrucción alcanzado	Categoría
Etnia	Autoidentificación étnica	Categoría
lugar_ocur	Lugar de ocurrencia del fallecimiento	Categoría
prov_fall	Provincia de fallecimiento	Categoría
cant_fall	Cantón de fallecimiento	Categoría
parr_fall	Parroquia de fallecimiento	Categoría
area_fall	Área de fallecimiento del fallecido	Categoría
causa4	Causa básica de defunción (categorías - 4 caracteres)	Categoría
causa	Causa básica de defunción (categorías - 3 caracteres)	Categoría
mor_mat	Mortalidad materna	Categoría
obser_mort_mat	Defunciones de mujeres en edad de 10 a 49 años	Categoría

CAMPO	SIGNIFICADO	TIPO
muj_fertil	En el caso de que la defunción corresponde a una mujer en edad fértil	Categoría
mor_viol	Tipo presuntivo de la muerte accidental o violenta	Categoría
lug_viol	Lugar donde ocurrió el hecho muerte accidental o violenta	Categoría
autopsia	Se realizó autopsia/necropsia	Categoría
cer_por	Certificación de defunción	Categoría
lc1	Lista corta de causas de defunción	Categoría
causa103	Lista condensada (103 causas)	Categoría
causa80	Lista de tabulación 2 para la mortalidad subcategorías (80 causas)	Categoría
causa67A	Lista condensada (67 causas) A	Categoría
causa67B	Lista condensada desagregada (67 causas) B	Categoría

Nota: Tomado de (Instituto Nacional de Estadística y Censos, 2015) e (Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo, 2016)

Tabla 17

Variables del archivo de defunciones generales del año 2017

CAMPO	SIGNIFICADO	TIPO
prov_insc	Provincia de inscripción	Categoría
cant_insc	Cantón de inscripción	Categoría
parr_insc	Parroquia de inscripción	Categoría
anio_insc	Año de inscripción	Numérico
mes_insc	Mes de inscripción	Numérico
día_insc	Día de inscripción	Numérico
fecha_insc	Fecha de inscripción	Cronológica
nac_fall	Nacionalidad	Categoría
cod_pais	Código del país	Categoría
Sexo	Sexo	Categoría
anio_nac	Año de nacimiento	Numérico
mes_nac	Mes de nacimiento	Numérico
día_nac	Día de nacimiento	Numérico
fecha_nac	Fecha de nacimiento	Cronológica
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
día_fall	Día de fallecimiento	Numérico
fecha_fall	Fecha de fallecimiento	Cronológica
Edad	Edad al fallecer	Numérico

CAMPO	SIGNIFICADO	TIPO
cod_edad	Código de edad al fallecer	Categórica
prov_res	Provincia de residencia habitual del fallecido	Categórica
area_res	Área de residencia habitual del fallecido	Categórica
est_civil	Estado civil y/o conyugal	Categórica
niv_inst	Nivel de instrucción alcanzado	Categórica
sabe_leer	Sabe leer y escribir	Categórica
Etnia	Autoidentificación étnica	Categórica
lugar_ocur	Lugar de ocurrencia del fallecimiento	Categórica
prov_fall	Provincia de fallecimiento	Categórica
cant_fall	Cantón de fallecimiento	Categórica
parr_fall	Parroquia de fallecimiento	Categórica
area_fall	Área de fallecimiento del fallecido	Categórica
causa4	Causa básica de defunción (categorías - 4 caracteres)	Categórica
muj_fertil	En el caso de que la defunción corresponde a una mujer en edad fértil	Categórica
mor_viol	Tipo presuntivo de la muerte accidental o violenta	Categórica
lug_viol	Lugar donde ocurrió el hecho muerte accidental o violenta	Categórica
Autopsia	Se realizó autopsia/necropsia?	Categórica
cer_por	Certificación de defunción	Categórica
lc1	Lista corta de causas de defunción	Categórica
Causa	Causa básica de defunción (categorías - 3 caracteres)	Categórica
mor_mat	Mortalidad materna	Categórica
causa103	Lista condensada (103 causas)	Categórica
causa80	Lista de tabulación 2 para la mortalidad subcategorías (80 causas)	Categórica
causa67A	Lista condensada (67 causas) A	Categórica
causa67B	Lista condensada desagregada (67 causas) B	Categórica

Nota: Tomado de (Instituto Nacional de Estadística y Censos, 2016) e (Instituto Nacional de Estadística y Censos - INEC, 2018)

Tabla 18

Variables del archivo de defunciones generales del año 2018

CAMPO	SIGNIFICADO	TIPO
Numeracion	Secuencial de registros	Numérico
prov_insc	Provincia de inscripción	Categórica
cant_insc	Cantón de inscripción	Categórica

CAMPO	SIGNIFICADO	TIPO
parr_insc	Parroquia de inscripción	Categórica
anio_insc	Año de inscripción	Numérico
mes_insc	Mes de inscripción	Numérico
dia_insc	Día de inscripción	Numérico
fecha_insc	Fecha de inscripción	Cronológica
nac_fall	Nacionalidad	Categórica
cod_pais	Código del país	Categórica
Sexo	Sexo	Categórica
fecha_nac	Fecha de nacimiento	Cronológica
anio_nac	Año de nacimiento	Numérico
mes_nac	Mes de nacimiento	Numérico
dia_nac	Día de nacimiento	Numérico
fecha_fall	Fecha de fallecimiento	Cronológica
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
dia_fall	Día de fallecimiento	Numérico
cod_edad	Código de edad al fallecer	Categórica
Edad	Edad al fallecer	Numérico
prov_res	Provincia de residencia habitual del fallecido	Categórica
area_res	Área de residencia habitual del fallecido	Categórica
est_civil	Estado civil y/o conyugal	Categórica
niv_inst	Nivel de instrucción alcanzado	Categórica
sabe_leer	Sabe leer y escribir	Categórica
Etnia	Autoidentificación étnica	Categórica
lugar_ocur	Lugar de ocurrencia del fallecimiento	Categórica
prov_fall	Provincia de fallecimiento	Categórica
cant_fall	Cantón de fallecimiento	Categórica
parr_fall	Parroquia de fallecimiento	Categórica
area_fall	Área de fallecimiento	Categórica
cer_por	Certificación de defunción	Categórica
muj_fertil	Defunciones de mujeres en edad de 10 a 49 años	Categórica
mor_viol	Tipo presuntivo de la muerte accidental o violenta	Categórica
lug_viol	Lugar donde ocurrió el hecho muerte accidental o violenta	Categórica
causa4	Causa básica de defunción (categorías - 4 caracteres)	Categórica
autopsia	Se realizó autopsia/necropsia?	Categórica
causa	Causa básica de defunción (categorías - 3 caracteres)	Categórica

CAMPO	SIGNIFICADO	TIPO
lc1	Lista corta de causas de defunción	Categórica
causa103	Lista condensada (103 causas)	Categórica
causa80	Lista de tabulación 2 para la mortalidad subcategorías (80 causas)	Categórica
causa67A	Lista condensada (67 causas) A	Categórica
causa67B	Lista condensada desagregada (67 causas) B	Categórica
residente	Variable creada para identificar si el fallecido es residente	Categórica

Nota: Tomado de (Instituto Nacional de Estadística y Censos - INEC, 2018)

No se consideran todas las variables para el análisis, pues no todas se repiten en los archivos, arrojando valores nulos para los años restantes y provocando una pérdida considerable de registros, por lo cual se contempla de forma inicial los siguientes campos disponibles en todos los años y relacionados con el caso de estudio para la exploración de datos de los egresos hospitalarios (ver Tabla 19) y de las defunciones generales (ver Tabla 20).

Tabla 19

Principales variables de los archivos de egresos hospitalarios

CAMPO	DESCRIPCIÓN	TIPO
prov_ubi	Provincia ubicación del establecimiento	Categórica
Sector	Sector al que pertenece el establecimiento	Categórica
Sexo	Sexo del paciente	Categórica
Edad	Edad del paciente	Numérico
prov_res	Provincia de residencia habitual del paciente	Categórica
con_egrpa	Condición del egreso	Categórica
cau_cie10	Causa de lista internacional detallada a 4 o más dígitos de la CIE-10	Categórica
causa3	Causa de lista internacional detallada a 3 dígitos de la CIE-10	Categórica
cau298rx	Lista de 298 causas	Categórica
cau221rx	Lista especial de 221 grupos	Categórica
cap221rx	Capítulo lista 221	Categórica
fecha_ingr	Fecha de ingreso	Cronológica
fecha_egr	Fecha de egreso	Cronológica

Tabla 20

Principales variables de los archivos de defunciones generales

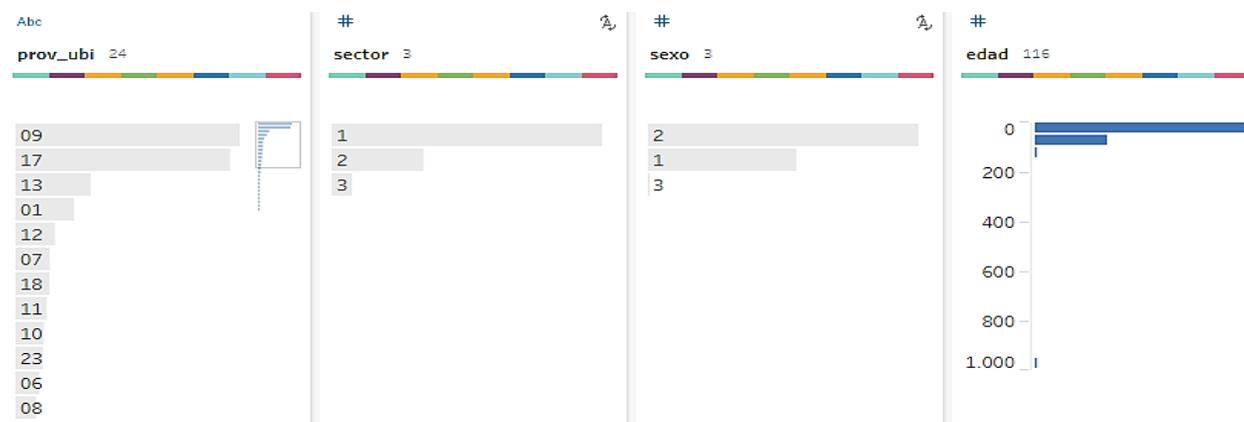
CAMPO	DESCRIPCIÓN	TIPO
prov_fall	Provincia de fallecimiento	Catagórica
anio_fall	Año de fallecimiento	Numérico
mes_fall	Mes de fallecimiento	Numérico
est_civil	Estado civil y/o conyugal	Catagórica
Sexo	Sexo	Catagórica
Edad	Edad al fallecer	Numérico
prov_res	Provincia de residencia habitual del fallecido	Catagórica
Causa	Lista corta de causas de defunción	Catagórica
mor_viol	Tipo presuntivo de la muerte accidental o violenta	Catagórica

Exploración de los datos

En la figura 9 se puede observar la representación gráfica de la distribución de frecuencias de los campos preliminares de todos los años de los egresos hospitalarios y en la figura 10 la de las defunciones generales.

Figura 9

Exploración de datos de egresos hospitalarios



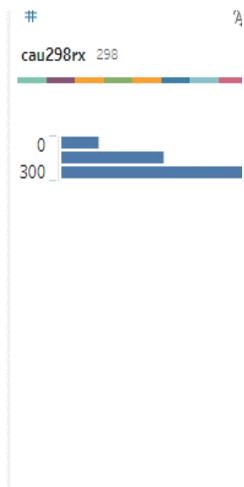
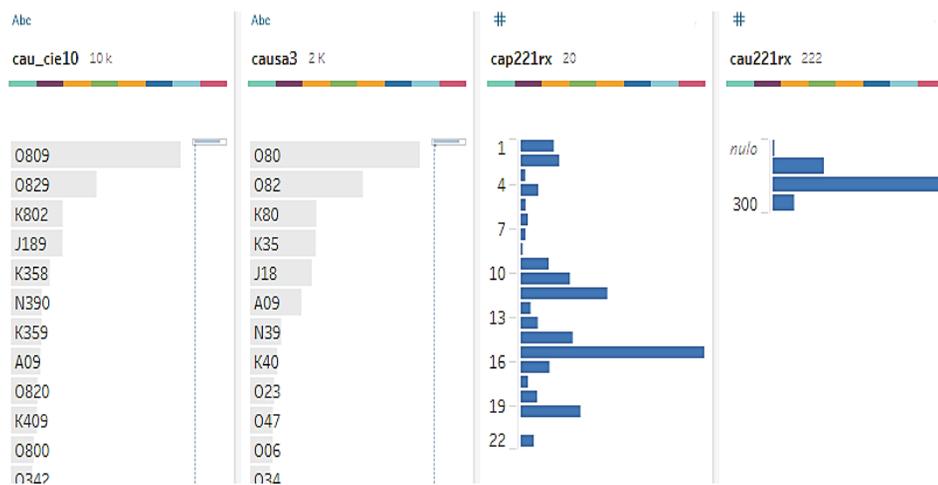
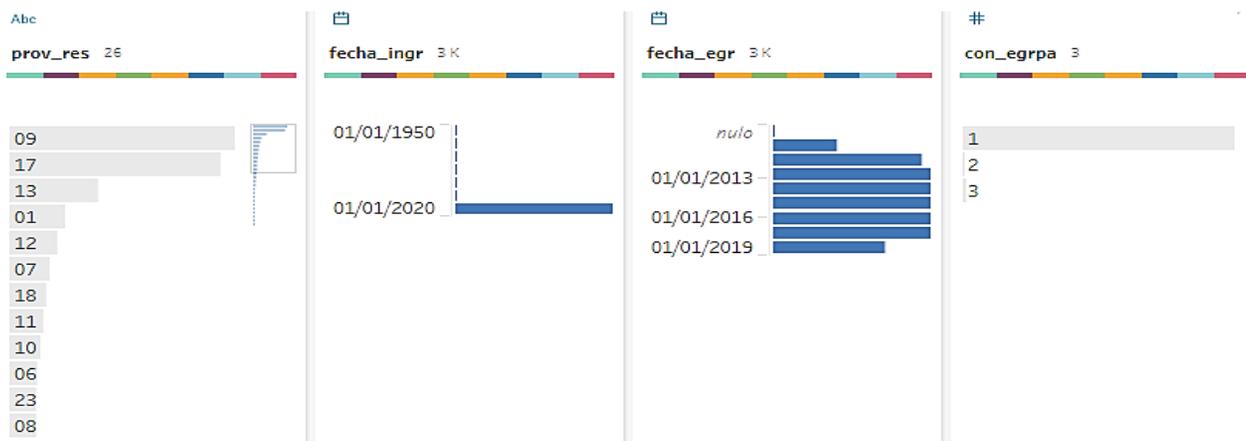


Figura 10*Exploración de datos de defunciones generales***Verificar la calidad de los datos**

En base a la exploración realizada de egresos hospitalarios se determina que los campos que no presentan inconsistencias son:

prov_ubi: contiene el código de las diferentes provincias del país a las que pertenece el establecimiento de salud, se puede observar que las que predominan son 09 Guayas, 17 Pichincha y 13 Manabí.

sector: el sector al que pertenecían los establecimientos de salud y que brindó más atención fue el 1 Público.

prov_res: contiene el código de las diferentes provincias del país en las que reside habitualmente el paciente, se puede observar que las que predominan son 09 Guayas, 17 Pichincha y 13 Manabí.

con_egrpa: la condición de egreso prevalente de un paciente es la 1 Vivo.

cau_cie10: dentro de las causas de la lista internacional CIE-10 las más prevalentes son O809, O829 y K802

causa3: entre las causas de la lista internacional CIE-10 de 3 dígitos, las más prevalentes son O 80, O82 y K80

cau298rx: En la lista de causas 298, las más frecuentes se encuentran incluidas en el intervalo de 200 a 300.

