



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA**

**ESCUELA PROFESIONAL DE INGENIERÍA DE  
SISTEMAS**

Sistema de pronóstico geolocalizado de atenciones de urgencia

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:**

Ingeniero de Sistemas

**AUTORES:**

Liñan Ruiz, William Wyler (ORCID: 0000-0002-3076-6782)

Reyes Taquire, Angello Jesus (ORCID: 0000-0003-0877-2737)

**ASESOR:**

Dr. Alfaro Paredes, Emigdio Antonio (ORCID: 0000-0002-0309-9195)

**LÍNEA DE INVESTIGACIÓN:**

Sistema de información y comunicaciones

LIMA – PERÚ

2020

### **Dedicatoria**

Dedicamos esta tesis a Dios por darnos vida y salud y a nuestros padres, quienes son la causa y la esencia de cada uno de nuestros esfuerzos.

### **Agradecimiento**

Agradecemos a nuestras familias por el apoyo que nos han brindado a lo largo de nuestra etapa universitaria para terminar con un buen trabajo de investigación y al profesor Emigdio Antonio Alfaro Paredes quien nos ayudó con el asesoramiento para conseguir una buena investigación.

## Índice de contenidos

I. INTRODUCCIÓN .....	1
II. MARCO TEÓRICO .....	13
III. MÉTODO.....	27
3.1 Tipo y diseño de investigación.....	28
3.2 Variables y operacionalización .....	29
3.3 Población, muestra y muestreo .....	29
3.4 Técnicas e instrumentos de recolección de datos.....	30
3.5 Procedimientos.....	31
3.6 Aspectos éticos .....	31
IV. RESULTADOS.....	33
V. DISCUSIÓN.....	104
VI. CONCLUSIONES.....	109
VII. RECOMENDACIONES .....	114
REFERENCIAS .....	118

## Índice de tablas

Tabla 1 Niveles de triaje .....	25
Tabla 2 Variables de la base de datos de atenciones de urgencias .....	39
Tabla 3 Exclusión de los datos en la base de atenciones de urgencia .....	45
Tabla 4 Resultados de totales atenciones de urgencia registrados por años .....	46
Tabla 5 Tiempo de carga o ejecución de la tarea ETL .....	59
Tabla 6 Requerimientos de hardware del proyecto .....	61
Tabla 7 Tabla resumen de residuos regresión lineal múltiple .....	69
Tabla 8 Tabla resumen de coeficientes de la regresión lineal múltiple de la variable cantidad de atenciones de urgencia .....	70
Tabla 9 Descripción de las variables regresión lineal múltiple .....	72
Tabla 10 Tabla resumen de interpretación de regresión lineal múltiple con la variable cantidad de atenciones .....	74
Tabla 11 Tabla resumen de la regresión lineal múltiple .....	79
Tabla 12 Tabla resumen ANOVA regresión lineal múltiple .....	79
Tabla 13 Total de Atenciones de urgencia de Antofagasta por meses .....	80
Tabla 14 Resultado de filtrado Holt-Winter's de atenciones de urgencias .....	82
Tabla 15 Proyección de atenciones Holt-Winter's en RStudio .....	84
Tabla 16 Resultados de datos de entrenamiento de Holt-Winter's en RStudio.....	85
Tabla 17 Resultados con la predicción Holt-Winter's .....	86
Tabla 18 Resultados de los datos de entrenamiento del método Naive en RStudio.....	88
Tabla 19 Resultados con la predicción método Naive en RStudio .....	89
Tabla 20 Resultados del entrenamiento exponencial de tendencia suavizado en RStudio.....	91
Tabla 21 Resultados con la predicción exponencial de tendencia suavizado en RStudio.....	91
Tabla 22 Margen de error porcentual comparación en RStudio.....	94
Tabla 23 Resumen de Coeficientes en la regresión ordinal logística en Rstudio.....	96
Tabla 24 Modelo Likelihood Ratio Test en la regresión ordinal logística en Rstudio .....	97
Tabla 25 Coeficiente de evaluación $a_j$ en la regresión ordinal logística en Rstudio .....	97
Tabla 26 Interpretación de las variables logísticas Regresión ordinal con la variable cantidad de atenciones de urgencias .....	99
Tabla 27 Matriz de operacionalización variable.....	134
Tabla 28 Matriz de consistencia .....	135
Tabla 29 Tabla de puntuación de metodología .....	153
Tabla 30 Tabla de puntuación Herramientas de minería de datos .....	154
Tabla 31 Tabla de puntuación de herramientas de visualización de datos .....	155
Tabla 32 Clasificación de metodologías de minería de datos .....	156
Tabla 33 Clasificación de herramientas de minería de datos.....	159
Tabla 34 Clasificación de herramientas de visualización de datos .....	160
Tabla 35 Dimensión H_AtencionUrgencia.....	167
Tabla 36 Dimensión D_Establecimiento.....	167
Tabla 37 Dimensión D_TipoEstablecimiento.....	168
Tabla 38 Dimensión D_Ubicacion .....	168
Tabla 39 Dimensión D_PeriodoTiempo.....	168
Tabla 40 Dimensión D_TipoAtencion .....	169
Tabla 41 Dimensión D_Campaña.....	169
Tabla 42 Dimensión D_Padecimiento .....	169
Tabla 43 Dimensión CEI10.....	169
Tabla 44 Dimensión D_Medicamento.....	170

## Índice de figuras

Figura 1 Diagrama de flujo de la atención de emergencia .....	36
Figura 2 Base de datos de atenciones de urgencias de Chile .....	38
Figura 3 Cantidad de atenciones por año.....	41
Figura 4 Numero de tratamiento por pacientes .....	41
Figura 5 Cantidad de atención al paciente de 0 a 11 meses.....	42
Figura 6 Cantidad de atención para los pacientes de 1 a 4 años.....	42
Figura 7 Cantidad de atención para los pacientes de 5 a 14 años de edad.....	43
Figura 8 Atención a pacientes de 15 a 64 años .....	43
Figura 9 Atenciones de pacientes de 65 años o más.....	44
Figura 10 Cantidad del tipo de medicamento utilizado en la atención .....	44
Figura 11 Modelo dimensional de atenciones de urgencias.....	47
Figura 12 Creación de proyecto ETL de atenciones de urgencias.....	48
Figura 13 Configuración del proyecto ETL .....	48
Figura 14 Selección de la fuente de destino de la base de datos .....	49
Figura 15 Localización de base de datos de atenciones de urgencia .....	49
Figura 16 Verificación de conexión de la base de datos .....	50
Figura 17 Modelo de flujo de control de minería de datos.....	50
Figura 18 Control de flujo de la dimensión establecimiento .....	51
Figura 19 Tarea de flujos de datos de la base de datos –Carga.....	52
Figura 20 Control de flujo de base de datos de MSAccess a SQLServer .....	52
Figura 21 Selección de la fuente de acceso a datos SQLServer .....	53
Figura 22 Selección del modo de acceso a los datos al SQLServer .....	53
Figura 23 Entrada de la sentencia SQL.....	54
Figura 24 Selección de la conexión a la base de datos SQLServer destino .....	54
Figura 25 Selección del modo de acceso a los datos del editor destino .....	55
Figura 26 Selección de la tabla destino en SQLServer .....	55
Figura 27 Resultado de la carga de la Dimensión D_ Establecimiento.....	56
Figura 28 Resultado de la carga de la Dimensión D_ Campaña.....	56
Figura 29 Resultado de la carga de la Dimensión D_ Medicamento.....	57
Figura 30 Resultado de la carga de la Dimensión D_ Padecimiento .....	57
Figura 31 Resultado de la carga de la Dimensión D_ PeriodoTiempo.....	57
Figura 32 Resultado de la carga de la Dimensión D_ TipoAtencion .....	58
Figura 33 Resultado de la carga de la Dimensión D_ TipoEstablecimiento.....	58
Figura 34 Resultado de la carga de la Dimensión D_ Ubicación .....	58
Figura 35 Resultado de la carga de la Dimensión H_ AtencionesUrgencia.....	59
Figura 36 Cargar de datos SQLServer en Rapiminer Tools.....	62
Figura 37 Cargar los datos SQLServer al RStudio.....	62
Figura 38 Selección de la tarea a ejecutar en Rapiminer Tools .....	63
Figura 39 Selección de las entradas de la base de datos .....	64
Figura 40 Selección del modelo de prediccion .....	64
Figura 41 Resultado de la carga de Atenciones de urgencia .....	65
Figura 42 Pantalla principal del Rstudio .....	65
Figura 43 Selección de la base de datos de atenciones de urgencia de Antofagasta .....	66
Figura 44 Configuración de la base de datos en RStudio .....	66
Figura 45 Conversión de variables de la base de datos a RStudio.....	67
Figura 46 Vista de los datos cargados en RStudio.....	67
Figura 47 Formula de regresión lineal multiple.....	68
Figura 48 Resultados de atenciones de urgencias por meses en RStudio.....	80

Figura 49 Diagrama de flujo de atenciones de urgencias en Antofagasta .....	81
Figura 50 Formula general del Holt-Winter's .....	81
Figura 51 Calcula el filtrado de pronostico del Holt-Winter's en RStudio .....	82
Figura 52 Gráfico de los datos filtrados de Holt-Winter's en RStudio.....	84
Figura 53 Código de predicción del Holt-Winter's en RStudio.....	84
Figura 54 Grafico de filtro de atenciones en Holt-Winter's en el RStudio.....	85
Figura 55 Resumen de resultado Holt-Winter's en RStudio .....	85
Figura 56 Grafica de predicción Holt-Winter's en RStudio .....	87
Figura 57 Formula general del método Naive .....	88
Figura 58 Código del método Navie en RStudio .....	88
Figura 59 Gráfico de predicción del método Naive estacional en RStudio.....	90
Figura 60 Formula general del Exponencial de tendencia suavizado .....	91
Figura 61 Código del exponencial de tendencia suavizado en RStudio.....	91
Figura 62 Grafica de predicción del exponencial de tendencia suavizado en RStudio .....	92
Figura 63 Formula de la media del porcentaje de error absoluto .....	93
Figura 64 Algoritmo Mape en RStudio.....	94
Figura 65 Fórmula general de la regresión ordinal logística.....	95
Figura 66 Formula general descompuesta de la regresión ordinal logística .....	95
Figura 67 Etapas de metodología CRISP-DM.....	161
Figura 68 Fases y tareas del CRISP DM.....	163
Figura 69 Modelo dimensional con variables dummy.....	166
Figura 70 Ícono Tableau.....	177
Figura 71 Ventana principal del Tableau .....	177
Figura 72 Selección del servidor de base de datos en Tableau.....	178
Figura 73 Entrada de datos para la conexión en Tableau.....	178
Figura 74 Tablas de ventanas de la base de datos en Tableau.....	179
Figura 75 Selección de datos de análisis en Tableau .....	179
Figura 76 Ventana principal del sistema de predicción en Antofagasta .....	180
Figura 77 Vista de la atención de emergencia de resultados.....	181
Figura 78 Vista superior del centro de salud con más atención .....	181
Figura 79 Visualización de etiquetas .....	181
Figura 80 Gráfico de flujo de atención por año.....	182
Figura 81 Gráfico de las campañas llevadas en el establecimiento.....	182
Figura 82 Mapa de Antofagasta y sus centros de salud.....	183
Figura 83 Información del centro de salud .....	184
Figura 84 Filtro de establecimientos.....	184
Figura 85 Remover el filtro .....	185
Figura 86 Flujo de enfermedades al año .....	185
Figura 87 Resultado de la edad de atención de emergencia por año .....	186
Figura 88 Pronostico de atenciones de urgencia .....	186
Figura 89 Estimación de atenciones de Urgencia 2020 .....	187
Figura 90 Filtrar la atención por fecha .....	187
Figura 91 Selección de un establecimientos para realizar filtros.....	188
Figura 92 Filtro de campaña del Centro de Salud Familiar de Chépica .....	188
Figura 93 Filtro de atención de enfermedad del Centro de Salud Familiar de Chépica ..	189
Figura 94 El filtro de atención del Centro de Salud Familiar de Chépica por edad .....	189
Figura 95 Pronóstico de atención en el centro de salud familiar Chépica.....	190

## Índice de anexos

Anexo 1: Matriz de operacionalización variable .....	134
Anexo 2: Matriz de consistencia .....	135
Anexo 3: Antecedentes adicionales a la investigación .....	136
Anexo 4: Teorías relacionadas adicionales a la investigación .....	143
Anexo 5: Tabla de puntuación de metodología .....	153
Anexo 6: Tabla de puntuación Herramientas de minería de datos.....	154
Anexo 7: Puntuación de herramientas de visualización de datos .....	155
Anexo 8: Clasificación de metodologías de minería de datos .....	156
Anexo 9: Clasificación de herramientas de minería de datos.....	159
Anexo 10: Clasificación de herramientas de visualización de datos .....	160
Anexo 11: Metodología CRISP-DM .....	161
Anexo 12: Modelo dimensional con variables dummy.....	166
Anexo 13: Diccionario de datos del modelo dimensional .....	167
Anexo 14: Sentencias SQL para ETL.....	171
Anexo 15: Código de RStudio .....	173
Anexo 16: Manual de usuario .....	177



## Índice de abreviaturas

ABREVIATURAS	SIGNIFICADO	PÁGINA
ACF1	Fof errors at lag 1.	p. 85
ANOVA	Analysis of variance	p. 79
BDA	Copyright Bulletin	p. 5
BDS	Boicot, Desinversión y Sanciones	p. 152
BI	Business Intelligence	p. 150
CDS	Interactive compact disc	p. 5
CRISP-DM	Cross Industry Standard Process for Data Mining	p. 2
CS	Case Sensitive	p. 139
DDKAT	Data-Driven Knowledge Acquisition Tool	p. 144
DF	Grados de libertad	p. 79
DM	Data modeling	p. 2
DT	Typical deviation	p. 5
ED	Emergency department	p. 151
EHR	European Hair Research	p. 16
EHRs	European Hair Research Society	p. 14
EMR	Electronic Medical Record	p. 137
EPU SVM	Universal Periodic Review of Support Vector Machine	p. 5
ESR	Socially responsible company	p. 8
ETL	Extract, Transform, and Load	p. 34
FFNN	Feed Forward Neural Network	p. 145
GBT	Gradient Boosted Trees	p. 5
GIS	Geographic information system	p. 116
GP	Geolocation Prediction	p. 140
GPS	Global Positioning System	p. 26
GUI	Graphical user interface	p. 148
ICD-10	International Statistical classification of diseases and related health problems	p. 25
IDE	Integrated Development Environment	p. 23
IG	Geolocated image	p. 139
IG/CS	International Gynecologic Cancer Societyigcs	p. 140
IG/GP	Geolocated image the Case Sensitive	p. 140
INAN	National Institute of Food and Nutrition	p. 137
IoT	Internet of Things	p. 158
KDD	Knowledge Discovery in Databases	p. 19
KEM-DT	Knowledge Engineering Methodology to Produce an Integrated Rules	p. 157
KMS	Key Management Service	p. 144
LM	Linear Model	p. 68
MAE	Mean Absolute Error	p. 85
MAPE	Mean Absolute Percentage Error	p. 7
MASE	Mean Absolute Scaled Error	p. 85
MBD	Mobile big data	p. 136
MD	Direct Message	p. 34
ME	Mean Error	p. 91

ABREVIATURAS	SIGNIFICADO	PÁGINA
MPE	Mean Percentage Error	p. 91
NB	Naive Bayes	p. 5
ODBC	Open Database Connectivity	p. 149
OLAP	On-Line Analytical Processing	p. 136
OLO-Bund	Blound prediction actions	p. 5
OLTP	OnLine Transaction Processing	p. 40
PPDARM	Atmospheric Decontamination and Prevention Plan	p. 139
QV	Quadrifoglio	p. 150
REST	Representational State Transfer	p. 149
RF	Random Forest	p. 5
RMSE	Root Mean Squared Error	p. 91
RSM	Received Sync Message	p. 147
SAP	Systeme Anwendungen und Produkte	p. 150
SAS	Stands for Statistical Analysis System	p. 149
SEMMA	Sample, Explore, Modify, Model, and Assess	p. 130
SMOTE	Synthetic Minority Over-sampling Technique	p. 18
SQL	Structured Query Language	p. 50
STAT	State-of-the-art statistical analysis	p. 26
Sum Sq	Suma de cuadrados	p. 79
SVM	Support Vector Machine	p. 5
TCGA	The Cancer Genome Atlas	p. 17
W2KM	Web 2.0 Knowledge management	p. 145
XML	Extensible Markup Language	p. 148

## Resumen

El objetivo de la investigación fue determinar a través de un nuevo modelo predictivo con la ayuda de la minería de datos y la técnica de predicción de minería de datos de regresión lineal, a través de un sistema de geolocalización para medir los impactos y cambios de atención de urgencia en los hospitales. Mediante una comparación de metodología, herramientas de minería de datos y software de modelos de proyección, se optará la mejor opción en base de su funcionalidad, escalabilidad, características, entre otros. Además, se recolectó datos en el período 2010 - 2019 del área de atención de urgencias. En un estudio transversal, las variables se identifican en un momento dado y se determinan las relaciones entre ellas.

Se concluyó las enfermedades respiratorias son las de mayor impacto las que generan más atenciones de urgencia en los hospitales de acuerdo a los modelos de regresión lineal múltiple y regresión logística ordinal. También se concluyó que la mejor técnica de series de tiempo el Holt-Winter's el cual nos arroja una precisión de 97.3% mostrando las cantidades que se esperan en las atenciones de urgencia del año 2020 en los hospitales.

Se recomienda que se realicen de vuelta con una cantidad de datos mayor para ver el comportamiento de la cantidad de atenciones de urgencia en los hospitales y ver cuál es la enfermedad que es la que más atenciones obtenidas en los siguientes años. Esto agregando más variables tomando otras normativas de clasificación de enfermedades otras técnicas para ver el impacto y saber cuál es la enfermedad más frecuente en los hospitales de cada localidad geográfica también tomar otras técnicas de series de tiempo para encontrar nuevos porcentajes de pronóstico de atenciones de urgencia.

**Palabras clave:** Data mining, diagnostico, técnicas Data mining, Sistema de información geográfica, geolocalización

## **Abstract**

The objective of the research was to determine through a new predictive model with the help of data mining and the predictive technique of linear regression data mining, through a geolocation system to measure the impacts and changes of emergency care in hospitals. Through a comparison of methodology, data mining tools and projection model software, the best option will be chosen based on its functionality, scalability, features, among others. In addition, data was collected in the period 2010 - 2019 from the area of emergency care. In a cross-sectional study, variables are identified at a given time and the relationships between them are determined.

It was concluded that respiratory diseases are the ones that generate more emergency care in hospitals according to multiple linear regression and ordinal logistic regression models. It was also concluded that the best time series technique, Holt-Winter's, gives us an accuracy of 97.3% showing the quantities expected in the emergency care of the year 2020 in hospitals.

It is recommended that they be performed again with a greater amount of data to see the behavior of the amount of emergency care in hospitals and to see which is the disease that is the most attended in the following years. This adds more variables by taking other disease classification norms and other techniques to see the impact and to know which is the most frequent disease in the hospitals of each geographic location.

**Keywords:** Data Mining, Diagnostics, Data Mining Techniques, Geographic Information System, Geolocation

# **I. INTRODUCCIÓN**

En el primer capítulo, la realidad problemática no se han encontrado estudio relacionado a pronóstico de atenciones de urgencia o emergencia en la actualidad que nos ayuden a determinar que enfermedad es la más recurrente que generan más atenciones y pronosticar cantidad de atenciones que pasaran en cada hospital, la justificación que indica que el estudio se realiza para determinar nuevos modelos de pronostico utilizando un estudio de tipo longitudinal de tendencia y correlacional-causal, el problema general no se han encontrado estudios que hagan el pronóstico de la atención de urgencia en los hospitales aplicando técnicas predictivas junto con la metodología CRISP-DM, teniendo como objetivo conseguir nuevos modelos de predicción con mejor exactitud que ayuden a determinar la cantidad de atenciones de urgencia en los hospitales y teniendo hipótesis con porcentajes que buscan ser superados mayores al 90% de exactitud en estudios similares.

La investigación tratara de conocer el impacto de cada enfermedad en las atenciones de urgencia y generar una tendencia para generar una serie de tiempo y pronosticar futuras cantidades de atenciones de urgencia en cada hospital por mes y año. Por otro lado, no hay estudio similar anterior que hable sobre los padecimientos que tiene una atención de urgencia adicionando que enfermedad genera más atenciones de urgencias. Solo hay estudios que se encargan de mencionar como logar que este servicio de atenciones de urgencia no tenga riesgos y se un éxito cuando se le da un paciente lo cual no genere las negligencias medias lo cual solo se conoce modelos de riesgos para evitar lo antes mencionado. Por ende, la investigación busca generar modelos correlacionales-causales y tendencias de tiempo.

En anteriores trabajos, se define que las tecnologías de la información y la comunicación son parte fundamental de las empresas y en el campo de la salud no es una excepción, ya que manejan grandes cantidades de información que se almacena, procesa y vuelve a utilizar como parte de los procesos que se desarrollan en estas instituciones. Teniendo en cuenta lo anterior y debido al creciente uso de la tecnología, en este trabajo se pretende desarrollar un sistema de información de apoyo, que permita hacer un seguimiento de los pacientes que ingresan al área de urgencias.

En los anteriores estudios se no explicaron sobre los pronósticos de cantidad de atenciones de urgencia por hospitales o el impacto que genera cada enfermedad, si no los riesgos que pueden perjudicar la atención, tampoco, cuáles son las enfermedades más comunes en las atenciones de urgencias. Zeng, Jian-Bang Lin y Tian Wang (2017) explicaron:

En este estudio se aplica la gran tecnología de datos en el campo financiero y se construye el nuevo modelo de riesgos en competencia para resolver el problema de las urgencias. El resultado mostró que el rendimiento del modelo de riesgos en competencia es mucho mejor que el modelo de predicción única (p. 988).

En el presente trabajo se realizará un estudio con la base de datos de atenciones de urgencia de Chile desde el año 2010 – 2019 en el cual se tiene como indicadores los padecimientos y las cantidades por mes y año de atenciones de urgencia. Con esta información se podrá determinar los padecimientos o enfermedades que más impacto poseen y nos dan una mayor cantidad de atenciones de urgencia. Luego se podrá determinar con unas técnicas de series de tiempo las tendencias y pronosticar cantidades de atenciones de urgencia en los próximos años.

Las consecuencias de no tener un estudio son las que nos comentan los autores Sidiq y Ahmad Khan (2017) explican que “la minería de datos ha proporcionado diferentes técnicas y herramientas para extraer valiosos patrones ocultos de datos de bases de datos médicas y se utilizan para diagnosticar con mayor precisión las enfermedades de un paciente” (p. 205). Nos ayudaran a determinar con mayor precisión las enfermedades que hay en localidades de acuerdo a la ubicación de los hospitales y cuáles son las que generan más atenciones de urgencias o el paciente porque ingresa constantemente también nos ayudara a tener patrones de tiempo para ver el incremento o descenso de atenciones de urgencia a través de los años para poder para generar pronósticos de cantidad de atenciones de urgencias por mes y años.

En esta parte justificaremos teórica, metodológica, tecnológica y financieramente. Mencionaremos las razones por las que la investigación es

importante, en las que se beneficiaría en los criterios mencionados, y cómo contribuiría nuestra investigación propuesta en los siguientes puntos.

El presente estudio tiene como justificación teórica mencionar el enfoque teórico mencionado a los siguientes autores Berente, Seidel y Safadi (2018) indicaron un enfoque general por el que los investigadores utilizan creativamente enfoques cualitativos según sea necesario, pero lo hacen en conjunción con una variedad de computaciones para triangular y validar las percepciones y conjeturas, lo que resulta en una teorización potencialmente más robusta y creativa. Sin embargo, su labor detectivesca no puede ignorar el conocimiento acumulado de la comunidad de científicos (p.8). En búsqueda de más conocimiento un investigador busca enfoques cualitativos cual él pueda tener percepciones de nuevas soluciones a una investigación potencial el cual pueda solucionar ese vacío de conocimiento que detecta la comunidad científica.

También agregaron Wang, Su, Lee, Hung y Chen (2018) quienes mencionaron: “Los desafíos en la aplicación clínica se han planteado una sobrecarga de información que es necesario combinar grandes cantidades de datos para aumentar la precisión y reducir al mismo tiempo los costos y el tiempo de espera” (p. 8). Por lo tanto, el pequeño tamaño de la muestra puede dar lugar a una sobreestimación de la exactitud de la predicción al sobre ajustarse a los datos. Hasta la fecha, no existe ninguna directriz para determinar un tamaño de muestra justo en el conjunto de la formación para obtener resultados fiables.

El presente estudio tiene una justificación metodológica que nos ayudara para conocer nuestros indicadores y que técnicas se puede aplicar en la minería de datos como mencionan Tobback, Naudts, Daelemans, Fortuny y Martens (2018) explicaron: Sus indicadores respaldan las técnicas aplicadas de minería de textos a un problema relacionado con las políticas y hemos tratado de mejorar el índice de incertidumbre política original construyendo un modelo de clasificación que sustituye las palabras clave autoseleccionadas de la metodología original (p.9).

También comentaron Galicia, Talavera, Troncoso, Koprinska y Martínez (2019) propusieron: “Un método de conjunto que calcula los pesos de cada miembro del conjunto utilizando un método de mínimos cuadrados, asignando



pesos más altos a los miembros del conjunto más precisos en base a su actuación pasada” (p.8). Además, Galicia et. Al (2019) agregaron lo siguiente: “se investigó dos estrategias para actualizar los pesos, dando como resultado un conjunto dinámico y estático. Donde combinar modelos de regresión basados en árboles (DT, GBT y RF), nuestro método puede utilizarse para combinar otro tipo de modelo de predicción” (p.8).

El estudio tiene como justificación tecnológica, el evaluar la base de datos de atenciones de urgencias que han sido recopiladas a lo largo de los años para conocer cómo va el avance de la cantidad de atenciones de urgencias en un recopilatorio de montañas de base de datos. Como explicaron Saggi y Jain (2018): Las siete características más importantes, a saber, Volumen, Velocidad, Variedad, Valencia, Veracidad, Variabilidad y Valor. Se está realizando una gran investigación para definir el BDA en términos de Vs para los desafíos de los datos. Nuestra respuesta a la segunda pregunta de investigación ha proporcionado una visión general de la arquitectura del marco BDA-DM, incluyendo seis componentes, a saber: (i) generación de datos, (ii) adquisición de datos, (iii) almacenamiento de datos, (iv) análisis avanzado de datos, (v) visualización de datos, y (vi) toma de decisiones para la creación de valor. (p.8).

Algoritmo de clasificación como SVM, NB y Decisión Tree considerado para la evaluación del desempeño en la predicción de enfermedades de desorden hepático. Según Kefelegn y Kamat (2018) explico: “La minería de datos es el procedimiento de recuperar un patrón de un gran conjunto de datos en relación con el aprendizaje automático, la base de datos y las estadísticas” (p.8). Por consiguiente, Kefelegn y Kamat (2018) concluyeron: “Una técnica de minería de datos como la agrupación, clasificación y asociación que es apropiada para el diagnóstico médico” (p.8).

Tiene como justificación financiera encontrar que la inclusión del indicador SVM de la EPU mejora significativamente la precisión del pronóstico al predecir el diferencial del OLO-Bund, el rendimiento de los bonos del gobierno a largo plazo, el diferencial de los CDS y la confianza del consumidor a corto plazo. Como indicaron Tobback et al. (2018) “un ejercicio de previsión sobre diez variables macroeconómicas y financieras demuestra que la metodología más avanzada tiene

un mayor poder de predicción y habla a favor del uso de un modelo de clasificación SVM” (p. 9). Por lo tanto, Tobback et al. (2018) en su estudio concluyo: “En promedio, el poder de predicción parece fallar cuando se pronostica con más de dos meses de antelación, y sólo el diferencial de CDS puede predecirse con tres meses de antelación” (p. 9).

También agregaron Bibri y Krogstie (2018) ellos explicaron que, “con un gran análisis de datos, y particularmente el uso y la combinación de diferentes tareas como parte de la minería de datos, seremos capaces de monitorear, entender y analizar mejor las ciudades inteligentes y sostenibles” (p.8). En particular, el uso y la combinación de diferentes tareas son parte de los procesos de minería de datos para utilizar y resolver diversos problemas de decisión relacionados con la sostenibilidad urbana. Además, Bibri y Krogstie (2018) concluyeron: “para ser operadas, administradas, planificadas, desarrolladas y gobernadas más inteligentemente en términos de mejorar y mantener su contribución a la sostenibilidad” (p.8).

Sobre la base de la problemática realidad presentada, el problema general de la investigación fue que no se han encontrado estudios que hagan el pronóstico de la atención de urgencia en los hospitales aplicando técnicas predictivas junto con la metodología CRISP-DM con la que se puede desarrollar un sistema de geolocalización para medir el impacto de la atención hospitalaria. Por consiguiente, no hay resultados que muestren cambios en el aumento o la disminución del número de cuidados urgentes en los hospitales.

El objetivo general se determinó mediante un nuevo modelo de predicción con la ayuda de la extracción de datos y las técnicas de predicción de la extracción de datos de regresión lineal múltiple, la regresión logística ordinal y las series cronológicas, mediante un instrumento de geolocalización para medir los impactos y los cambios de la atención de urgencia en los hospitales. Los objetivos específicos son los siguientes:

- OE1: Determinar un mejor cálculo de pronóstico donde podamos identificar el impacto, la precisión de la atención urgente en los hospitales.

- OE2: Determinar demográficamente qué hospital recibe la atención más urgente renovando anualmente los datos de entrada para identificar con precisión aquellos que reciben más atención que otros.

Se definirá las hipótesis que serán nuestros resultados anteriores o lo que se espera obtener una vez terminada la investigación, que serán rechazadas o aceptadas al final de nuestra investigación según los resultados de la misma.

HE1: La precisión del pronóstico de casos de atenciones de urgencia con la regresión lineal múltiple fue por lo menos 60%.

Huang et al. (2020) explicaron:

El resultado es un deterioro de la función y un mayor riesgo de asma bronquial, EPOC, enfisema y otras enfermedades respiratorias. Para evaluar la eficiencia se utilizó el porcentaje de error absoluto medio (MAPE). Encontramos que el PM<sub>2,5</sub> y el SO<sub>2</sub> son los factores más importantes que contribuyen a mejorar la precisión de la predicción. El modelo de regresión lineal múltiple ponderada localmente (LWLR) basado en el marco del núcleo integrado con el algoritmo K-means demostró un error de predicción mínimo del 9,03% (p.2).

Explique su modelo de agrupación utilizando la regresión lineal múltiple para validar el resultado con mayor riesgo de enfermedades respiratorias, como indica su modelo de predicción en el marco del núcleo integrado con el algoritmo k-means de demostración, un error de predicción de 9,03% (Huang et al., 2020, p. 2).

Según Villani et al. (2017) concluyeron:

Los resultados de este estudio demuestran la necesidad de una planificación adecuada de los servicios de salud y de la supervisión de los recursos en el futuro, con una mayor prevalencia de la diabetes y presiones sobre los servicios médicos de emergencia pre hospitalarios previstos. Como se predijo, la utilización del horizonte de previsión de 1 mes dio lugar a una mayor precisión que el horizonte de previsión de 3 meses y 12 meses; sin embargo, para otros tipos de modelos de series cronológicas, SARIMA (0,1,0,12) (un paso) generó un MAPE de 7. 3%, mejor rendimiento b)

SARIMA + tendencia temporal (MAPE = 9,1%), d) suavizado exponencial (MAPE = 8,8%) y e) tendencia temporal lineal + estacionalidad (MAPE = 7,6%), y fue comparable a c) ARIMA + estacionalidad (MAPE = 7,2%).

Explicaron que su modelo de pronóstico para los casos de aumento de la diabetes en el servicio de urgencias pre hospitalario utilizando diversos métodos para tener la mejor precisión, pero en el caso de la regresión lineal, arrojó un error potencial del 7,6%, presionando sobre una duración de 12 meses y teniendo sus variables agrupadas y utilizando diversos modelos de series temporales (Villani et al., 2017, p.5).

HE2: La precisión del pronóstico de casos de atenciones de urgencia con la regresión logística ordinal fue por lo menos 70%.