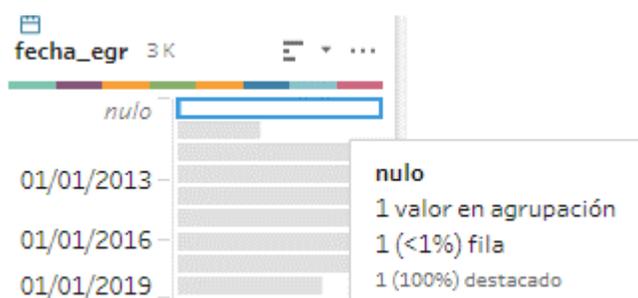
cap221rx: según los capítulos de la lista 221, los más habituales son el 13, 14 y 16

Son dos los campos que presentan valores nulos, y cuya cantidad se considera insignificante al ser un minúsculo porcentaje:

fecha_egr: dentro de la fecha de egreso se encuentra un registro nulo y los registros restantes mantienen consistencia con el tiempo previsto para el análisis (ver figura 11).

Figura 11

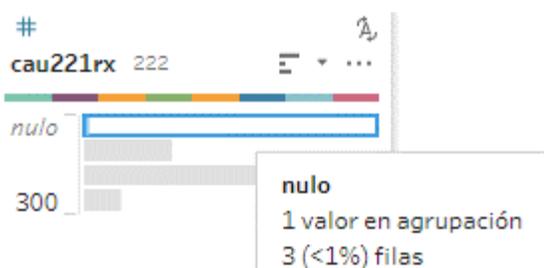
Campo fecha_egr de los egresos hospitalarios



cau221rx: en la lista especial de 221 grupos se observa 3 registros nulos, y entre los más reiterados se encuentran los incluido el intervalo de grupos 100 a 200 (ver figura 12).

Figura 12

Campo cau221rx de los egresos hospitalarios

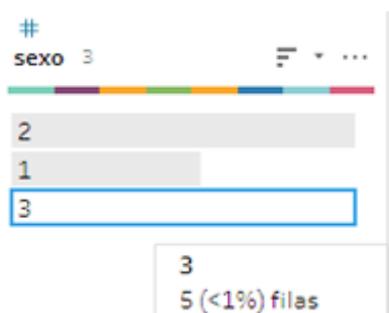


Se considera valores fuera de rango y que pueden constituirse en ruido a los siguientes:

sexo: se encuentran 5 registros diferentes a 1 Masculino y 2 Femenino (ver figura 13).

Figura 13

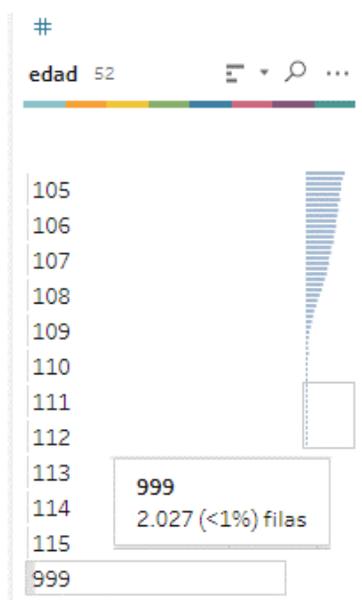
Campo sexo de los egresos hospitalarios



edad: se localizan 2027 registros con edades irreales iguales a 999 (ver figura 14).

Figura 14

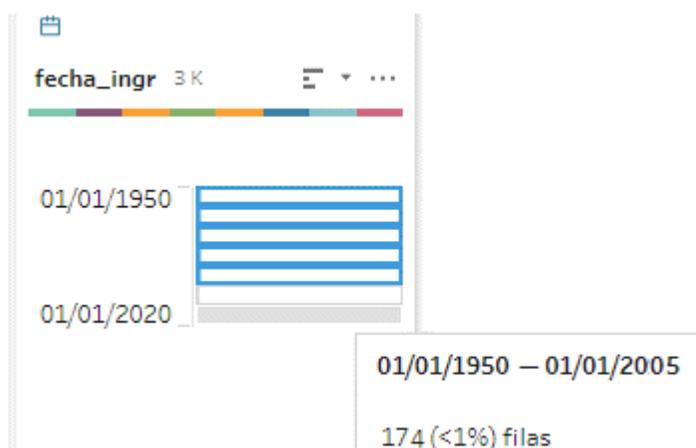
Campo edad de los egresos hospitalarios



fecha_ingr: aunque una persona puede estar internada de forma indefinida dependiendo de la enfermedad y la evolución que tenga, se revelan 174 registros de fechas de ingreso que son anteriores al año 2005 (ver figura 15).

Figura 15

Campo fecha_ingr de los egresos hospitalarios



En referencia a la exploración realizada en los datos de las defunciones generales los campos que no presentan inconsistencias son:

prov_fall: Provincia en la que se produjo el fallecimiento contiene el código de las diferentes provincias del país en las que se puede observar que las que predominan son 09 Guayas, 17 Pichincha y 13 Manabí.

mes_fall: Mes del fallecimiento, se puede percibir que se registran con mayor frecuencia en los meses de 1 enero, 3 marzo y 4 abril

sexo: Identifica el sexo del fallecido, se puede distinguir que han fallecido más hombres (1) que mujeres (2).

prov_res: Provincia de residencia habitual del fallecido, contiene el código de las diferentes provincias del país, las más relevantes son 09 Guayas, 17 Pichincha y 13 Manabí.

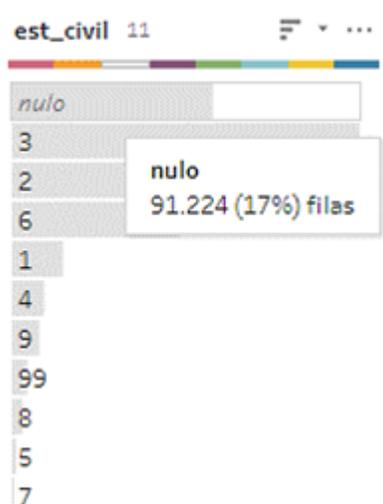
causa: entre las causas básicas de defunción de la lista internacional CIE-10 de 3 dígitos prevalecen la I21, J18 y R99.

Son dos los campos que presentan valores nulos, y cuya cantidad se considera significativa:

est_civil: en este campo se registran 91.224 valores nulos que equivalen a 17% (ver figura 16).

Figura 16

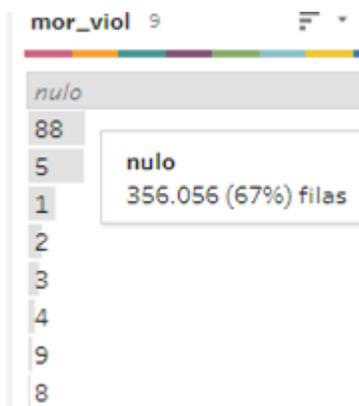
Campo est_civil de las defunciones generales



mor_viol: en el campo muerte violenta se encuentran 356.056 que representan el 67% (ver figura 17).

Figura 17

Campo mor_viol de las defunciones generales

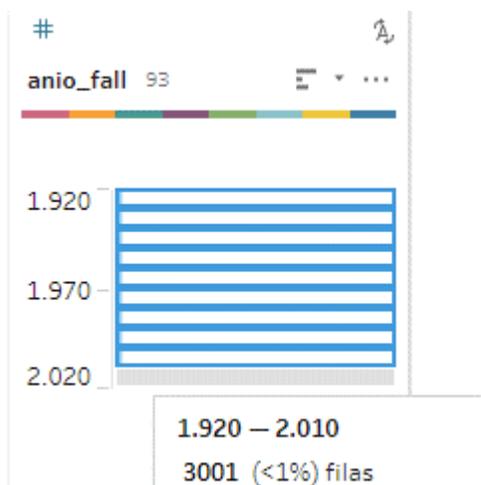


Se considera valores fuera de rango y que pueden constituirse en ruido a los siguientes:

anio_fall: en el año de fallecimiento se observa 3001 registros fuera del periodo de estudio menores al 1% (ver figura 18).

Figura 18

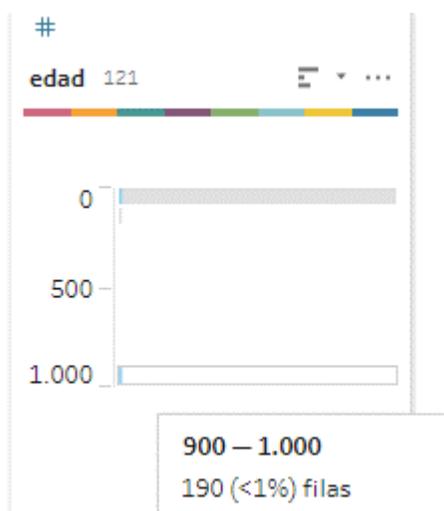
Campo anio_fall de las defunciones generales



edad: entre la edad que se registró al fallecer, se localizan 190 registros con edades irreales iguales a 999 (ver figura 19).

Figura 19

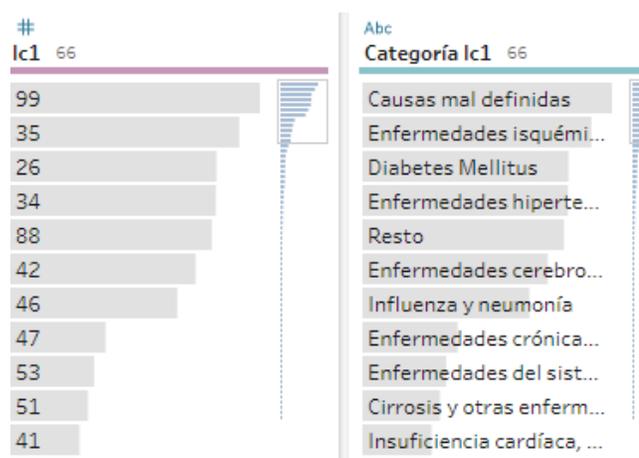
Campo edad de las defunciones generales



Ic1: en la lista corta de causas de defunción, se omitirá las causas Nro 99 y 55 ya que se consideran ruido, además se puede visualizar que las que predominan son 35 Enfermedades isquémicas del corazón, 26 Diabetes Mellitus y 34 Enfermedades hipertensivas (ver figura 20).

Figura 20

Campo Ic1 de las defunciones generales



Preparación de datos

Selección

De todos los datos recopilados se eligen aquellos que corresponden al caso de estudio en los que la edad es mayor a 64 años, reduciéndose el volumen original de datos en el caso de los egresos hospitalario de 8'130.999 a 1'236.172 y en el caso de las defunciones generales de 532.765 a 312.237 registros válidos durante el período 2011-2018 (ver tabla 21).

Tabla 21

Número de egresos y defunciones periodo 2011-2018

Año	Población adulta mayor	Nro. Egresos	Nro. Defunciones
2011	1'003.156	137.247	34.743
2012	1'024.425	143.482	36.336
2013	1'049.321	150.884	37.305
2014	1'077.587	155.484	37.649
2016	1'108.991	155.607	39.277
2015	1'143.494	156.707	40.910
2017	1'180.944	163.955	43.829
2018	1'221.286	172.806	42.188
Total		1.236.172	312.237

Los campos que se consideran a analizar de forma preliminar se describen en la tabla 22 y 23.

Tabla 22

Campos preliminares de los egresos hospitalarios

Campo	Descripción	Tipo
fecha	Año de egreso del paciente	Cronológica
sexo	Sexo del paciente	Categórica
edad	Edad del paciente	Numérico
provincia	Provincia de residencia habitual del paciente	Categórica
provincia establecimiento	Provincia del establecimiento	Categórica
sector	Sector al que pertenece el establecimiento de salud	Categórica
capitulo	Capítulos del Código Internacional de Enfermedades CIE10	Categórica
causa298	Lista especial de tabulación para la morbilidad de 298 causas del Código Internacional de Enfermedades CIE10	Categórica

Tabla 23

Campos preliminares de las defunciones generales

Campo	Descripción	Tipo
fecha	Año de fallecimiento	Numérico
sexo	Sexo	Categórica
edad	Edad al fallecer	Numérico
provincia	Provincia de fallecimiento	Categórica
provincia fallecimiento	Provincia de residencia habitual del fallecido	Categórica
causa	Lista corta de 99 causas de defunción, del Código Internacional de Enfermedades CIE10	Categórica

Limpieza

Se determinó que los registros atípicos y que no corresponde al periodo 2011-2018, así como los valores nulos serán excluidos.

Estructuración, integración y formateo

En base a lo definido en la selección y limpieza, con el apoyo de la herramienta Tableau Prep se procedió a agrupar los archivos de los de los egresos hospitalarios (ver figura 21) y de las defunciones generales (ver figura 22) del período establecido, a unificar el nombre de los campos, y a cambiar el tipo de dato que no correspondía como por ejemplo las fechas y edad que en lugar de encontrarse con tipo fecha y entero respectivamente se encontraban en tipo texto.

Figura 21

ETL de los egresos hospitalarios con Tableau Prep

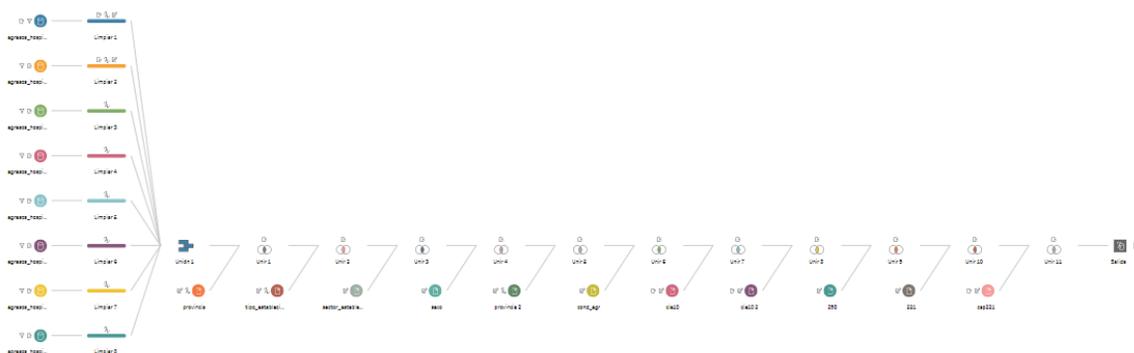
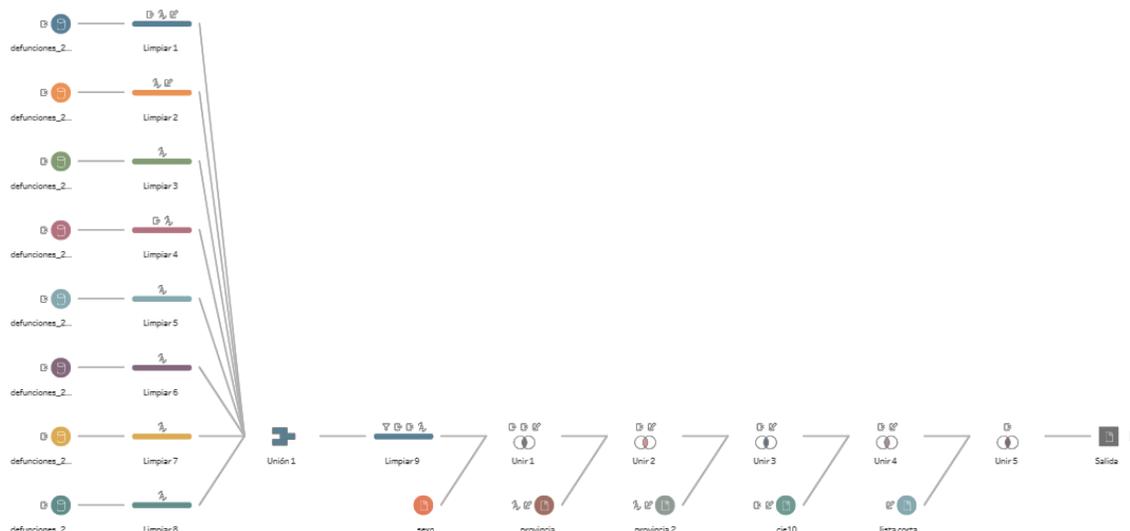


Figura 22

ETL de las defunciones generales con Tableau Prep



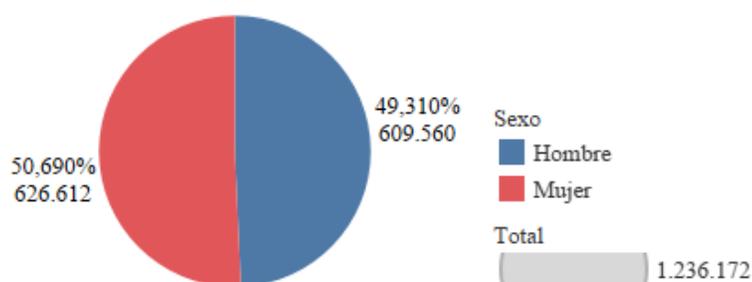
Análisis preliminar

Con apoyo de Tableau Public y en base a los campos seleccionados del periodo 2011 – 2018 se obtuvo información relevante, referente a los pacientes asistidos por los diferentes establecimientos de salud y de los que fallecieron por diferentes causas.

En el sistema nacional de salud, durante el periodo de estudio fueron atendidos alrededor de 1'236.172 adultos mayores. De los cuales la mayoría un 50,69% fueron mujeres, mientras que el 49,31% fueron hombres (ver figura 23).

Figura 23

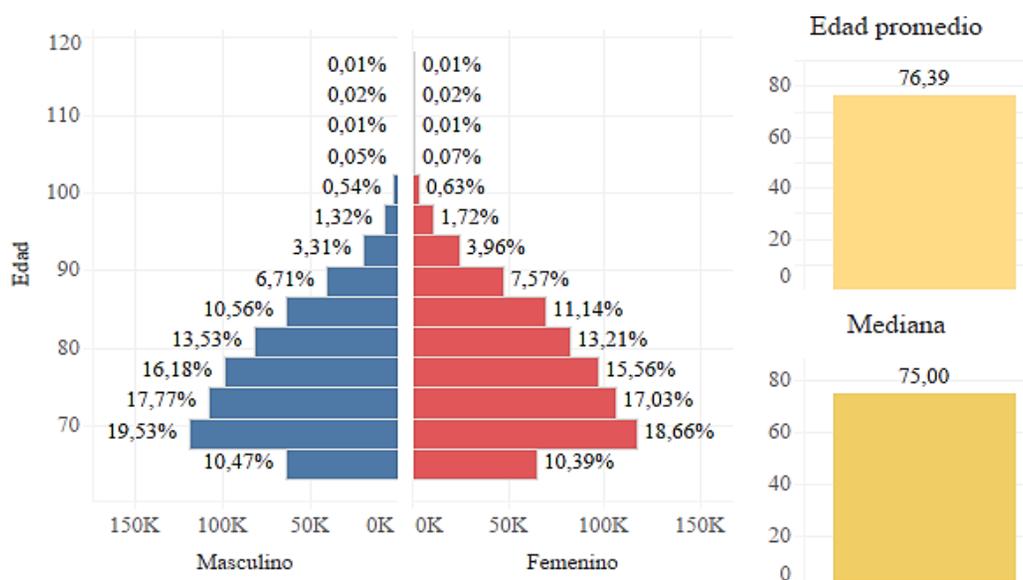
Número de egresos de adultos mayores según su sexo



El rango de edad reiterativo se encuentra entre los 65 y 70 años. La edad promedio del adulto mayor fue de 76 años y la mediana fue de 75 años (ver figura 24).

Figura 24

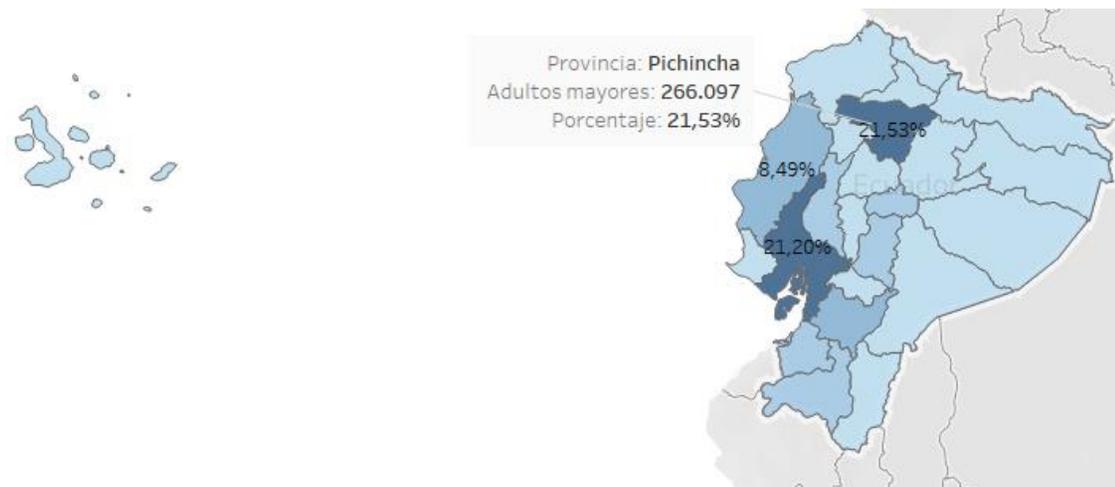
Pirámide de edad del adulto mayor atendido



Los adultos mayores atendidos en la red pública integral de salud y complementaria, residían principalmente en las provincias de Pichincha con un 21,53%, Guayas con un 21,20%, Manabí con un 8,49%, Azuay con un 7,64% y el Oro con un 4,63% (ver figura 25).

Figura 25

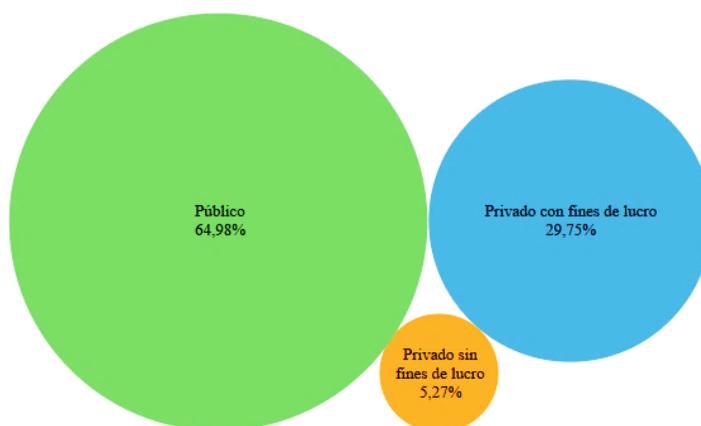
Mapa de la residencia de los adultos mayores egresados y porcentajes



Los establecimientos de salud del sector públicos albergaron el 64.98% de los pacientes atendidos, los privados con fines de lucro 29,75% y los privados sin fines de lucro asistieron al 5,27% (ver figura 26).

Figura 26

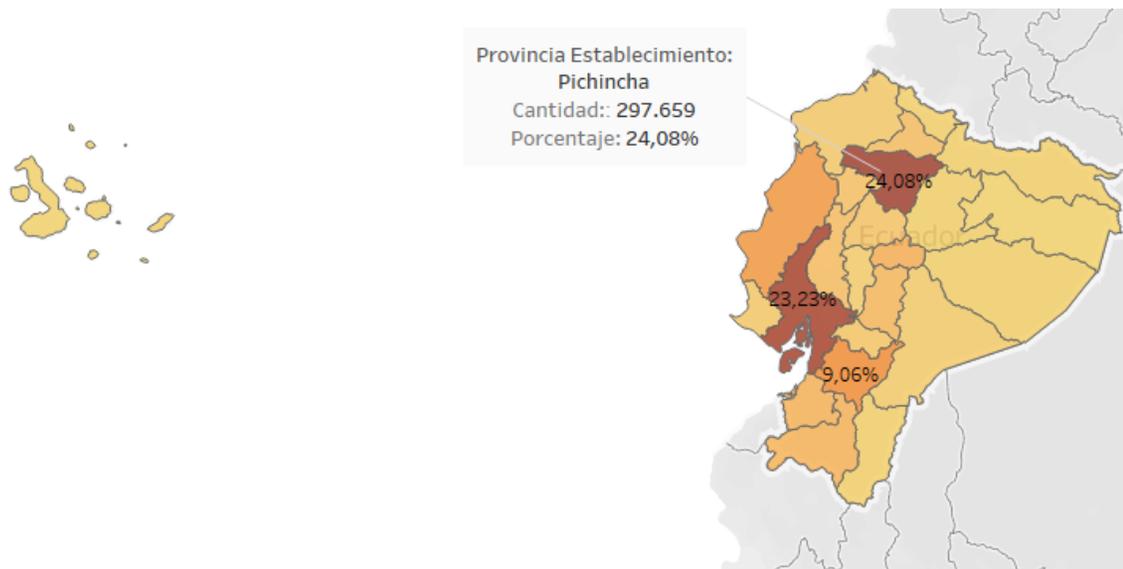
Sectores de los establecimientos de salud



Los establecimientos de salud se ubican principalmente en las provincias de Pichincha con un 24,08%, Guayas con 23,23%, Azuay con 9,06%, Manabí con 7,63% y Tungurahua con 4,68% personas atendidas (ver figura 27).

Figura 27

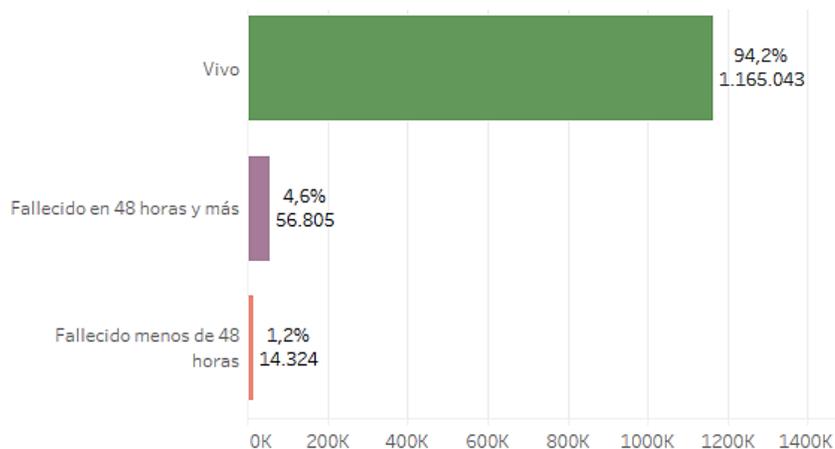
Mapa de la ubicación de los establecimientos de salud y porcentajes



Un 94,2% de adultos mayores egresaron de las diferentes instituciones de salud en calidad de vivos, un 4,6% fallecieron en 48 horas y más y el 1,2% fallecieron en menos de 48 horas (ver figura 28).