Según Innala et al. (2016) concluyeron:

De los pacientes, el 53,2% de los pacientes tenían una o más comorbilidades en la línea de base. En el modelo de regresión logística múltiple, la ESR ( $p = 0,036$ ) en el momento de la inclusión se asoció con una nueva comorbilidad después de 5 años, ajustada según la edad, el sexo, el tabaquismo y el tratamiento de GC. En un modelo similar, Ex-RA ( $p < 0,05$ ) se asoció con una nueva comorbilidad en T5. En un tercer modelo, ajustado por edad y sexo, una nueva comorbilidad pulmonar se asoció con antecedentes de tabaquismo en el momento de la inclusión ( $p < 0,01$ ), pero no con ESR ( $p.1$ ).

Definieron tener una precisión del 53,2% en su regresión logística múltiple ya que esta regresión tomó como una nueva variable la morbilidad que se ajusta a la edad, el sexo, el tabaquismo y el tratamiento, esto se asoció con Ex-Ra y con otra que sólo tenía el tabaquismo como antecedente en el momento de la admisión (Innala et al., 2016, p.1).

Gholipour et al. 2016 explicaron:

Para el modelo de regresión logística, la edad, la historia familiar de DM2, y la residencia fueron los predictores más importantes de DM2, seguidos por el tabaquismo actual, la hipertensión controlada, la hipertensión no

controlada, el IMC, la TC raeisd ( $> 200$  mg / dl) y el género. con una predicción del 85,4% (pp. 773 - 774).

Concluyeron que en un modelo de regresión logística en el que tenía la edad, un historial de T2DM, el tabaquismo actual, la hipertensión controlada y no controlada, el IMC, el Tc raeisd y el género, la predicción era del 85,4% (Gholipour et al., 2016, pp. 773-774).

HE3: La precisión del pronóstico de casos de atenciones de urgencia con la técnica de Holt Winter fue por lo menos 90%.

Según explicaron Zor y Cebi (2018):

los resultados de Holt-Winter, el modelo gris GM (1, 1) y el modelo triangular gris difuso (1, 1) se muestran en la Tabla IX. Para el modelo de Holt - Winters, los resultados de las mediciones de evaluación de MAPE, MAD y MSE son 11,81 por ciento, 7.297,12 y 59.072.686,63, respectivamente. Para el GM (1, 1), los resultados de las mediciones de evaluación de MAPE, MAD y MSE son 8,01 por ciento, 4.933,04 y 29.148.087,33 respectivamente. Para el TFGM (1, 1), son 7,64 por ciento, 4.754,71 y 29.064.897,93, respectivamente. Las tasas de error muestran que la precisión del pronóstico es aceptable (ASM, 8,01 por ciento 10 por ciento y 7,64 por ciento 10 por ciento), y el rendimiento del pronóstico del TFGM (1, 1) es mejor que el método de Holt-Winter y GM (1, 1) en este caso (sf).

Concluyeron que evaluando la misma evaluación para probar quién nos daría el mejor modelo de rendimiento para saber cuál es el mejor en el pronóstico en el que los Holt-Winter evalúan su modelo más fiable en la evaluación del 11,81% de error que perdieron contra su modelo. GM que es el mejor en el pronóstico de rendimiento en el caso demostrado (Zor y cebi, 2018).

Según Hernández et al. (2019) concluyeron:

El análisis de la demanda de cesáreas obtuvo mejores resultados con el modelo de Holt-Winter ( $\alpha = 0,5$ ,  $\beta = 0,05$ , MAPE = 0,119, MAD = 27,8), la combinación de Clean y Six sigma parece ser un enfoque eficaz para aplicar en los procesos de atención de la salud con el fin de identificar, evaluar y

reducir los residuos, así como aumentar la eficiencia y, por tanto, contribuir a la sostenibilidad.

Segundos después del resultado indicado, se obtuvieron buenos resultados ya que se hizo una agrupación para el análisis de la cesárea, pero la técnica de limpieza y seis sigmas se utilizará para identificar los deseos, lo que dio un error de 0,119%, dando una probabilidad muy alta de 99,881%. para su combinación de limpio y seis sigmas (Hernández et al., 2019).

HE4: La precisión del pronóstico de casos de atenciones de urgencia con el método de Naives fue por lo menos 90%.

Según Spreco (2017) concluyó:

Para el estudio, los códigos de diagnóstico de la gripe (Clasificación Internacional de Enfermedades, Décima Revisión (CIE-10) / Conferencia Internacional sobre Drogas y Clasificación Farmacológica (CIPF) J10.0, J10.1, J10.8, J11.0, J11.1, J11. 8 y J11.0-P) y las quejas del jefe de tele-enfermería potencialmente asociadas a la gripe, es decir, fiebre (adulto, niño), tos (adulto, niño), dolor de cabeza (adulto, niño)), disnea, dolor de garganta, vértigo, letargo y síncope se utilizaron. Los algoritmos evaluados, la regresión de Serfling como algoritmo de detección y la regresión logarítmica no adaptativa como algoritmo de predicción, mostraron un rendimiento satisfactorio cuando se aplicaron a los datos de diagnóstico durante las temporadas de gripe de invierno en un entorno de salud pública local. También se mostró un rendimiento aceptable para las predicciones de 2 semanas para el método ingenuo ( $r = 0 - 74$  y  $MAPE = 0 - 17$ ). Para el brote pandémico de 2009, los algoritmos produjeron correlaciones más bajas; la correlación más alta fue  $= 0 - 52$  para el método ingenuo para las predicciones de 2 semanas de antelación (p. 2173).

Concluyo que utiliza el método de navegación para hacer una predicción diagnóstica de enfermedades correlativas a la CIE-10 que influye en las enfermedades respiratorias en 2 semanas predijo que tendrían un margen de error del 0,17% y una fiabilidad del 99,83% (Spreco, 2017, p. 2173).

Chand, Rao, Tekabu y Khan (2018) concluyeron:

Las estadísticas de medición del error del modelo LR (1) y el método del Pronóstico Ingenuo. El resultado muestra que el modelo óptimo, LR (1), supera al método de previsión ingenua, ya que el modelo LR (1) produce pequeños valores de error con respecto al método de previsión ingenua en el conjunto de datos de formación, lo que da lugar a que se sigan MAPE Naive 23,83 y LR 21,73 (p. 288).

Se definió que para los datos de capacitación el margen de error con los modelos del modelo LR y el pronóstico ingenuo tenía un error estándar de 21,73% y 23,83% respectivamente, luego se definió que el método LR es mejor que el pronóstico ingenuo (Chand, Rao, Tekabu y Khan, 2018, p. 288).

HE5: La precisión del pronóstico de casos de atenciones de urgencia con el método de tendencia exponencial suavizada fue por lo menos 90%.

Zang, et al. (2020) concluyeron:

Las elevadas tasas de registro de OPA activas en marzo, abril, mayo y junio mostraron variaciones estacionales. El modelo de multiplicación exponencial de invierno suavizado fue seleccionado como el mejor modelo. Los valores pronosticados fueron aproximadamente coherentes con los observados entre 2017 y 2019. Se pronosticaron tasas de registro de PPT activas anuales de 49,1 (IC del 95%: 36,2-62,0) y 34,4 (IC del 95%: 18,6-50,2) por cada 100.000 poblaciones en 2020 y 2025, respectivamente. En comparación con la tasa de registro de PPT activa en 2015, la reducción alcanzará el 23,7% (IC 95%: 3,2% -44,1%) y el 46,8% (IC 95%: 21,4% - 72,1%) en 2020 y 2025, respectivamente (p. 22).

Los autores dijeron que el exponencial predice por la cantidad de población a hacer registros de PTB que van en la variación de las variables de la estación a medida que se registran los datos, la precisión disminuirá hasta llegar al 23,7% terminado en un modelo que tiene una estimación de 5 años (Zang, et al., 2020, p.22).

Según Ke et al. (2016) concluyeron:

Los resultados revelaron que los NIF en China han sido continuos desde 2006 hasta 2015, y la tendencia de variación podría dividirse en tres secciones; los casos de incidencia disminuyeron notablemente de 16129 a 9203 de 2006 a 2009, con un cambio porcentual de - 42,94%; aumentaron a 13918 a partir de 2012, con un cambio porcentual de 51,23% en relación con el de 2009 (p. 4).

Como se ha indicado, el método exponencial se utiliza para ver la variación porcentual de los datos a lo largo del tiempo, la estimación de que durante los años 2006 a 2009, se predijo un aumento que sería la reducción para 2012, que logra una precisión del 51,23% y que refleja un aumento de 13918 con la enfermedad HFRS (Ke, 2016, p. 4).



## **II. MARCO TEÓRICO**

En este capítulo se hablará sobre los antecedentes donde mencionaremos los trabajos de investigación donde se menciona un tema similar o relacionado para poder llevar a cabo nuestra discusión después de realizar nuestra investigación y comparar con los resultados de diferentes autores que tocaron el mismo tema o compartieron una relación en el mismo campo de estudio con el fin de tener una base para comparar los resultados. También hablaremos de los trabajos relacionados que nos ayudan como base a las definiciones necesarias en todo el conocimiento necesario para realizar la investigación y así poder relacionar los términos de minería de datos y atenciones de urgencia, que nos ayudarán la definición de las herramientas necesarias para desarrollar las nuevas propuestas que se harán en esta investigación.

Se buscaron anteriores trabajos en el campo de la medicina realizadas con la investigación de predicción de atenciones de urgencias, para realizar comparación o nos ayuden a sustentar las variables y metodologías que se utilizaran en la investigación poder buscar un nuevo enfoque que nos den mejores resultados en el desarrollo de la nueva predicción de cantidad de atenciones de urgencia.

Livieris, Kotsilieris, Dimopoulos y Pintelas (2018) indicaron: "Sistema de apoyo a la decisión de fácil manejo para la predicción de hospitalizados que incorpora un clasificador de aprendizaje automático de dos niveles. La técnica de clasificación propuesta presenta una mayor precisión de clasificación en comparación con algunos de los algoritmos de clasificación individual más populares y comúnmente utilizados" (p.10). Luego, Livieris et al. (2018) también indicaron: "Las ventajas significativas del programa informático presentado son el empleo de una interfaz sencilla y fácil de usar, su escalabilidad debido a su carácter modular de diseño y aplicación y su neutralidad del sistema operativo". (p.10). Por último, Livieris et al. (2018) señalaron: "servir de referencia para la adopción de decisiones en el proceso de admisión y reforzar el sistema de servicios en los hospitales ofreciendo una asistencia personalizada en función del tiempo de hospitalización previsto de los pacientes". (p.10).

Islam, Hasan, Wang y Germac (2018) exhibieron: "El desarrollo de un marco de toma de decisiones informadas se deriva de la creciente preocupación por

garantizar un sistema de atención de la salud de alto valor y centrado en el paciente. Al mismo tiempo, la disponibilidad de grandes datos sobre la atención de la salud ha creado una vía de investigación prometedora para los académicos y los profesionales". (p.35). Además de Islam et al. (2018) señalaron: "el aumento del número de publicaciones en los últimos años corrobora la importancia de los análisis de la atención de la salud para construir mejores sistemas de atención de la salud en todo el mundo. El objetivo es facilitar sistemas de atención de salud coordinados y bien informados capaces de garantizar la máxima satisfacción del paciente" (p. 35).

Chen (2019) explicó: "Utilizó el área de tecnología de minería de datos para analizar las reglas de medicación de enfermedades pediátricas comunes, y ejemplificó el estudio de casos de tos en niños". (p.11). Además, Chen (2019) indicó: "El campo de la medicina china también está utilizando cada vez más este método para promover su propio desarrollo. Existen diversos métodos de explotación minera, y el contenido de la investigación también se centra en ellos. Cada vez se utilizan más resultados mineros útiles para orientar la clínica". (p.11).

Huang, Zhao, Zhou y Jiang (2018) indicaron: "La previsión del índice de calidad de la salud cerebral inteligente basada en la tecnología de la información popular, y propone un espacio vacío basado en el análisis de grandes datos. El modelo de predicción de calidad tiene cierta viabilidad a partir de la verificación experimental". (p.10). Además, Huang et al. (2018) Explicaron: El modelo de vigilancia predictiva propuesto en este documento combina grandes datos, la Internet de las Cosas y dispositivos inteligentes, de modo que los usuarios puedan detectar y predecir fácilmente las condiciones de calidad de la salud cerebral, mejorar los métodos tradicionales de prueba de calidad de la salud cerebral única y hacer más exactos los recibos de detección. (p.10).

Wang, Kung y Byrd (2018) explicaron: "La gobernanza de los datos es una extensión de la gobernanza de la tecnología de la información que se centra en el aprovechamiento de los recursos de datos de toda la empresa para crear valor comercial. De hecho, el gran análisis de datos es un arma de doble filo para la inversión en TI, lo que puede suponer una enorme carga financiera para las organizaciones sanitarias". (p.8). De modo que, Wang et al. (2018) mencionaron:

Por otra parte, con una gobernanza de datos apropiada, el análisis de grandes datos tiene el potencial de equipar a las organizaciones para aprovechar las montañas de datos, información y conocimientos heterogéneos de una compleja serie de aplicaciones internas (por ejemplo, EHR de pacientes hospitalizados y ambulatorios) y aplicaciones de las redes de atención de salud (por ejemplo, sistemas de información de laboratorios y farmacias)". (p.8). Wang et al. (2018) concluyeron: "El éxito en el gobierno de los datos requiere una serie de cambios organizativos en los procesos empresariales, ya que todos los datos deben ser bien comprendidos, confiables, accesibles y seguros en un entorno impulsado por los datos". (p. 8).

Kumar (2018) indico: "Hay numerosas investigaciones de diversos ámbitos que trabajan continuamente para desarrollar Predice el flujo de pacientes en el sector de la atención de la salud. El objetivo de esta encuesta era resumir las investigaciones recientes y sus deméritos en la predicción del flujo de pacientes". (p. 5). Entonces, Kumar (2018) definió: "Este trabajo da los méritos y deméritos de las técnicas recientes y se estudian sus capacidades. Este trabajo concluye que no hay ningún método de predicción efectivo que no se aplique y se concentra en la predicción del flujo de pacientes con gran precisión". (p. 5).

Gepp, Linnenluecke, O'Neill y Smith (2018) mencionados: "Salir de los orígenes de las técnicas de grandes datos en la literatura estadística multivariante y luego categorizar la contabilidad de grandes datos y la investigación financiera en varias agrupaciones de investigación". (p. 26). A continuación, Gepp et al. (2018): "Su análisis muestra que, además de la auditoría, hay documentos influyentes sobre la modelización de las dificultades financieras, la modelización del fraude financiero y la predicción del mercado de valores y la modelización cuantitativa". (p. 26). Gepp et al. (2018) concluyeron: La auditoría se beneficiaría enormemente si se adoptara el uso de grandes técnicas de datos, independientemente de que las empresas clientes las utilicen o no. Las conclusiones de la investigación contable y financiera sugieren combinar múltiples modelos de grandes datos en lugar de aplicar un modelo individual, y utilizar modelos de grandes datos para complementar a los expertos humanos. (p.27).

Chen, Li, Rong, Bilal, Yang y Li (2018) explicaron: "El sistema propuesto primero agrupa los informes de inspección para obtener los síntomas de la enfermedad como centros de agrupación. Luego, identifica fuertes vínculos entre las enfermedades y los tratamientos aplicando un algoritmo de análisis de asociación" (p. 28). Así mismo, Chen et al. (2018) indicaron: "El documento proporcionó además un análisis cualitativo del sistema propuesto demostrando cómo un paciente y un médico obtienen recomendaciones de tratamiento basadas en los informes de inspección del paciente". (p.28). Por lo tanto, Chen et al. (2018) concluyeron: "Los resultados experimentales y de aplicación indicaron que el sistema logra efectivamente los objetivos requeridos y proporciona recomendaciones de alta calidad con una respuesta de baja latencia" (p. 28).

Ghaibeh, Kasem, Ng, Nair, Hirose y Thiruchelvam (2018) explicaron: Árboles de decisión para analizar los datos retrospectivos en un intento de identificar las tendencias y patrones relacionados con la recurrencia de las CPN. Curiosamente, el modelo demuestra varios hallazgos que se corresponden con los conocimientos médicos actuales (por ejemplo, la dosis de radioterapia, el efecto del retraso del tratamiento y la edad del paciente). (p.5).

Modhukur et al. (2018) concluyeron: "Creemos que nuestra herramienta será un recurso inestimable para la comunidad científica que trabaja en la evaluación de los biomarcadores del cáncer. MethSurv se mantendrá y actualizará continuamente para abarcar los nuevos datos de metilación disponibles en el TCGA" (p.9). Explicaron Modhukur et al. (2018): A medida que la cantidad de datos sobre el metiloma del cáncer crece constantemente, se necesitan herramientas bioinformáticas que faciliten la extracción y visualización de la información. Hasta donde sabemos, MethSurv es la primera herramienta de la web para evaluar el potencial de pronóstico de los candidatos a biomarcadores de metilación de ADN mediante el análisis de supervivencia multivariable" (p. 9).

Mirza, Mittal y Zaman (2018) explicaron: "Como ha habido una mejora sustancial en los sistemas expertos y las herramientas de aprendizaje automático, su efecto ha invadido cada vez más dominios de aplicación con cada día que pasa y el campo de la medicina no es una excepción. La toma de decisiones en el campo de la medicina a veces puede ser muy problemática" (p. 5). Mirza et al. (2018)

decidieron: "Los sistemas de clasificación utilizados para la toma de decisiones médicas están provistos de datos médicos que examinan de forma más exhaustiva, pero en un plazo más corto. En este estudio de investigación propusimos un sistema que se basaba en el árbol de decisión y el SMOTE"(p.5). Mirza et al. (2018) concluyeron: "aplicaron este sistema en el diagnóstico de la diabetes y se evaluó uno de los mecanismos de aprendizaje precisos. El estudio sugiere firmemente que la reducción del desequilibrio de clases puede resultar en una mejora de la tasa de predicción"(p.5).

Así, Kaur y Sharma (2018) definieron: "Parece que las técnicas híbridas son más eficaces y precisas en lo que respecta a la tasa de precisión para el diagnóstico de la diabetes. La utilización de técnicas de pre procesamiento de datos puede mejorar aún más la tasa de exactitud de las predicciones" (p. 18). Definición de Kaur y Sharma (2018): Por lo tanto, es necesario prestar atención al desarrollo de un sistema de diagnóstico de la diabetes inteligente e híbrido que utilice el aprendizaje por máquina, la informática blanda, la Internet de las cosas, las ontologías y la teoría de la información para despertar y salvar a las masas humanas de un amplio espectro decisivo de este trastorno humano más mortífero. (p.18).

Mediktor con respecto al diagnóstico final de los pacientes fue del 76,5%, superior al publicado en programas similares (Arancibia et al., 2019). Los resultados de los diagnósticos sugeridos por Mediktor con respecto al diagnóstico final de los pacientes fue 76,5%, superior al publicado en programas similares. Su precisión, como la de otros programas comparables, puede mejorarse con la retroalimentación de los diagnósticos finales reales y los comentarios médicos. (p.5).

El-Jerjawi y Abu-Naser (2018) explicaron: "La red neural artificial fue utilizada para predecir la diabetes. Usando el modelo de redes neuronales artificiales podemos diseñar e implementar procesos médicos complejos usando software". (p. 9). En consecuencia, El\_Jerjawi y Abu-Naser (2018) comentan: "Los sistemas de software son más eficaces y eficientes en varios campos de la medicina, incluyendo la predicción, el diagnóstico, el tratamiento y la ayuda a los cirujanos, los médicos y la población en general. Estos sistemas pueden aplicarse de manera paralela y

se distribuyen en diferentes medidas" (p. 9). Concluyeron El-Jerjawi y Abu-Naser (2018) concluyeron: "La red neuronal artificial es un sistema de procesamiento paralelo que se utiliza para detectar patrones complejos en los datos". (p. 9).

Amirabadizadeh, Nezami, Vaughn, Nakhaee y Mehrpour (2018) concluyeron: "el modelo de árbol de decisión ofrece una mayor exactitud de predicción que la regresión logística. Esto demuestra que el modelo de árbol de decisión puede ser más útil para los médicos en la identificación y selección de adictos". (p. 10), por lo tanto, Amirabadizadeh et al. (2018): "Se recomienda utilizar el árbol de decisiones para identificar los factores asociados con el uso de drogas en la literatura. En el presente estudio, los factores más importantes para identificar los opiáceos naturales y el uso de drogas a base de heroína fueron la primera sustancia y la edad" (p. 10).

Se explicará todos los temas aplicados en la investigación definición sobre la minería de datos, modelamiento, implementación y técnicas, también se explicará que es una atención de urgencia y un sistema de geolocalización que es lo que se tendrá como producto en la investigación esto se detallara con las opiniones de diferentes autores que ayudarán a argumentar el significado y sus componentes que es el proceso principal de nuestra investigación.

Minería de datos es uno de los procesos más conocidos para la extracción de conocimientos en un área de exploración en el manejo de la información como la agrupación, la regresión y la técnica de agrupación utilizado para extraer, transformar y analizar una gran base de datos. Según Zulkurnain y Shah (2017) explicaron: "La minería de datos es uno de los pasos bien conocidos del proceso de KDD y es un área de investigación centrada debido a su importancia" (p. 1). La minería de datos busca explicar una variable utilizando métodos, para Zulkurnain y Shah (2017) explicaron "La suma de tareas como la clasificación, la agrupación, la regresión y la regla de asociación en la minería de datos se utiliza para extraer conocimientos de grandes bases de datos" (p. 1).

La minería de datos es un tema global en la informática y la estadística durante su aparición. Los cálculos de prueba y evolución son subcapas de las estructuras evolutivas que estudian colectivamente los Grandes Datos en un grupo

de diferentes tipos de elementos. Además, la evolución de la minería de datos aportó mucho al aspecto académico de la informática y estadística según explicaron Tan y Shi (2016) la minería de datos ha sido un tema académico popular en la informática y la estadística durante décadas. La inteligencia de enjambre y los algoritmos evolutivos son subcapas de las técnicas de computación evolutiva que estudian la inteligencia colectiva de la Computación Evolutiva y Big Data en un grupo de individuos simples (pp. 8-9).

El gran almacén de datos para aplicar la minería de datos puede ser muy útil para extraer conjuntos de datos específicos de estudios más grandes dentro de un gran almacén de información recopilada a través de los años. Entonces para mejorar ese procedimiento, explica un poco que todas técnicas se pueden aplicar si se implementa en el estudio la metodología CRISP-DM que según los siguientes autores Huber, Wiemer, Schneider e Ihlenfeldt (2019) propusieron: "En el proyecto se utilizó el Detact® del Tablero de mandos de la IO, que proporciona las funciones necesarias desde la conexión de las fuentes de datos y el análisis de los datos hasta la visualización de los resultados" (p. 405). Esto hizo énfasis que se visualizará un mapeo de cómo está compuesta el desarrollo de la metodología definiendo las siguientes fases: Comprensión de los negocios, comprensión de los datos y preparación de los datos, modelización, evaluación y despliegue los resultados oscilaron entre el 90 % y el 100 % de precisión de clasificación para los diferentes rangos de frecuencia.

Se explicarán todas las técnicas que se pueden aplicar a nuestra investigación, para crear los nuevos modelos de predicción de atención de emergencias, que serán nuestra fuente de lo que vamos a obtener con nuestras técnicas de predicción.

En el presente estudio se utilizó como primer método la técnica de regresión lineal la cual es una técnica de predicción dándole valores a las variables teniendo una dependiente y muchas dependientes para ver la variación y así poder generar una variable de predicción según como explica Field (2009) expone que extender esta relación básica para responder a una pregunta como si faltan 10 minutos para que alguien tenga que dar una charla, ¿cuán ansioso estará? Es la esencia del análisis de regresión: ajustamos un modelo a nuestros datos y lo usamos para



predecir los valores de la Variación dependiente (VD) de una o más variables independientes (IV). El análisis de regresión es una forma de predecir una variable de resultado a partir de una variable predictor (regresión simple) o de varias variables predictores (regresión múltiple) (p.230).

Como segundo método se aplicó el Holt-Winter's es una mejora de la suavización exponencial simple ya que esta detalla la tendencia, estacionalidad y constancia que esta evalúa con puntos medios que son dados de acuerdo al nivel de precisión que se quiere dar para el pronóstico de una serie temporal. También puede utilizarse para comparar entre modelos de suavización exponencial. Además, según Tanyavutti y Tanlamai explican que “el alisamiento exponencial de Holt-Winter's mejora la técnica básica de alisamiento exponencial con más capacidad para apoyar las tendencias y los datos estacionales” (2018, p. 2).

Por lo tanto, “la técnica aplica tres fórmulas de suavizado exponencial a la serie. En primer lugar, se suaviza el nivel (o la media) para obtener un valor medio local de la serie; en segundo lugar, se suaviza la tendencia; y, por último, se suaviza cada sub serie estacional” (Tanyavutti, 2018, p. 2). Además, existes dos técnicas additive y multiplicative, “Fórmulas de alisamiento exponencial que se aplican a una serie con tendencia y componente estacional constante utilizando la técnica aditiva de Holt-Winter” (Tanyavutti, 2018, p. 2).

Como tercer método se aplicó la regresión ordinal logística como la forma exacta de la ecuación puede disponerse de varias maneras, pero la versión que he elegido expresa la ecuación en términos de la probabilidad de que ocurra Y (es decir, la probabilidad de que un caso pertenezca a una determinada categoría). El valor resultante de la ecuación, por lo tanto, varía entre 0 y 1. Un valor cercano a 0 significa que es muy improbable que, Y se haya producido, y un valor cercano a 1 significa que es muy probable que Y se haya producido. Según Field utilizo:

Para medir el impacto de las variables se utilizará una técnica de regresión logística, la cual explica que el modelo de regresión logística predice la probabilidad de que un evento ocurra - anillo para una persona dada (podríamos denotar esto como  $P(Y_i)$  la probabilidad de que Y ocurra para la  $i$ -ésima persona), basado en observaciones de si el evento ocurrió o no para

esa persona (podríamos denotar esto como  $Y_i$ , el resultado real para la  $i$ -ésima persona). Así pues, para una persona determinada,  $Y$  será 0 (el resultado no ocurrió) o 1 (el resultado sí ocurrió), y el valor predicho,  $P(Y)$ , será un valor entre 0 (no hay posibilidad de que el resultado ocurra) y 1 (el resultado ciertamente ocurrirá) (2009, p. 230).

El cuarto método Naive es una técnica ingenua que busca sólo la predicción, pero con una constante igual para predecir los cambios futuros, por otra parte, si se añade la estacionalidad aparte puede indicarse como algo externo y funcionaría con medios o niveles es un algoritmo muy común en las series temporales simples, pero puede utilizarse de otras maneras. Agregando que el método naive, según Messias et. al., explicando: “el método ingenuo es muy simple porque se basa en el supuesto de que el último valor observado se producirá en el siguiente intervalo de tiempo. Este método requiere que se aplique un único punto de la serie temporal” (2016, p. 4).

Como quinto método la exponencial tendencia suavizado que evalúa el promedio ponderado con nivel y tendencia pertenece a las técnicas de alisamiento exponencial ya que es similar al método predictivo sin tendencia con la única variante que se aplica con promedio para cada predicción y busca la mejora constante en la estimación para poder obtener mejores resultados ya que este funciona de la forma de mejora bajo. Así mismo se menciona que el Exponencial de tendencia suavizado según Jain y Mallick definieron:

“Cuando la serie temporal predijo el método de suavizado exponencial (ETS), Taylor estableció la auto correlación de primer orden en los residuos. Cada método de alisamiento exponencial (ETS) se relaciona de manera equivalente con uno o más modelos estocásticos. También sigue la propiedad de robustez y se trata como un modelo de extrapolación crítica.” (2017, p. 5).

Las herramientas de análisis de datos que sirven para la gestión de los datos y nos permitirán hacer nuestro Data Mining para nuestra investigación y poder aplicar todas las técnicas antes mencionadas para relacionar nuestro nuevo modelo de predicción. En el presente estudio se utilizó el RStudio Server le permite

proporcionar una interfaz basada en un navegador (el IDE de RStudio) a una versión de R que se ejecuta en un servidor Linux remoto. Desplegar R y RStudio en un servidor tiene algunos beneficios, entre ellos:

- La capacidad de acceder a las sesiones de R desde cualquier computadora en cualquier lugar.
- El fácil intercambio de códigos, datos y otros archivos con los colegas.
- Permitir que múltiples usuarios compartan el acceso a los recursos de computación más poderosos (memoria, procesadores, etc.) disponibles en un servidor bien equipado; y
- Instalación y configuración centralizada de los paquetes R, R, TeX y otras bibliotecas de apoyo. (Rstudio, 2016, p.4).

Se define que es un sistema de geolocalización para poder visualizar una información gráfica que es lo que mostrará, que se llama SIG que podrá ser detallada para la fase de muestra donde se demostrará de manera visual el resultado de toda la investigación. Indicando que Adeyemi, Markus, Gbolahan y Edeki (2018) detallo: Las aplicaciones de la información geoespacial en la solución de los problemas de gestión de los desechos contribuirían en gran medida a mejorar la toma de decisiones cualitativas y sanitarias adecuadas en nuestras sociedades, salvaguardando la salud humana mediante la prevención de enfermedades y dolencias". (p. 36). Además, Adeyemi et al. (2018): "una adecuada sensibilización del público y una planificación eficaz por parte de las autoridades competentes aportarían soluciones al vertido ilegal de desechos en lugares equivocados" (p. 36).

Definición de cuánto se considera atención de urgencia en los hospitales para tener la teoría y definición de lo que se evaluará, ya que la investigación consiste en conocer una previsión de la cantidad de atención de urgencia. Como comentan Silva, Pessoa y Menezes (2016) indicaron: "Como problema de salud emergente, la comprensión de los traumas en la población de edad avanzada permite al equipo multiprofesional, del que forma parte la enfermería, planificar y aplicar estrategias para una atención geriátrica más específica y, por tanto"(p. 2). Además, Silva et al. (2016) explicaron: "contribuir a la reducción de las posibilidades

de secuelas temporales o permanentes, así como a la prevención de esos acontecimientos, con miras a equilibrar el sistema de salud en su conjunto" (p. 2).

Para poder entender que es una atención de urgencia se debe ver que es una atención de emergencia ya que estos son niveles de clasificación de acuerdo a un nivel crítico de la complejidad de la atención de salud que se debe de tener los niveles definidos según los servicios que brinda el centro de salud, Además Cassiani y De Bortoli (2014) indicaron: En los servicios de salud de emergencia, la enfermería desempeña un papel fundamental para lograr el acceso universal a esos servicios aplicando sus conocimientos y fundamentos en el uso del razonamiento crítico y reflexivo sobre esta atención, teniendo en cuenta su presencia en los diferentes niveles de complejidad de la atención de salud (p. 3).

Comprendió que son las atenciones de en un centro salud es necesitamos definir los niveles para poder saber en que nivel se encuentra una atención de urgencia la cual es una atención de nivel II que son con dolores intensos como son las situaciones de riesgo vital y que el paciente este inestable También existen otros niveles de atención en un centro de salud las cuales según Vásquez, Muñoz y Ramos (2019) definieron los niveles de triaje en lo siguiente:

- Nivel I: prioridad absoluta con atención inmediata y sin demora.
- Nivel II: situaciones muy urgentes de riesgo vital, inestabilidad o dolor muy intenso. Retraso de la asistencia médica de hasta 10 minutos.
- Nivel III: urgentes, pero hemodinámicamente estables, con riesgo vital potencial que probablemente requiera pruebas diagnósticas y/o terapéuticas. Retraso de atención de 60 minutos.
- Nivel IV: urgencia menor, potencialmente sin riesgo vital para el paciente. Retraso máximo de 120 minutos.
- Nivel V: Sin urgencia. Poca complejidad de la patología o asuntos administrativos, citas, etc. Retraso de hasta 240 minutos (p. 16).

Tabla 1

*Niveles de triaje*

<b>Nivel</b>	<b>Color</b>	<b>Categoría</b>	<b>Tiempo de atención</b>
<b>I</b>	Azul	Reanimación	Inmediato
<b>II</b>	Rojo	Emergencia	10 minutos
<b>III</b>	Naranja	Urgente	60 minutos
<b>IV</b>	Verde	Menos urgente	120 minutos
<b>V</b>	Negro	No urgente	240 minutos

Para clasificar los padecimientos encontrados en la base de datos de atenciones de urgencias se utilizará la clasificación internacional de enfermedades (CIE-10), según la Organización Mundial de la Salud (2015) define:

Una clasificación de enfermedades puede definirse como un sistema de categorías a las que se asignan entidades mórbidas según criterios establecidos. La finalidad de la CIE es permitir el registro, el análisis, la interpretación y la comparación sistemáticos de los datos de mortalidad y morbilidad recogidos en diferentes países o zonas y en diferentes momentos (p. 3).

Explica la definición de lo que es un pronóstico o lo que debe evaluarse para definir qué es la atención de emergencia o qué características o indicadores debe tener este pronóstico para conocer el impacto. En el caso de un individuo con un determinado estado de salud, un modelo de pronóstico convierte la asociación de valores de predicción en una evaluación de la probabilidad de examinar un punto final específico de la atención de emergencia. Steyerberg et al. (2013) afirmaron: “el modelo de pronóstico es una combinación formal de múltiples predictores a partir de los cuales se pueden calcular los riesgos de un criterio de valoración específico para cada paciente” (p. 1). Por lo tanto, Steyerberg et al. (2013) definieron: “Otros nombres para un modelo de pronóstico incluyen índice o regla de pronóstico (o de predicción), modelo de predicción de riesgo (o clínico) y modelo de predicción” (p. 1).

Definición de geolocalización, es uno de los principales complementos de la utilización de un GIS, por lo que es necesario mostrar los resultados de nuestra

investigación para ver el impacto del hospital con la mayor cantidad de cuidados de atenciones de urgencia. Como indicaron Luna, Sapio, Sawyer y Bloechl (2019) explicaron: "El Sistema de Posicionamiento Global (GPS) es un sistema de navegación basado en satélites que proporciona información sobre la ubicación y la hora en cualquier lugar de la Tierra o cerca de ella en el que haya una línea de visión sin obstáculos hacia cuatro o más satélites GPS" (p.13). Además, también menciona que para un sistema de posicionamiento según Luna et al. (2019): "El GPS fue creado y es mantenido por el gobierno de los Estados Unidos. El GPS es de libre acceso para cualquier persona con un receptor GPS. Entre los muchos dispositivos que utilizan el GPS están los aviones" (p.13).

### **III. MÉTODO**

En este capítulo mostraremos el tipo, diseño de nuestra investigación además escoger la población y muestras para hacer del desarrollo del proyecto y los aspectos éticos. También marcaremos las fórmulas que permiten el desarrollo de nuestro proyecto de la misma manera que especificaremos cómo se ve afectada cada variable y su relación mostrará diferentes resultados en este estudio.

### **3.1 Tipo y diseño de investigación**

La investigación es la búsqueda de la aplicación o la utilización de conocimientos teniendo en cuenta que, a partir de sus resultados y descubrimientos, pueden surgir nuevos productos y avances científicos.

La investigación aplicada referido al uso del conocimiento en la práctica, según el autor Vargas (2009) explica: “como la utilización de los conocimientos en la práctica, para aplicarlos en provecho de los grupos que participan en esos procesos y en la sociedad en general, además del bagaje de nuevos conocimientos que enriquecen la disciplina” (p. 159). Además, se mencionan dos tipos de investigación, “en las ciencias puras y la investigación básica se busca indagar cómo funcionan las cosas para un uso posterior, mientras en las ciencias prácticas la investigación aplicada tiene como propósito hacer un uso inmediato del conocimiento existente” (Vargas, 2009, p. 159)

Enfoque cuantitativo es la elección del método o método que llevará a obtener de la investigación resultados válidos que respondan a los objetivos inicialmente planteados. Se caracteriza por “refleja la necesidad de medir y estimar magnitudes de los fenómenos o problemas de investigación; además, el investigador plantea un problema de estudio delimitado y concreto sobre el fenómeno. Sus preguntas de investigación versan sobre cuestiones específicas” (Hernández, Fernández, & Baptista, 2014, p. 5).

El diseño de la investigación es no experimental, de tendencia longitudinal. En esta investigación se seleccionará los datos para un rango temporal 2010-2019 del área de atención de urgencias, considerada adecuada para el análisis y desarrollo de un nuevo modelo de pronóstico de la información seleccionada, tal y como indican Hernández, Fernández y Baptista (2014): "Es longitudinal, cuando el interés del investigador es analizar los cambios a lo largo del tiempo en determinadas variables



o en las relaciones entre ellas” y Además Hernández, Fernández y Baptista (2014) agregaron: “el tipo de tendencia porque indica un cambio a lo largo del tiempo seleccionado en la variable de forma significativa”.

Así mismo, se utilizó el diseño correlacional causal porque se estudiaron las relaciones entre las variables en una determinada fecha o periodo de un lugar en específico, “porque describe las relaciones entre las variables en un momento determinado en función de la relación causa-efecto. Es importante precisar que, la causalidad implica correlación, pero no toda correlación significa causalidad” (Hernández, 2014, p.157).

### **3.2 Variables y operacionalización**

El impacto o efecto de la utilización de la previsión, tiene importantes implicaciones para los administradores, ya que proporcionan orientación sobre cómo aprovechar las diferentes variables que se quieren prever. Para Xu, Gao, Daneshmand, Wang y Liu (2017) expuesta:

La precisión de la predicción de la geolocalización aprovechando la movilidad de los grandes datos. El objetivo final es proporcionar un buen GBS (servicios basados en la geolocalización) a los usuarios en cualquier momento y en cualquier lugar. Como tal, la cuestión esencial es cómo extraer la información oculta de los usuarios mediante el análisis y el procesamiento de los grandes datos de movilidad y cómo hacer uso de estos recursos para retroalimentar a los usuarios (p.9).

Según Ñaupas et al. (2014), "la operacionalización de las variables es un procedimiento lógico que consiste en transformar las variables teóricas en variables intermedias, luego éstas en variables o indicadores empíricos y finalmente se producen los índices" (p. 191). Ver anexo 1.

### **3.3 Población, muestra y muestreo**

A continuación, se detallan los conceptos relacionados con la población, la muestra, el muestreo y la unidad de análisis:

- A. Población: Está constituida por un grupo de personas o elementos que tienen características comunes para esta investigación nuestra población refleja las causas que causan una atención urgente. Por lo tanto, la población estará constituida por la base de datos, que cuenta con una estimación de 33 millones de registros de atención de urgencia recogidos entre el año 2010 y 2019. Hernández et al. (2014) indicaron que la población es la "que se va a estudiar y sobre la que se van a generalizar los resultados" (p. 174).
- B. Muestra: Son los elementos de una población; se explica el tamaño de la muestra y cómo se determinó. Se realizó la selección de una región de Chile como muestra de la cantidad de registros de atención de emergencia. Según indicaron Ñaupas et al. (2014): "la muestra es el subconjunto, o parte del universo o de la población, seleccionado por diversos métodos, pero siempre teniendo en cuenta la representatividad del universo" (p. 246).

### **3.4 Técnicas e instrumentos de recolección de datos**

Los datos secundarios, son datos ya recopilados y fácilmente disponibles de otras fuentes cuya información es data real extraída, por lo tanto, para la investigación se recurrió al DEIS de atenciones de urgencias de Chile. Definición de González Río (1997): "Los datos secundarios aluden a la información que no ha sido producida explícitamente para los objetivos de la investigación, pero que preexiste y que el investigador hace rentable" (Citado en Pulido, 2015, p. 244).

Validez, significa que un instrumento de encuesta o estadístico mide lo que tiene que medir. Según indicaron Ñaupas et al. (2014):

La validez es la pertinencia de un instrumento de medición, se refiere a la exactitud con que el instrumento mide lo que se propone medir, en otras palabras, y es la eficacia de un instrumento para representar o describir el atributo de interés para el examinador (p. 215).

En cuanto a la validez del contenido, Naúpas et al. (2014) señalaron que se refiere al "grado en que un instrumento refleja un dominio o contenido particular, y para que haya validez de contenido, el instrumento debe estar relacionado con los objetivos, competencias y contenidos del curso" (p. 215). Además, Ñaupas et al.

(2014) mencionaron que "la validez de constructo se refiere al grado de correspondencia entre los resultados de una prueba y los conceptos teóricos en los que se basan los temas que se van a medir" (p. 216); Ñaupas et al. explicaron que "este tipo de validez se denomina también validez de hipótesis de trabajo y se determina sobre la base del juicio de los expertos" (p. 216).

No se están empleando cuestionarios, sino al ser información almacenada en base de datos, por lo que no hay cálculo de confiabilidad. Sin embargo, cabe resaltar que estamos utilizando un nivel de confianza de 95% en la aplicación de las pruebas estadísticas.

### **3.5 Procedimientos**

El método de análisis de datos utilizado es de naturaleza cuantitativa, por lo que los datos pueden ser analizados numéricamente. La recopilación de datos se utilizó un a base de datos obtenida de Departamento de Estadística e Información Sanitaria de Chile (DEIS), cuyos datos son de acceso libre para después utilizarlo con la Metodología CRISP-DM (Anexo 11) y poder probar la hipótesis basada en la medición numérica y el análisis estadístico. Utilizando algoritmos predictivos como la regresión lineal, ordinal, Naive, Holt-winter's, ETS.

### **3.6 Aspectos éticos**

Esta investigación utiliza documentación y datos proporcionados por el departamento de estadística e información sanitaria de Chile, que nos proporciona la base de datos, por lo que la información mostrada es genuina y veraz. "La Ley 19.628, de 18/08/1999, de protección de la vida privada, protege a los ciudadanos de la publicación de información o datos personales" (Departamento de Estadística e Información Sanitaria, 2018, p. 9) por lo que se ha tratado con el debido cuidado. Cabe destacar que se respetó la autoría intelectual de las fuentes, citando y refiriéndose a ellas según las normas establecidas por la Universidad César Vallejo.

Además, se respeta la autoría de las fuentes de información. Esto se logró citando adecuadamente los estilos internacionales en el estilo de la APA y cumpliendo con los aspectos relevantes del código de ética de la investigación de la universidad o institución que autorizó la investigación. De acuerdo al artículo

15,16 del código de ética de investigación de la UCV que se están cumpliendo correspondiente a la política de anti plagio y los derechos de autor. Además, los artículos 37, 42,44 del código de ética del Colegio de Ingenieros del Perú referidos a la divulgación de información o la omisión de autor o coautores que intervienen en la investigación las cuales se están cumpliendo. Ñaupas et al. (2014) explico: “No se trata sólo de preocuparse por la dignidad de los sujetos que participan en los procesos de investigación” (p. 462), Ñaupas et al. (2014) agrego: “sino de preocuparse por las políticas estatales de investigación y sobre todo por los mismos investigadores que deben cumplir un código de ética” (p. 462).

## **IV. RESULTADOS**

En este capítulo se explicará todo el proceso de la metodología seleccionada, en este caso CRISP-DM, en el que se explicará cada fase y con su respectivo entregable que nos ayudará a obtener los resultados. Comenzaremos con la primera parte que es el flujo de atención empresarial que nos proporcionará la carga de datos, luego el proceso de limpieza de datos se conoce como ETL que nos permitirá obtener los datos ideales y adecuados para ejecutar nuestra herramienta que nos permitirá crear nuevos modelos de predicción y probabilidad.

Como segunda y última parte aplicaremos la herramienta de modelo predictivo para hacer modelos con los datos previamente elaborados para obtener tanto los resultados como la predicción o probabilidad con cada uno de los diferentes modelos aplicados y una vez terminado buscar cuál es el más óptimo con exactitud y poder interpretar los comportamientos de los datos para conseguir nuestro objetivo una predicción de la cantidad de atención urgente.

## **4.1. FASE 1: COMPRENSIÓN DEL NEGOCIO**

### **4.1.1. Determinar el objetivo del negocio**

La primera actividad de esta fase es determinar los objetivos de los centros de salud de Chile, esto debido a que nuestra solución de predicción debe apoyar directamente estos objetivos generando valor a los mismos, pero segmentados en las regiones donde se investiga. Los objetivos son:

- Proporcionar servicios de salud a las personas de la comunidad.
- Eficiencia y eficacia del Sistema Nacional de Salud.
- Vigilancia, regulación, promoción y prevención de la salud.

Por lo tanto, de acuerdo con las metas de los centros de salud, los objetivos relacionados con nuestra investigación y donde nuestra herramienta ayudará.

### **4.1.2. Evaluación de la situación**

Esta actividad de la fase de conocimiento del negocio de la metodología, cubre una visión general del proceso al que atacaremos en el negocio donde se implementará la solución MD. En los capítulos anteriores se describió la situación de los centros de salud, además de las estadísticas relativas a la atención de emergencias en las

regiones, por lo que en esta etapa sólo se detallarán las actividades del proceso de "Atención de Urgencias" que es el que se busca mejorar con nuestro sistema. A partir del conocimiento de los procesos de los centros de salud, y de la información recogida, hemos podido describir el proceso de Atención de Urgencias, que tiene una serie de actividades bien diferenciadas.

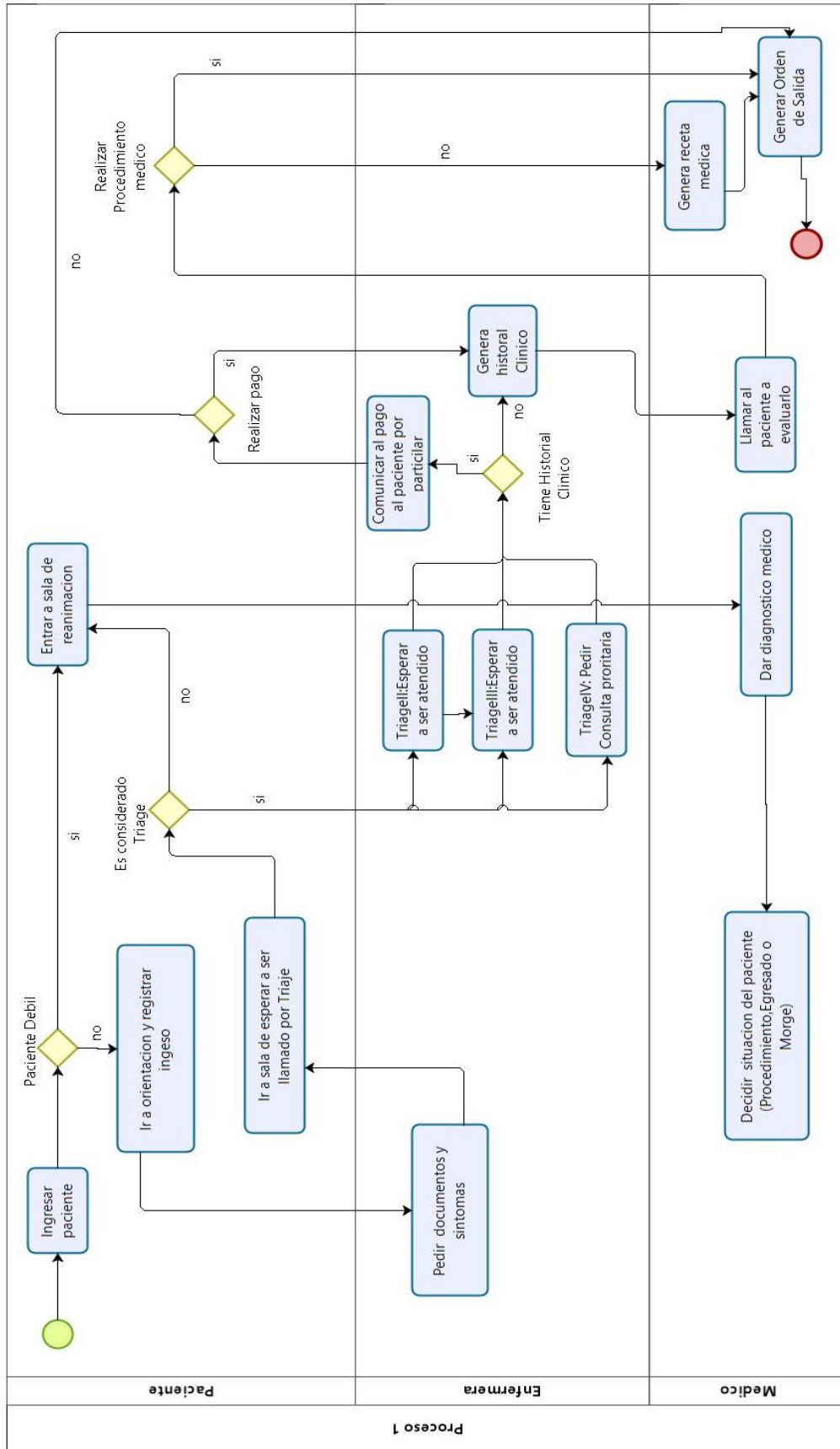


Figura 1 Diagrama de flujo de la atención de emergencia



Como se puede ver en la figura, la solución cubrirá sólo la parte de atención de emergencia; es decir, servirá como herramienta de información en actividades relacionadas cuando no haya atención. Cuando no hay controles para la atención específica en el proceso. Lo primero que hacen los médicos encargados de la unidad de estadística es analizar y detallar la frecuencia de estas atenciones para determinar cuán recurrentes son, si se producen con una frecuencia moderada o si se producen de forma esporádica.

#### **4.1.3. Determinación de los objetivos de la minería de datos.**

Se propusieron los siguientes objetivos:

- Determinar un modelo eficaz de extracción de datos para predecir la atención de emergencia en las regiones de Chile.
- Determinar las regiones en las que podría producirse la atención de emergencia sobre la base de una selección de variables de entrada.
- Mostrar los resultados de la predicción en un mapa geográfico.

Criterios clave para el éxito:

Dependiendo de los objetivos del negocio, y de las atenciones que se registren mensualmente podemos tener como criterio de éxito que:

- En las predicciones se obtiene al menos un 85% de posibilidades de éxito.
- Se tiene un detalle de las regiones con las atenciones más comunes con su probabilidad de ocurrencia.
- El modelo puede ser reeditado con más datos históricos.
- Se pueden predecir más atenciones de emergencia a medida que pasa el tiempo (no sólo para uno o dos meses anteriores, sino que pueden llegar a ser continuas en el futuro).

#### **4.1.4. Elaborar el plan del proyecto.**

Esta actividad comprende dos bloques importantes e independientes. El primero se refiere a la elaboración de un plan general para el proyecto que se va a emprender, y el segundo es la preparación de una evaluación preliminar del instrumento que se utilizará en el proyecto de extracción de datos. La evaluación preliminar del

instrumento esta parte se reflejará en las siguientes fases de la metodología CRISP-DM. De acuerdo con los criterios identificados en las fases 3 y 4, el instrumento que se utilizará será Rapidminer.

## 4.2. Fase 2: Comprensión de los datos.

### 4.2.3. Recolección de los datos iniciales.

Como primer paso, antes de la recolección de datos, se debe contar con una base de datos donde se haya almacenado toda la información que se obtendrá de los centros de salud y al mismo tiempo apoyar las funcionalidades del sistema a desarrollar (incluyendo el modelo de minería), después de eso, algunas de ellas serán utilizadas en el modelo de minería, mientras que otras sólo se utilizarán para la visualización del sistema o para los informes que se construirán en fases posteriores. Para cualquiera de los casos anteriores, es necesario contar con un soporte de base de datos.

A partir de los datos de Access, se podría elaborar un modelo de base de datos para apoyar todos los datos que se reúnan y el modelo que se desarrolle en las fases posteriores. La imagen que se muestra a continuación es el modelo físico desarrollado tras la ingeniería inversa a partir de la información.

IdEstablecim	NEstablecim	IdCausa	GlosaCausa	Total	Menores de	1-4 años	5-14 años	15-64 años	65 años y m.	fecha	semana
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 265		0	49	41	126	49	01/01/2016	52	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 255		0	53	43	120	39	02/01/2016	52	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 230		0	57	37	106	30	03/01/2016	1	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 311		0	51	51	170	39	04/01/2016	1	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 248		0	41	36	125	46	05/01/2016	1	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 225		0	53	31	102	39	06/01/2016	1	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 249		0	54	33	121	41	07/01/2016	1	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 264		2	54	39	125	44	08/01/2016	1	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 259		0	41	42	126	50	09/01/2016	1	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 247		0	46	27	142	32	10/01/2016	2	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 272		0	46	45	134	47	11/01/2016	2	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 256		3	50	48	122	33	12/01/2016	2	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 227		0	55	28	106	38	13/01/2016	2	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 240		0	32	38	130	40	14/01/2016	2	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 227		0	48	36	108	35	15/01/2016	2	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 234		0	39	39	125	31	16/01/2016	2	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 232		0	38	35	131	28	17/01/2016	3	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 249		1	45	33	128	42	18/01/2016	3	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 221		3	37	31	107	43	19/01/2016	3	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 233		0	42	27	119	45	20/01/2016	3	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 233		1	49	28	122	33	21/01/2016	3	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 241		0	46	34	115	46	22/01/2016	3	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 223		1	34	25	135	28	23/01/2016	3	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 209		0	41	27	114	27	24/01/2016	4	
01-100	Hospital Dr. Ju. 1	SECCIÓN 1. TO 270		1	52	42	134	41	25/01/2016	4	

Figura 2 Base de datos de atenciones de urgencias de Chile

#### 4.2.4. Descripción de los datos.

A continuación, la definición de los cuadros con sus respectivos atributos. Véase Anexo 13. La base de datos construida contiene más de 13 millones de atenciones de urgencia desde el 1º de enero de 2008 hasta el 30 de noviembre de 2019, que se registraron en los centros de salud de la región. Según la base de datos, la tabla principal es Atención\_Médicas, en el que cada registro indica las circunstancias de la Atención. Según nuestra opinión, los cuadros secundarios que se relacionan con los datos personales de los pacientes, los médicos, las marcas de medicamentos y las prácticas son poco detallados y pueden excluirse del análisis.

#### 4.2.5. Exploración de datos

Después de tener la descripción de los datos, se debe saber en qué cantidades están representados; sin embargo, esto tomaría un tiempo significativo si se hiciera con cada uno de los atributos de los cuadros (donde no todos son significativos para nuestro problema), por lo que fundamentalmente esta actividad se debe hacer en forma de un informe inicial de listado que muestre los datos más relevantes y luego poder seleccionarlos en la fase 3, ya que en esa etapa se mostrarán gráficos de frecuencia y distribución de estos datos. A continuación, se presentan los datos en forma de informe preliminar, detallando por qué se consideraron relevantes.

Tabla 2

*Variables de la base de datos de atenciones de urgencias*

#	VARIABLES	TIPO	DEFINICIÓN
1	IDESTABLECIMIENTO	CHAR(10)	Número de registro del lugar donde se encuentra el hospital
2	IDTEstablecimiento	CHAR(10)	Número de tipo de registro del establecimiento
3	IDUBICACION	CHAR(10)	Número de lugares
4	IDTipodeAtencion	CHAR(10)	Número de tipos de atención
5	IDPadeamiento	CHAR(10)	Número de la enfermedad
6	IDMedicamento	CHAR(10)	Número de medicamento
7	IDCampaña	CHAR(10)	Número de la campaña
8	Año	CHAR(10)	Año de atención
9	Mes	CHAR(10)	Mes de la atención
10	Día	CHAR(10)	Día de la atención

#	VARIABLES	TIPO	DEFINICIÓN
11	TotalAtenciones	int	Cantidad de atención
12	TatencionEdad0	Int	Cantidad de atención para los pacientes de 0 años
13	TatencionesEdad1a4	Int	Cantidad de atención para los pacientes de 1 a 4 años
14	TatencionesEdad5a14	Int	Cantidad de atención para los pacientes de 5 a 14 años
15	TatencionesEdad15a64	Int	Número de pacientes de 15 a 64 años
16	TatencionesEdad65	Int	Cantidad de cuidados para pacientes de 65 años
17	TotalMedicUso	int	Cuidado de los resultados

#### 4.2.6. Verificar la calidad de los datos.

La información obtenida es de calidad porque nuestro ETL cargó la información necesaria para tener una base de datos OLTP, sobre la cual se elabora el conjunto de datos que se utilizará en el modelo.

#### 4.3. Fase 3: Preparando los datos.

##### 4.3.3. Selección de los datos.

Los campos de interés se convertirán entonces en nuestras variables de entrada del modelo de predicción, éstas deben ser elegidas de acuerdo a un juicio en el que se obtenga una importante relevancia de cada una de ellas. A continuación, se describe cada uno de estos campos. Los datos que se recogerán y que servirán para crear el modelo de predicción de Urgencias son los siguientes:

##### 4.3.4. Frecuencia

###### 4.3.4.1. Cantidad de Atenciones por año.

```
SELECT Año, Sum(CAtenciones)AS CAtenciones FROM
dbo.H_AtencionesUrgencia GROUP BY Año ORDER BY Año ASC;
```

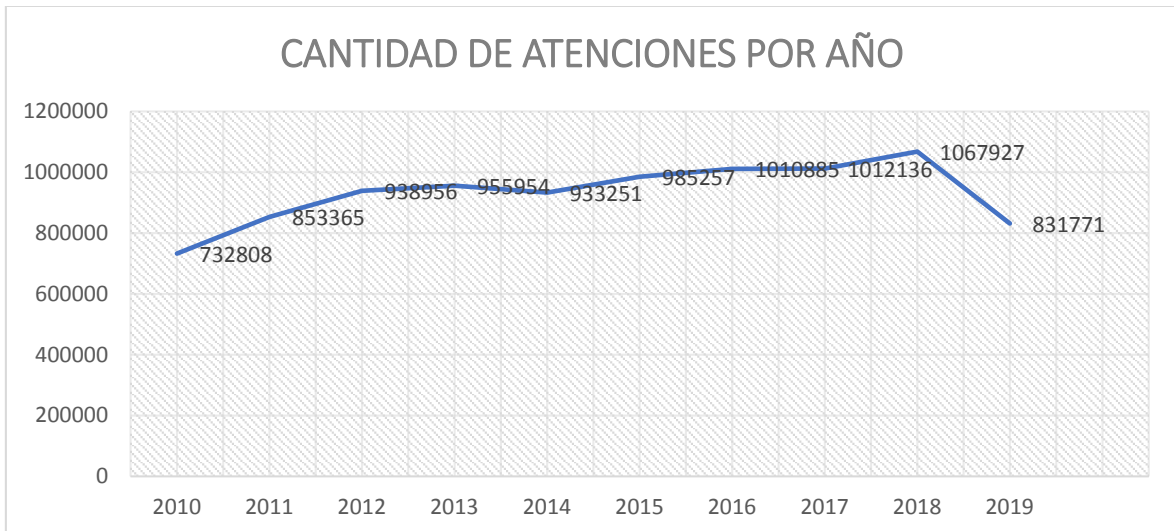


Figura 3 Cantidad de atenciones por año

#### 4.3.4.2. Número de tratamientos por paciente.

Select IDPadecimiento, count(CAtenciones) as Numero\_de\_Atenciones  
 from AtencionesUrgencia group by IDPadecimiento ORDER BY IDPadecimiento,  
 Number\_of\_Atenciones DESC

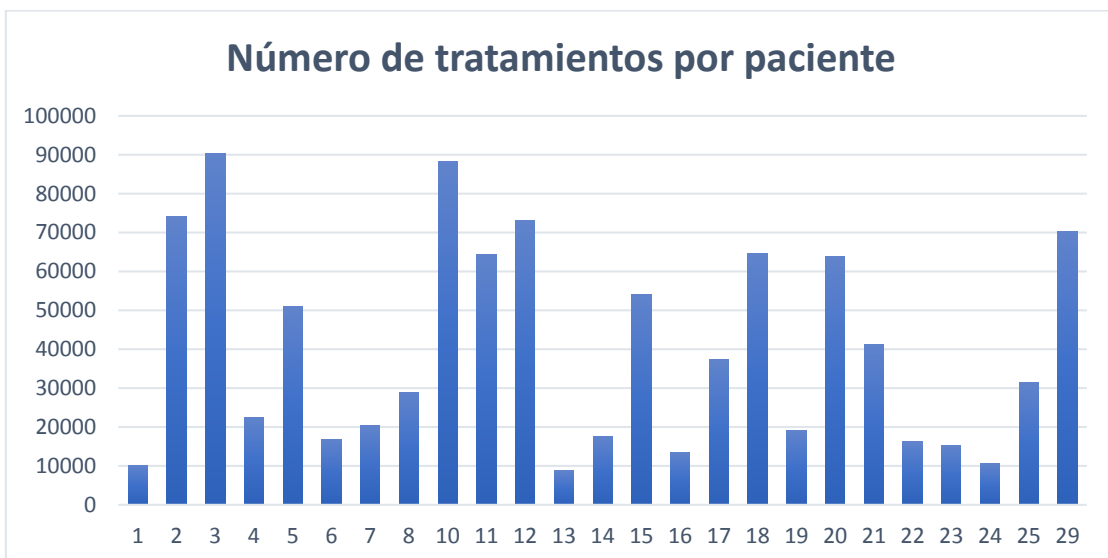


Figura 4 Número de tratamiento por pacientes

#### 4.3.4.3. Cantidad de atenciones al paciente de 0 a 11 meses sobre una base anual.

Select Año, sum(CAtenciones0) as  
 Numero\_Pacientes\_Atencion\_0\_11\_Mes from AtencionUrgencia group by Año  
 ORDER BY Año, Numero\_Pacientes\_Atencion\_0\_11\_Mes DESC

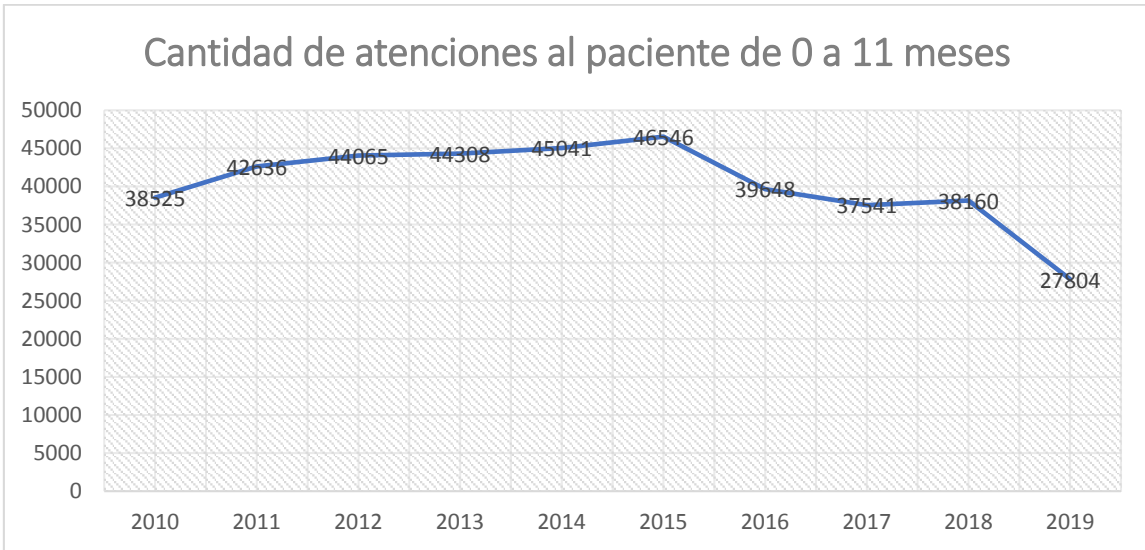


Figura 5 Cantidad de atención al paciente de 0 a 11 meses

**4.3.4.4.** Cantidad de atenciones para los pacientes de 1 a 4 años sobre una base anual.

```

Select      Año,          sum(CAtenciones14)          as
Numero_Pacientes_Atencion_1_4_Años  from AtencionUrgencia group by Año
ORDER BY Año, Numero_Pacientes_Atencion_1_4_Años DESC

```

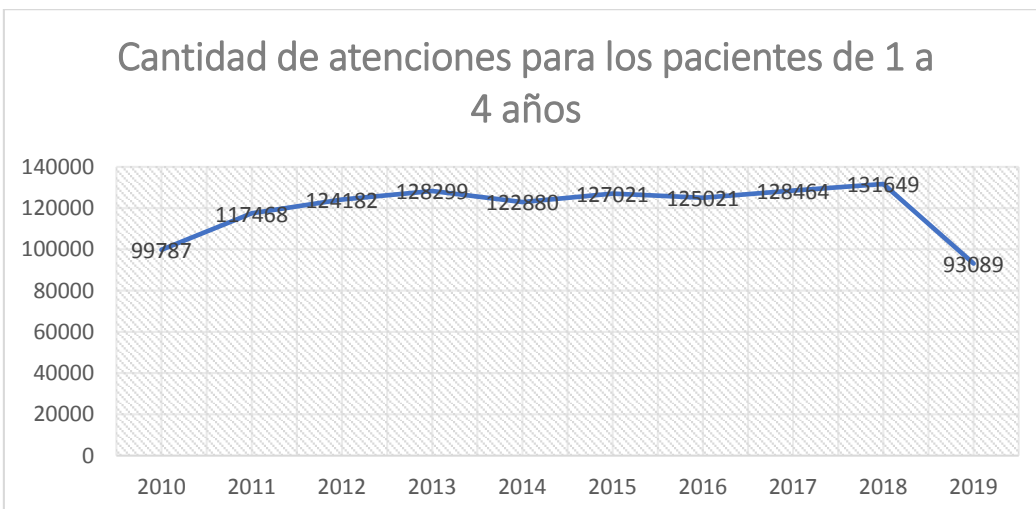


Figura 6 Cantidad de atención para los pacientes de 1 a 4 años

**4.3.4.5.** Cantidad de atenciones para los pacientes de 5 a 14 años de edad sobre una base anual.

```

Select      Año,          sum(CAtenciones514)          as
Numero_Pacientes_Atencion_5_14_Años from AtencionUrgencia where group by Año
ORDER BY Año, Numero_Pacientes_Atencion_5_14_Años DESC