Figura 28

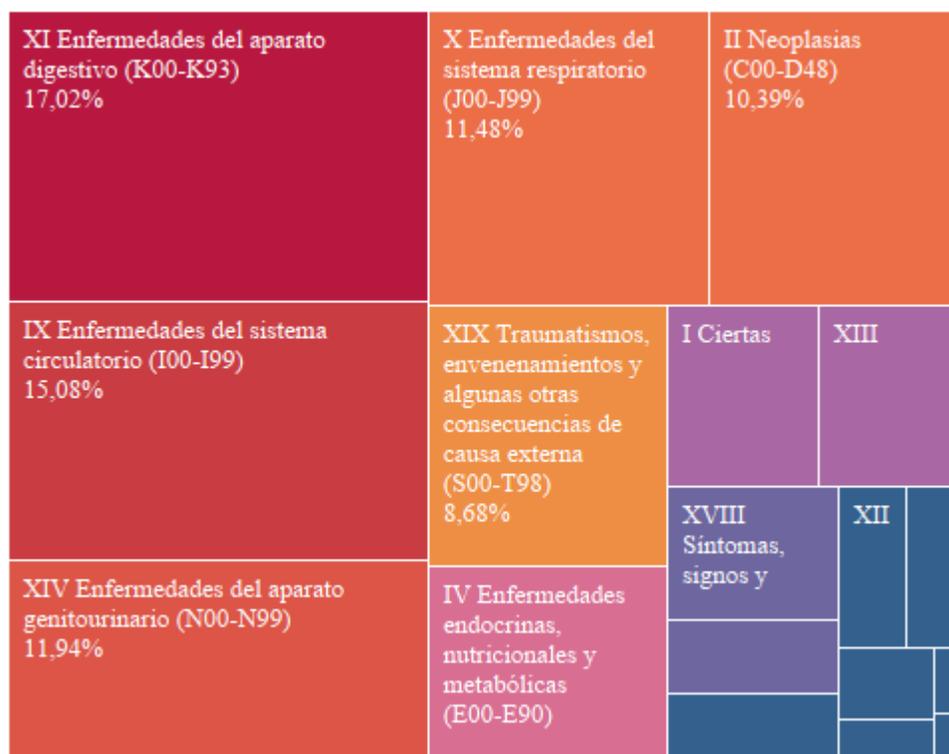
Condición de egreso hospitalario



Los grupos de enfermedades más frecuentes presentes en los adultos mayores corresponden a los capítulos XI Enfermedades del aparato digestivo (K00-K93) con el 17,02%, IX Enfermedades del sistema circulatorio (I00-I99) con el 15,08%, XIV Enfermedades del aparato genitourinario (N00-N99) con el 11,94%, X Enfermedades del sistema respiratorio (J00-J99) con el 11,48% y II Neoplasias (C00-D48) 10,39% (ver figura 29).

Figura 29

Grupos de enfermedades del adulto mayor



Entre las diversas enfermedades presentes en la lista de 298 causas de morbilidad en el periodo estudiado (ver figura 30), se evidencia a la Neumonía como la principal causa con un 6,2%, seguida de la Diabetes mellitus con un 4,8%, la Colelitiasis y colecistitis afecta al 4,7% los adultos mayores, la Hiperplasia de la Próstata a un 3% (aunque esta se aplica sólo a hombres) y Otras enfermedades del sistema urinario con el 2,6%. En la figura 31, se puede observar la evolución de las causas de morbilidad.

Figura 30

Porcentaje de las causas de morbilidad

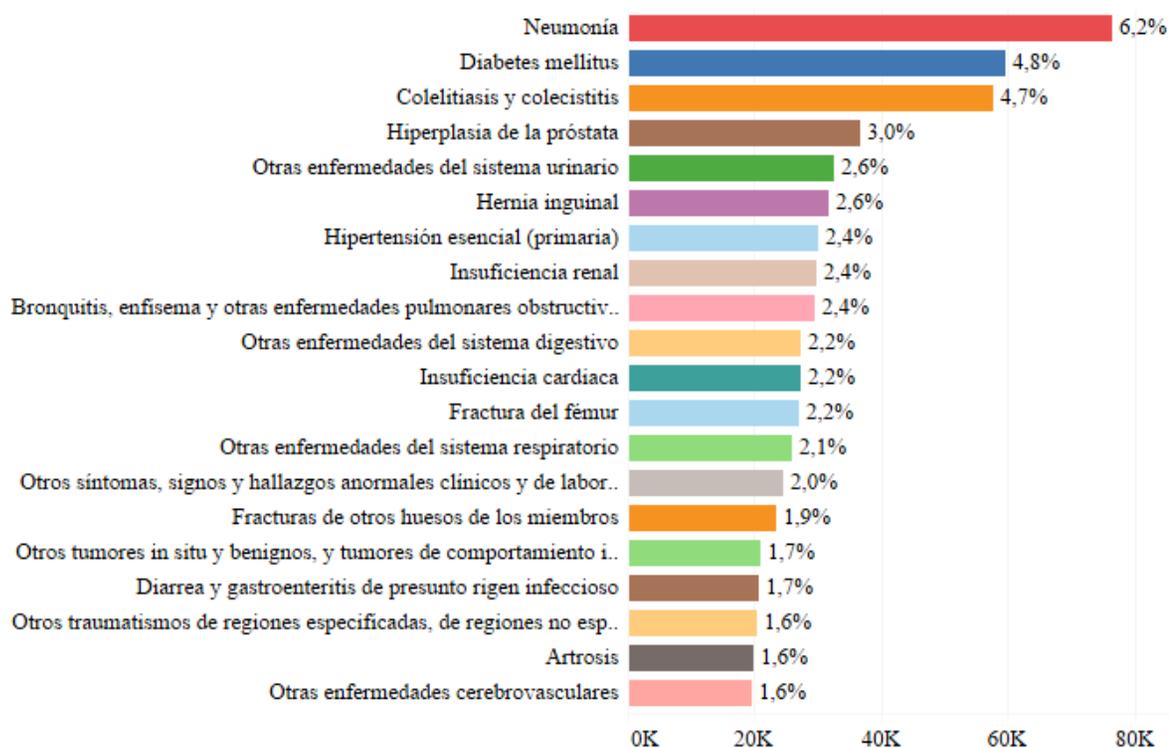
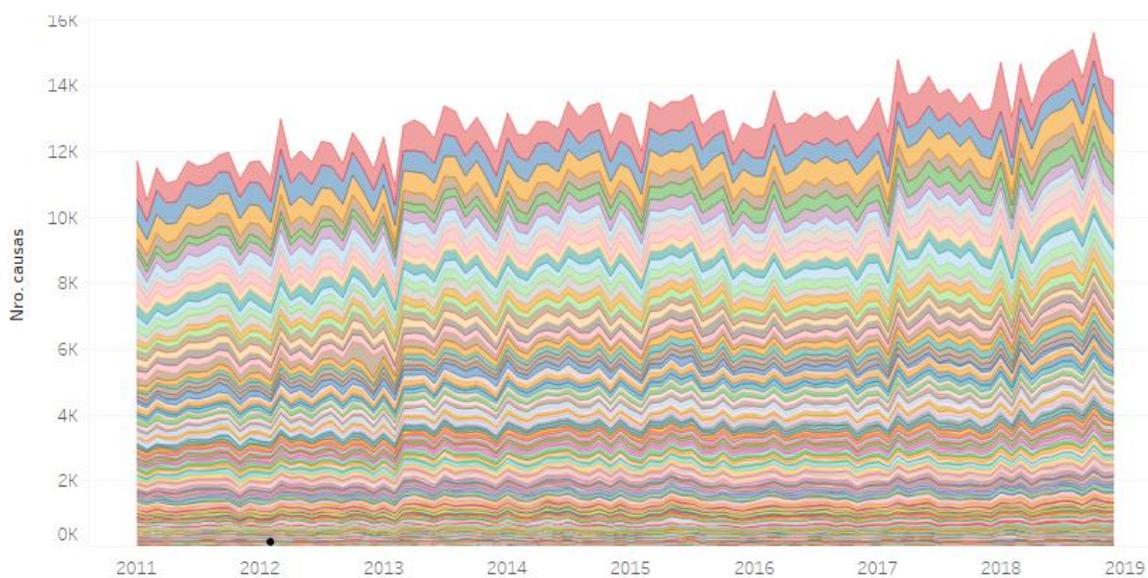


Figura 31

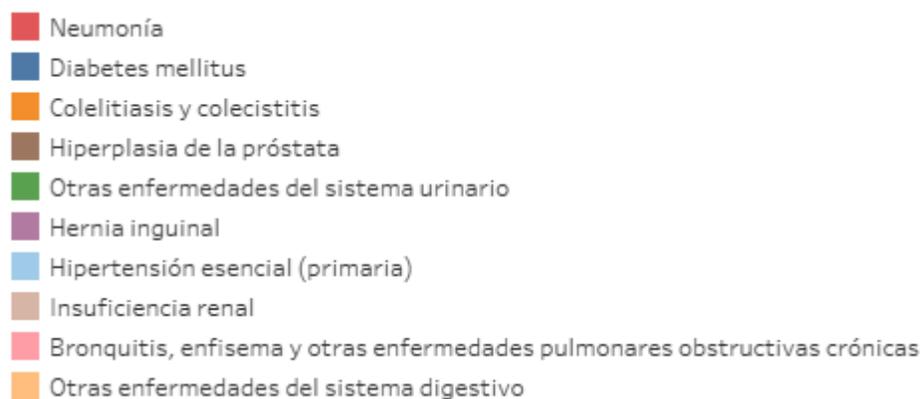
Evolución de las causas de morbilidad

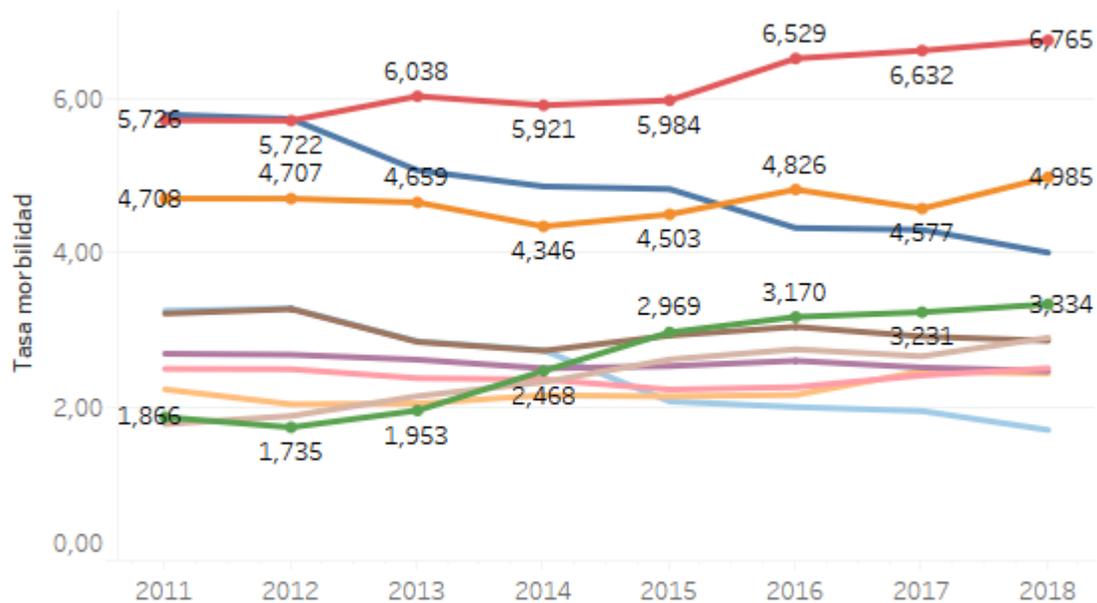


De las diez principales causas de morbilidad expresadas como porcentaje durante cada año, las que predominan son: Neumonía, Diabetes mellitus, y Colelitiasis y colecistitis (ver figura 32).

Figura 32

Diez principales causas de morbilidad

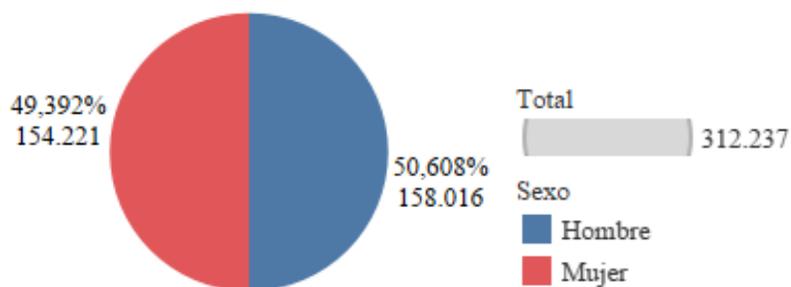




En la información recopilada de las defunciones generales desde el 2011 al 2018 se registra en el país alrededor de 312.237 fatalidades referentes a adultos mayores. De los cuales un mayor porcentaje del 50,61% fueron mujeres, mientras que el 49,39% fueron hombres (ver figura 33).

Figura 33

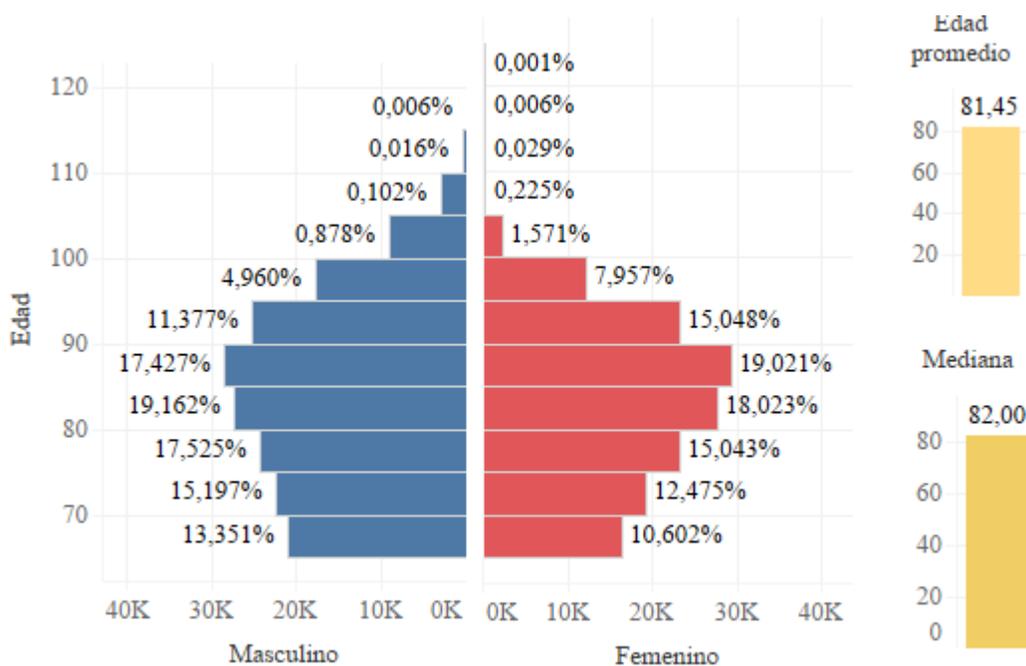
Adultos mayores fallecidos según su sexo



El rango de edad de los fallecidos se encuentra entre los 80 y 85 años en hombres y entre los 85 y 90 en mujeres. La edad promedio del adulto mayor fallecido fue de 81 años y la mediana fue de 82 años (ver figura 34).

Figura 34

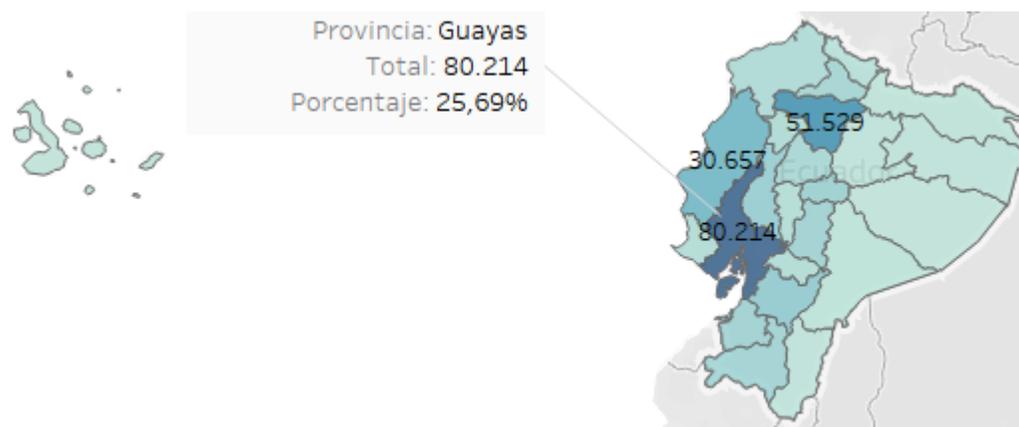
Pirámide de edad del adulto mayor fallecido



Los fallecidos habitaban principalmente en las provincias de Guayas con 25,69%, Pichincha con un 16,5%, Manabí con el 9,82%, Azuay 5,7% y Los Ríos 5,15% (ver figura 35).

Figura 35

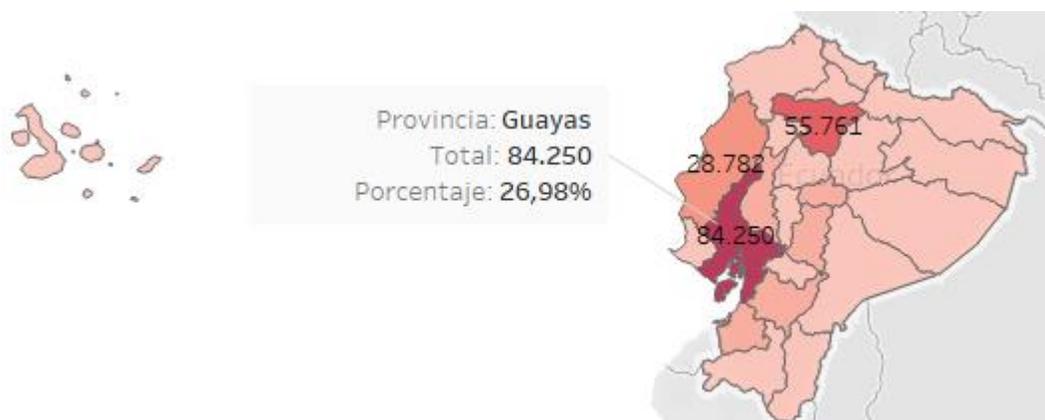
Mapa de la residencia de los adultos mayores fallecidos y porcentajes



Los perecimientos se evidencian principalmente en las provincias del Guayas con 26,98%, Pichincha con 17,86%, Manabí con 9,22%, Azuay con 6,06% y el Oro con 4,03% (ver figura 36).

Figura 36

Mapa según el lugar de fallecimiento y porcentaje



Dentro de la lista corta de la categoría de tres caracteres según la codificación del Manual de clasificación estadística internacional de enfermedades de la OMS (CIE - 10), en el periodo estudiado, se presenta como principal causa de muerte a las Enfermedades isquémicas del corazón con 11,24%, Diabetes Mellitus con 10,12%,

Enfermedades hipertensivas con 10,09%, Enfermedades cerebrovasculares con 9,10% e Influenza y neumonía con 8,19% (ver figura 37). En la figura 38, se puede observar la evolución de las causas de mortalidad.

Figura 37

Porcentaje de causas de mortalidad

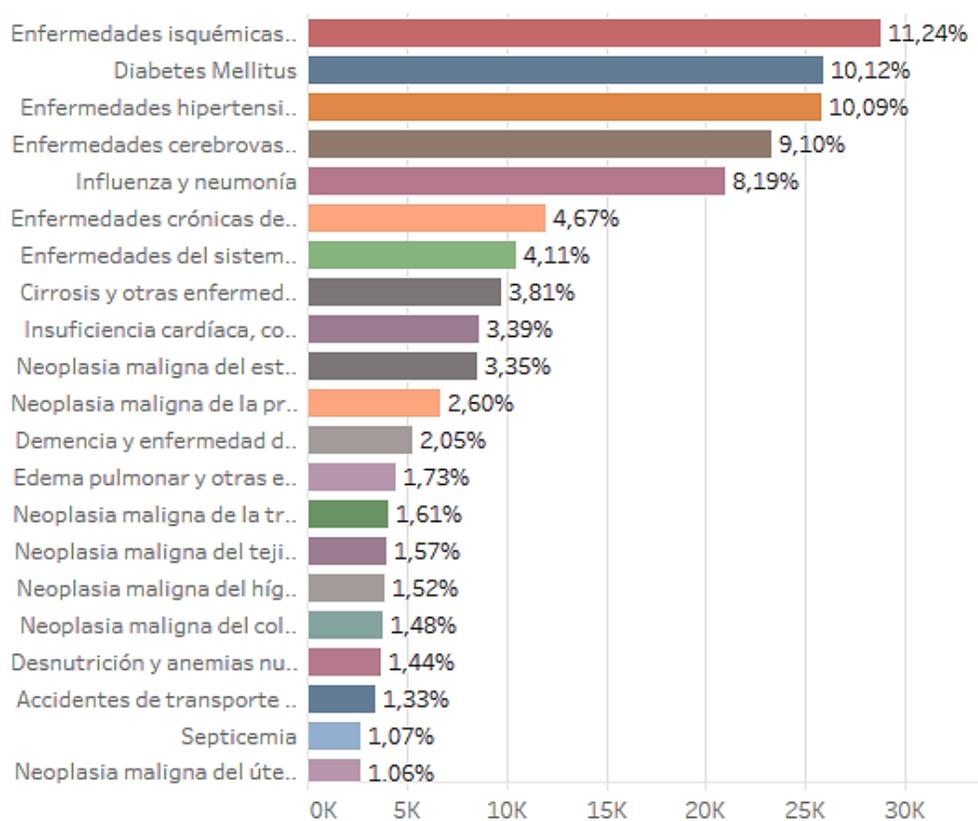
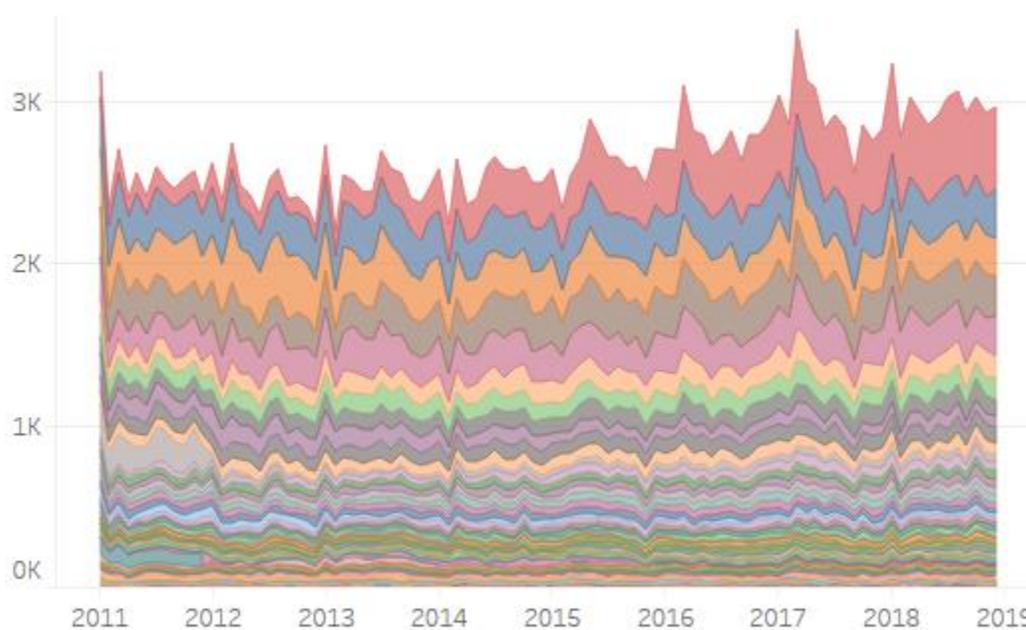


Figura 38

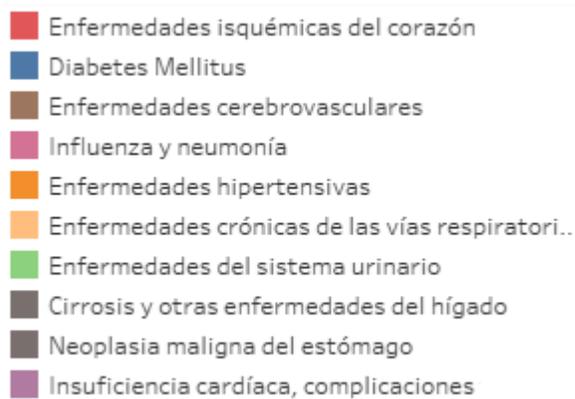
Evolución de las causas de mortalidad

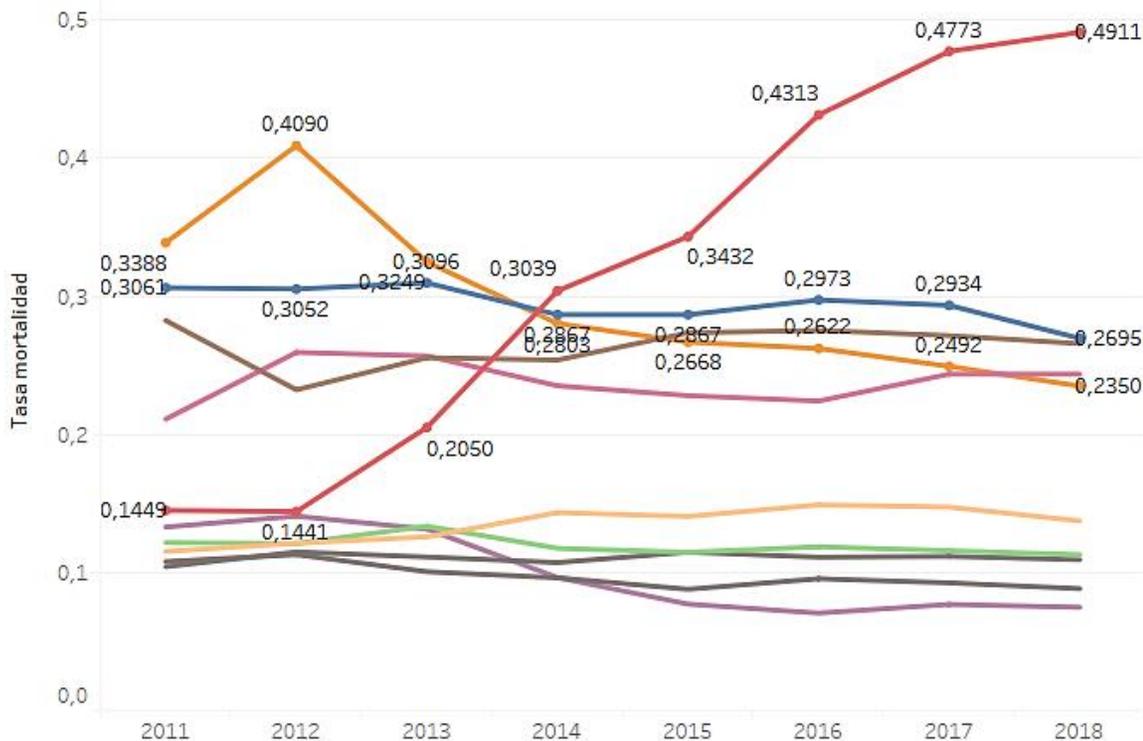


Finalmente, de las diez principales causas de mortalidad, expresadas como porcentaje durante cada año, las más relevantes son: Enfermedades isquémicas del corazón (con tendencia alcista desde el 2014), Diabetes Mellitus y Enfermedades hipertensivas. (ver figura 39).

Figura 39

Diez principales causas de mortalidad





Modelado

Selección de la técnica de modelado

En función de las variables usadas para el cálculo de la morbilidad y mortalidad se establece hacer uso de series temporales y la técnica de suavizado exponencial elegida Holt y Winters, sin considerar la tendencia y con temporada multiplicativa, excepto en el caso de prevalencia de Colelitiasis y colecistitis y las Enfermedades hipertensivas se usó tendencia aditiva y se excluyó la temporada.