```

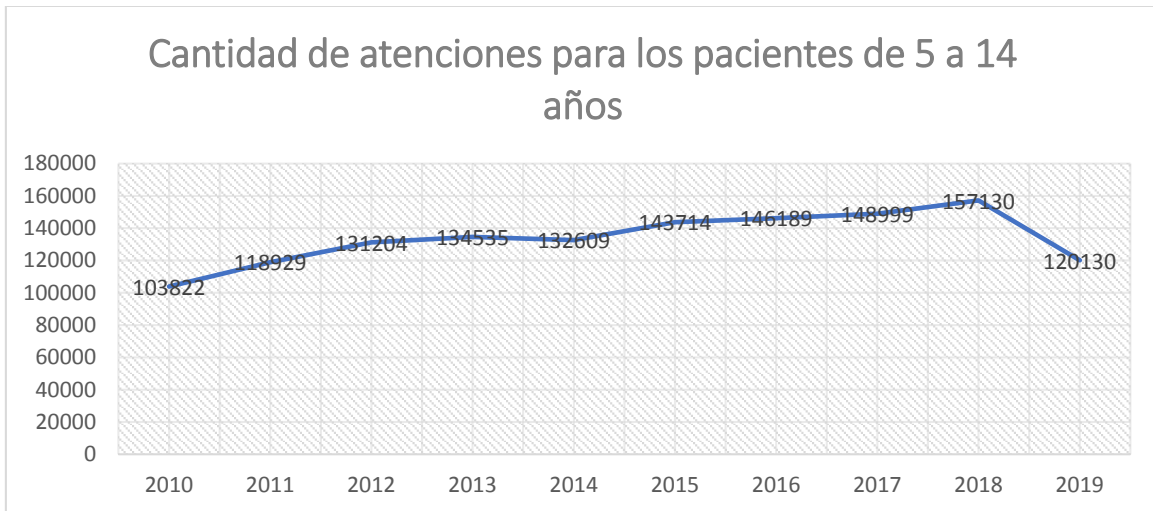


Figura 7 Cantidad de atención para los pacientes de 5 a 14 años de edad

**4.3.4.6.** Cantidad de atenciones para pacientes de 15 a 64 años anualmente.

```
Select Año, sum(CAtenciones1564) as Numero_Pacientes_Atencion
_15_64_Años from AtencionUrgencia group by Año ORDER BY Año,
Numero_Pacientes_Atencion _15_64_Años DESC
```

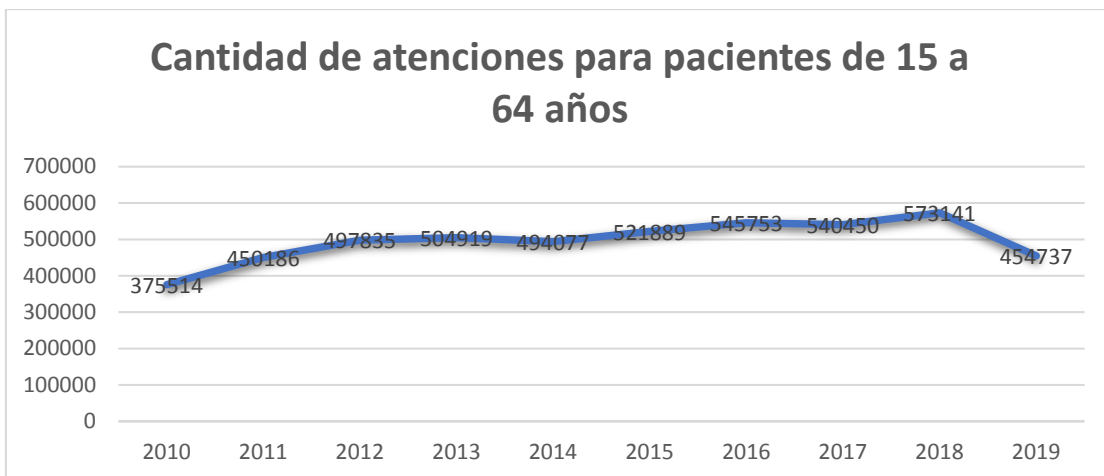


Figura 8 Cantidad de atención a pacientes de 15 a 64 años

**4.3.4.7.** Cantidad de atenciones para los pacientes de 65 años o más sobre una base anual.

```
Select Año, sum(CAtenciones65) as Numero_Pacientes_Atencion
_65_Años from AtencionUrgencia group by Año ORDER BY Año,
Numero_Pacientes_Atencion _65_Años DESC
```

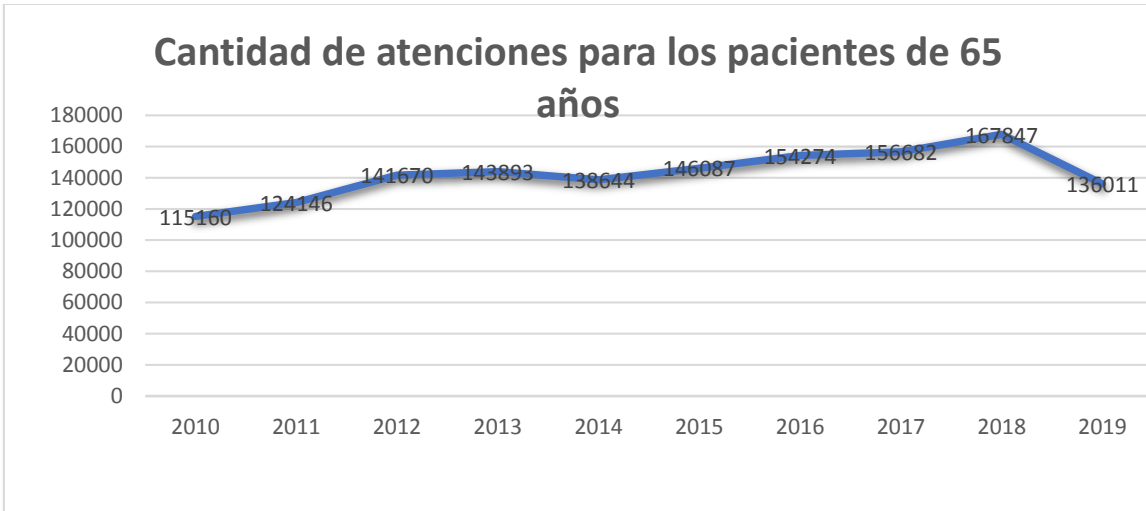


Figura 9 Cantidad de atenciones de pacientes de 65 años o más

**4.3.4.8.** Cantidad de la medicación por IDMedicación.

```
select IDMedicamento, sum(CMedication) as Total_Medicacion from
AtencionUrgencia group by IDMedicamento ORDER BY IDMedicamento DESC
```

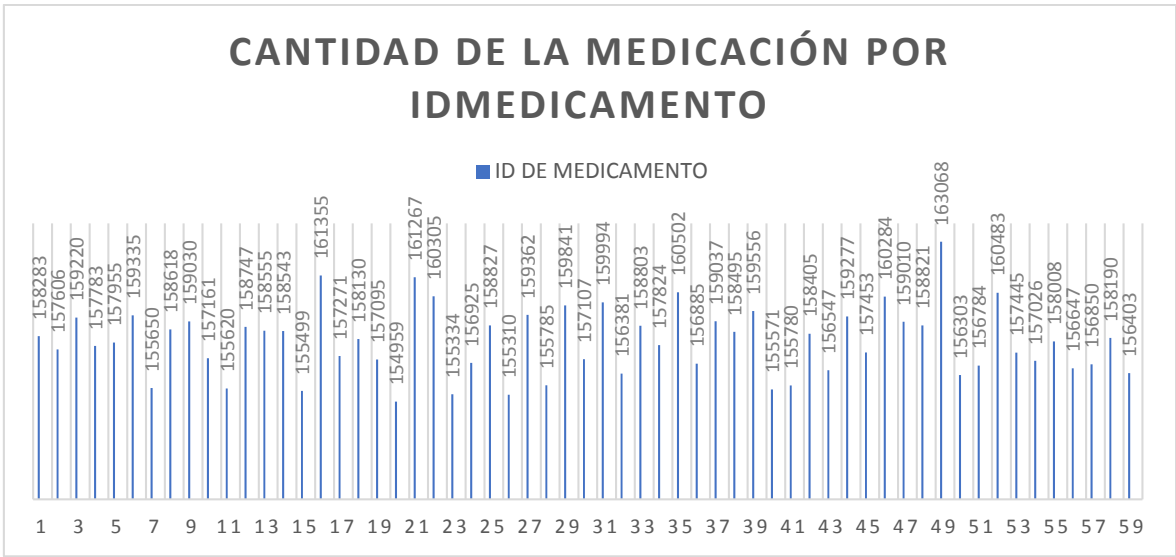


Figura 10 Cantidad del tipo de medicamento utilizado en la atención

**4.3.5. Depuración de datos.**

Como se puede ver, algunos registros tienen campos incompletos o que no coinciden con el atributo específicamente, hay cuidados de emergencia en los que no se especifica el atributo o condición de causa, y hay registros en sus campos como nulos. Es necesario reemplazar estos caracteres para que la herramienta pueda interpretar nuestro conjunto de datos sin errores. A continuación, mostramos los términos de exclusión de los datos proporcionados:



Tabla 3

*Exclusión de los datos en la base de atenciones de urgencia*

#	DATOS	REFERENCIA
1	- LAS DEMÁS CAUSAS	Campo GlosaCausa
2	- TRAUMATISMOS Y ENVENENAMIENTOS	Campo GlosaCausa
3	- CAUSAS SISTEMA CIRCULATORIO	Campo GlosaCausa
4	CAUSAS SISTEMA RESPIRATORIO	Campo GlosaCausa
5	Pacientes en espera de hospitalización	Campo GlosaCausa
6	Pacientes en espera de hospitalización que esperan menos de 12 horas para ser trasladados a cama hospitalaria	Campo GlosaCausa
7	SECCIÓN 1. TOTAL ATENCIONES DE URGENCIA	Campo GlosaCausa
8	SECCIÓN 2. TOTAL DE HOSPITALIZACIONES	Campo GlosaCausa
9	TOTAL CAUSAS SISTEMA CIRCULATORIO	Campo GlosaCausa
10	TOTAL CAUSAS SISTEMA RESPIRATORIO	Campo GlosaCausa
11	TOTAL DEMÁS CAUSAS	Campo GlosaCausa
12	TOTAL TRAUMATISMOS Y ENVENENAMIENTO	Campo GlosaCausa
13	Valores NULL	Campo GlosaCausa, GLOSATIPOATENTION, fecha, Glosacausa, NEstablecimiento, GLOSATIPOESTABLECIMIENTO, GlosaTipoCampaña, GlosaMedicamento,

Chu, Ilyas, Krishnan, Wang (2016) explicaron: "Detectar y reparar los datos sucios es uno de los desafíos perennes en el análisis de datos, y el no hacerlo puede resultar en análisis inexactos y decisiones poco fiables". (p. 2204), También Chu al et. (2016) definieron "Este tutorial se centró en la limpieza de datos cualitativos que utiliza restricciones, reglas o patrones para detectar errores". (p. 2204).

A continuación, se muestra el número de registros que se excluirán que contienen valores atípicos, valores NULL. Por lo tanto, no hay valores que contengan letras mayúsculas o valores con caracteres que se manejen en la ETL.

Tabla 4

*Resultados de totales atenciones de urgencia registrados por años*

GLOSA-CAUSA	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Resultados
<b>1</b>	116442	129763	140600	149611	153115	160756	164285	165005	166258	126969	1472804
<b>2</b>	116442	129763	140600	149611	153115	160756	164282	165005	166258	126969	1472801
<b>3</b>	116442	129763	140599	149610	153114	160756	164282	165005	166258	126969	1472798
<b>4</b>	116442	129763	140599	149609	153114	160756	164282	165005	166258	126969	1472797
<b>5</b>	116442	129763	140600	149613	153115	160756	164285	165005	166259	126970	1472808
<b>6</b>	116442	129763	140600	149614	153117	160756	164285	165005	166259	126969	1472810
<b>7</b>	116452	129763	140599	149607	153110	160756	164279	165004	166256	126969	1472795
<b>8</b>	116443	129763	140599	149609	153114	160756	164282	165005	166257	126970	1472798
<b>9</b>	116448	129763	140599	149608	153112	160756	164280	165005	166256	126969	1472796
<b>10</b>	116452	129763	140599	149607	153111	160756	164280	165004	166256	126970	1472798
<b>11</b>	116443	129763	140599	149609	153113	160756	164282	165005	166257	126970	1472797
<b>12</b>	116444	129763	140599	149609	153112	160756	164280	165005	166257	126969	1472794
<b>13</b>	659008	773751	839221	909826	925166	982794	994823	982316	973235	726715	8766855
<b>TOTAL</b>	2056342	2330907	2526413	2705143	2762528	2911866	2966207	2962374	2968324	2250347	26440451

#### 4.3.6. Construcción del modelo de dimensional.

Después de que tenga la base de datos modelada, debería obtener la información de los cuidados de emergencia. Esta información fue recogida en formato Access, porque la plataforma nos proporciona la información. Teniendo toda esta información en la tabla de Access, tuvimos que desarrollar un ETL para clasificar y cargar esta información.

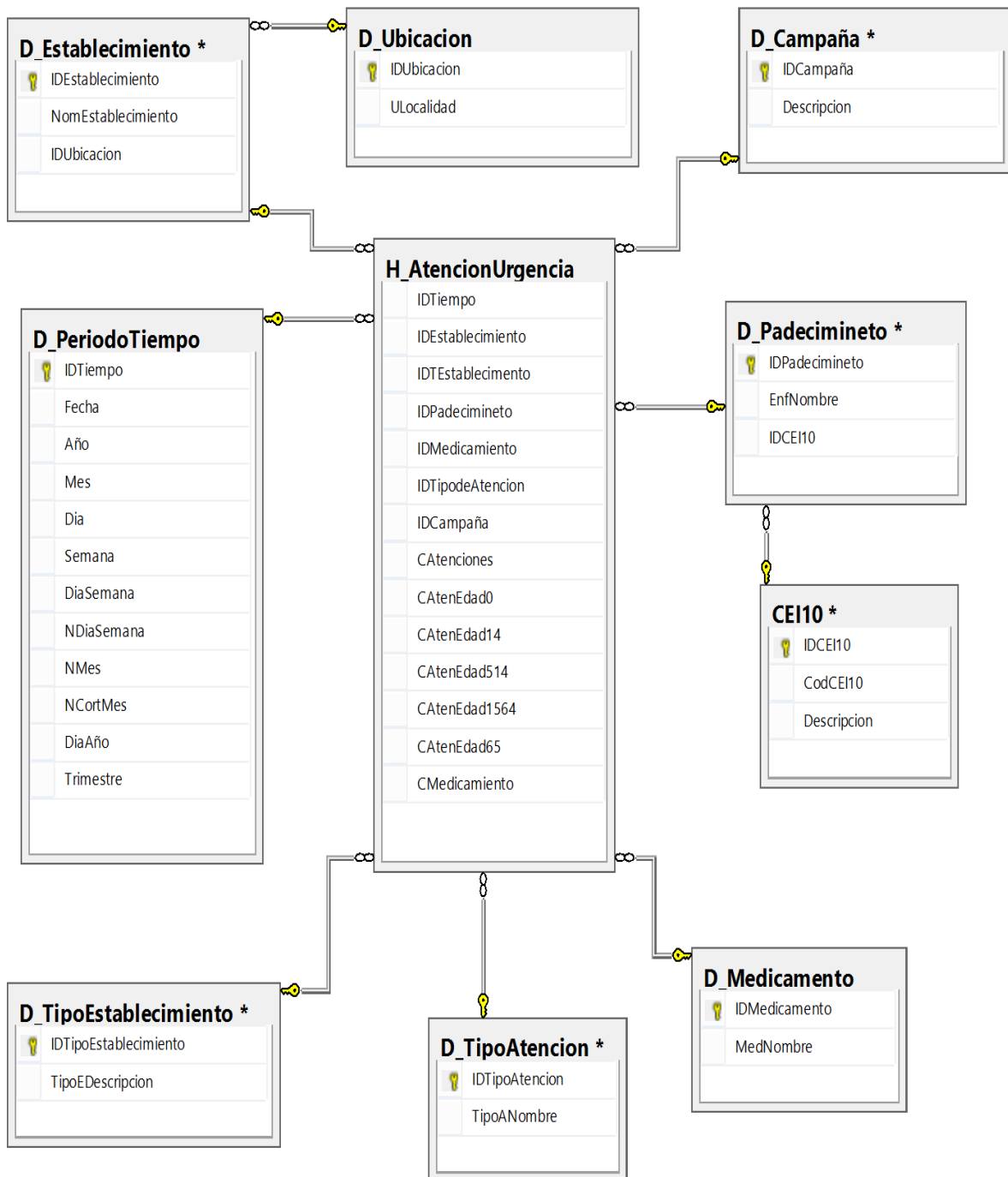


Figura 11 Modelo dimensional de atenciones de urgencias

Finalmente, el ETL (usando técnicas de inteligencia de negocios) se desarrolla para cargar las tablas.

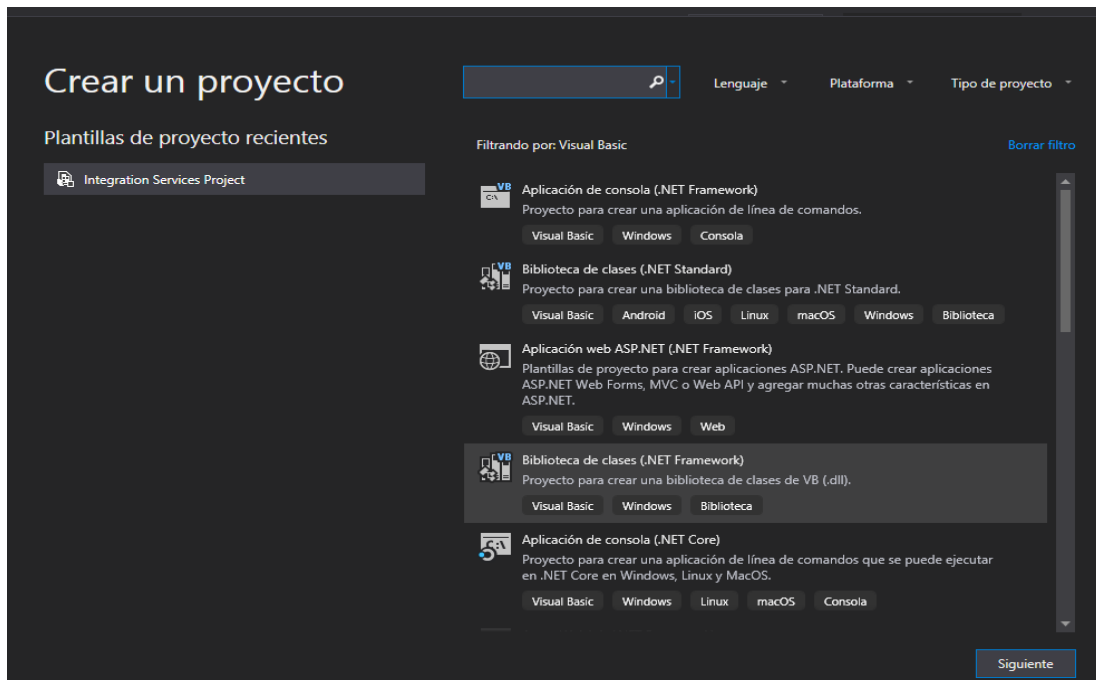


Figura 12 Creación de proyecto ETL de atenciones de urgencias

Buscamos el Proyecto de Servicios de Integración y lo damos a continuación.

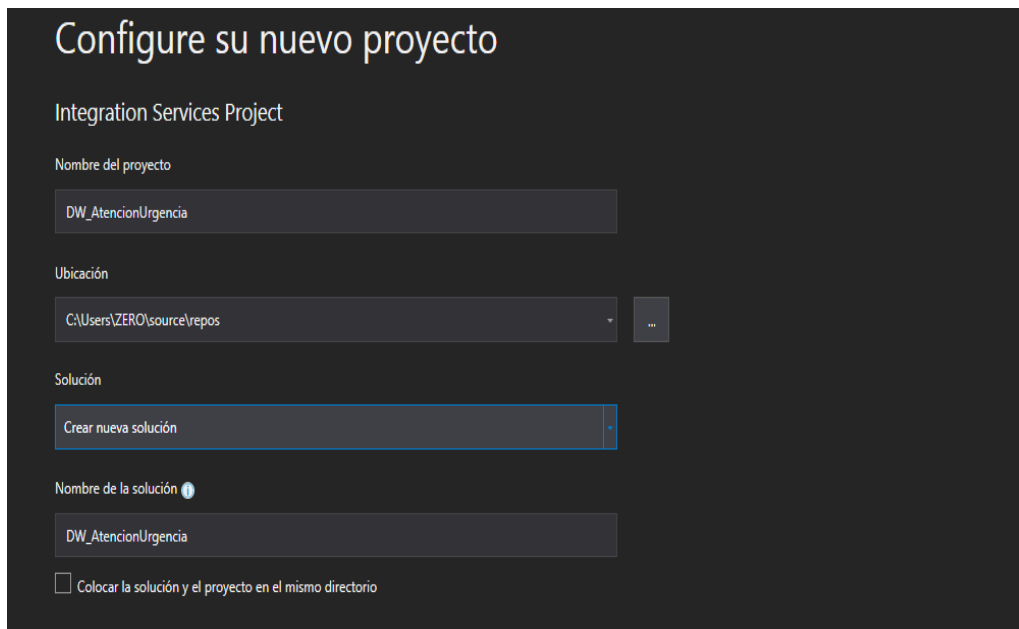


Figura 13 Configuración del proyecto ETL

A continuación, configura las conexiones que tendrá como fuentes de origen y destino. En este caso, la fuente de la Fuente será la base de datos y la fuente de

destino será el Almacén de Datos de Atención de Emergencias. El servidor debe ser colocado, en este caso, uno local. Luego prepare la conexión de la Fuente indicando el nombre de la base de datos.

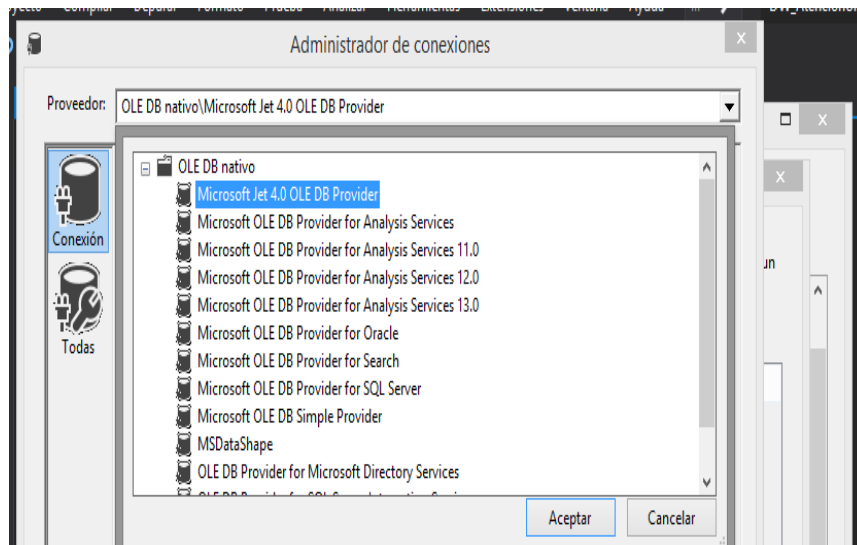


Figura 14 Selección de la fuente de destino de la base de datos

Entonces localizamos donde se encuentra la base de datos.

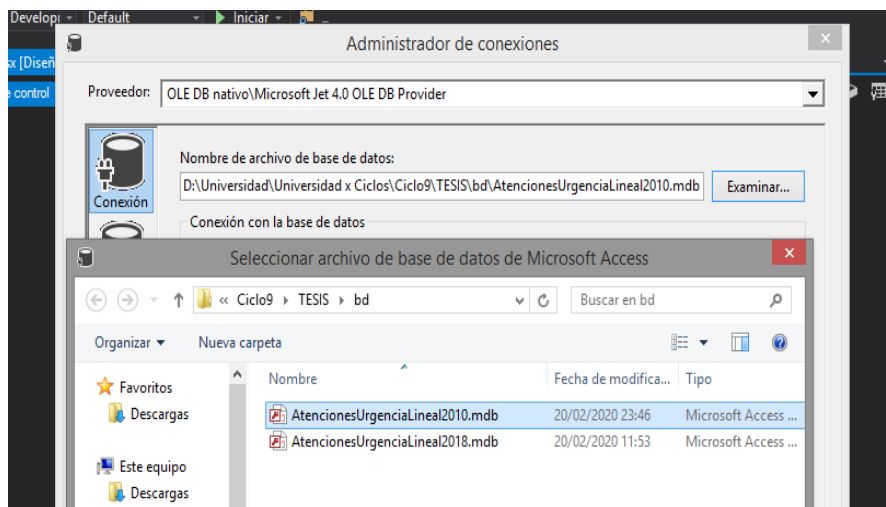


Figura 15 Localización de base de datos de atenciones de urgencia

Entonces comprobamos si había una conexión correcta.

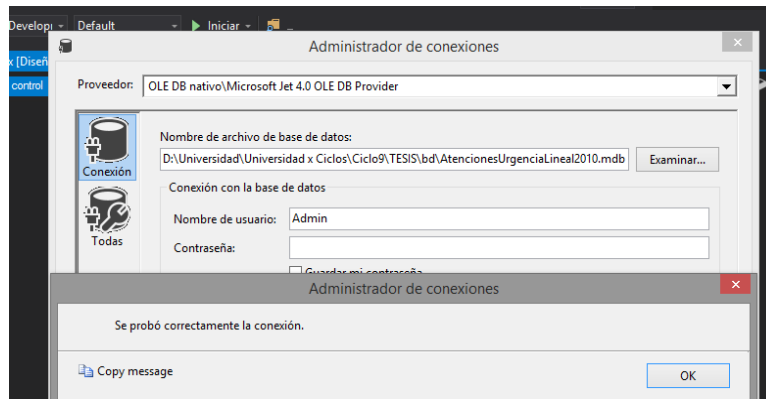


Figura 16 Verificación de conexión de la base de datos

Nos muestra el mensaje de que la conexión es exitosa. Por lo tanto, la fuente de la base de datos de la Fuente ya está configurada. Realiza el mismo paso con la base de datos de destino. Aquí, preparamos la conexión con el Destino, que viene a ser nuestro Almacén de Datos.

Por lo tanto, ya tenemos las conexiones configuradas para poder realizar la integración y a través de un proceso ETL para cargar los datos en el Data Warehouse. Lo siguiente es colocar las Tareas SQL o Tareas de Flujo de Datos (con tareas internas). En este caso, se tuvo en cuenta la siguiente estructura para el ETL:

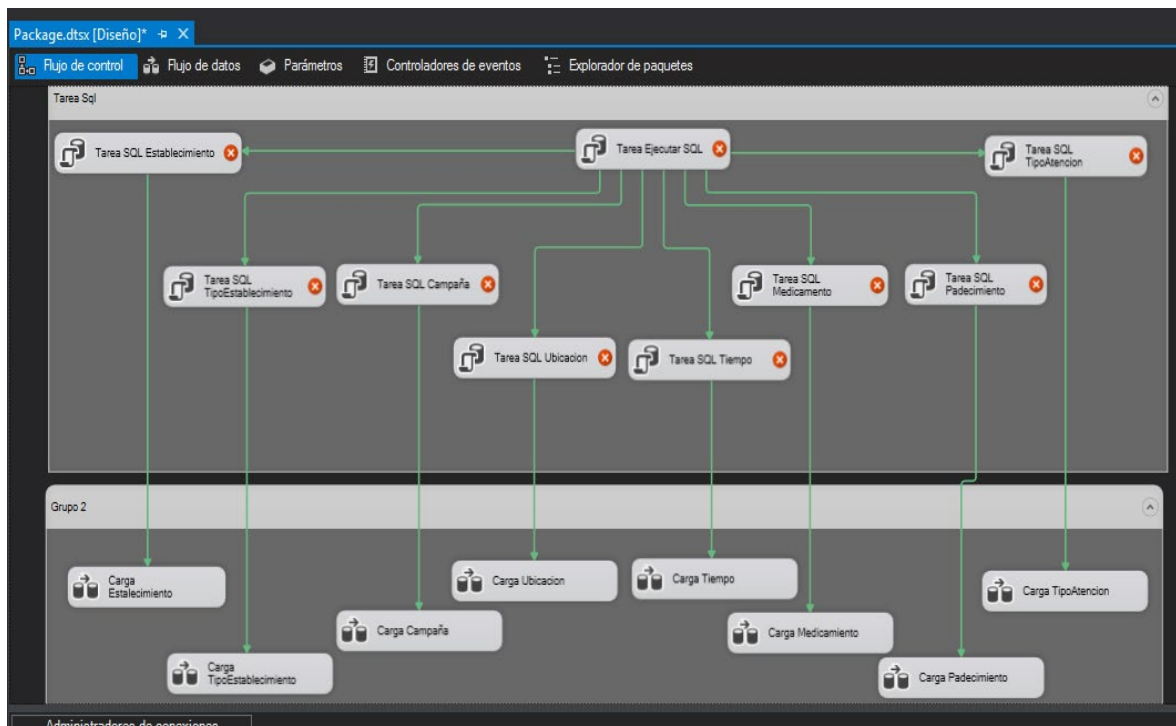


Figura 17 Modelo de flujo de control de minería de datos

Una vez conectadas todas las tareas que se ejecutarán en el orden indicado, se cargarán las siguientes consultas a la base de datos en cada proceso, empezando por la fuente:

#### 4.3.7. Flujo de control.

##### 4.3.7.1. Dimensión Establecimiento

##### 4.3.7.1.1. Ejecución de tareas SQL– Limpieza

A través de una consulta se descartan las filas que contienen los datos arriba indicados por la irrelevancia de la información que es necesaria para el proyecto.