Selección de atributos para el análisis del algoritmo y construcción del modelo

a) *Pronosticar la prevalencia de las tres principales causas de morbilidad del adulto mayor durante el 2019 y 2020 por cada 1.000 pacientes egresados.*

Tabla 24

Variables seleccionadas, pronóstico de causas de morbilidad

Variable	Descripción
Lista de 298 causas	Lista especial de tabulación para la morbilidad de 298 causas del Código Internacional de Enfermedades CIE10
Fecha	Año de egreso del paciente

En base a las variables seleccionadas para la elaboración del modelo (ver tabla 24), se calculó la prevalencia de las enfermedades (causas) más relevantes por año usando la fórmula:

$$\frac{\text{nro de egresos de adultos mayores por la causa (x) en el año t} * 1000(\text{habitantes})}{\text{total de adultos mayores que presentaron morbilidad en el año t}}$$

Las opciones usadas para crear el pronóstico son:

Serie temporal: Anual (ver figura 40)

Medidas: Prevalencia Morbilidad

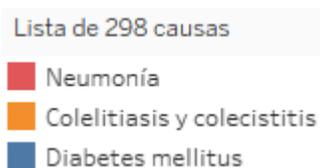
Pronóstico hacia delante: 2 períodos (2019 – 2020)

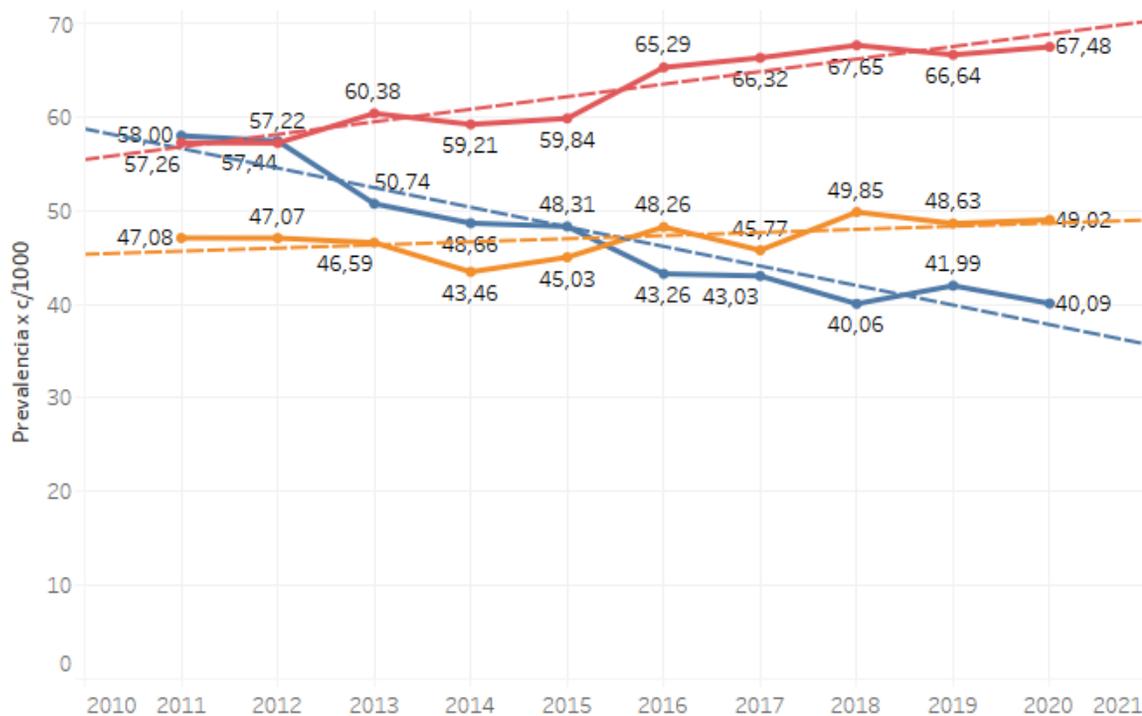
Pronóstico basado en: 2011 – 2018

Ignorar últimos: No se ignoraron períodos

Figura 40

Pronóstico principales causas de morbilidad obtenido en Tableau





En la tabla 25 se describe el tipo de modelo usado para cada causa de morbilidad y los coeficientes de suavizado respectivamente.

Tabla 25

Modelo usado para cada causa de morbilidad

Lista de 298 causas	Modelo			Coeficientes de suavizado		
	Nivel	Tendencia	Temporada	Alfa	Beta	Gamma
Neumonía	Multiplicativo	Ninguno	Multiplicativo	0,50	0	0,437
Colelitiasis y colecistitis	Aditivo	Aditivo	Ninguno	0,22	0	0
Diabetes mellitus	Multiplicativo	Ninguno	Multiplicativo	0,50	0	0,50

b) Pronosticar la prevalencia de las tres principales causas de mortalidad durante el 2019 y 2020 por cada 10.000 adultos mayores.

Tabla 26

Variables seleccionadas, pronóstico de causas de mortalidad

Variable	Descripción
Lista corta LC1	Lista corta de 99 causas de defunción, del Código Internacional de Enfermedades CIE10.
Fecha	Año de fallecimiento

En base a las variables seleccionadas (ver tabla 26) para la elaboración del modelo, se calculó la prevalencia de las causas de muerte más importantes por año usando la fórmula:

$$\frac{\text{nro de muertes de adultos mayores por la causa (x) en el año } t * 10000(\text{habitantes})}{\text{total de adultos mayores en el año } t}$$

Las opciones usadas para crear el pronóstico se describen a continuación:

Serie temporal: Anual (ver figura 41)

Medidas: Prevalencia Mortalidad

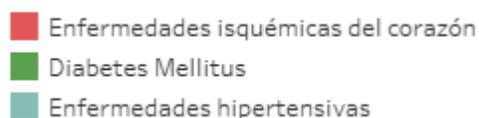
Pronóstico hacia delante: 2 períodos (2019 – 2020)

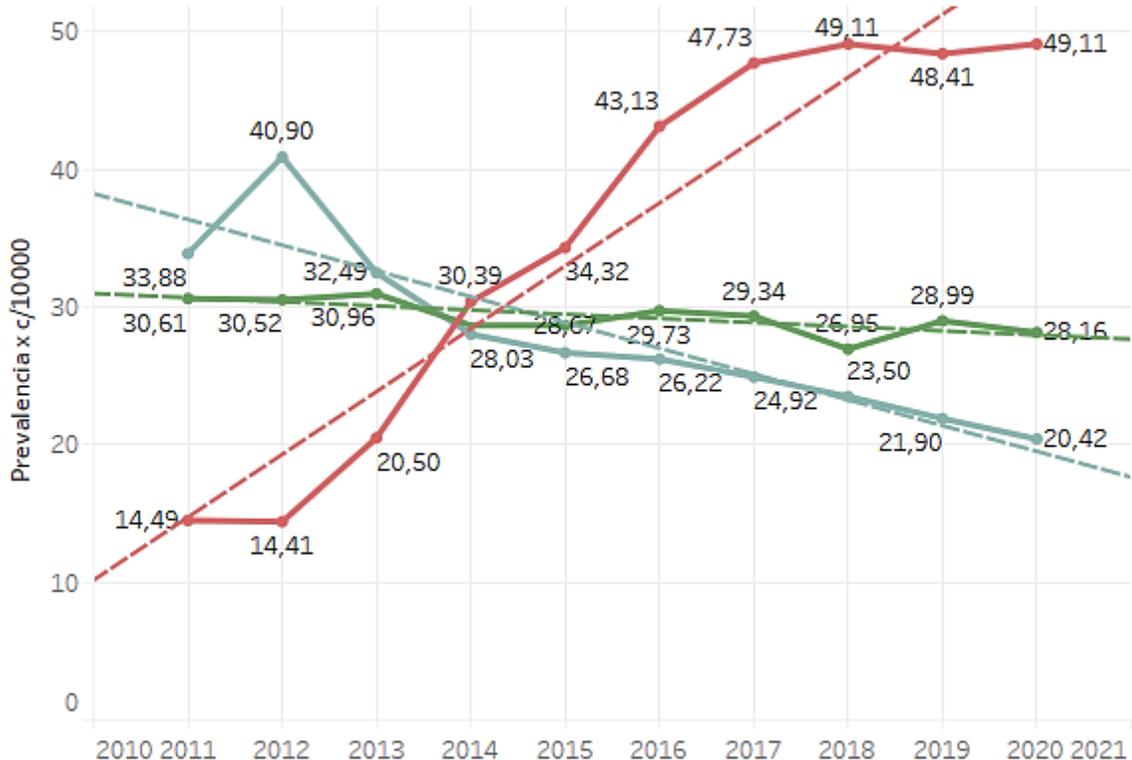
Pronóstico basado en: 2011 – 2018

Ignorar últimos: No se ignoraron períodos

Figura 41

Pronóstico principales causas de mortalidad obtenido en Tableau





En la tabla 27 se describe el tipo de modelo usado para cada causa de mortalidad y los coeficientes de suavizado respectivamente.

Tabla 27

Modelo usado para cada causa de mortalidad

Lista de 99 causas	Modelo			Coeficientes de suavizado		
	Nivel	Tendencia	Temporada	Alfa	Beta	Gamma
Enfermedades isquémicas del corazón	Multiplicativo	Ninguno	Multiplicativo	0,5	0	0,5
Diabetes Mellitus	Multiplicativo	Ninguno	Multiplicativo	0,5	0	0
Enfermedades hipertensivas	Aditivo	Aditivo	Ninguno	0,154	0	0

Evaluación del modelo

Al evaluar un modelo en Tableau, los posibles valores jerárquicos que se pueden encontrar son: Aceptar, Bueno y Pobre. En este caso todos los modelos tanto de morbilidad (ver tabla 28) como de mortalidad (ver tabla 29) son aceptables.

Tabla 28

Evaluación modelos morbilidad

Lista de 298 causas	Efecto de temporada				Contribución		
	Alto		Bajo		Tendencia	Temporada	Calidad
Neumonía	2020	1,06	2019	1,04	0,00%	100,00%	Aceptar
Colelitiasis y colecistitis			Ninguno		100,00%	0,00%	Aceptar
Diabetes mellitus	2019	0,91	2020	0,87	0,00%	100,00%	Aceptar

Tabla 29

Evaluación modelos mortalidad

Lista de 99 causas	Efecto de temporada				Contribución		
	Alto		Bajo		Tendencia	Temporada	Calidad
Enfermedades isquémicas del corazón	2020	1,51	2019	1,48	0,00%	100,00%	Aceptar
Diabetes Mellitus	2019	1,02	2020	0,98	0,00%	100,00%	Aceptar
Enfermedades hipertensivas			Ninguno		100,00%	0,00%	Aceptar

Capítulo V: Análisis y Discusión de Resultados

Evaluación de resultados obtenidos

Tableau muestra información estadística sobre las diferentes métricas de calidad como error cuadrático medio (RMSE), error absoluto medio (MAE), errores escalado absoluto medio (MASE), y error porcentual absoluto medio (MAPE). Al basarse en la interpretación de MAPE se observa que en el caso de la morbilidad se obtuvieron pronósticos altamente precisos siendo todos los valores menores al 10% (ver tabla 30), en cambio en los modelos de mortalidad (ver tabla 31) se obtuvo una previsión razonable en la predicción del Infarto agudo del miocardio, un buen pronóstico para el caso de la Neumonía, organismo no especificado y una alta precisión en los valores de la Diabetes mellitus no insulino dependiente

Tabla 30

Métricas de calidad Morbilidad

Lista de 298 causas	RMSE	MAE	MASE	MAPE	Interpretación MAPE
Neumonía	2,78	1,93	0,56	3,00%	Pronóstico altamente preciso <10%
Colelitiasis y colecistitis	2,015	1,635	0,76	3,60%	Pronóstico altamente preciso <10%
Diabetes mellitus	3,94	3,02	0,56	6,50%	Pronóstico altamente preciso <10%

Tabla 31

Métricas de calidad Mortalidad

Lista de 99 causas	RMSE	MAE	MASE	MAPE	Interpretación MAPE
Enfermedades isquémicas del corazón	6,92	5,72	0,51	19,30%	Buen pronóstico (entre 10% y 20%)
Diabetes Mellitus	1,2	1,1	0,73	3,80%	Pronóstico altamente preciso <10%
Enfermedades hipertensivas	3,64	2,03	0,58	5,70%	Pronóstico altamente preciso <10%

Para evaluar los resultados también se hizo uso de modelos de la regresión lineal. Se tomaron en cuenta principalmente dos medidas, una de las medidas de calidad es la bondad del ajuste del modelo es decir el valor p , el modelo puede ser significativo si $p \leq 0,05$. La segunda medida fue el coeficiente de determinación R cuadrado, es decir cómo se ajustaron los datos al modelo, si los valores se aproximan a 1, mejor será la predicción.

En base al cálculo de la prevalencia anual de las tres principales causas de morbilidad, se obtuvo como pronóstico los valores de los períodos 2019-2020 descritos en la tabla 32.

Tabla 32

Resultados del pronóstico de morbilidad 2019-2020

Causa298		2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Neumonía	Prevalencia x c/1000	57,26	57,22	60,38	59,21	59,84	65,29	66,32	67,65	66,64	67,48
	Intervalo de predicción inferior									59,30	60,29
	Intervalo de predicción superior									73,97	74,68
Colelitiasis y colecistitis	Prevalencia x c/1000	47,08	47,07	46,59	43,46	45,03	48,26	45,77	49,85	48,63	49,02
	Intervalo de predicción inferior									43,12	43,42
	Intervalo de predicción superior									54,14	54,63
Diabetes mellitus	Prevalencia x c/1000	58	57,44	50,74	48,66	48,31	43,26	43,03	40,06	41,99	40,09
	Intervalo de predicción inferior									36,06	34,62
	Intervalo de predicción superior									47,93	45,56

Se calcula un modelo de tendencia lineal para la prevalencia de las causas de morbilidad dado registro natural del año de egreso. A nivel general se obtiene como resultado un R cuadrado favorable de 0,96. El análisis de discordancia, indica que el valor de $p < 0,0001$ también es significativo.

Fórmula de modelo:	Causa298*(ln Año)+intercepto)
Número de observaciones modeladas:	30
Número de observaciones filtradas:	0
Grados de libertad de modelo:	6
Grados de libertad residual (DF):	24
SCE (suma de cuadrados del error):	84,1975
MSE (error cuadrático medio):	3,50823
R cuadrado:	0,962186
Error estándar (StdErr):	1,87303
valor p (Significancia):	< 0,0001

Análisis de discordancia:

<u>Field</u>	<u>DF</u>	<u>SSE</u>	<u>MSE</u>	<u>F</u>	<u>Valor p</u>
Causa298	4	21.376.372	534,409	152,33	< 0,0001

A continuación, en la tabla 33, se describe las ecuaciones del modelo exponencial para la tasa de prevalencia las principales causas de morbilidad.