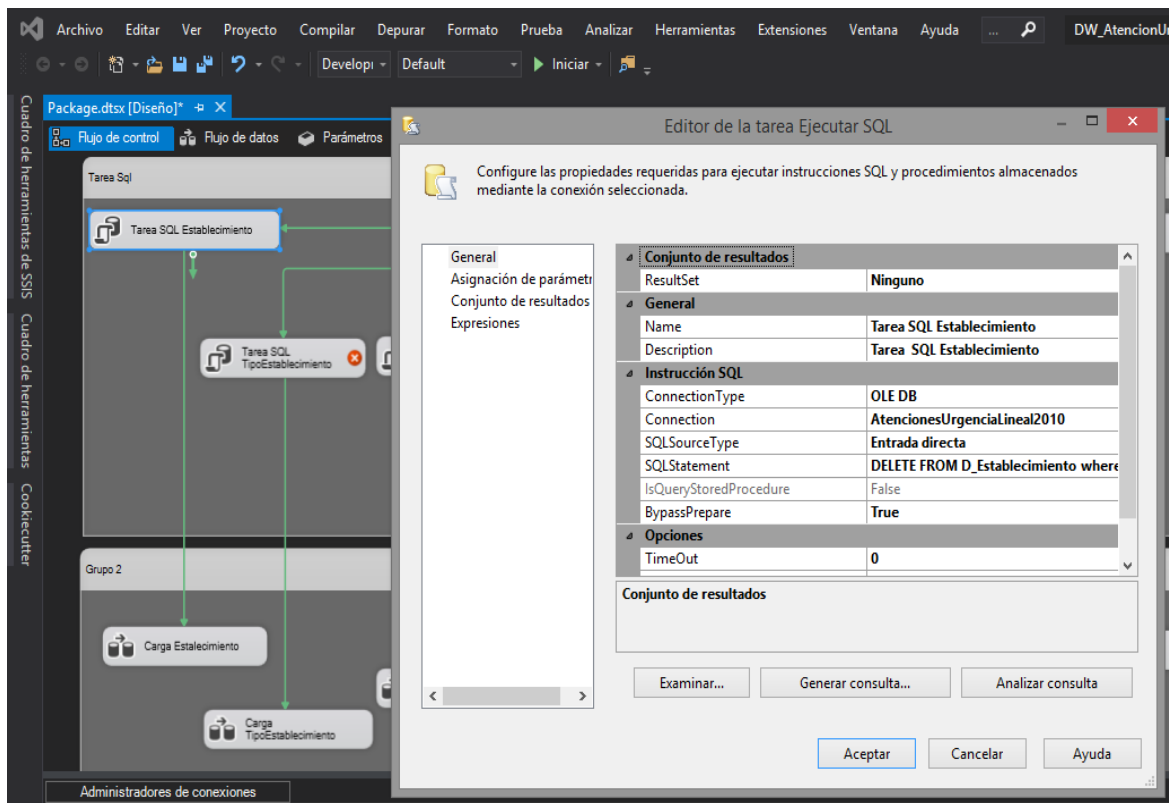


Figura 18 Control de flujo de la dimensión establecimiento

#### Sentencia SQL

```
DELETE FROM AtencionUrgenciaLineal2010 where GlosaCausa in ('TOTAL TRAUMATISMOS Y ENVENENAMIENTO', 'TOTAL DEMÁS CAUSAS','TOTAL CAUSAS SISTEMA RESPIRATORIO', 'TOTAL CAUSAS SISTEMA CIRCULATORIO', 'SECCIÓN 2. TOTAL DE HOSPITALIZACIONES', 'SECCIÓN 1. TOTAL ATENCIONES DE URGENCIA', 'Pacientes en espera de hospitalización
```

que esperan menos de 12 horas para ser trasladados a cama hospitalaria', 'Pacientes en espera de hospitalización', ' - LAS DEMÁS CAUSAS',)

DELETE FROM AtencionUrgenciaLineal2010 where GlosaCausa is null or GLOSATIPOATENCION is null or Fecha is null or Glosacausa is null or Nestablecimiento is null or GLOSATIPOESTABLECIMIENTO is null or GlosaTipoCampana is null or GlosaMedicina is null

#### 4.3.7.1.2. Tarea de flujos de datos –Carga

Localizamos y hacemos doble clic en el DataFlowTask Establecimiento, donde pasaremos a realizar el flujo de datos de origen y destino.

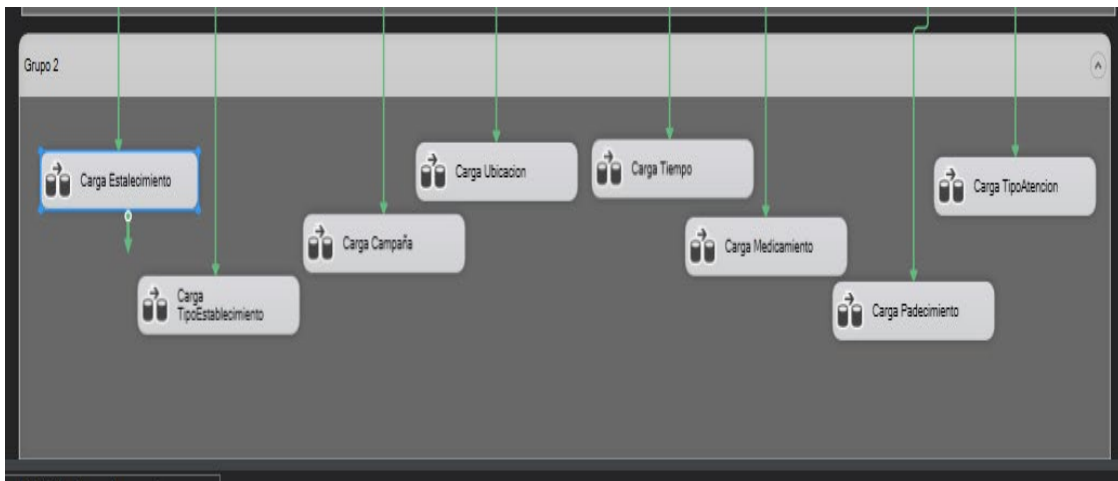


Figura 19 Tarea de flujos de datos de la base de datos –Carga

A continuación, arrastramos un elemento de origen y destino de los datos y los unimos.

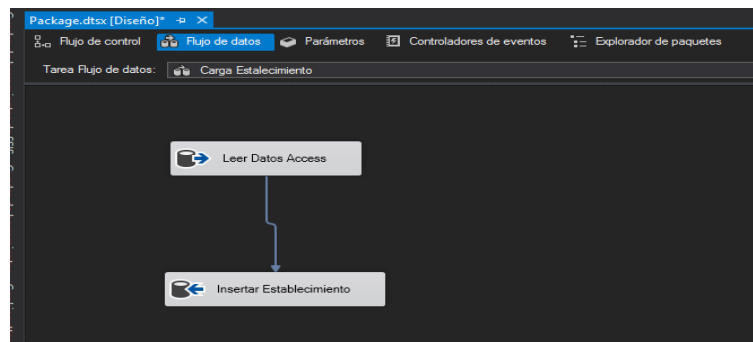


Figura 20 Control de flujo de base de datos de MS Access a SQL Server



Luego, al hacer clic en "Leer los datos de acceso" elegimos la fuente de datos que pertenece a los datos de acceso.

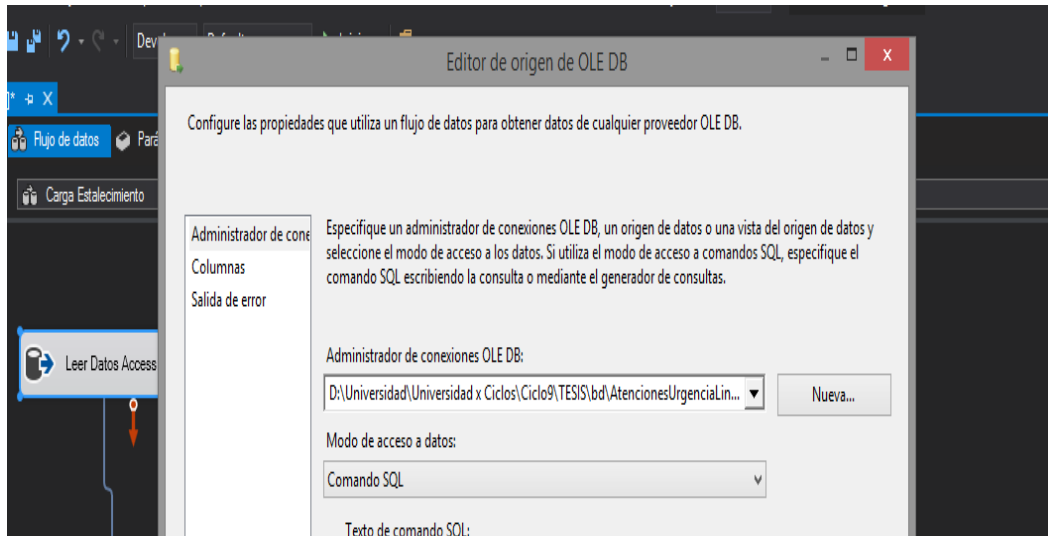


Figura 21 Selección de la *fuentes de acceso a datos SQL Server*

Luego elegimos el modo de acceso a los datos que usaremos "Comando SQL"

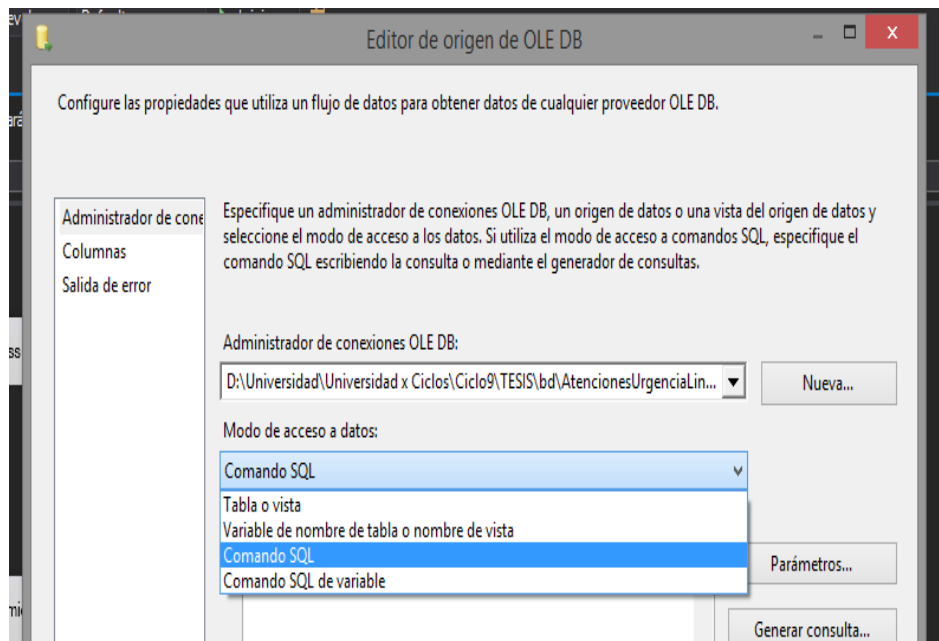


Figura 22 Selección del modo de acceso a los datos al *SQL Server*

Finalmente insertamos la declaración SQL y aceptamos. Ir el anexo 14

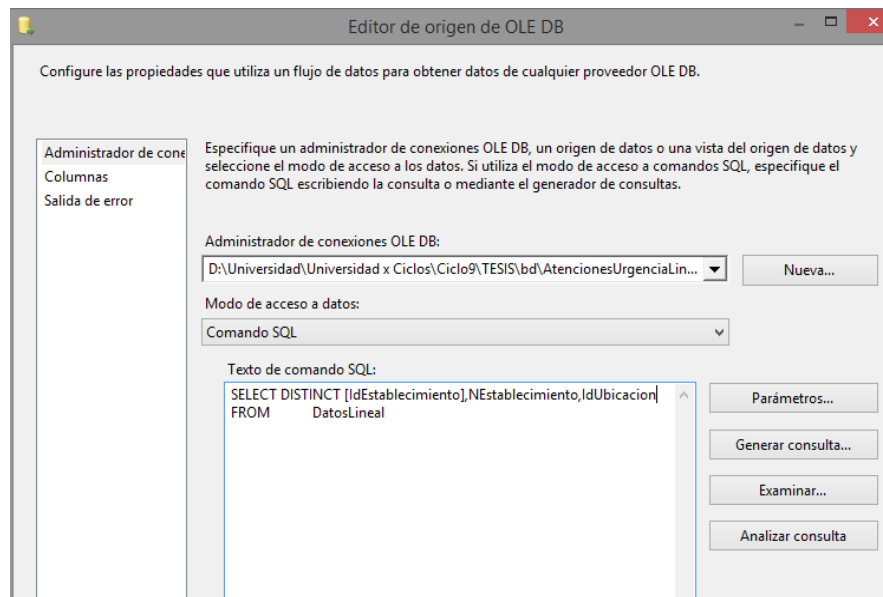


Figura 23 *Entrada de la sentencia SQL*

Ahora pasamos a "Insertar Establecimiento" y elegimos la conexión del objetivo.

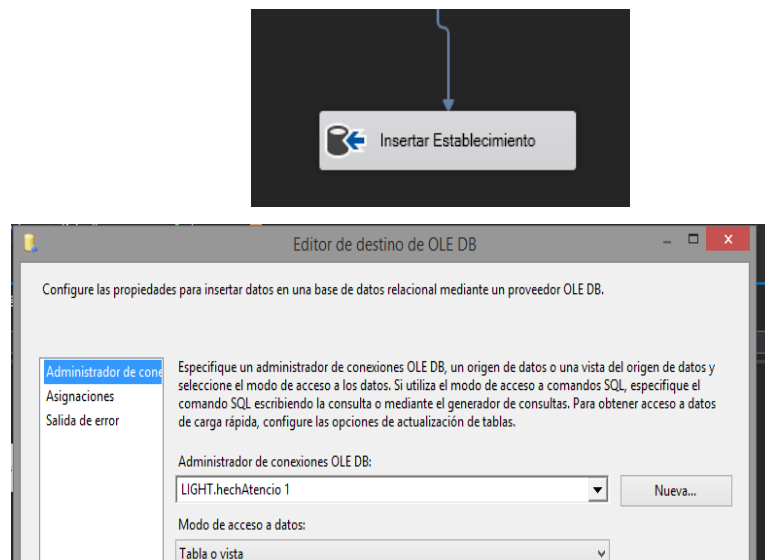


Figura 24 *Selección de la conexión a la base de datos SQL Server destino*

Luego elegimos el modo de acceso a los datos que usaremos "Tabla o Vista"

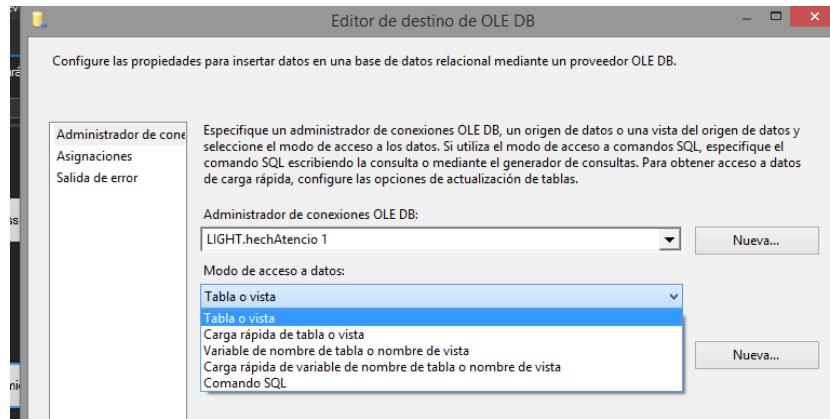


Figura 25 Selección del modo de acceso a los datos del editor destino

Finalmente elegimos la tabla donde se cargarán los datos en este caso "D\_Establecimiento" y aceptamos.

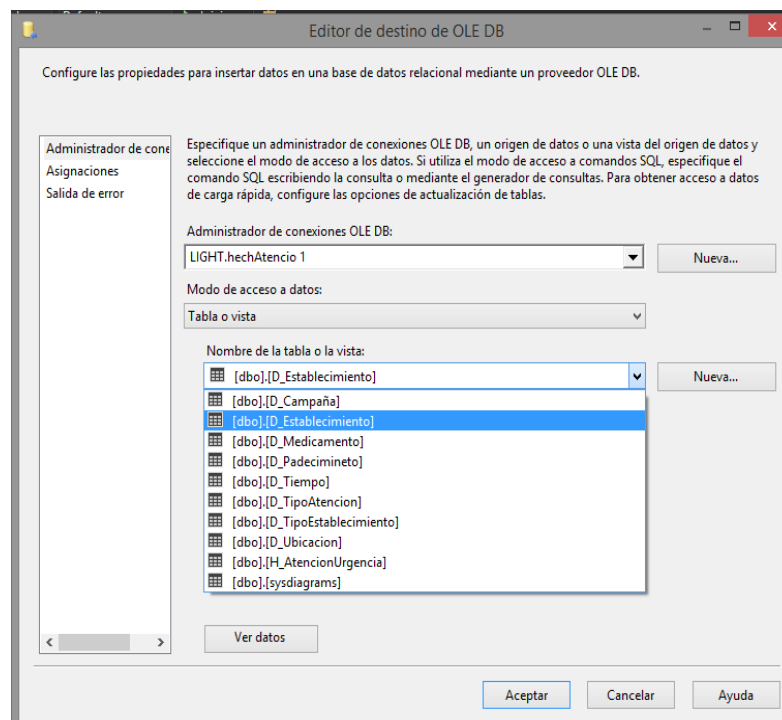


Figura 26 Selección de la tabla destino en SQL Server

Por lo tanto, para lo siguiente, se hará el mismo proceso para cada carga de datos, que aplica para el año 2010 a 2019 en los demás la declaración SQL de "Read Access Data", donde se cambia la fuente de datos. Véase el anexo 14 que contiene el resto de las declaraciones SQL.

#### 4.3.7.2. Ejecución del proceso.

Entonces ejecute el proceso ETL y la prueba que se ejecutó con éxito. Aquí podemos ver cómo los datos programados en el ETL fueron cargados en las tablas:

- Dimensión D\_Establecimiento

IDEstablecimiento	NEstablecimiento	IDUbicacion
1	Hospital Dr. Juan Noé Crevanni	5
2	Centro de Salud Familiar Iris Véliz Hume	11
3	SAPU Dr. Amador Neghme de Arica	7
4	SAR EU Iris Véliz Hume	33
5	Hospital Dr. Ernesto Torres Galdames	36
6	Hospital Comunitario de Salud Familia...	1
7	Consultorio General Rural Dr. Amador...	3
8	SAPU-Cirujano Aguirre	7

Figura 27 Resultado de la carga de la Dimensión D\_ Establecimiento

- Dimensión D\_Campaña

IDCampaña	CDescripcion
1	Adulto Metropolitana
2	Infantil Metropolitana
3	Infantil y Adulto M...
4	Invierno Regiones
5	Ninguna
6	Sapu Metropolitana

Figura 28 Resultado de la carga de la Dimensión D\_ Campaña

- Dimensión D\_Medicamento

IDMedicamento	MedNombre
1	trimetafán (afonad)
2	Verapamilo
3	Valproato Sódico
4	Tramadol (Cloridrato)
5	Tiapride
6	Sulfato de Magnesio
7	Succinilcolina

Figura 29 Resultado de la carga de la Dimensión D\_ Medicamento

- Dimensión D\_Padecimiento

IDPadecimiento	CANombre
1	Shock séptico
2	Astoria
3	Bronquitis/bronquiolitis aguda (J20-J21)
4	Influenza (J09-J11)
5	Neumonía (J12-J18)
6	Otra causa respiratoria (J22, J30-J39, J47, J60-...
7	quemaduras

Figura 30 Resultado de la carga de la Dimensión D\_ Padecimiento

- Dimensión D\_PeriodoTiempo

IDTiempo	Fecha	Año	Mes	Dia	Semana	DiaSemana	NDiasemana	NMes	NCortMes	DiaAño	Trimestre
1	20100103	2010	1	3	1	7	Domingo	Enero	Ene	3	1
2	20100104	2010	1	4	2	1	Lunes	Enero	Ene	4	1
3	20100105	2010	1	5	2	2	Martes	Enero	Ene	5	1
4	20100106	2010	1	6	2	3	Miércoles	Enero	Ene	6	1
5	20100107	2010	1	7	2	4	Jueves	Enero	Ene	7	1
6	20100108	2010	1	8	2	5	Viernes	Enero	Ene	8	1
7	20100109	2010	1	9	2	6	Sábado	Enero	Ene	9	1

Figura 31 Resultado de la carga de la Dimensión D\_ PeriodoTiempo

- Dimensión D\_TipoAtencion

```
select * from [dbo].[D_TipoAtencion]
```

IDTipoAtencion	TNombre
1	Adulto
2	Indiferenciado
3	Infantil

Consulta ejecutada correctamente. LIGHT (12.0 RTM) | LightZERO (53) | hechAtencio | 00:00:00 | 3 filas

Figura 32 Resultado de la carga de la Dimensión D\_TipoAtencion

- Dimensión D\_TipoEstablecimiento

```
select * from [dbo].[D_TipoEstablecimiento]
```

IDTipoEstablecimiento	TEDescripcion
1	Hospital
2	SAPU
3	SAR

Consulta ejecutada correctamente. LIGHT (12.0 RTM) | LightZERO (53) | hechAtencio | 00:00:00 | 3 filas

Figura 33 Resultado de la carga de la Dimensión D\_TipoEstablecimiento

- Dimensión D\_Ubicacion

```
select * from [dbo].[D_Ubicacion]
```

IDUbicacion	ULocalidad
1	Antofagasta
2	Arica
3	Calama
4	Casablanca
5	Cauquenes
6	Chañaral
7	Chilán

Consulta ejecutada correctamente. LIGHT (12.0 RTM) | LightZERO (53) | hechAtencio | 00:00:00 | 36 filas

Figura 34 Resultado de la carga de la Dimensión D\_Ubicación

- Dimensión H\_AtencionesUrgencia

IDTiempo	IDEstablecimiento	IDTipoEstablecimiento	IDPadecimiento	IDMedicamiento	IDTipoAtencion	IDCampaña	CAtenciones	CAtenEstado0	CAtenEstado14	CAtenEstado514	CAtenEstado1564	CAtenEstado55	CMedicamiento
20100402	22-304	2	10	42	2	5	6	0	1	1	4	0	7
20100403	22-304	2	10	15	2	5	18	2	6	2	8	0	23
20100404	22-304	2	10	29	2	5	17	2	2	3	10	0	21
20100405	22-304	2	10	21	2	5	21	2	4	4	11	0	23
20100406	22-304	2	10	17	2	5	10	1	4	1	4	0	11
20100407	22-304	2	10	21	2	5	21	1	2	5	12	1	25
20100408	22-304	2	10	34	2	5	3	0	1	1	1	0	5

Figura 35 Resultado de la carga de la Dimensión H\_AtencionesUrgencia

Después del proceso exitoso, se indican los tiempos de ejecución en la migración y la carga de datos.

Tabla 5

Tiempo de carga o ejecución de la tarea ETL

TIEMPO DE CARGA DE LA TAREA				
#	DIMENSION	INICIO	FINAL	
1	D_Establecimiento	3:07:23	05:43:03	Elapsed time: 02:35:55.328
2	D_Campaña	5:43:18	6:23:48	Elapsed time: 00:10:30.558
3	D_Medicamiento	6:23:48	7:28:58	Elapsed time: 01:05:10.301
4	D_Padecimientos	7:28:58	8:24:32	Elapsed time: 00:55:34.120
5	D_PeriodoTiempo	8:24:32	10:24:56	Elapsed time: 02:00:24.520
6	D_TipoAtencion	10:24:56	11:05:17	Elapsed time: 00:40:21.348
7	D_TipoEstablecimiento	11:05:17	11:48:37	Elapsed time: 00:43:20.311
8	D_Ubicacion	11:48:37	12:49:18	Elapsed time: 01:00:41.332
9	H_AtencionesUrgencia	12:49:18	16:44:28	Elapsed time: 03:55:10.600

#### 4.3.8. Integración de los datos.

En esta actividad debemos integrar los datos construidos en la actividad anterior ("Construir Datos"), si alguno de los atributos seleccionados debe combinarse con otro para formar una variable de entrada del modelo.

#### **4.3.9. Formato de los datos.**

El siguiente paso, para llegar a la construcción del modelo, es la elaboración del conjunto de datos. El conjunto de datos se llama el conjunto de datos que se analizará con el software de minería de datos. Nuestro conjunto de datos debe estar en el formato en que se utilizó la herramienta Rapidminer por su amplia conectividad con los motores de base de datos y también por su intuitivo Tablero de mandos. Para el proceso de depuración y selección de algoritmos integrados para aplicar la validación y remodelación de la fórmula de predicción de la atención de urgencia.

Además, R Studio es uno de los programas informáticos que cuenta con una diversidad de herramientas de análisis estadístico univariante y multivariante, que se utilizan no sólo en las ciencias sociales sino también en otras disciplinas. Al ser un programa de contribución libre (código abierto), no requiere la adquisición de una licencia de pago para su uso; también hay una gran variedad de paquetes descargables gratuitos.

La base de datos de atención de emergencias de Chile tiene un resultado de 13, 226, 208 es demasiada información que si es soportada por el rapidminer pero no los requerimientos de hardware en el laptop donde fue implementado.



Tabla 6

*Requerimientos de hardware del proyecto*

Rapidminer	R Studio Desktop	Hardware de laptop donde se implementara
<p>Mínimo</p> <ul style="list-style-type: none"> <li>• Dual core</li> <li>• 2GHz processor</li> <li>• 4GB RAM</li> <li>• &gt;1GB free disk space</li> <li>• Resolución: 1280x1024</li> </ul> <p>Recomendado</p> <ul style="list-style-type: none"> <li>• Quad core</li> <li>• 3GHz or faster processor</li> <li>• 16GB RAM</li> <li>• &gt;100GB free disk space</li> </ul> <p>Sistema operativo</p> <ul style="list-style-type: none"> <li>• Windows 7, Windows 8, Windows 8.1, Windows 10 (64-bit highly recommended)</li> <li>• Linux (64-bit only)</li> <li>• MacOS X 10.10 - 10.14</li> </ul> <p>Java plataforma</p> <ul style="list-style-type: none"> <li>• 64-bit highly recommended</li> <li>• Oracle Java 8 (GmbH, s.f.)</li> </ul>	<p>Supported Windows Versiones Windows 7+</p> <p>Recommended system requirements</p> <p>4+ CPUs 8+ GB of RAM 100+ GB of disk storage</p> <p>(RStudio Server Pro Administrator's Guide, 2016)</p>	<p>MODELO 1: Laptop Toshiba Satellite C845-SP4334SL Procesador: INTEL (R) CORE(TM) i3-2330M CPU @ 2.20 GHz RAM instalada: 4.00 GB WINDOWS:10 x64</p> <p>MODELO 2: Laptop ASUS X550L Procesador: INTEL (R) CORE(TM) i5-M CPU @ 1.60 GHz RAM instalada: 6.00 GB WINDOWS:8.1 x64</p>

#### 4.4. Fase 4: Modelado de datos.

##### 4.4.1. Selección de técnicas de modelado

Dependiendo de nuestro estudio hemos tratado de utilizar Rapiminer y R Studio con sus respectivos algoritmos de predicción para realizar nuestro modelado para manejar un nuevo modelo de predicción que son la Regresión Lineal para encontrar la probabilidad de Holt-Winter's, la Serie de tiempo y la Regresión Logística.

##### 4.4.2. Elaboración del diseño de la prueba.

El diseño de nuestra prueba a realizar tendrá las siguientes consideraciones: Será de Chile donde se apoyó previamente la cantidad de datos para 2010 a 2019, esto para comprobar hasta qué punto nuestro modelo puede predecir la cantidad de atención de emergencia. A continuación, se cargaron los 13 millones de registros en el software, por lo que se mostró el siguiente mensaje.

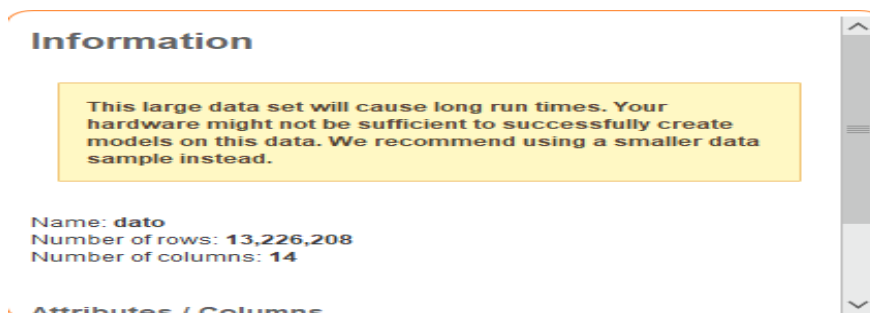


Figura 36 Cargar de datos SQL Server en Rapiminer Tools

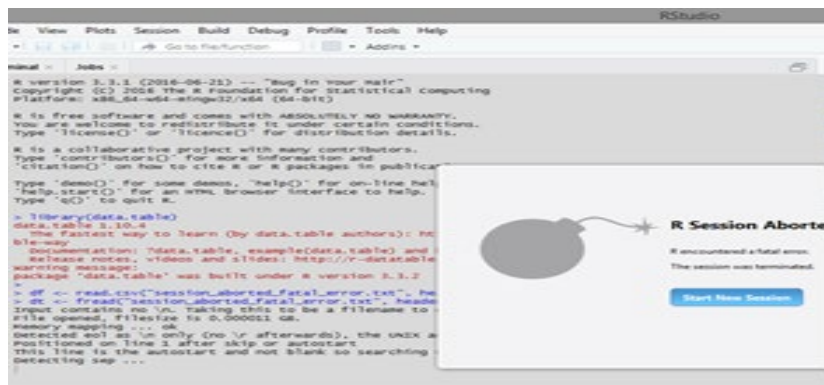


Figura 37 Cargar los datos SQL Server al RStudio

### 4.4.3. Construcción del modelo.

Después de haber hecho la comprensión y la selección de los datos, y de haber establecido el algoritmo a utilizar, exportan los datos de nuestra base de datos. El nuevo modelo cubre la cantidad de 2014 al 2018 correspondiente a la región de "Antofagasta", estos datos serán utilizados para crear el modelo, es decir, el aprendizaje. El resto se utilizará como datos para la validación del modelo. Comenzaremos explicando lo que sucederá internamente cuando insertemos nuestros datos en el Rapiminer y lo que resultará.

A continuación, nos muestra las variables de nuestra base de datos y pasamos a la opción de predecir, que nos pregunta ¿cuál es la columna de predicción? donde la variable a predecir es CAtenciones.

**Attributes / Columns**

IDTiempo, IDEstablishimiento, IDPadecciminetto, IDMedicamiento, IDTipodeAtencion, IDCampaña, CAtenciones, CAtenEdad0, CAtenEdad14, CAtenEdad514, CAtenEdad65, CMedicamiento

Load Data | **Select Task** | Prepare Target | Select Inputs | Model Types | Results

« RESTART < BACK > NEXT

**Predict**  
Want to predict the values of a column?

**Clusters**  
Want to identify groups in your data?

**Outliers**  
Want to detect outliers in your data?

IDTiempo Number	IDEstablishim... Category	IDEstablishim... Number	IDPadecciminetto Number	IDMedicamiento Number	IDTipodeAtencion Number	IDCampaña Number	<b>CAtenciones Number</b>	CAtenEdad0 Number	CAtenEdad14 Number	CAtenEdad514 Number	CAtenEdad65 Number
20100402	22-304	2	10	42	2	5	6	0	1	1	4
20100403	22-304	2	10	15	2	5	18	2	6	2	8
20100404	22-304	2	10	29	2	5	17	2	2	3	10
20100405	22-304	2	10	21	2	5	21	2	4	4	11
20100406	22-304	2	10	17	2	5	10	1	4	1	4
20100407	22-304	2	10	21	2	5	21	1	2	5	12
20100408	22-304	2	10	34	2	5	3	0	1	1	1
20100409	22-304	2	10	20	2	5	10	1	2	0	16

13,226,268 rows - 14 columns (1 nominal, 13 numerical)

Figura 38 Selección de la tarea a ejecutar en Rapiminer Tools

A continuación, se muestra cada variable con su grado de aceptación entre las variables.

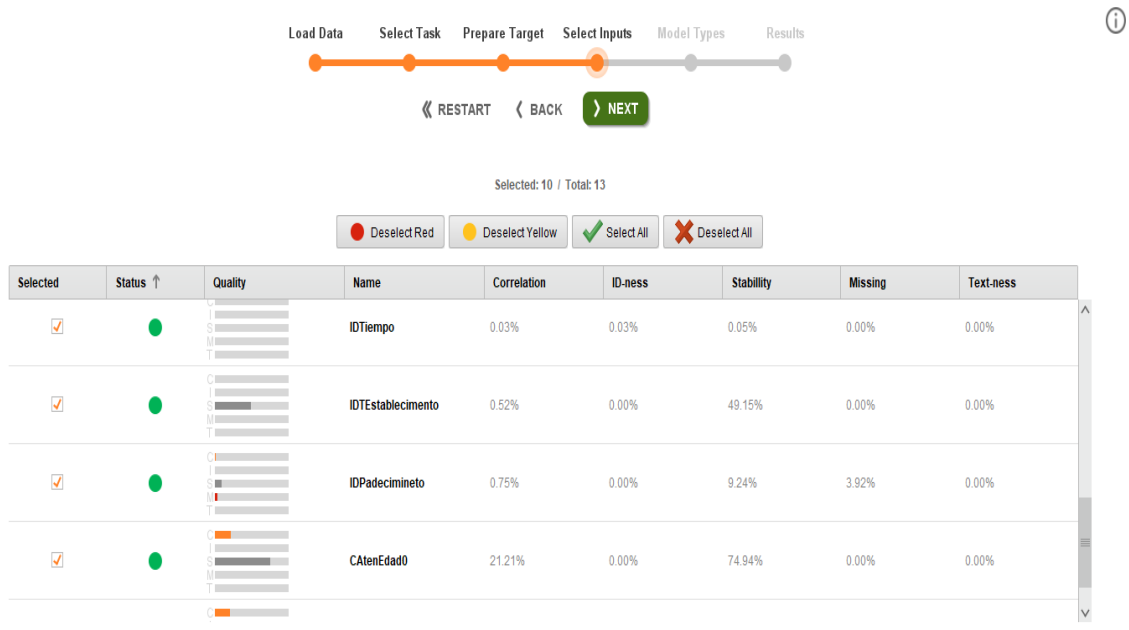


Figura 39 Selección de las entradas de la base de datos

Entonces se elige el modelo de regresión lineal.

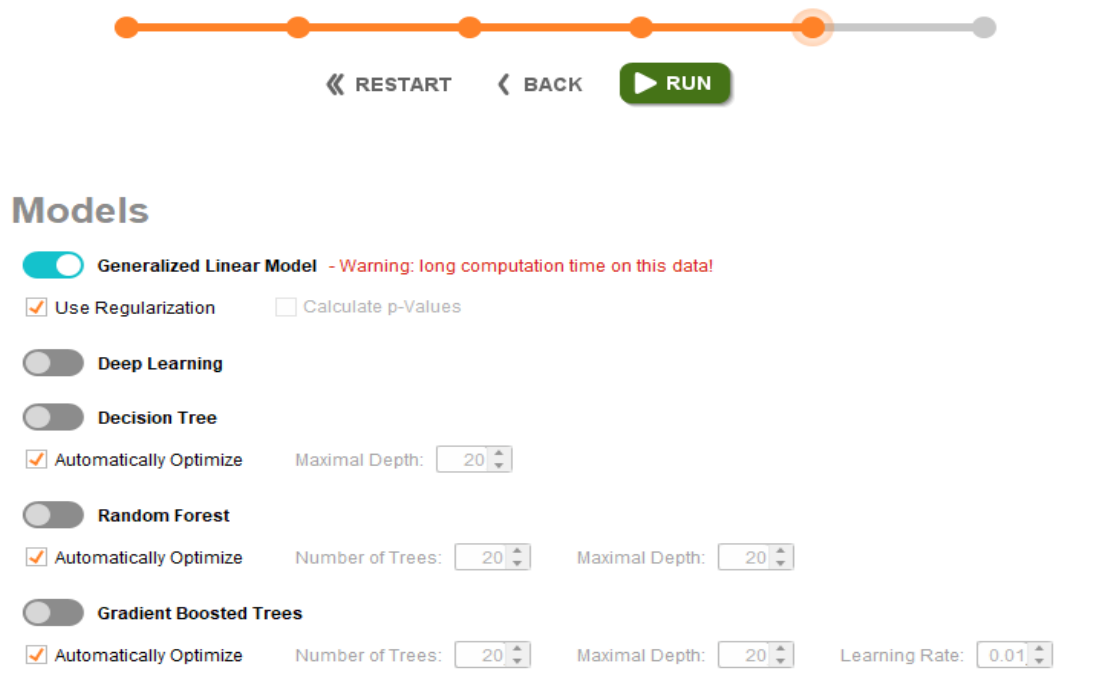


Figura 40 Selección del modelo de predicción

La conclusión nos muestra que no hay suficiente memoria para ejecutar el modelo.

## Generalized Linear Model - Model

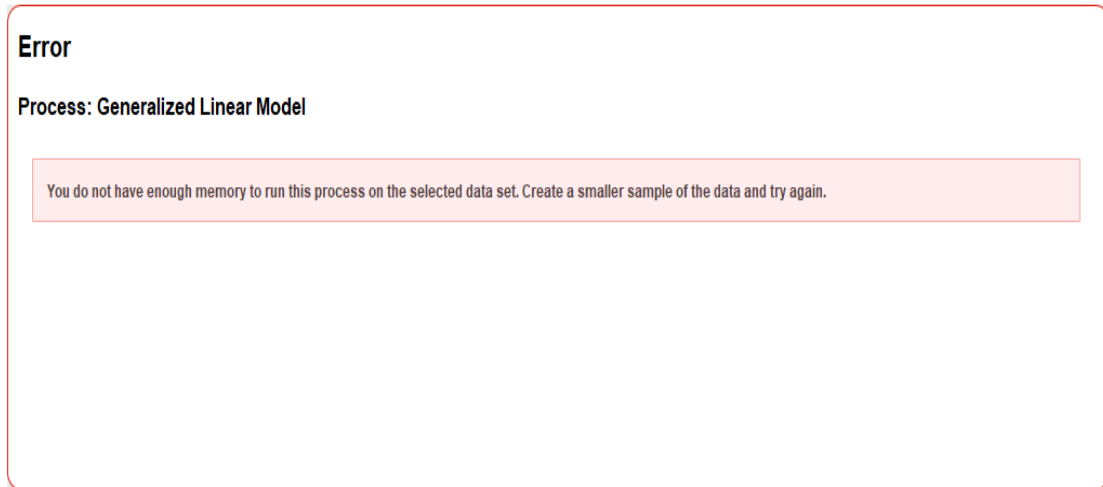


Figura 41 Resultado de la carga de Atenciones de urgencia

Ahora usamos la siguiente herramienta de RStudio para generar nuestro modelo.

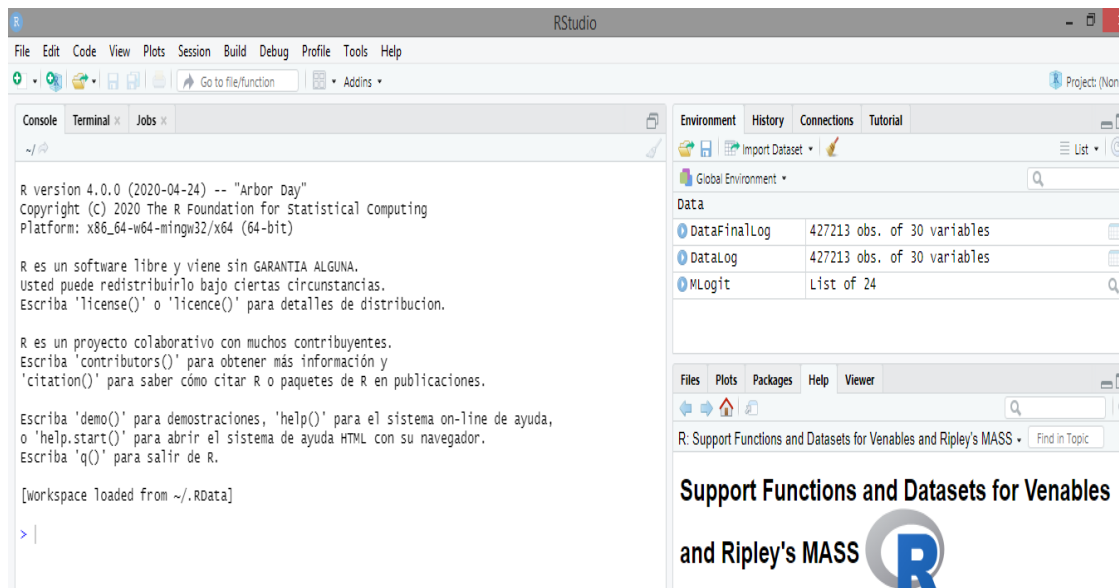


Figura 42 Pantalla principal del Rstudio

Importamos nuestra base de datos en formato .CSV previamente extraído de SQL Server para mejorar el rendimiento del procesamiento.

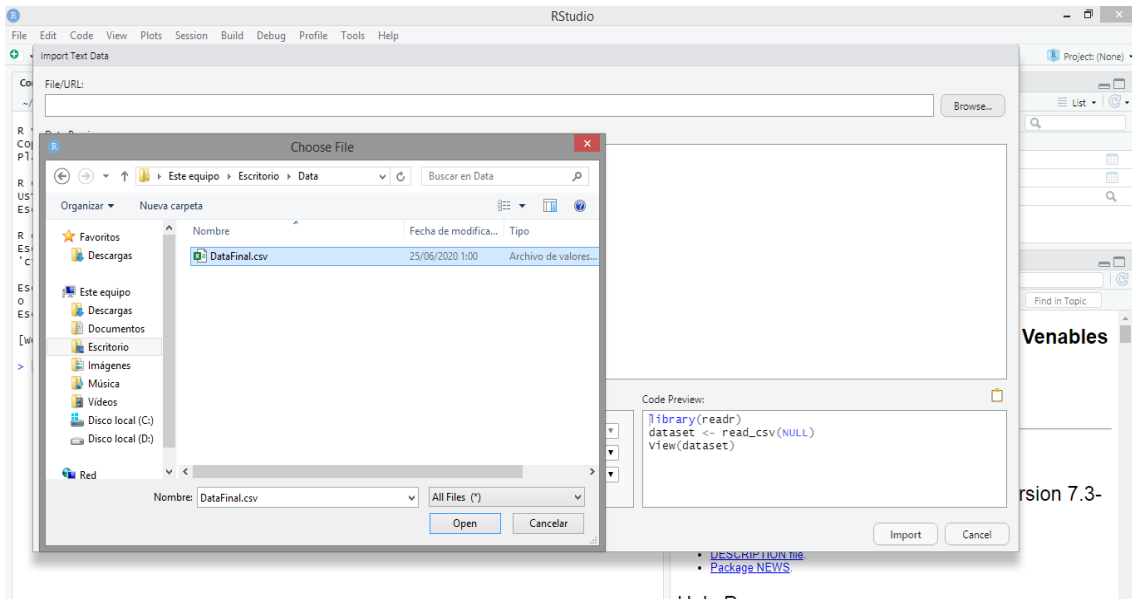


Figura 43 Selección de la base de datos de atenciones de urgencia de Antofagasta

Luego nos muestra una ventana que carga la base de datos en una sola columna delimitada por un punto y coma, para ello cambiamos la opción de delimitación.

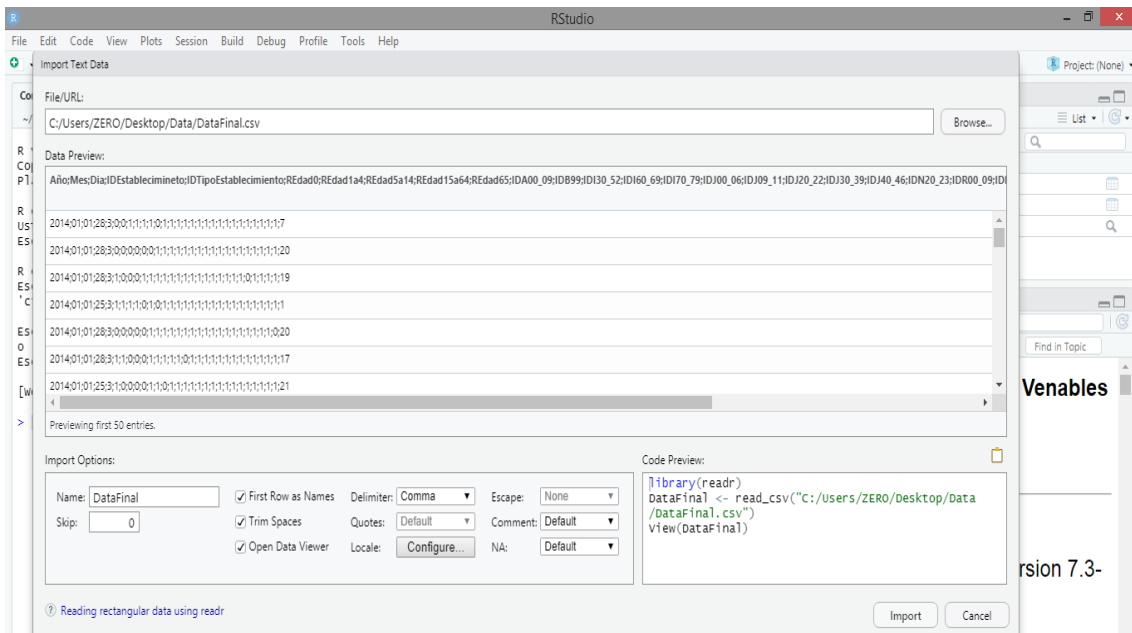


Figura 44 Configuración de la base de datos en RStudio

Luego cambia la variable de doble a numérica y la importa.

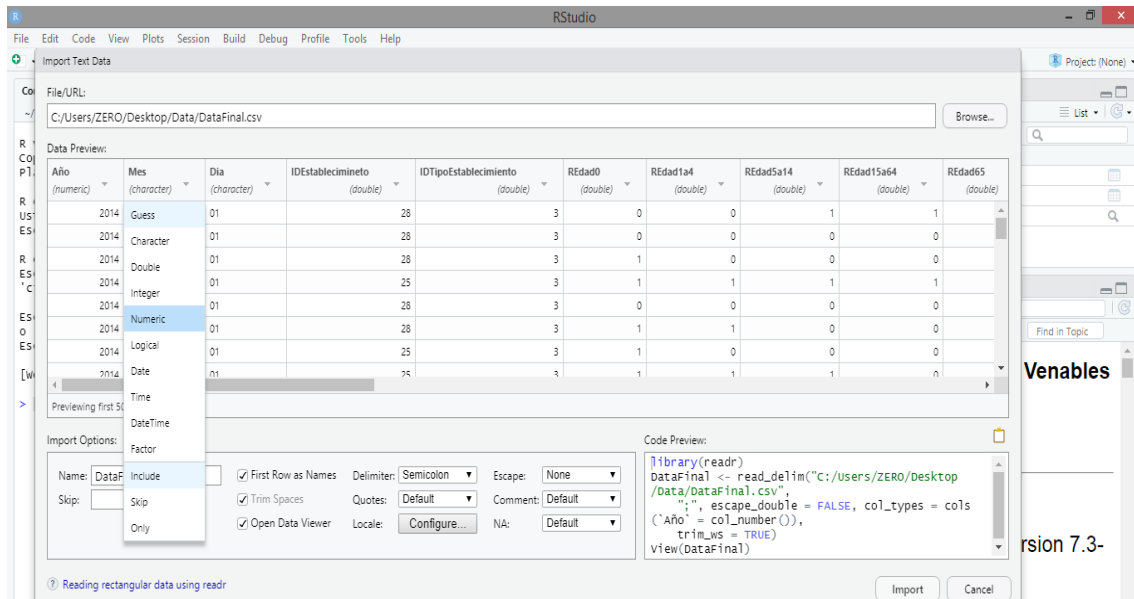


Figura 45 Conversión de variables de la base de datos a RStudio

Para concluir, poseemos los datos importados para generar nuestros modelos.

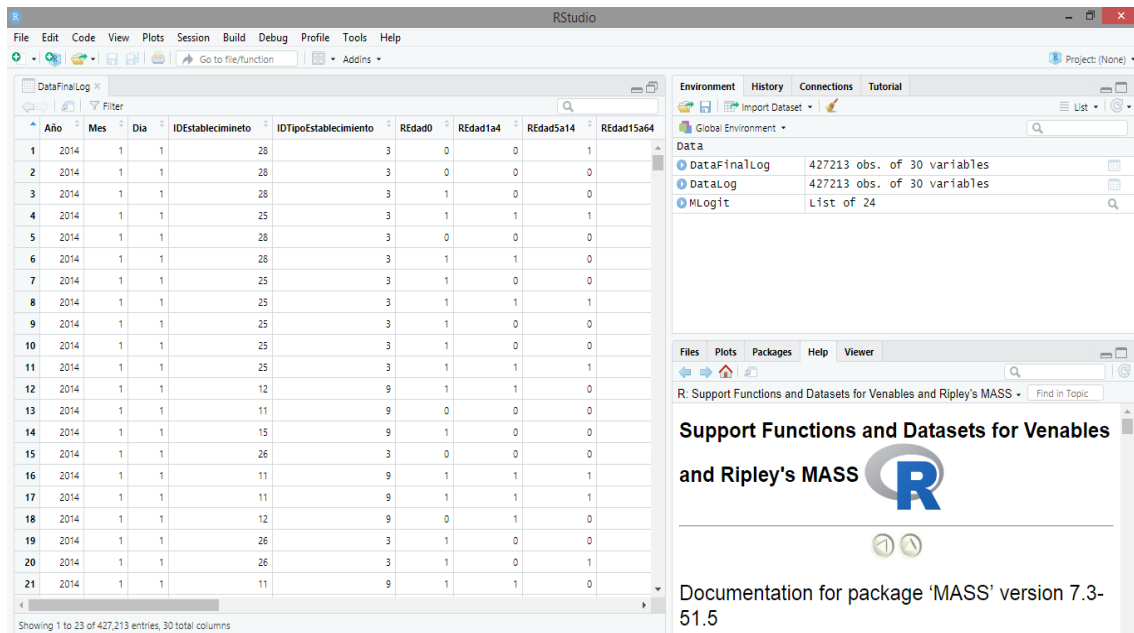


Figura 46 Vista de los datos cargados en RStudio

Continuando con la investigación, se desarrollaron divisiones técnicas para construir y generar un modelo de predicción para la atención de emergencias en la región de Antofagasta.

#### 4.4.4. Regresión lineal múltiple

El comando básico es lm (modelos lineales). El primer argumento de este comando es una fórmula  $y \sim x$  en la que se especifica cuál es la respuesta o variable dependiente (y) y cuál es la variable de regresión o independiente (x). El segundo argumento, llamado datos, especifica el archivo en el que se encuentran las variables. El resultado se almacena en un objeto llamado regresión. Este objeto es una lista que contiene toda la información relevante sobre el análisis.

Los modelos lineales múltiples siguen la siguiente ecuación:

$$Y = \beta_0 + \sum_{i=1}^7 \beta_i X_i + \varepsilon$$

Figura 47 *Formula de regresión lineal multiple*

Nota: Fórmula utilizada para medir el crecimiento de cada variable utilizada en la regresión lineal múltiple. Tomado de comparación de las capacidades de predicción de varios algoritmos de minería de datos y de regresión lineal múltiple en la predicción del peso corporal por medio de mediciones del cuerpo en la cabra Beetal autóctona del Pakistán, por Eyduran et al., 2017, Pakistan Journal of Zoology.

$\beta_0$ : es el orden en el origen, el valor de la variable dependiente Y cuando todos los predictores son cero.

$\beta_i$ : es el efecto promedio que el aumento de una unidad de la variable predictora  $X_i$  tiene sobre la variable dependiente Y, manteniendo constante el resto de la variable. Se conocen como coeficientes de regresión parcial.

$\varepsilon_i$ : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

Interpretación:

$NAtencion = (\beta_0 + \beta_1Año + \beta_2Mes + \beta_3Dia + \beta_4IDEstablecimiento + \beta_5IDTipoEstablecimiento + \beta_6Edad0 + \beta_7Edad1a4 + \beta_8Edad5a14 + \beta_9Edad15a64 + \beta_{10}Edad65 + \beta_{11}A00_09 + \beta_{12}IDB99 + \beta_{13}IDI30_52 + \beta_{14}IDI60_69 + \beta_{15}IDI70_79 + \beta_{16}IDJ00_06 + \beta_{17}IDJ09_11 + \beta_{18}IDJ20_22 +$



$$\beta_{19}IDJ30_{39} + \beta_{20}IDJ40_{46} + \beta_{21}IDN20_{23} + \beta_{22}IDR00_{09} + \beta_{23}IDR50_{69} + \beta_{24}IDS00_{09} + \beta_{25}IDT14 + \beta_{26}IDT20_{31} + \beta_{27}IDT51_{65} + \beta_{28}IDT66_{78} + \beta_{29}IDV00_{99}) + e_i$$

$\beta_0$ : es el orden en el origen, el valor de la variable dependiente NAtenciones cuando todos los predictores son cero.

$\beta_i$ : es el efecto promedio que tiene el incremento en una unidad de las variables predictoras Año, Mes, Día, IDEstablecimiento, IDTipoEstablecimiento, REdad0, REdad1a4, REdad5a14, REdad15a64, REdad65, IDA00\_09, IDB99, IDI30\_52, IDI60\_69, IDI70\_79, IDJ00\_06, IDJ09\_11, IDJ20\_22, IDJ30\_39, IDJ40\_46, IDN20\_23 + IDR00\_09, IDR50\_69, IDS00\_09, IDT14, IDT20\_31, IDT51\_65 + IDT66\_78, IDV00\_99. tiene en la variable dependiente CAtenciones, manteniendo el resto de la variable constante. Estos se conocen como coeficientes de regresión parcial.

$e_i$ : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

Usando el comando summary, obtenemos un resumen de los principales resultados:

la biblioteca lmtest es solicitada ya que contiene todos los algoritmos de aplicación para nuestro modelo de regresión lineal múltiple:

```
modeloRegresion <- lm (NAtenciones ~ Año + Mes + Dia
+IDestablecimiento +IDTipoEstablecimiento + REdad0 +REdad1a4
+REdad5a14 +REdad15a64 +REdad65 + IDA00_09 + IDB99 + IDI30_52 +
IDI60_69 + IDI70_79 + IDJ00_06 + IDJ09_11 + IDJ20_22+ IDJ30_39 + IDJ40_46
+ IDN20_23 + IDR00_09 + IDR50_69 + IDS00_09 + IDT14 + IDT20_31 +
IDT51_65 + IDT66_78, data = DataFinal)
```

Tabla 7

*Tabla resumen de residuos regresión lineal múltiple*

<b>Residuales:</b>				
<b>Min</b>	1Q	Median	3Q	Max
<b>-21.268</b>	-3.257	-0.280	2.046	38.616

Tabla 8

*Tabla resumen de coeficientes de la regresión lineal múltiple de la variable cantidad de atenciones de urgencia*

	Estimate	Beta	Std. Error	t value	Pr(> t )	Sig	VIF
(Intercept)	-1.4594869	0.0000	12.6007118	-0.116	0.90779	0.1	
Año	0.0178814	0.0025	0.0062436	2864	0.00418	0.001	1.006194
Mes	0.0198681	0.0068	0.0025607	7759	8.59e-15	0.0001	1.003133
Día	0.0009847	0.0009	0.0009953	0.989	0.32250	0.1	1.000242
IDEstablecimiento	0.0325991	0.0241	0.0013206	24685	< 2e-16	0.0001	1.236074
IDTipoEstablecimiento	-0.1604615	-0.0474	0.0034540	-46456	< 2e-16	0.0001	1.343587
REdad0	-4.5149389	-0.1865	0.0247558	-182379	< 2e-16	0.0001	1.350510
REdad1a4	-4.5537918	-0.2256	0.0229107	-198.762	< 2e-16	0.0001	1.665012
REdad5a14	-5.0601985	-0.2524	0.0224630	-225.268	< 2e-16	0.0001	1.621358
REdad15a64	-4.7678425	-0.1901	0.0238145	-200.207	< 2e-16	0.0001	1.164551
REdad65	-4.8856365	-0.2434	0.0189241	-258.170	< 2e-16	0.0001	1.148175
IDA00_09	-0.8827321	-0.0334	0.0378449	-23.325	< 2e-16	0.0001	2.644981
IDB99	2.5485826	0.0764	0.0422447	60.329	< 2e-16	0.0001	2.072265
IDI30_52	-6.8661402	-0.2292	0.0418494	-164.068	< 2e-16	0.0001	2.521315
IDI60_69	0.3859697	0.00598	0.0659129	5.856	4.75e-09	0.0001	1.346076
IDI70_79	-4.1835000	-0.0694	0.0620182	-67.456	< 2e-16	0.0001	1.366559
IDJ00_06	0.9855268	0.0268	0.0440388	22.379	< 2e-16	0.0001	1.849315
IDJ09_11	0.5340990	0.0023	0.2060414	2.592	0.00954	0.001	1.022328
IDJ20_22	-5.4595582	-0.0927	0.0612016	-89.206	< 2e-16	0.0001	1.395306
IDJ30_39	-0.8485834	-0.0120	0.0705741	-12.024	< 2e-16	0.0001	1.286873
IDJ40_46	0.0625249	0.0008	0.0760538	0.822	0.41101	0.1	1.198670

	Estimate	Beta	Std. Error	t value	Pr(> t )	Sig	VIF
IDN20_23	0.2045455	0.0023	0.0841610	2.430	0.01508	0.05	1.184069
IDR00_09	-0.6720849	-0.0168	0.0477315	-14.081	< 2e-16	0.0001	1.834480
IDR50_69	0.8748128	0.0223	0.0466082	18.769	< 2e-16	0.0001	1.819056
IDS00_09	0.6889223	0.0029	0.2086284	3.302	0.00096	0.0001	1.021707
IDT14	-3.6703185	-0.1046	0.0445986	-82.297	< 2e-16	0.0001	2.087613
IDT20_31	1.3841662	0.0352	0.0464795	29.780	< 2e-16	0.0001	1.802542
IDT51_65	0.5243555	0.0022	0.2090553	2.508	0.01213	0.05	1.021908
IDT66_78	0.0204965	0.0002	0.0803020	0.255	0.79854		1.206117

Terminando la regresión lineal múltiple, la interpretación de las variables evaluando el coeficiente nos da los siguientes resultados.

$$\begin{aligned}
 \text{NAtenciones} = & -1.4594869 + 0.0178814 * \text{Año} + 0.0198681 * \text{Mes} + \\
 & 0.0009847 * \text{Days} + 0.0325991 * \text{IDEstablecimiento} - 0.1604615 * \\
 & \text{IDTipoEstablecimiento} - 4.5149389 * \text{REdad0} - 4.5537918 * \text{REdad1a4} - \\
 & 5.0601985 * \text{REdad5a14} - 4.7678425 * \text{REdad15a64} - 4.8856365 * \\
 & \text{REdad65} - 0.8827321 * \text{A00}_09 + 2.5485826 * \text{IDB99} - 6.8661402 * \\
 & \text{IDI30}_52 + 0.3859697 * \text{IDI60}_69 - 4.1835000 * \text{IDI70}_79 + 0.9855268 * \\
 & \text{IDJ00}_06 + 0.5340990 * \text{IDJ09}_11 - 5.4595582 * \text{IDJ20}_22 - 0.8485834 * \\
 & \text{IDJ30}_39 + 0.0625249 * \text{IDJ40}_46 + 0.2045455 * \text{IDN20}_23 - 0.6720849 \\
 & * \text{IDR00}_09 + 0.8748128 * \text{IDR50}_69 + 0.6889223 * \text{IDS00}_09 - \\
 & 3.6703185 * \text{IDT14} + 1.3841662 * \text{IDT20}_31 + 0.5243555 * \text{IDT51}_65 + \\
 & 0.0204965 * \text{IDT66}_78
 \end{aligned}$$

Tabla 9

*Descripción de las variables regresión lineal múltiple*

VARIABLES	DESVIACION ESTANDAR
NAtenciones	9.957531
Año	1.407437
Mes	3.426455
Dia	8.802594
IDEstablecimiento	7.375278
IDTipoEstablecimiento	2.939884
REdad0	0.4112398
REdad1a4	0.4933928
REdad5a14	0.4965873
REdad15a64	0.396973
REdad65	0.4960347
IDA00_09	0.3764676
IDB99	0.2985206
IDI30_52	0.3323902
IDI60_69	0.1542011
IDI70_79	0.1651271
IDJ00_06	0.2705163
IDJ09_11	0.04298967
IDJ20_22	0.1690812
IDJ30_39	0.1408141
IDJ40_46	0.1261109

<b>VARIABLES</b>	<b>DESVIACIÓN ESTÁNDAR</b>
IDN20_23	0.1132664
IDR00_09	0.2485847
IDR50_69	0.2535034
IDS00_09	0.0424437
IDT14	0.2838094
IDT20_31	0.2530492
IDT51_65	0.04236119
IDT66_78	0.1198097

Tabla 10

*Tabla resumen de interpretación de regresión lineal múltiple con la variable cantidad de atenciones*

VARIABLE	B	BETA	INTERPRETACIÓN
<b>B<sub>0</sub></b>	-14.594.869	0.0000	Es el punto donde la ecuación cruza el eje Y
<b>Año</b>	0.0178814	0.0025	Cuando se produce un incremento de un año con una desviación estándar (1.407), las atenciones de urgencias aumenta en un 0,25% de desviación estándar de la cantidad de atención 9.958, entonces esto constituye un cambio de 14.011 (1.407 * 9.958) manteniendo constantes las otras condiciones.
<b>Mes</b>	0.0198681	0.0068	Cuando se produce un incremento de un mes con una desviación estándar (3.426), las atenciones de urgencias aumenta en un 0,68% de desviación estándar del número de atenciones 9.958, entonces esto constituye un cambio de 34.116 (3.426 * 9.958) manteniendo constantes las otras condiciones.
<b>Día</b>	0.0009847	0.0009	Cuando hay un incremento de un día con una desviación estándar (8.803), las atenciones de urgencias aumenta en un 0,09% de desviación estándar de la cantidad de atención 9.958, entonces esto constituye un cambio de 87.6603 (8.803 * 9.958) manteniendo las otras condiciones constantes.
<b>IDestablecimiento</b>	0.0325991	0.0241	Cuando se produce un aumento en un establecimiento con una desviación estándar (7.375), las atenciones de urgencias aumenta en un 2,4% de desviación estándar de la cantidad de atención 9.958, entonces esto constituye un cambio de 73.4403 (7.375 * 9.958) manteniendo constantes las otras condiciones.
<b>IDTipoEstablecimiento</b>	-0.1604615	-0.0474	Cuando se produce un aumento por tipo de centro con una desviación estándar (2.9399), las atenciones de urgencias disminuye en un 4,7% las desviaciones estándar del número de pacientes 9.958, entonces esto constituye un cambio de 29.276 (2.9399 * 9.958) manteniendo constantes las otras condiciones.
<b>REdad0</b>	-45.149.389	-0.1865	Cuando existe un rango de edad inferior a 11 meses con una desviación estándar (0,411), las atenciones de urgencias disminuye en un 18% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 4.391 (0,441 * 9.958), manteniendo las demás condiciones constantes.

VARIABLE	B	BETA	INTERPRETACIÓN
<b>REdad1a4</b>	-45.537.918	-0.2256	Cuando existe un rango de edad de 1 a 4 años con una desviación estándar (0,493), las atenciones de urgencias disminuye en un 22,5% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 4.909 (0,493 * 9.958) manteniendo las otras condiciones constantes.
<b>REdad5a14</b>	-50.601.985	-0.2524	Cuando existe la presencia de un rango de edad de 5 de 14 años con una desviación estándar (0,497), las atenciones de urgencias disminuye en un 25,2% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 4.949 (0,497 * 9.958) manteniendo las otras condiciones constantes.
<b>REdad15a64</b>	-47.678.425	-0.1901	Cuando existe un rango de edad de 15 a 64 años con una desviación estándar (0,397), las atenciones de urgencias disminuye en un 19% las desviaciones estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 3.953 (0,397 * 9.958) manteniendo las otras condiciones constantes.
<b>REdad65</b>	-48.856.365	-0.2434	Cuando existe un rango de edad de 65 años y más con una desviación estándar (0,496), la atenciones de urgencias disminuye en un 24,3% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 4.939 (0,496 * 9.958), manteniendo las demás condiciones constantes.
<b>IDA00_09</b>	-0.8827321	-0.0334	Cuando se produce un aumento de la condición IDA00_09 con una desviación estándar (0,3765), las atenciones de urgencias disminuye en un 3,34% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 3,7492 (0,3765 * 9,958) manteniendo constantes las demás condiciones.
<b>IDB99</b>	25.485.826	0.0764	Cuando se produce un aumento de la enfermedad IDB99 con una desviación estándar (0,299), las atenciones de urgencias aumenta en un 7,64% las desviaciones estándar del número de pacientes 9.958, entonces esto constituye un cambio de 2.9774 (0,299 * 9.958) manteniendo constantes las otras condiciones.
<b>IDI30_52</b>	-68.661.402	-0.2292	Cuando se produce un aumento de la enfermedad IDI30_52 con una desviación estándar (0,3324), las atenciones de urgencias disminuyen en un 22,9% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 3,31004 (0,3324 * 9,958) manteniendo constantes las otras condiciones.