Tabla 33

Modelo de líneas de tendencia para las causas de morbilidad

Ecuación	Causa298	Término	Valor	StdErr	valor t	valor p
Prevalencia = 2700,07*ln(Año) + (-20481,1)	Neumonía	ln(Año) intercepto	2700,07 -20481,1	350,58 2667,43	7,70171 -7,67819	< 0,0001 < 0,0001
Prevalencia = - 4211,19*ln(Año) + 32088,5	Colelitiasis y colecistitis	ln(Año) intercepto	669,561 -5047,36	395,061 3005,87	1,69483 -1,67917	0,128555 0,131634
Prevalencia = 669,561*ln(Año) + (-5047,36)	Diabetes mellitus	ln(Año) intercepto	-4211,19 32088,5	489,129 3721,6	-8,60956 8,62223	< 0,0001 < 0,0001

En base al cálculo de la prevalencia anual de las tres principales causas de mortalidad, se obtuvo como pronóstico los valores de los períodos 2019-2020 descritos en la tabla 34.

Tabla 34

Resultados del pronóstico de mortalidad 2019-2020

Lista de 99 causas		2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Enfermedades isquémicas del corazón	Prevalencia x c/10000	14,49	14,41	20,50	30,39	34,32	43,13	47,73	49,11	48,41	49,11
	Intervalo de predicción inferior									73,39	74,06
	Intervalo de predicción superior									23,43	24,16
Diabetes Mellitus	Prevalencia x c/10000	30,61	30,52	30,96	28,67	28,67	29,73	29,34	26,95	28,99	28,16
	Intervalo de predicción inferior									31,29	30,24
	Intervalo de predicción superior									26,70	26,07
Enfermedades hipertensivas	Prevalencia x c/10000	33,88	40,90	32,49	28,03	26,68	26,22	24,92	23,50	21,90	20,42
	Intervalo de predicción inferior									29,18	27,82
	Intervalo de predicción superior									14,62	13,01

Se calcula un modelo de tendencia lineal para para la prevalencia de las causas de mortalidad dado el registro natural del año de fallecimiento. A nivel general se obtiene como resultado un R cuadrado favorable de 0,95 y una significancia $p < 0,0001$.

Fórmula de modelo: Lista de 99 causas $\cdot (\ln (\text{Año}) + \text{intercepto})$

Número de observaciones modeladas: 30

Número de observaciones filtradas: 0

Grados de libertad de modelo: 6

Grados de libertad residual (DF): 24

SCE (suma de cuadrados del error): 229,325

MSE (error cuadrático medio): 9,5552

R cuadrado: 0,909904
 Error estándar (StdErr): 3,09115
 valor p (Significancia): < 0,0001

Análisis de discordancia:

<u>Field</u>	<u>DF</u>	<u>SSE</u>	<u>MSE</u>	<u>E</u>	<u>Valor p</u>
Lista 99 causas	4	21.585.764	539,644	56,4765	< 0,0001

En la tabla 35, se describe las ecuaciones del modelo exponencial para la tasa de prevalencia de las principales causas de mortalidad.

Tabla 35

Modelo de líneas de tendencia para las causas de mortalidad

<u>Ecuación</u>	<u>Lista 99 causas</u>	<u>Término</u>	<u>Valor</u>	<u>StdErr</u>	<u>valor t</u>	<u>valor p</u>
Prevalencia = $9205,12 \cdot \ln(\text{Año}) + -70003,1$	Enfermedades isquémicas del corazón	ln(Año)	9205,12	997,753	9,22585	< 0,0001
		Intercepto	-70003,1	7591,52	-9,22122	< 0,0001
Prevalencia = $-608,405 \cdot \ln(\text{Año}) + 4658,39$	Diabetes Mellitus	ln(Año)	-608,405	198,26	-3,06873	0,0153775
		Intercepto	4658,39	1508,48	3,08813	0,0149319
Prevalencia = $-3774,05 \cdot \ln(\text{Año}) + 28743,2$	Enfermedades hipertensivas	ln(Año)	-3774,05	613,719	-6,14948	0,0002742
		intercepto	28743,2	4669,56	6,15545	0,0002724

Despliegue

Tableau Public, brinda un servicio en la nube gratuito para la publicación de los trabajos realizados, por lo cual los resultados se pueden visualizar de forma interactiva en la web, pudiendo encontrar el producto de la presente investigación en el siguiente enlace:

<https://public.tableau.com/profile/catalina2916#!/vizhome/PrevalenciadelaMorbimortalidaddeladultomayorecuatoriano/AdultoMayor>

Discusión de los resultados

En el presente estudio se hizo uso de los datos abiertos que ofrece el INEC en la página web, esta información histórica correspondió a las estadísticas anuales de los egresos hospitalarios y de las defunciones generales del periodo 2011-2018.

Empleando Tableau Prep se realizó la preparación de los datos (extracción, transformación y carga), utilizando Tableau Public se realizó un análisis preliminar de las variables, así como el diseño y generación de los modelos.

En el sistema nacional de salud, durante el periodo 2011-2018 de estudio fueron atendidos alrededor de 1'236.172 adultos mayores. De los cuales la mayoría un 50,69% fueron mujeres, mientras que el 49,31% fueron hombres. La edad promedio del adulto mayor fue de 76 años y la mediana fue de 75 años. Los establecimientos de salud del sector públicos albergaron el 64,98% de los pacientes atendidos, los privados con fines de lucro 29,75% y los privados sin fines de lucro asistieron al 5,27%. Estos establecimientos se ubican principalmente en las provincias de Pichincha con un 24,08%, Guayas con 23,23% y Azuay con 9,06%. Un 94,2% de adultos mayores egresaron de las diferentes instituciones de salud en calidad de vivos. Los grupos de enfermedades más frecuentes presentes en los adultos mayores corresponden a los capítulos XI Enfermedades del aparato digestivo (K00-K93) con el 17,02%. De las diez principales causas de morbilidad las que predominan son: Neumonía, Diabetes mellitus, y Colelitiasis y colecistitis.

En la información recopilada de las defunciones generales en el periodo estudiado se registra en el país alrededor de 312.237 fatalidades referentes a adultos mayores. De los cuales un mayor porcentaje del 50,61% fueron mujeres, mientras que el 49,39% fueron hombres. La edad promedio del adulto mayor fallecido fue de 81 años

y la mediana fue de 82 años. Los perecimientos se evidencian principalmente en las provincias del Guayas con 26,98%, Pichincha con 17,86% y Manabí con 9,22%. De las diez principales causas de mortalidad las más relevantes son: Enfermedades isquémicas del corazón (con tendencia alcista desde el 2014), Diabetes Mellitus y Enfermedades hipertensivas.

En la evaluación de los resultados se consideraron la calidad del pronóstico usando el valor del error porcentual absoluto medio (MAPE), y en los modelos de regresión lineal, las medidas de calidad como la bondad del ajuste del modelo (es decir el valor p), y el coeficiente de determinación (R cuadrado). En referencia al pronóstico de la prevalencia de las tres principales causas de morbilidad del adulto mayor durante el 2019 y 2020 por cada 1000 pacientes egresados se obtiene valores aceptables con un R cuadrado igual a 0,962186 y una significancia de $p < 0,0001$ (ver tabla 36).

Tabla 36

Pronóstico y evaluación, principales causas de morbilidad

R cuadrado		0,962186										
valor p (Significancia)		< 0,0001										
Causa298	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	MAPE	Interpretación MAPE
Neumonía	57,26	57,22	60,38	59,21	59,84	65,29	66,32	67,65	66,64 ±7,33	67,48 ±7,19	3,00%	Pronóstico altamente preciso
Colelitiasis y colecistitis	47,08	47,07	46,59	43,46	45,03	48,26	45,77	49,85	48,63 ±5,51	49,02 ±5,6	3,60%	Pronóstico altamente preciso
Diabetes mellitus	58	57,44	50,74	48,66	48,31	43,26	43,03	40,06	41,99 ±5,93	40,09 ±5,47	6,50%	Pronóstico altamente preciso

En cuanto al pronóstico de la prevalencia de las tres principales causas de mortalidad del adulto mayor durante el 2019 y 2020 por cada 10000 habitantes, también se obtiene valores aceptables con un R cuadrado igual a 0,909904 y una significancia de $p < 0,0001$ (ver tabla 37).

Tabla 37

Pronóstico y evaluación de las principales causas de mortalidad

R cuadrado		0,909904										
valor p (Significancia)		< 0,0001										
Lista 99 causas	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	MAPE	Interpretación MAPE
Enfermedades isquémicas del corazón	14,49	14,41	20,50	30,39	34,32	43,13	47,73	49,11	48,41 ±24,98	49,11 ±24,95	19,30%	Buen pronóstico
Diabetes Mellitus	30,61	30,52	30,96	28,67	28,67	29,73	29,34	26,95	28,99 ±2,3	28,16 ±2,08	3,80%	Pronóstico altamente preciso
Enfermedades hipertensivas	33,88	40,90	32,49	28,03	26,68	26,22	24,92	23,50	21,90 ±7,28	20,42 ±7,4	5,70%	Pronóstico altamente preciso

En referencia a la hipótesis *la creación del modelo de series temporales permitirá determinar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano*, fue demostrada, ya que los modelos creados tanto para la morbilidad como para la mortalidad proporcionan una prevalencia anual aceptable con un MAPE apropiado entre 19,30% y 3,00% inferior al 20%, con una bondad de ajuste significativa $p < 0,0001$, y un considerable coeficiente de determinación R igual a 0,9.

Capítulo VI: Conclusiones y Recomendaciones

Conclusiones

Al efectuar la revisión inicial de literatura no se encontraron estudios que describan específicamente la aplicación de series temporales en la identificación de patrones de comportamiento de la morbimortalidad del adulto mayor ecuatoriano, sin embargo, en forma general respecto a la morbilidad y mortalidad a establecer a través de una apropiada cadena de búsqueda se pudo evidenciar información relevante que sirvió como referencia en el desarrollo del proyecto. De acuerdo a diferentes investigaciones las series temporales han sido de gran apoyo en el área de la salud, ya que han permitido múltiples estudios sobre el diagnóstico y tratamiento de enfermedades, permitiendo el análisis de los datos médicos y predicciones de gran precisión que han cobrado una mayor notoriedad.

La divulgación de información oficial fue de mucha utilidad para el desarrollo del presente caso de estudio, ya que a través de los datos proporcionados del INEC se pudo recolectar los archivos del periodo 2011-2018 necesarios para el desarrollo del proyecto.

Para realizar el ETL es necesario verificar las inconsistencias de los campos y registros, y así poder determinar si se promedian o eliminan los valores fuera de rango o nulos. La herramienta Tableau Prep fue seleccionada para el proceso de ETL, bajo los criterios de acceso libre, procesamiento veloz de datos en memoria, estadísticas de rápida visualización, conexión a múltiples fuentes de datos y archivos. Además, proporcionó las operaciones necesarias para el análisis de los datos de una forma rápida y sencilla.

La plataforma Tableau Public de uso gratuito permitió incorporar todos los datos depurados para el diseño y generación del modelo de series temporales ya que internamente hace uso del lenguaje R y ofrece un adecuado procesamiento de los datos, a diferencia de otras herramientas que se pretendieron usar, ya que al ser de pago presentaban limitaciones.

Con referencia a la evaluación del pronóstico de los modelos que hicieron uso del método estacional de Holt-Winters, se obtuvo una prevalencia anual aceptable y un apropiado Error Porcentual Absoluto Medio (MAPE) inferior al 19.30%; pese a que existen algunas métricas de evaluación se considera el MAPE como el más apropiado.

Se pudo determinar que los modelos de series temporales elaborados para la morbilidad y la mortalidad permitieron determinar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano.

Recomendaciones

Al momento de realizar la revisión inicial de literatura se recomienda seguir el proceso respectivo, para obtener una adecuada cadena de búsqueda y en consecuencia artículos que proporcionen información relevante del tema.

En la fase de comprensión de los datos, es recomendable tener presente el formato o extensión que posee la información recolectada para poder determinar la mejor herramienta.

Se sugiere que, al presente estudio, se pueden aplicar otros métodos de análisis y pronósticos tales como Fuzzy, Naive, CUmulative SUM (CUSUM), el Early

Aberration Reporting System (EARS), Box-Jenkins o el promedio móvil integrado autorregresivo (ARIMA), haciendo uso de las herramientas y métricas que se adapten mejor a estos métodos.

Se propone que la información obtenida en este estudio pueda ser utilizada por instituciones públicas, privadas, individuos y profesionales de la salud para evaluar posibles riesgos, y tomar medidas preventivas o establecer políticas respecto a este sector vulnerable.