<b>VARIABLE</b>	<b>B</b>	<b>BETA</b>	<b>INTERPRETACIÓN</b>
<b>IDI60_69</b>	0.3859697	0.00598	Cuando se produce un aumento de la enfermedad IDI60_69 con una desviación estándar (0,15420), las atenciones de urgencias aumenta en un 0,598% de desviación estándar para el número de pacientes 9,958, por lo que esto constituye un cambio de 1,5355 (0,15420 * 9,958) manteniendo constantes las demás condiciones.
<b>IDI70_79</b>	-41.835.000	-0.0694	Cuando se produce un aumento de la enfermedad IDI70_79 con una desviación estándar (0,16513), las atenciones de urgencias disminuye en un 6,94% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 1.64436 (0,16513 * 9.958) manteniendo constantes las demás condiciones.
<b>IDJ00_06</b>	0.9855268	0.0268	Cuando se produce un aumento de la condición IDJ00_06 con una desviación estándar (0,2705), las atenciones de urgencias aumenta en un 2,68% las desviaciones estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 2,6936 (0,2705 * 9,958) manteniendo constantes las otras condiciones.
<b>IDJ09_11</b>	0.5340990	0.0023	Cuando se produce un aumento de la condición IDJ09_11 con una desviación estándar (0,04299), las atenciones de urgencias aumenta en un 0,23% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 0,42809 (0,04299 * 9.958) manteniendo constantes las demás condiciones.
<b>IDJ20_22</b>	-54.595.582	-0.0927	Cuando se produce un aumento de la condición IDJ20_22 con una desviación estándar (0,1691), las atenciones de urgencias disminuye en un 9,27% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 1.6839 (0,1691 * 9.958) manteniendo constantes las demás condiciones.
<b>IDJ30_39</b>	-0.8485834	-0.0120	Cuando se produce un aumento de la condición IDJ30_39 con una desviación estándar (0,14081), las atenciones de urgencias disminuye en un 1,20% las desviaciones estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 1,4022 (0,14081 * 9.958) manteniendo constantes las demás condiciones.
<b>IDJ40_46</b>	0.0625249	0.0008	Cuando se produce un aumento de la condición IDJ40_46 con una desviación estándar (0,1261), las atenciones de urgencias aumenta en un 0,08% las desviaciones estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 1,2557 (0,1261 * 9,958) manteniendo constantes las demás condiciones.



VARIABLE	B	BETA	INTERPRETACIÓN
<b>IDN20_23</b>	0.2045455	0.0023	Cuando se produce un aumento de la enfermedad IDN20_23 con una desviación estándar (0,1133), las atenciones de urgencias aumenta en un 0,23% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 1.1282 (0,1133 * 9.958) manteniendo constantes las demás condiciones.
<b>IDR00_09</b>	-0.6720849	-0.0168	Cuando se produce un aumento de la condición IDR00_09 con una desviación estándar (0,2486), las atenciones de urgencias disminuye en un 1,68% las desviaciones estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 2.47556 (0,2486 * 9.958) manteniendo constantes las demás condiciones.
<b>IDR50_69</b>	0.8748128	0.0223	Cuando se produce un aumento de la enfermedad IDR50_69 con una desviación estándar (0,2535), las atenciones de urgencias aumenta en un 2,23% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 2.5244 (0,2535 * 9.958) manteniendo constantes las demás condiciones.
<b>IDS00_09</b>	0.6889223	0.0029	Cuando se produce un incremento de la condición IDS00_09 con una desviación estándar (0,0424), las atenciones de urgencias aumenta en un 0,29% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 0,4222 (0,0424 * 9.958) manteniendo constantes las demás condiciones.
<b>IDT14</b>	-36.703.185	-0.1046	Cuando hay un aumento de la enfermedad IDT14 con una desviación estándar (0,2838 las atenciones de urgencias disminuye en un 10,46% las desviaciones estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 2.8261 (0,2838 * 9.958) manteniendo constantes las otras condiciones.
<b>IDT20_31</b>	13.841.662	0.0352	Cuando se produce un aumento de la enfermedad IDT20_31 con una desviación estándar (0,2530), las atenciones de urgencias aumenta en un 3,52% las desviaciones estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 2.5194 (0,2530 * 9.958) manteniendo constantes las demás condiciones.
<b>IDT51_65</b>	0.5243555	0.0022	Cuando se produce un aumento de la enfermedad IDT51_65 con una desviación estándar (0,0424), las atenciones de urgencias aumenta en un 0,22% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 0,4222 (0,0424 * 9.958) manteniendo constantes las demás condiciones.

VARIABLE	B	BETA	INTERPRETACIÓN
IDT66_78	0.0204965	0.0002	Cuando se produce un aumento de la enfermedad IDT66_78 con una desviación estándar (0,1198), las atenciones de urgencias aumenta en un 0,02% de desviación estándar para el número de pacientes 9.958, por lo que esto constituye un cambio de 1.1282 (0,1198 * 9.958) manteniendo constantes las demás condiciones.

Tabla 11

*Tabla resumen de la regresión lineal múltiple*

<b>Residual standard error:</b>	<b>5.726 on 427184 degrees of freedom</b>
<b>Multiple R-squared:</b>	0.6694
<b>Adjusted R-squared:</b>	0.6693
<b>F-statistic:</b>	3.089e+04 on 28 and 427184 DF
<b>p-value:</b>	< 2.2e-16
<b>Durbin-Watson test</b>	19.842
<b>p-value =</b>	1,23E-04
<b>data:</b>	DataFinal

### Tabla ANOVA

Cuando se informe de los resultados de un ANOVA, debe indicarse el valor obtenido para la estadística F, los grados de libertad, el valor p y el tamaño del efecto ( $\eta^2$ ). Puede ocurrir que un análisis del ANOVA indique que hay diferencias significativas entre las medias y entonces las pruebas t no encuentren ninguna comparación que lo sea. No es contradictorio; es simplemente porque se trata de dos técnicas diferentes.

Tabla 12

*Tabla resumen ANOVA regresión lineal múltiple*

	DF	Sum Sq	Mean Sq	F value	Pr(>F)	Sig
<b>Año</b>	1	21	21	6.500e-01	0.4200	0.01
<b>Mes</b>	1	4258	4258	1.299e+02	< 2e-16	0.0001
<b>Día</b>	1	61	61	1.871e+00	0.1714	0.01
<b>IDEstablecimiento</b>	1	44384	44384	1.354e+03	< 2e-16	0.0001
<b>IDTipoEstablecimiento</b>	1	689414	689414	2.103e+04	< 2e-16	0.0001
<b>REdad0</b>	1	9689132	9689132	2.955e+05	< 2e-16	0.0001
<b>REdad1a4</b>	1	6019386	6019386	1.836e+05	< 2e-16	0.0001
<b>REdad5a14</b>	1	3494002	3494002	1.066e+05	< 2e-16	0.0001
<b>REdad15a64</b>	1	1805942	1805942	5.508e+04	< 2e-16	0.0001
<b>REdad65</b>	1	3864798	3864798	1.179e+05	< 2e-16	0.0001
<b>IDA00_09</b>	1	93	93	2.823e+00	0.0929	0.1
<b>IDB99</b>	1	603379	603379	1.840e+04	< 2e-16	0.0001
<b>IDI30_52</b>	1	1133718	1133718	3.458e+04	< 2e-16	0.0001
<b>IDI60_69</b>	1	12192	12192	3.719e+02	< 2e-16	0.0001
<b>IDI70_79</b>	1	131981	131981	4.026e+03	< 2e-16	0.0001
<b>IDJ00_06</b>	1	103758	103758	3.165e+03	< 2e-16	0.0001
<b>IDJ09_11</b>	1	1451	1451	4.426e+01	2.88e-11	0.0001

	DF	SUM SQ	MEAN SQ	F VALUE	PR(>F)	SIG
IDJ20_22	1	285403	285403	8.705e+03	< 2e-16	0.0001
IDJ30_39	1	4214	4214	1.285e+02	< 2e-16	0.0001
IDJ40_46	1	1223	1223	3.731e+01	1.01e-09	0.0001
IDN20_23	1	556	556	1.695e+01	3.84e-05	0.0001
IDR00_09	1	9378	9378	2.860e+02	< 2e-16	0.0001
IDR50_69	1	42091	42091	1.284e+03	< 2e-16	0.0001
IDS00_09	1	1591	1591	4.853e+01	3.26e-12	0.0001
IDT14	1	379423	379423	1.157e+04	< 2e-16	0.0001
IDT20_31	1	31338	31338	9.558e+02	< 2e-16	0.0001
IDT51_65	1	204	204	6.236e+00	0.0125	0.05
IDT66_78	1	2	2	6.500e-02	0.7985	0.1
<b>Residuales:</b>	427184	14005714	33			

La tabla presentada nos da como resultado un Resultado Sig de 0.00015 que es la significación de todas las variables aplicadas en el modelo de regresión lineal múltiple y esto nos da el error estándar de 5.72630

### Preparando los datos para la frecuencia Tseries:

Cargando la base de datos de la cantidad de cuidados de emergencia, para hacer la agrupación año por mes.

```
#TIME SERIES
DA=ts(timeseries[,2], start = c(2014, 1), end = c(2018, 12), frequency = 12)
DA
autoplot(DA) # Histograma
```

Figura 48 Resultados de atenciones de urgencias por meses en RStudio.

Nota. Se agrupa la data correspondiente a cada Año, para generar la cantidad de atenciones de urgencia por Año.

Tabla 13

Total de Atenciones de urgencia de Antofagasta por meses

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
<b>2014</b>	47699	42000	48358	48555	49475	49106	52287	53400	50897	52917	49825	48754
<b>2015</b>	46645	43273	50922	49523	50733	49709	53885	50466	47191	50095	51958	52245
<b>2016</b>	48823	43541	50388	47209	49621	49758	52494	53537	51476	54078	50066	49597
<b>2017</b>	49077	44897	48837	49452	53350	52442	53963	56512	56005	54760	54243	52386
<b>2018</b>	45860	42163	51493	51290	53495	51939	54626	53314	53191	54580	51854	52805

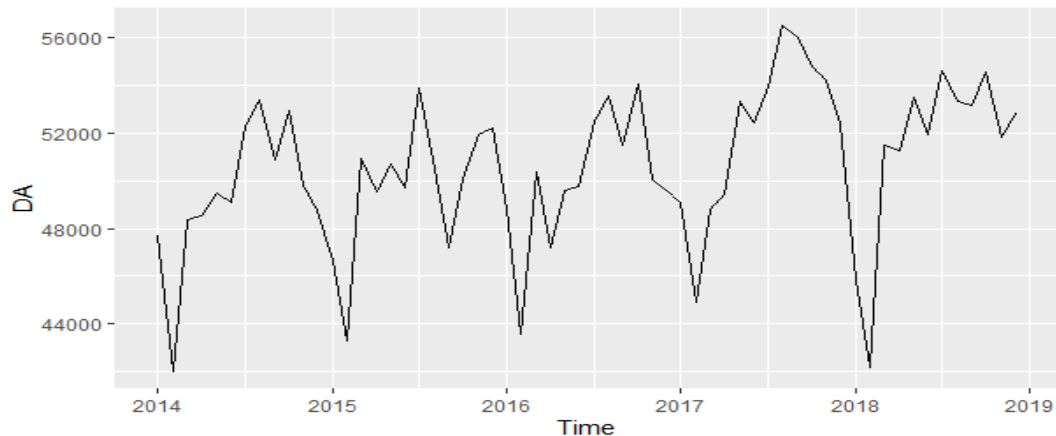


Figura 49 Diagrama de flujo de atenciones de urgencias en Antofagasta

#### 4.4.5. Predicciones de las series de tiempo

##### 4.4.5.1. Método de predicción: Holt-Winter's

Según Ardthawee y Uthai (2018): El suavizado exponencial de Holt-Winter's mejora la técnica básica de suavizado exponencial con capacidad adicional para soportar datos de tendencias y estacionales. La técnica aplica tres fórmulas de suavizado exponencial a la serie. En primer lugar, se suaviza el nivel (o la media) para obtener un valor medio local para la serie; en segundo lugar, se suaviza la tendencia; por último, se suaviza cada subserie estacional (p. 512).

Concluyo que el modelo de Holtwinter aplica la tendencia y la estacionalidad para dibujar series temporales que ayudarán a visualizar la cantidad por medio de los niveles o la media que se espera en la serie temporal.

La suavización exponencial de Holt-Winter con la tendencia y el componente estacional aditivo.

$$F_t = \alpha(\text{Demanda real del periodo anterior}) + (1 - \alpha)(\text{Pronóstico del periodo anterior} + \text{tendencia estimada para el periodo anterior})$$

$$T_t = \delta(\text{Pronóstico de este periodo} - \text{Pronóstico del último periodo}) + (1 - \delta)(\text{Tendencia estimada para el último periodo})$$

$$FIT_t = \text{Pronóstico suavizado exponencialmente } (F_t) + \text{tendencia suavizada exponencialmente } (T_t)$$

Figura 50 Formula general del Holt-Winter's

Nota: La fórmula de Holtwinters explica la frecuencia del tiempo, el tiempo de evaluación y el cálculo medio de la atención de las fórmulas mencionadas. Tomado de Negocios IT: New Revolution ,Betancourt, 2016, Obtenido de <https://ingenioempresa.com/suavizacion-exponencial-doble/>

Según Betancurt (2016):

Donde:

- Ft= Pronóstico suavizado exponencialmente con la serie de datos del período t
- Tt= Tendencia suave para el período t
- At= Demanda real del período t
- Alfa  $\alpha$ = Constante de suavizado para el promedio
- Delta  $\delta$ = Constante de la tendencia de alisamiento
- FITt= Previsión de la demanda con tendencia (p.10)

```
#Holtwinters
modeloholt = Holtwinters(DA,seasonal = "additive")
fitted(modeloholt)# Filter interpretation
#xhat = Average forecast
#Level = Level of interaction
#Trend = Data training
#Season = Parking
plot(fitted(modeloholt))
autoplot(fitted(modeloholt))
```

Figura 51 Calcula el filtrado de pronóstico del Holt-Winter's en RStudio

Nota: Calcula el filtrado de Holt-Winters de una serie temporal determinada. Los parámetros desconocidos se determinan minimizando el error de predicción al cuadrado.

Tabla 14

Resultado de filtrado Holt-Winter's de atenciones de urgencias

	XHAT	LEVEL	TREND	SEASON
Jan 2015	46025.51	49629.15	0.3939394	-3604.03472
feb-15	42823.41	49743.38	0.3939394	-6920.36806
mar-15	50832.08	49826.39	0.3939394	1005.29861
Apr 2015	49722.00	49843.30	0.3939394	-121.70139
may-15	50924.53	49807.13	0.3939394	1117.00694
jun-15	49631.40	49772.33	0.3939394	-141.32639
jul-15	52403.51	49786.99	0.3939394	2616.13194
Aug 2015	53780.01	50059.61	0.3939394	3720.00694
sep-15	50508.57	49451.04	0.3939394	1057.13194
oct-15	51772.19	48841.83	0.3939394	2929.96528
nov-15	48279.65	48534.04	0.3939394	-254.78472
Dec 2015	47807.40	49210.33	0.3939394	-1403.32639
Jan 2016	46498.29	50026.14	0.3939394	-3528.25360
feb-16	43588.73	50453.71	0.3939394	-6865.36968
	XHAT	LEVEL	TREND	SEASON

<b>mar-16</b>	51462.02	50445.33	0.3939394	1016.29839
<b>Apr 2016</b>	50102.72	50248.37	0.3939394	-146.04452
<b>may-16</b>	50811.01	49717.04	0.3939394	1093.57685
<b>jun-16</b>	49367.33	49498.77	0.3939394	-131.83359
<b>jul-16</b>	52368.70	49570.95	0.3939394	2797.36104
<b>Aug 2016</b>	52909.37	49594.37	0.3939394	3314.60787
<b>sep-16</b>	50361.78	49710.09	0.3939394	651.29680
<b>oct-16</b>	52640.41	49915.22	0.3939394	2724.79643
<b>nov-16</b>	50375.35	50179.78	0.3939394	195.18463
<b>Dec 2016</b>	49263.24	50123.32	0.3939394	-860.47946
<b>Jan 2017</b>	46941.57	50185.05	0.3939394	-3243.87353
<b>feb-17</b>	43707.02	50577.83	0.3939394	-6871.20885
<b>mar-17</b>	51682.19	50796.89	0.3939394	884.91407
<b>Apr 2017</b>	49774.83	50274.47	0.3939394	-500.03061
<b>may-17</b>	51163.94	50215.54	0.3939394	948.00403
<b>jun-17</b>	50533.98	50617.63	0.3939394	-84.04302
<b>jul-17</b>	53781.71	50968.63	0.3939394	2812.68851
<b>Aug 2017</b>	54394.11	51002.33	0.3939394	3391.38557
<b>sep-17</b>	52179.88	51391.89	0.3939394	787.59839
<b>oct-17</b>	54996.21	52095.16	0.3939394	2900.65511
<b>nov-17</b>	52209.88	52052.15	0.3939394	157.34171
<b>Dec 2017</b>	51606.87	52426.13	0.3939394	-819.65083
<b>Jan 2018</b>	49587.44	52569.69	0.3939394	-2982.64827
<b>feb-18</b>	45159.92	51885.16	0.3939394	-6725.63928
<b>mar-18</b>	51872.13	51334.87	0.3939394	536.86425
<b>Apr 2018</b>	50726.47	51265.60	0.3939394	-539.52262
<b>may-18</b>	52585.36	51369.54	0.3939394	1215.42239
<b>jun-18</b>	51686.84	51537.08	0.3939394	149.36313
<b>jul-18</b>	54419.07	51583.81	0.3939394	2834.86581
<b>Aug 2018</b>	55273.09	51622.23	0.3939394	3650.46478
<b>sep-18</b>	52518.55	51262.64	0.3939394	1255.52107
<b>oct-18</b>	54258.75	51386.59	0.3939394	2871.76026
<b>nov-18</b>	51852.46	51446.02	0.3939394	406.05086
<b>Dec 2018</b>	50722.75	51446.70	0.3939394	-724.34104

La interpretación de la serie con el modelo

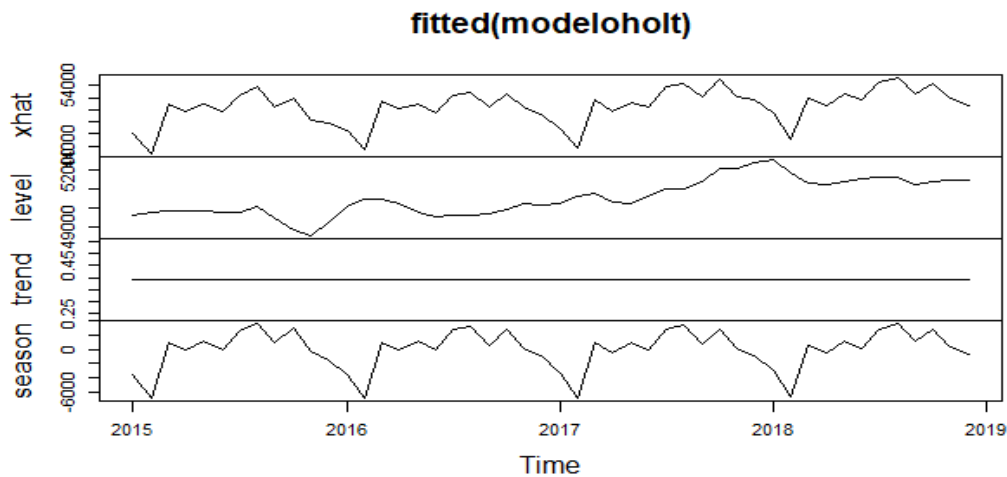


Figura 52 Gráfico de los datos filtrados de Holt-Winter's en RStudio

Nota: Gráfico de la predicción reflejada en la línea roja de todo el aumento y la predicción de la cantidad de atención de urgencia.

```
#prediccion
predholt=predict(modeloholt,12,prediction.interval = TRUE)
predholt
#Fit = estimated quantity
#Upr = maximum acceptance margin
#Lwr = minimum declination margin
plot(modeloholt,predholt)
```

Figura 53 Código de predicción del Holt-Winter's en RStudio

Nota: La proyección se realiza después de 12 meses, lo que le da la capacidad de ser estacionario para estimar las cantidades

Tabla 15

*Proyección de atenciones Holt-Winter's en RStudio*

	<b>FIT</b>	<b>UPR</b>	<b>LWR</b>
<b>Jan 2019</b>	48391.48	51984.23	44798.73
<b>feb-19</b>	44738.25	48391.15	41085.34
<b>mar-19</b>	52321.37	56033.45	48609.30
<b>Apr 2019</b>	51360.70	55131.02	47590.37
<b>may-19</b>	53158.37	56986.06	49330.69
<b>jun-19</b>	52012.28	55896.48	48128.08
<b>jul-19</b>	54692.64	58632.55	50752.74
<b>Aug 2019</b>	55243.67	59238.50	51248.84
<b>Sep 2019</b>	53171.03	57220.05	49122.02
<b>oct-19</b>	54744.70	58847.18	50642.23
<b>nov-19</b>	52240.28	56395.54	48085.02
<b>Dec 2019</b>	51364.81	55572.19	47157.44



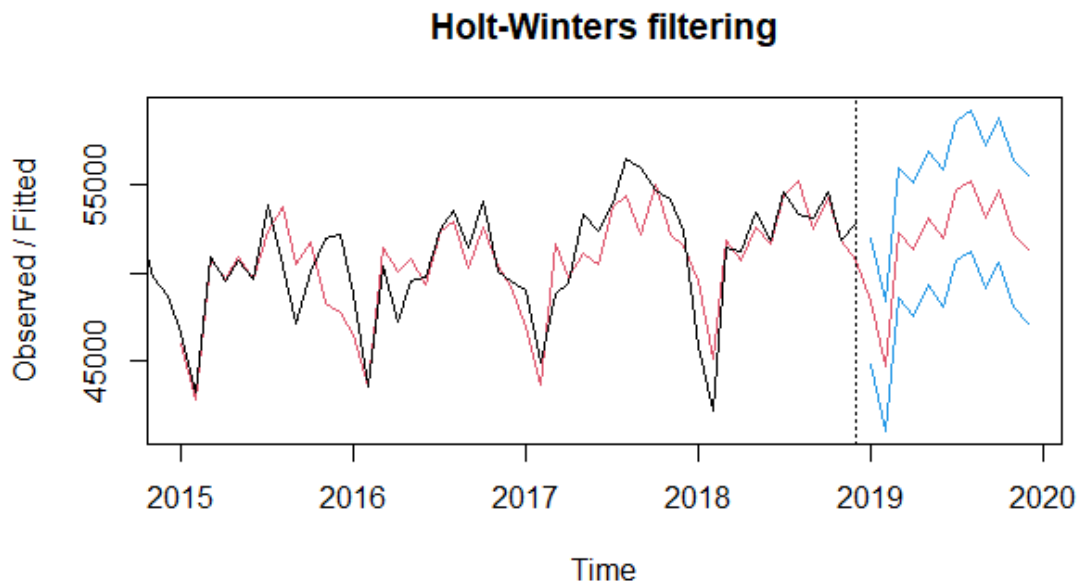


Figura 54 Grafico de filtro de atenciones en Holt-Winter's en el RStudio

Nota: El gráfico refleja la estimación de la atención de emergencia para el año siguiente mostrando el rango de atención.

```
#Holtwinters
summary(forecast(modeloholt))
autoplot(forecast(modeloholt))
```

Figura 55 Resumen de resultado Holt-Winter's en RStudio

Nota: Comando en Rstudio para mostrar la tabla de resultados y el gráfico del modelo de predicción.

Tabla 16

Resultados de datos de entrenamiento de Holt-Winter's en RStudio

Error measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	247.3495	1830.663	1359.055	0.3782752	2.697769	0.6702902	0.3792879

Point Forecast = Punto medio

Lo 80 = 80% variación

Hi 80 = Atenciones al 80%

Lo 95 = 95% variación

Hi 95 = Atenciones al 95%

Tabla 17

*Resultados con la predicción Holt-Winter's*

<b>Forecasts:</b>					
	<b>POINT FORECAST</b>	<b>LO 80</b>	<b>HI 80</b>	<b>LO 95</b>	<b>HI 95</b>
<b>Jan 2019</b>	48391.48	46042.30	50740.65	44798.73	51984.23
<b>feb-19</b>	44738.25	42349.74	47126.75	41085.34	48391.15
<b>mar-19</b>	52321.37	49894.18	54748.57	48609.30	56033.45
<b>Apr 2019</b>	51360.70	48895.41	53825.98	47590.37	55131.02
<b>may-19</b>	53158.37	50655.58	55661.16	49330.69	56986.06
<b>jun-19</b>	52012.28	49472.54	54552.02	48128.08	55896.48
<b>jul-19</b>	54692.64	52116.48	57268.81	50752.74	58632.55
<b>Aug 2019</b>	55243.67	52631.59	57855.75	51248.84	59238.50
<b>sep-19</b>	53171.03	50523.53	55818.54	49122.02	57220.05
<b>oct-19</b>	54744.70	52062.24	57427.17	50642.23	58847.18
<b>nov-19</b>	52240.28	49523.30	54957.26	48085.02	56395.54
<b>Dec 2019</b>	51364.81	48613.76	54115.87	47157.44	55572.19
<b>Jan 2020</b>	48396.21	45552.74	51239.67	44047.50	52744.92
<b>mar-20</b>	52326.10	49417.84	55234.36	47878.30	56773.90
<b>Apr 2020</b>	51365.42	48425.30	54305.54	46868.90	55861.95
<b>may-20</b>	53163.10	50191.46	56134.74	48618.37	57707.83
<b>jun-20</b>	52017.01	49014.18	55019.84	47424.58	56609.44
<b>jul-20</b>	54697.37	51663.67	57731.07	50057.73	59337.01
<b>Aug 2020</b>	55248.40	52184.14	58312.65	50562.02	59934.77
<b>sep-20</b>	53175.76	50081.25	56270.27	48443.12	57908.40
<b>oct-20</b>	54749.43	51624.96	57873.90	49970.97	59527.90
<b>nov-20</b>	52245.01	49090.86	55399.15	47421.15	57068.86
<b>Dec 2020</b>	51369.54	48185.99	54553.09	46500.72	56238.36

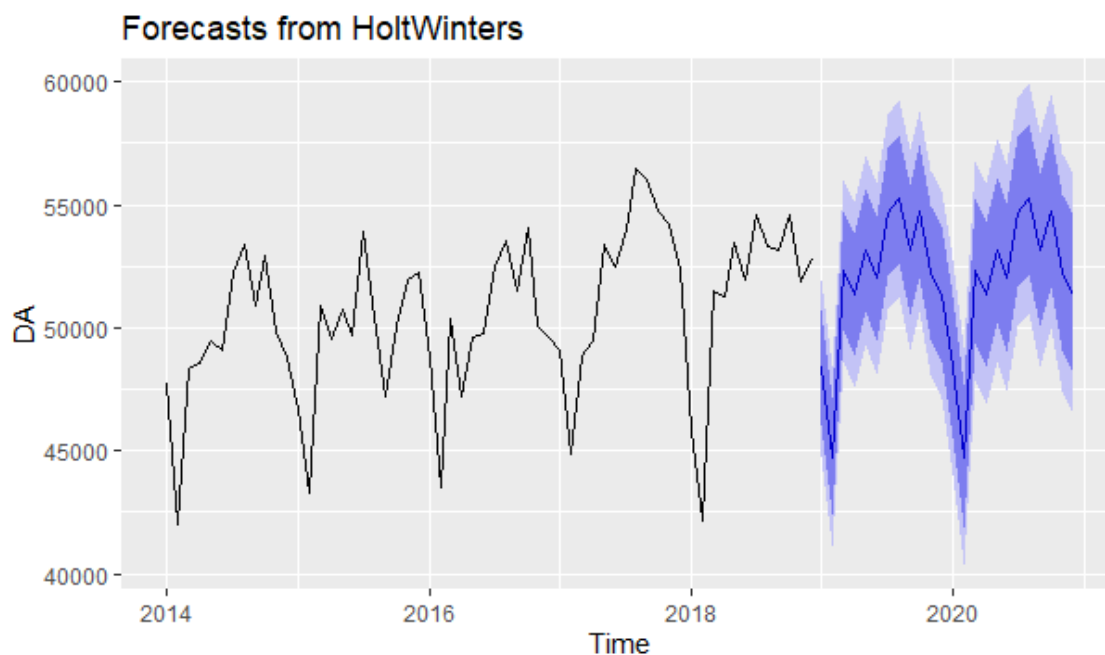


Figura 56 Grafica de predicción Holt-Winter's en RStudio

Nota: El gráfico muestra la estimación de la cantidad de atención de emergencia con una precisión del 95% muestra una línea azul que representa el promedio estimado y la sombra azul es la desviación en la cantidad que puede tomar.

Interpretación de la fórmula:

$$\text{modeloholt} = 0.1837517 (51364.81) + (1-0.1837517) (54115.87+12)$$

$$\text{modeloholtTrend} = 54115.87 (48185.99-54115.87+(1- 54934.29) (12)$$

Final formula:

$$\text{ModeloholtPrognosis} = 72558.55 (\text{modeloholt})+ 24(\text{modeloholtTrend})$$

Significado:

- modeloholt: Predicción suavizada exponencialmente con la serie de datos del período t
- modeloholtTrend = Tendencia suave para el período t
- $A_t = 54115.87$
- Alfa  $\alpha = 0.1837517$
- $\delta = 48185.99$
- Modeloholtprostituto = Previsión de la demanda con tendencia

#### 4.4.5.2. Método de predicción: Método Naive

El método de predicción más sencillo es utilizar la observación más reciente como previsión para la siguiente observación. Se llama pronóstico ingenuo y

puede ser implementado usando la función "ingenuo ()". Este método puede no ser la mejor técnica de predicción, pero a menudo proporciona un punto de referencia útil para otros métodos de previsión más avanzados.

$$\hat{Y}_{t+h|t} = Y_t$$

Figura 57 *Formula general del método Naive*

Nota: El método Naives es el modelo que calcula de la manera más sencilla utilizando la demanda real del último período, ya que se repetirá la demanda prevista para el período futuro con una suposición del pasado. Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting, Catal et al. ,2019, Balkan Journal of Electrical and Computer Engineering.

El método de la Fórmula Naive:

$$\hat{y}_{T+h|T} = y_T.$$

Dónde:

$\hat{y}_T$ : Los datos temporales de la serie de tiempo

$h|T$ : h corresponde a los meses de previsión sobre el tiempo estimado

$y_T$ : Medición de los datos del resultado en y con la frecuencia del tiempo  
(Catal et al. ,2019, p. 23).

```
#Naives Method
modelonaive = naive(DA, h =24)
plot(modelonaive)
prednaive=predict(modelonaive,12,prediction.interval = TRUE)
prednaive
autoplot(forecast(modelonaive))
summary(forecast(modelonaive))
```

Figura 58 *Código del método Navie en RStudio*

Nota: Nos muestra la creación del método del modelo naives dando una predicción de 2 años y luego creamos la predicción para los próximos 12 meses que seguirá la misma tendencia para el próximo año.

Tabla 18

*Resultados de los datos de entrenamiento del método Naive en RStudio*

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	486.1875	2373.265	2027.562	0.8358906	3.977533	1	0.4039885

Tabla 19

*Resultados con la predicción método Naive en RStudio*

<b>Forecasts:</b>					
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
<b>Jan 2019</b>	45860	42818.54	48901.46	41208.49	50511.51
<b>feb-19</b>	42163	39121.54	45204.46	37511.49	46814.51
<b>mar-19</b>	51493	48451.54	54534.46	46841.49	56144.51
<b>Apr 2019</b>	51290	48248.54	54331.46	46638.49	55941.51
<b>may-19</b>	53495	50453.54	56536.46	48843.49	58146.51
<b>jun-19</b>	51939	48897.54	54980.46	47287.49	56590.51
<b>jul-19</b>	54626	51584.54	57667.46	49974.49	59277.51
<b>Aug 2019</b>	53314	50272.54	56355.46	48662.49	57965.51
<b>sep-19</b>	53191	50149.54	56232.46	48539.49	57842.51
<b>oct-19</b>	54580	51538.54	57621.46	49928.49	59231.51
<b>nov-19</b>	51854	48812.54	54895.46	47202.49	56505.51
<b>Dec 2019</b>	52805	49763.54	55846.46	48153.49	57456.51
<b>Jan 2020</b>	45860	41558.72	50161.28	39281.77	52438.23
<b>feb-20</b>	42163	37861.72	46464.28	35584.77	48741.23
<b>mar-20</b>	51493	47191.72	55794.28	44914.77	58071.23
<b>Apr 2020</b>	51290	46988.72	55591.28	44711.77	57868.23
<b>may-20</b>	53495	49193.72	57796.28	46916.77	60073.23
<b>jun-20</b>	51939	47637.72	56240.28	45360.77	58517.23
<b>jul-20</b>	54626	50324.72	58927.28	48047.77	61204.23
<b>Aug 2020</b>	53314	49012.72	57615.28	46735.77	59892.23
<b>sep-20</b>	53191	48889.72	57492.28	46612.77	59769.23
<b>oct-20</b>	54580	50278.72	58881.28	48001.77	61158.23
<b>nov-20</b>	51854	47552.72	56155.28	45275.77	58432.23
<b>Dec 2020</b>	52805	48503.72	57106.28	46226.77	59383.23

En el que se evalúa el posible aumento, pero teniendo el mismo rango de forma de rango 80% como valor mínimo ah 95% como máximo la cantidad de antecedentes de ese mes varían en los rangos de hi 80 a hi 95.