Bibliografía

- Abdullah, S., Sapii, N., Dir, S., & Jalal, T. M. (2012). Application of univariate forecasting models of tuberculosis cases in Kelantan. *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), 2012*, 1-7. <https://doi.org/10.1109/ICSSBE.2012.6396582>
- Al-Sakkaf, A., & Jones, G. (2013). Comparison of Time Series Models for Predicting Campylobacteriosis Risk in New Zealand. *Zoonoses and Public Health*, 61(3), 167-174. <https://doi.org/10.1111/zph.12046>
- Amini, P., Ghaleiha, A., Zarean, E., Sadeghifar, M., Ghaffari, M. E., Taslimi, Z., & Yazdi-Ravandi, S. (2018). Modelling the Frequency of Depression using Holt-Winters Exponential Smoothing Method. *Journal of Clinical and Diagnostic Research* 12(10), 24-27. <https://doi.org/10.7860/JCDR/2018/35765.12085>
- Anguera, Lara, Lizcano, Martínez, & Pazos. (2012). Sensor-Generated Time Series Events: A Definition Language. *Sensors*, 12(9), 11811–11852. <https://doi.org/10.3390/s120911811>
- Archivo Nacional de Datos y Metadatos Estadísticos - ANDA. (2012). *Ecuador - Estadística de Defunciones Generales y Fetales 2012*. Instituto Nacional de Estadística y Censos. <https://anda.inec.gob.ec/anda/index.php/catalog/389/vargrp/VG10>
- Betancourt, S. (2014). Densidad mineral ósea, calcio dietético y factores presuntivos de riesgo de osteoporosis en mujeres ecuatorianas de la tercera edad. *Nutrición Hospitalaria*, 30(2), 372-384. <http://dx.doi.org/10.3305/nh.2014.30.2.7563>
- Borja, M., & Castaño, V. (2017). *Minería de datos de salud: estudio de los factores personales, familiares y vivienda que influyen en las enfermedades de diabetes e hipertensión a partir de la encuesta de atención primaria en salud del Área Metropolitana del Valle de Aburrá*. [Trabajo de grado, Universidad Pontificia Bolivariana]. <http://hdl.handle.net/20.500.11912/3571>
- Buczak, A. L., Baugher, B., Guven, E., Ramac-Thomas, L. C., Elbert, Y., Babin, S. M., & Lewis, S. H. (2015). Fuzzy association rule mining and classification for the prediction of malaria in South Korea. *BMC Medical Informatics and Decision Making*, 15(1), 1-17 <https://doi.org/10.1186/s12911-015-0170-6>
- Ceballos, C., & Gutierrez, S. (2014). *Humanización de la atención en salud* [Archivo PDF]. <https://pdfs.semanticscholar.org/ac5b/e9442577b6126eb05e11064d8a992dd6ee86.pdf>
- Consejo de Igualdad Intergeneracional - CNII. (2019). *Personas adultas mayores, situación y derechos*. <https://www.igualdad.gob.ec/personas-adultas-mayores-situacion-y-derechos/>
- Constitución de la República del Ecuador. (2008). TÍTULO II: DERECHOS. Capítulo tercero: Derechos de las personas y grupos de atención prioritaria. Sección primera: Adultas y adultos mayores, Art. 36. https://www.oas.org/juridico/mla/sp/ecu/sp_ecu-int-text-const.pdf
- Crowson, M. G., Schulz, K., & Tucci, D. L. (2016). National Utilization and Forecasting of Otological Antibiotics: Medicaid Data Versus “Dr. Google”. *Otology & Neurotology*, 37(8), 1049-1058. <https://doi.org/10.1097/MAO.0000000000001115>
- Dávila Hernández, F., & Sánchez Corales, Y. (2012). Data mining techniques applied to diagnosis of clinical entities. *Revista Cubana de Informática Médica*. 4(2)174-

183. <https://www.medigraphic.com/cgi-bin/new/resumenl.cgi?IDARTICULO=37857>
- Feng, L., & Shi, Y. (2018). Forecasting mortality rates: multivariate or univariate models? *Journal of Population Research*, 2018, 289–318. <https://doi.org/10.1007/s12546-018-9205-z>
- Gartner, Inc. (2020). *Magic Quadrant for Analytics and Business Intelligence Platforms*. https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb&ocid=mkto_eml_EM526600A1LA1
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Herrera, A. (2008). *Psicogeriatría: Vejez y envejecimiento, Universidad de Chile* [Archivo PDF]. https://www.ucursos.cl/medicina/2008/1/MPSIQUI5/2/material_docente/detalle?id=157478
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice. 2nd edition*. OTexts. <https://otexts.org/fpp2/>
- Huenchuan, S. (2018). Envejecimiento, personas mayores y Agenda 2030 para el Desarrollo Sostenible: perspectiva regional y de derechos humanos. (Libros de la CEPAL, N° 154 (LC/PUB.2018/24-P), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL).
- Instituto Nacional de Estadística y Censos. (2015). *¿Cómo crecerá la población en Ecuador?* [Archivo PDF]. http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Proyecciones_Poblacionales/presentacion.pdf
- Instituto Nacional de Estadística y Censos. (2018). *Encuesta Nacional De Empleo, Desempleo y Subempleo - ENEMDU. Indicadores Laborales Marzo 2018* [Archivo PDF]. http://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Marzo-2018/032018_Presentacion_M_Laboral.pdf
- Instituto Nacional de Estadística y Censos. (2015). *CRITERIOS DE VALIDACIÓN* [Tabla]. Recuperado de: <https://anda.inec.gob.ec/anda/index.php/catalog/573/download/8733>
- Instituto Nacional de Estadística y Censos. (2016). *CRITERIOS DE VALIDACIÓN* [Tabla]. Recuperado de: <https://anda.inec.gob.ec/anda/index.php/catalog/594/download/9714>
- Instituto Nacional de Estadística y Censos. (2018). *Nacimientos y Defunciones*. [Tabla]. Recuperado de: https://www.ecuadorencifras.gob.ec/nacimientos_y_defunciones/
- Instituto Nacional de Estadística y Censos. (2018). *Registro Estadístico de Camas y Egresos Hospitalarios 2017* [Tabla]. Recuperado de: http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/Camas_Egresos_Hospitalarios/Cam_Egre_Hos_2017/Presentacion_CEH_2017.pdf
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2011). *Ecuador - Estadística de Defunciones Generales y Fetales 2011*. [Tabla]. Recuperado de: <https://anda.inec.gob.ec/anda/index.php/catalog/359/datafile/F6/V226#page=F6&tab=data-dictionary>
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2011). *Ecuador - Estadísticas Hospitalarias Camas y Egresos 2011* [Tabla]. Recuperado de: <https://anda.inec.gob.ec/anda/index.php/catalog/608/datafile/F6>

- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2012). *Ecuador - Estadística de Defunciones Generales y Fetales 2012* [Tabla]. Recuperado de:
<https://anda.inec.gob.ec/anda/index.php/catalog/389/datafile/F7/V296#page=F7&tab=data-dictionary>
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2012). *Ecuador - Estadísticas Hospitalarias Camas y Egresos 2012* [Tabla]. Recuperado de:
<https://anda.inec.gob.ec/anda/index.php/catalog/609/datafile/F8>
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2013). *Ecuador - Estadística de Defunciones Generales y Fetales 2013* [Tabla]. Recuperado de:
<https://anda.inec.gob.ec/anda/index.php/catalog/544/datafile/F11>
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2013). *Ecuador - Estadísticas Hospitalarias Camas y Egresos 2013* [Tabla]. Recuperado de:
<https://anda.inec.gob.ec/anda/index.php/catalog/555/datafile/F12>
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2014). *Ecuador - Estadística de Defunciones Generales y Fetales 2014* [Tabla]. Recuperado de:
https://anda.inec.gob.ec/anda/index.php/catalog/545/data_dictionary#page=F14&tab=data-dictionary
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2014). *Ecuador - Estadísticas Hospitalarias Camas y Egresos 2014* [Tabla]. Recuperado de:
<https://anda.inec.gob.ec/anda/index.php/catalog/561/datafile/F14>
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2016). *Ecuador - Estadística de Defunciones Generales y Fetales 2016* [Tabla]. Recuperado de:
<https://anda.inec.gob.ec/anda/index.php/catalog/594/datafile/F22>
- Instituto Nacional de Estadística y Censos - Secretaría Nacional de Planificación y Desarrollo. (2017). *Ecuador - Estadísticas Hospitalarias Camas y Egresos 2017* [Tabla]. Recuperado de:
<https://anda.inec.gob.ec/anda/index.php/catalog/753/datafile/F30>
- Jain, N., & Srivastava, V. (2013). Data mining techniques: a survey paper. *International Journal of Research in Engineering and Technology*, 02(11), 116-119.
<https://doi.org/10.15623/ijret.2013.0211019>
- Kotu, V., & Deshpande, B. (2015). *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Amsterdam: Morgan Kaufmann. ISBN: 978-0-12-801460-8
- Li & Wang. (2018). Similarity Measure Based on Incremental Warping Window for Time Series Data Mining. *IEEE Access*, 07(99), 3909-3917.
<https://doi.org/10.1109/ACCESS.2018.2889792>
- Ministerio de Inclusión Económica y Social - MIES. (2013). *Población Adulta Mayor*.
<https://www.inclusion.gob.ec/direccion-poblacion-adulta-mayor/>
- Montaño Moreno, J. J., Palmer Pol, A., Sesé Abad, A., & Cajal Blasco, B. (2013). *Universidad de Oviedo, Using the R-MAPE index as a resistant measure of forecast accuracy* [Tabla]. Recuperado de:
<https://www.redalyc.org/pdf/727/72728554013.pdf>

- Nancy, J., Khanna, N., & Kannanb, A. (2017). A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals. *Expert Systems With Applications (2017)*.
<https://doi.org/10.1016/j.eswa.2017.01.056>
- Navaz, A. N., Mohammed, E., Serhani, M. A., & Zaki, N. (2016). The use of data mining techniques to predict mortality and length of stay in an ICU. *2016 12th International Conference on Innovations in Information Technology (IIT)*, 2016, 1-5. <https://doi.org/10.1109/INNOVATIONS.2016.7880045>
- Organización Mundial de la Salud. (2013). *Mortalidad*.
<https://www.who.int/topics/mortality/es/>
- Organización Mundial de la Salud. (2015). *Envejecimiento y salud*.
<https://www.who.int/es/news-room/fact-sheets/detail/envejecimiento-y-salud>
- Organización Mundial de la Salud. (2015). *Informe mundial sobre envejecimiento y salud* [Archivo PDF].
https://apps.who.int/iris/bitstream/handle/10665/186466/9789240694873_spa.pdf?sequence=1
- Organización Mundial de la Salud. (2017). *Clasificación Internacional de Enfermedades (CIE)*.
https://www.paho.org/hq/index.php?option=com_content&view=article&id=3561:2010-clasificacion-internacional-enfermedades-cie&Itemid=2560&lang=es
- Organización Panamericana de la Salud. (2003). *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud* [Archivo PDF].
<http://ais.paho.org/classifications/Chapters/pdf/Volume2.pdf>
- Organización Panamericana de la Salud. (2008). *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud* [Archivo PDF].
<http://ais.paho.org/classifications/Chapters/pdf/Volume1.pdf>
- Organización Panamericana de la Salud. (2018). *INDICADORES DE SALUD: Aspectos conceptuales y operativos*.
https://www.paho.org/hq/index.php?option=com_content&view=article&id=14402:health-indicators-conceptual-and-operational-considerations-section-2&Itemid=0&showall=1&lang=es
- Pérez, C. (2007). *Minería de datos: técnicas y herramientas*. Madrid: Paraninfo.
- Piedra, D., Ferrer, A., & Gea, J. (2013). *Text Mining and Medicine: Usefulness in Respiratory Diseases*. Piedra, D., Ferrer, A., & Gea, J. (2014). *Text Mining and Medicine: Usefulness in Respiratory Diseases*. *Archivos de Bronconeumología (English Edition)*, 50(3), 113–119. <https://doi.org/10.1016/j.arbr.2014.02.008>
- Pinlac, P. A., & Soonthornworasiri, N. (2016). *Descriptive and Predictive Time Series Analysis of Premature Mortality from Noncommunicable Disease among Filipinos*. <https://hrdo.upm.edu.ph/index.php/acta/article/view/805/709>
- Ríos, G. (2008). *Series de Tiempo* [Archivo PDF]. https://www.u-cursos.cl/ingenieria/2010/1/CC52A/1/material_docente/detalle?id=296003
- Romero, H. S. (2015). *Derecho constitucional de los adultos mayores: su estado de abandono y su implicación socio económica* [Archivo PDF].
<http://www.dspace.uce.edu.ec/bitstream/25000/5583/1/T-UCE-0013-Ab-368.pdf>
- Seneviratne, M., Kahn, M., & Hernandez-Boussard, T. (2018). Merging heterogeneous clinical data to enable knowledge discovery. *Biocomputing 2019*, 439-443.
https://doi.org/10.1142/9789813279827_0040
- Serhani, M. A., Benharref, A., & Nujum, A. R. (2014). An Adaptive Expert System for Automated Advices Generation-Based Semi-continuous M-Health Monitoring.

- Brain Informatics and Health*, 8609, 388-399. https://doi.org/10.1007/978-3-319-09891-3_36
- Sistema de Indicadores Sociales del Ecuador. (2015). *Diez principales causas de morbilidad hospitalaria en adultos mayores*.
http://www.siise.gob.ec/siiseweb/PageWebs/SALUD/ficsal_S57.htm
- Sistema de Indicadores Sociales del Ecuador. (2015). *Diez principales causas de mortalidad en el adulto mayor*.
http://www.siise.gob.ec/siiseweb/PageWebs/SALUD/ficsal_S109.htm
- Sistema de Indicadores Sociales del Ecuador (2015). *Los objetivos del SIISE*.
http://www.siise.gob.ec/siiseweb/PageWebs/SIISE/macsii_objeti.htm
- Tian, Z., Li, S., Zhang, J., Jaakkola, J., & Guo, Y. (2012). *Ambient temperature and coronary heart disease mortality in Beijing, China: A time series study*. *Environmental health*, 11(1), 1-7. <https://doi.org/10.1186/1476-069X-11-56>
- Trivedi, S. K., Dey, S., Kumar, A., & Panda, T. K. (2017). *Handbook of Research on Advanced Data Mining Techniques and Applications for Business Intelligence*. pp. 68-69. Hershey, Pensilvania: IGI Global.
- Villena, J. (2016). *CRISP-DM*. <https://data.sngular.com/es/art/26/crisp-dm-fase-i-comprension-del-negocio-business-understanding>
- Waters, W. F., & Gallegos, C. A. (2012). *Asociación Latinoamericana de Población - ALAP. Envejecimiento, Salud y Bienestar en los Adultos Mayores Indígenas del Ecuador* [Archivo PDF].
http://www.alapop.org/Congreso2012/DOCSFINAIS_PDF/ALAP_2012_FINAL211.pdf
- Wu, Y. (27 de 12 de 2018). Por qué China, el país más poblado del mundo, necesita incentivar a sus mujeres a tener más hijos. *British Broadcasting Corporation – BBC*. <https://www.bbc.com/mundo/noticias-46682872>
- Yang, E., Park, H. W., Choi, Y. H., Kim, J., Munkhdalai, L., Musa, I., & Ryu, K. H. (2018). A Simulation-Based Study on the Comparison of Statistical and Time Series Forecasting Methods for Early Detection of Infectious Disease Outbreaks. *International Journal of Environmental Research and Public Health*, 15(5), 966. <https://doi.org/10.3390/ijerph15050966>