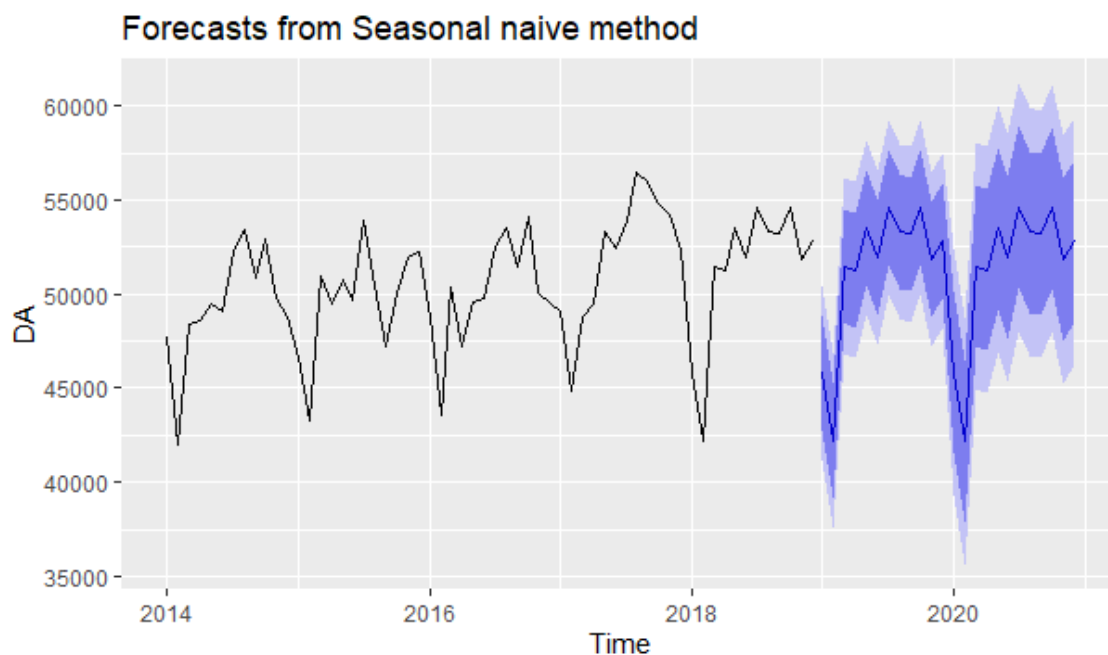


Figura 59 Gráfico de predicción del método Naive estacional en RStudio

Note: The graph shows us in blue line the forecast of the amount of emergency care the estimated means and the blue part is the interval that can take 95%

Se auto-acumula a medida que maneja rangos, por lo que toma una base promedio

Fórmula de interpretación Método Navie:

$$DAT +24|DA = 45860$$

DAT: la variable de frecuencia con los datos aplicados con los Ts se refleja en ts (DA, Frecuencia = 24)

24|DA: donde 24 es los siguientes meses sobre los datos del resultado

45860: estándar para hacer nuestras predicciones forecasts

#### 4.4.5.3. Método de predicción: Exponencial de Tendencia Suavizado

Exponencial de Tendencia Suavizado son una extensión del método ingenuo, en el que las previsiones se producen utilizando promedios ponderados de observaciones pasadas, con los pesos decayendo exponencialmente a medida que las observaciones envejecen. En palabras sencillas, se dan mayores pesos a las observaciones más recientes y viceversa.

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1}$$

Figura 60 *Formula general del Exponencial de tendencia suavizado*

Nota: La fórmula explica la serie total más la tendencia temporal para obtener la cantidad media de atención de emergencia, tomada de Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting, Catal et al. ,2019, Balkan Journal of Electrical and Computer Engineering.

St -1: Variable constante para encontrar el punto de partida

α: alfa para la iniciación

st: la variable de la serie de tiempo (Catal et al. ,2019, p. 22)

El valor del parámetro de suavizado para el nivel se decide por el parámetro "alfa".

```
#Exponential Trend Smoothing
modeloets<- ets(DA)
prediccionets<- forecast(fit_ets_default, h = 24)
plot(prediccionets)
summary(forecast(prediccionets))
```

Figura 61 *Código del exponencial de tendencia suavizado en RStudio*

Nota: Muestra la creación de nuestro modelo Exponencial de Tendencia Suave para el cual damos la proyección de 2 años estacionaria para ver la estimación de la cantidad de atención.

Tabla 20

*Resultados del entrenamiento exponencial de tendencia suavizado en RStudio*

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	90.52731	1432.727	1131.353	0.1026939	2.9655	0.557987	0.3341373

Tabla 21

*Resultados con la predicción exponencial de tendencia suavizado en RStudio*

Forecasts:	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2019	49175.87	47117.76	51233.97	46028.26	52323.47
feb-19	44396.42	42475.52	46317.33	41458.65	47334.20
mar-19	51292.89	49003.31	53582.48	47791.27	54794.52
Apr 2019	50331.96	48018.34	52645.58	46793.59	53870.34

Forecasts:	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
may-19	52676.12	50186.67	55165.57	48868.84	56483.41
jun-19	51704.83	49196.24	54213.43	47868.27	55541.40
jul-19	54807.81	52081.46	57534.17	50638.21	58977.41
Aug 2019	54785.06	51994.26	57575.86	50516.91	59053.21
sep-19	52923.71	50165.84	55681.59	48705.91	57141.51
oct-19	54529.02	51625.12	57432.92	50087.89	58970.15
nov-19	52465.70	49612.93	55318.48	48102.76	56828.65
Dec 2019	52100.94	49210.82	54991.06	47680.88	56520.99
Jan 2020	49175.87	46395.04	51956.69	44922.97	53428.77
feb-20	44396.43	41838.98	46953.87	40485.16	48307.70
mar-20	51292.90	48284.98	54300.81	46692.69	55893.10
Apr 2020	50331.96	47329.10	53334.82	45739.48	54924.44
may-20	52676.12	49480.62	55871.63	47789.02	57563.23
jun-20	51704.84	48517.27	54892.40	46829.87	56579.80
jul-20	54807.81	51375.76	58239.87	49558.94	60056.69
Aug 2020	54785.06	51302.07	58268.06	49458.28	60111.84
sep-20	52923.72	49509.22	56338.21	47701.70	58145.74
oct-20	54529.02	50960.34	58097.70	49071.20	59986.85
nov-20	52465.70	48984.04	55947.37	47140.96	57790.45
Dec 2020	52100.94	48596.44	55605.44	46741.27	57460.61

**Forecasts from ETS(M,N,M)**

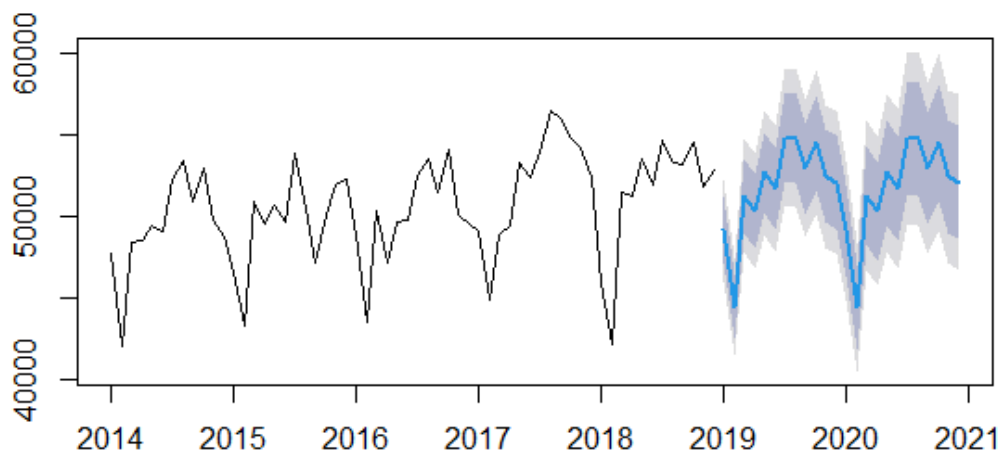


Figura 62 Grafica de predicción del exponencial de tendencia suavizado en RStudio

Nota: En el presente gráfico nos muestra en una línea azul la previsión de la cantidad de atención a lo largo de 2 años para y el color de plomo es el intervalo en el que esta cantidad puede cambiar teniendo una precisión del 95%



Es un modelo sencillo que no encuentra ningún foco de impacto si no parte de una cantidad estimada y ya está manipulado por el rango, ya que no toma valores que puedan alterar su frecuencia, por lo que su nivel de rango no es recomendable para el tipo de estudios en los que queremos una mayor precisión.

Interpretación de la fórmula:

$$50364.8565 + 1 = 0.2621 * 24 + (1 - 0.2621) 50364.8565$$

Ft = 50364.8565 cuidado promedio

α: 0.2621

yt: tiempo estimado en este caso 24 meses

**El mejor fue evaluado por el siguiente comando para los márgenes de error**

Según Prayudani y otros (2019): La media del porcentaje de error absoluto (MAPE) se calcula utilizando los errores absolutos de cada período divididos por los valores reales de observación de ese período. Luego, el porcentaje de error absoluto medio. Este enfoque es útil cuando el tamaño o la magnitud de las variables de predicción es importante para evaluar la exactitud de las predicciones. MAPE indica cuánto error de predicción se compara con el valor real de la serie. MAPE también puede utilizarse para comparar la exactitud de los mismos o diferentes métodos en dos series diferentes y medir la exactitud del valor estimado del modelo expresado en términos del promedio del porcentaje de error absoluto. Cuando a es un dato real, b es un dato de resultado, n es la cantidad de datos (p.4)

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{a - b}{a} \right|}{n} \times 100\%$$

Figura 63 *Fórmula de la media del porcentaje de error absoluto*

Nota: n es el número de casos en un conjunto, k es el número de parámetros en el modelo, yi es el valor real de una variable de salida (BW), yip es el valor predicho de una variable de salida (BW), sm es la desviación estándar de los errores en el modelo, tomo la precisión del análisis de la técnica de medición de la predicción sobre el vecino K más cercano (RKNN) usando ASM y MSE. En el Journal of Physics: Conference Series, por Prayudani et al.,2017, IOP Publishing.

Prueba Evaluar el comportamiento y el error de la predicción de la predicción.

```

#MAPE test
mape<- function(Actualt, pred){
  mape<- mean(abs((Actualt - pred)/actual))*100
  return (mape)
}
#Actualt = the estimated average
#pred = the estimate ranges from hi 80 to hi 95 this is going
#       to change according to the model used

```

Figura 64 Algoritmo Mape en RStudio

Nota: Es la forma en que el Rstudio refleja el algoritmo Mape para ver el porcentaje de error del modelo de pronóstico.

Tabla de márgenes de error con la función Mape:

Tabla 22

*Margen de error porcentual comparación en RStudio*

	Holt-Winter's	Naive Método	Exponential Trend Smoothing
<b>Margen de Error:</b>	2.7%	3.98%	2.97%
<b>Porcentaje de exactitud:</b>	97.3%	96.02%	97.03%

#### 4.4.6. Regresión ordinal logística

La regresión ordinal logística nos permite moldear la dependencia de una respuesta ordinal logística o politómicas en un conjunto de predictores, que pueden ser factores o covariables. Además, supone reducir al mínimo las diferencias en la suma de los cuadrados entre una variable de respuesta (la dependiente) y una combinación ponderada de las variables predictoras (las independientes). Para Rodríguez et al., 2016 definido:

Al aplicar un modelo de regresión logística, en lugar de construir un modelo de regresión para estimar los valores reales de la variable de interés, se construye una función basada en el cálculo de la probabilidad de que la variable de interés adopte el valor del evento previamente definido. (p. 217).

$$p = P(Y = 1) = \frac{e^{\hat{z}}}{1 + e^{\hat{z}}}$$

Figura 65 *Fórmula general de la regresión ordinal logística*

Nota: General ordinal regression equation that calculates the probability of variables from: Introduction to applied multivariate statistical analysis. Experience and cases in the Colombian Caribbean. Universidad del Norte by Rodriguez, et. al, 2016, Universidad del Norte.

También, Rodríguez et al., 2016 expusieron "lo que se presenta en la regresión logística es exponer la probabilidad de ocurrencia del evento ecuestre en función de ciertas variables, que se presumen relevantes o influyentes". (p. 217)

$$p = P(Y = 1) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}$$

Figura 66 *Fórmula general descompuesta de la regresión ordinal logística*

Nota: Ecuación descompuesta en variables de la regresión ordinal. De: Introducción al análisis estadístico multivariado aplicado. Experiencia y casos en el Caribe colombiano. Universidad del Norte por Rodriguez, al et., 2016, Universidad del Norte.

Las tablas de frecuencias, así como las representaciones gráficas de las observaciones son útiles para intuir si las variables independientes elegidas están relacionadas con la variable de respuesta y, por lo tanto, son buenos predictores.

```
> library(rms)
> MLogit<- lrm (formula = RangAtenciones ~ Año + Mes + Days +
IDEstablecimiento + IDTipoEstablecimiento + REdad0 + REdad1a4 +
REdad5a14 + REdad15a64 + REdad65 + IDA00_A09 + IDB99 + IDI30_52 +
IDI60_69 + IDI70_79 + IDJ00_06 + IDJ09_11 + IDJ20_22 + IDJ30_39 +
IDJ40_46 + IDN20_23 + IDR00_09 + IDR50_69 + IDS00_09 + IDT14 + IDT20_31
+ IDT51_65 + IDT66_78, data = DataLog, tol=1e-12)
```

Según la fórmula, se obtuvieron los siguientes resultados.

Tabla 23

*Resumen de Coeficientes en la regresión ordinal logística en Rstudio*

	Coef	S.E.	wald	gl	Wald Z	Pr(> Z )
y>=2	1.1037	6.0947	0,0324		0,18	0.8563
y>=3	-0.8950	6.0946	0,0225		-0,15	0.8833
y>=4	-2.5174	6.0946	0,1681		-0,41	0.6796
Año	0.0003	0.0030	0.01	1	0.10	0.9205
Mes	0.0135	0.0012	121.05	1	11.00	<0.0001
Dia	0.0003	0.0005	0.44	1	0.67	0.5056
IDEstablecimiento	0.0131	0.0006	432.87	1	20.81	<0.0001
IDTipoEstablecimiento	-0.0651	0.0017	1463.98	1	-38.26	<0.0001
REdad0	-1.1564	0.0099	13777.15	1	-117.38	<0.0001
REdad1a4	-1.5525	0.0117	17753.81	1	-133.24	<0.0001
REdad5a14	-1.8527	0.0123	22758.80	1	-150.86	<0.0001
REdad15a64	-2.9596	0.0340	7568.50	1	-87.00	<0.0001
REdad65	-1.6069	0.0100	25719.09	1	-160.37	<0.0001
IDA00_09	-0.4622	0.0182	641.46	1	-25.33	<0.0001
IDB99	1.4003	0.0237	3479.91	1	58.99	<0.0001
IDI30_52	-1.5206	0.0176	7486.24	1	-86.52	<0.0001
IDI60_69	0.4589	0.0476	93.00	1	9.64	<0.0001
IDI70_79	-2.0166	0.0261	5947.91	1	-77.12	<0.0001
IDJ00_06	0.3583	0.0212	285.41	1	16.89	<0.0001
IDJ09_11	0.0285	0.1104	0.07	1	0.26	0.7962
IDJ20_22	-2.4012	0.0267	8058.20	1	-89,77	<0.0001
IDJ30_39	-0.5052	0.0471	115.11	1	-10.73	<0.0001
IDJ40_46	0.8040	0.0619	168.96	1	13.00	<0.0001
IDN20_23	1.0171	0.1079	88.85	1	9.43	<0.0001
IDR00_09	-0.7558	0.0280	730.91	1	-27.04	<0.0001
IDR50_69	1.2829	0.0399	1035.79	1	32.18	<0.0001
IDS00_09	0.1782	0.1187	2.25	1	1.50	0.1332
IDT14	-0.9739	0.0185	2764.74	1	-52.58	<0.0001
IDT20_31	1.1332	0.0323	1228.25	1	35.05	<0.0001
IDT51_65	0.0757	0.1209	0.39	1	0.63	0.5312
IDT66_78	0.7725	0.0911	71.92	1	8,48	<0.0001

El modelo de probabilidad es el núcleo de la estimación de la probabilidad. Dado el modelo científico y un conjunto de valores específicos para sus parámetros, existe una cierta probabilidad de observar los datos reales.

Tabla 24

*Modelo Likelihood Ratio Test en la regresión ordinal logística en Rstudio*

Modelo Likelihood Ratio Test	
<b>Obs</b>	427213
<b>max  deriv </b>	1,00E-07
<b>LR chi2</b>	326549.16
<b>d.f.</b>	28
<b>Pr(&gt; chi2)</b>	<0.0001
<b>Cox y Snell (r2ML)</b>	0,534
<b>Nagelkerke (r2CU)</b>	0,636
<b>McFadden</b>	0,416

El cuadrado de la R en Nagelkerke significa que el poder explicativo del modelo, que nos dio un 0,636. Por lo tanto, la ecuación general para la regresión ordinal en la que A es el coeficiente del rango de edad a calcular es la siguiente:

Tabla 25

*Coefficiente de evaluación  $\alpha_j$  en la regresión ordinal logística en Rstudio*

<b>RANGE DE ATENCIÓN</b>	<b>A</b>	
<b>11 a 20</b>	$\alpha_1$	1.1037
<b>21 a 30</b>	$\alpha_2$	-0.8950
<b>31 a 40</b>	$\alpha_3$	-2.5174

El modelo obtenido es:

$$\begin{aligned}
 Z = & \alpha_j + 0.0003 \cdot \text{Año} + 0.0135 \cdot \text{Mes} + 0.0003 \cdot \text{Days} + \\
 & 0.0131 \cdot \text{IDEstablecimiento} - 0.0651 \cdot \text{IDTipoEstablecimiento} - 1.1564 \cdot \text{REdad0} - \\
 & 1.5525 \cdot \text{REdad1a4} - 1.8527 \cdot \text{REdad5a14} - 2.9596 \cdot \text{REdad15a64} - \\
 & 1.6069 \cdot \text{REdad65} - 0.4622 \cdot \text{IDA00\_09} + 1.4003 \cdot \text{IDB99} - 1.5206 \cdot \text{IDI30\_52} + \\
 & 0.4589 \cdot \text{IDI60\_69} - 2.0166 \cdot \text{IDI70\_79} + 0.3583 \cdot \text{IDJ00\_06} + 0.0285 \cdot \text{IDJ09\_11} - \\
 & 2.4012 \cdot \text{IDJ20\_22} - 0.5052 \cdot \text{IDJ30\_39} + 0.8040 \cdot \text{IDJ40\_46} + 1.0171 \cdot \text{IDN20\_23} - \\
 & 0.7558 \cdot \text{IDR00\_09} + 1.2829 \cdot \text{IDR50\_69} + 0.1782 \cdot \text{IDS00\_09} - 0.9739 \cdot \text{IDT14} + \\
 & 1.1332 \cdot \text{IDT20\_31} + 0.0757 \cdot \text{IDT51\_65} + 0.7725 \cdot \text{IDT66\_78}
 \end{aligned}$$

Por lo tanto, nuestra ecuación sería la siguiente:

$$\text{Ecuación P: } (e^Z) / (1 + e^Z)$$

Dónde:

Z : suma de variables según el coeficiente del rango de atención.

Terminando con la regresión ordinal, la interpretación de las variables mediante la elevación de e al coeficiente nos da los siguientes resultados.

Tabla 26

*Interpretación de las variables logísticas Regresión ordinal con la variable cantidad de atenciones de urgencias*

VARIABLES	B	EXP(B)	lo	INTERPRETACIÓN
Año	0,0003	1,0003	1	Por cada incremento de la atención en la unidad en el Año, las posibilidades de ser más probable se multiplican por 1.0003, es decir, se incrementan en un 0003%.
			2	Por cada unidad de disminución de cuidados en el Año, la probabilidad de ser menos probable se multiplica por 1.0003 veces, es decir, disminuye en un 0,003%.
Mes	0,0135	1,0136	1	Por cada unidad de aumento de los cuidados durante el mes, la probabilidad de ser más probable se multiplica por 1.0135 veces, es decir, aumenta en un 0,0135%.
			2	Por cada unidad de disminución de la atención en el mes, la probabilidad de ser menos probable se multiplica por 1.0135, es decir, disminuye en 0,0135%.
Dia	0,0003	1,0003	1	Por cada unidad de incremento de cuidados en el día, la probabilidad de ser más probable se multiplica por 1.0003 veces, es decir, un 0,0003% de incremento
			2	Por cada unidad de disminución de la atención en el día, las posibilidades de ser menos probable se multiplican por 1.0003, es decir, la disminución por 1.0003%.
IDestablecimiento	0,0131	1,0132	1	Por cada unidad de aumento de la atención en los establecimientos, las probabilidades de ser más probable que sea la atención de emergencia son 1.013 veces mayores que las de la atención no de emergencia de 0.013%.
			2	Por cada unidad de disminución de la atención en los establecimientos, la probabilidad de ser menos probable que sea una emergencia se multiplica por 1,013 más que si pertenece al 0,013%.
IDTipoEstablecimiento	-0,0651	0,9370	1	Para la atención según el tipo de establecimiento, la probabilidad de realizar una atención de emergencia es un 6,3% menor que en otros tipos de establecimientos
			2	En el caso de la atención en otro tipo de instalaciones, la probabilidad de recibir atención de emergencia es 1,06 veces mayor que la de un paciente registrado.
REdad0	-1,1564	0,3146	1	Para la atención en la RAE0, la probabilidad de recibir atención de emergencia es 68,53% menos que la atención fuera del rango de edad 0
			2	Para los cuidados que no pertenecen a la REdad0, la probabilidad de recibir cuidados de emergencia es 3,17 veces la del rango de edad 0.
REdad1a4	-1,5525	0,2117	1	Para la atención entre la ER1-4, la probabilidad de recibir atención de emergencia es 78,82% menos que la atención fuera del rango de edad 1-4
			2	Para la atención de no-ERD1 a 4, la probabilidad de recibir atención de emergencia es 4,72 veces la del rango de edad de 1 a 4

VARIABLES	B	EXP(B)	lo	INTERPRETACIÓN
REdad5a14	-1,8527	0,1568	1	En el caso de la atención en el rango de edad de 5 a 14 años, la probabilidad de recibir atención de urgencia es 84,318% menor que la atención fuera del rango de edad de 5 a 14 años
			2	En el caso de la atención no perteneciente al grupo de edad 5-14, la probabilidad de recibir atención de urgencia es 6.377 veces mayor que la del grupo de edad 5-14.
REdad15a64	-2,9596	0,0518	1	Para la atención en el rango de edad de 15-64, la probabilidad de recibir atención de urgencia es 94,816% menos que la atención fuera del rango de edad de 15-64
			2	En el caso de la atención de personas sin rango de edad 15-64, la probabilidad de recibir atención de urgencia es 19,29 veces mayor que la del rango de edad 15-64.
REdad65	-1,6069	0,2005	1	Para la atención entre los 65 años, la probabilidad de recibir atención de urgencia es 79,949% menos que la atención fuera del rango de los 65 años
			2	Para la atención no pertenece al rango de edad mayores de 65, la probabilidad de recibir atención de urgencia es 4.987 veces mayor que la del rango de edad de 65 años.
IDA00_09	-0,4622	0,6299	1	Para los cuidados cuyas enfermedades pertenecen al grupo IDA00_09, la probabilidad de padecerlas es 37,010% menor que los cuidados que no pertenecen al grupo IDA00_09
			2	Para los cuidados que no pertenecen al grupo de enfermedades IDA00_09, la probabilidad de que se sufran es 1.587 veces mayor que la de aquellos en el
IDB99	1,4003	4,0564	1	Para la atención con la enfermedad IDB99, la probabilidad de padecerla es 4.056 mayor que la atención no BID99.
			2	Para la atención no IDB99, la probabilidad de ser menos afectado es de 4.056 más que para la atención no BPD99
IDI30_52	-1,5206	0,2186	1	Para los cuidados cuyas enfermedades pertenecen al grupo IDI30_52, la probabilidad de padecerlas es 78,141% menor que los cuidados que no pertenecen al
			2	Para los cuidados que no pertenecen al grupo de enfermedades del IDI30_52, la probabilidad de que los sufran es de 4.574 veces a los que son del grupo
IDI60_69	0,4589	1,5823	1	Para aquellos cuya enfermedad IDI60_69, la probabilidad de sufrirla es 1.582 mayor que para aquellos cuya enfermedad IDI60_69
			2	En el caso de los cuidados cuya enfermedad no es de IDI60_69, la probabilidad de ser menos afectados es de 1,582 más que los cuidados que pertenecen a IDI60_69
IDI70_79	-2,0166	0,1331	1	Para los cuidados cuyas enfermedades pertenecen al grupo IDI70_79, la probabilidad de padecerlas es un 86,689% menor que los cuidados que no pertenecen al
			2	Para los cuidados que no pertenecen al grupo de enfermedades del IDI70_79, la probabilidad de que se sufran los cuidados es 7.512 veces la de los que son del grupo
IDJ00_06	0,3583	1,4309	1	Para aquellos cuya enfermedad IDJ00_06, la probabilidad de sufrirla es 1.430 mayor que para aquellos cuya enfermedad no es IDJ00_06
			2	En el caso de los cuidados cuya enfermedad no es IDJ00_06, la probabilidad de estar menos afectados es de 1,4308 más que los cuidados que pertenecen al IDJ00_06



VARIABLES	B	EXP(B)	lo	INTERPRETACIÓN
IDJ09_11	0,0285	1,0289	1	Para el cuidado cuya enfermedad IDJ09_11, la probabilidad de sufrirla es 1.0289 mayor que el cuidado que no es IDJ09_11
			2	En el caso de los cuidados cuya enfermedad no es IDJ09_11, la probabilidad de ser menos afectados es de 1,0289 más que los cuidados que pertenecen a la IDJ09_11
IDJ20_22	-2,4012	0,0906	1	Para los cuidados cuyas enfermedades pertenecen al grupo IDJ20_22, las probabilidades de que el cuidado sufra es 90,939% menos que los cuidados que no pertenecen al IDJ20_22
			2	Para los cuidados que no pertenecen al grupo de enfermedades IDJ20_22, la probabilidad de que se sufran es 11.036 veces mayor que la de los del grupo IDJ20_22
IDJ30_39	-0,5052	0,6034	1	Para los cuidados cuyas enfermedades pertenecen al grupo IDJ30_39, las probabilidades de que el cuidado sufra es 39,66% menos que los cuidados que no pertenecen al IDJ30_39
			2	Para los cuidados que no pertenecen al grupo de enfermedades IDJ30_39, la probabilidad de que se sufran es 1.657 veces mayor que la de los del grupo IDJ30_39
IDJ40_46	0,804	2,2345	1	Para aquellos cuya enfermedad IDJ40_46, la probabilidad de sufrirla es 2.234 mayor que para aquellos cuya enfermedad no es IDJ40_46
			2	En el caso de los cuidados cuya enfermedad no es IDJ40_46, la probabilidad de ser menos afectados es de 2.234 más que los cuidados que pertenecen al IDJ40_46
IDN20_23	1,0171	2,7652	1	Para aquellos cuya enfermedad IDN20_23, la probabilidad de sufrirla es 2.765 mayor que para aquellos cuya enfermedad no es IDN20_23
			2	En el caso de los cuidados cuya enfermedad no es IDN20_23, la probabilidad de ser menos afectados es de 2.765 más que los cuidados que pertenecen al IDN20_23
IDR00_09	-0,7558	0,4696	1	Para los cuidados cuyas enfermedades pertenecen al grupo IDR00_09, las probabilidades de padecerlas son 53,0365% menores que los cuidados que no pertenecen al IDR00_09
			2	Para los cuidados que no pertenecen al grupo de enfermedades del IDR00_09, la probabilidad de que los sufran es de 2.129 veces a los que son del grupo IDR00_09
IDR50_69	1,2829	3,6071	1	Para aquellos cuya enfermedad IDR50_69, la probabilidad de sufrirla es 3.607 mayor que para aquellos cuya enfermedad no es IDR50_69
			2	En el caso de los cuidados cuya enfermedad no es de la IDR50_69, la probabilidad de ser menos afectados es de 3.607 más que los cuidados que pertenecen a la IDR50_69
IDS00_09	0,1782	1,1951	1	Para aquellos cuya enfermedad IDS00_09, la probabilidad de sufrirla es 1,195 mayor que para aquellos cuya enfermedad no es IDS00_09
			2	En el caso de los cuidados cuya enfermedad no es del IDS00_09, la probabilidad de estar menos afectados es de 1,195 más que los cuidados del IDS00_09

VARIABLES	B	EXP(B)	lo	INTERPRETACIÓN
IDT14	-0,9739	0,3776	1	Para aquellos cuyas enfermedades pertenecen al grupo IDT14, la probabilidad de padecerlas es 62,239% menor que para aquellos que no pertenecen al grupo IDT14
			2	Para los cuidados que no pertenecen al grupo de enfermedades de la RTD14, la probabilidad de padecerlas es 2.648 veces mayor que la de los del grupo
IDT20_31	1,1332	3,1056	1	Para los cuidados cuya enfermedad IDT20_31, la probabilidad de sufrirla es 3.105 mayor que para los cuidados que no son IDT20_31
			2	En el caso de los cuidados cuya enfermedad no es IDT20_31, la probabilidad de ser menos afectados es de 3,1055 más que los cuidados que pertenecen a la IDT20_31
IDT51_65	0,0757	1,0786	1	Para aquellos cuya enfermedad IDT51_65, la probabilidad de padecerla es 1,078 mayor que la de no tenerla.
			2	En el caso de los cuidados cuya enfermedad no es IDT51_65, la probabilidad de ser menos afectados es de 1,0786 más que los cuidados que pertenecen al IDT51_65
IDT66_78	0,7725	2,1652	1	Para aquellos cuya enfermedad IDT66_78, la probabilidad de sufrirla es 2.165 mayor que para aquellos cuya enfermedad no es IDT66_78
			2	En el caso de los cuidados cuya enfermedad no es IDT66_78, la probabilidad de ser menos afectados es de 2.165 más que los cuidados que pertenecen a la IDT66_78

Explicación de los resultados de la comprobación de la hipótesis:

- Considerando que el nivel de precisión de la atención de urgencia en la regresión lineal múltiple es del 66,9% que se ha obtenido, se acepta la hipótesis nula  $HE1_0$  y se rechaza la hipótesis alternativa  $HE1_1$ .
- Considerando el nivel de precisión de los servicios de urgencia en la regresión logística ordinal del 63,6% que se ha obtenido, se rechaza la hipótesis nula  $HE2_0$  y se acepta la hipótesis alternativa  $HE2_1$ .
- Considerando el nivel de precisión de los servicios de urgencia con el 97,3% obtenido por Holt-Winter's, se acepta la hipótesis nula  $HE3_0$  y se rechaza la hipótesis alternativa  $HE3_1$ .
- Considerando el nivel de precisión de la atención de urgencia con el método Naive del 96,02% obtenido, se acepta la hipótesis nula  $HE4_0$  y se rechaza la hipótesis alternativa  $HE4_1$ .
- Considerando el nivel de precisión de la atención de urgencia con el método exponencial de tendencia suavizada del 97,03% que se ha obtenido, se acepta la hipótesis nula  $HE5_0$  y se rechaza la hipótesis alternativa  $HE5_1$ .

## **V. DISCUSIÓN**

En este capítulo discutiremos los resultados obtenidos en nuestros 5 modelos aplicados, que son la regresión lineal múltiple, la regresión ordinal logística, Holtwinters, el método Naives y el suavizado exponencial de tendencias, comentaremos con resultados similares o aproximados o temas relacionados con la investigación. Debido a que nos muestran por qué nuestro modelo es mejor o en los otros trabajos de investigación, salieron mejores resultados de predicción, la razón por la que era superior o el nuestro mostró una mayor precisión.

Los resultados del estudio utilizaron la regresión lineal múltiple con una precisión del 66,9% en la predicción de la atención de urgencia en la región de Antofagasta, que fue inferior a los resultados de los estudios de Huang et al. (2020) y Villani et al. (2017), que encontraron en sus resultados utilizando el modelo de regresión lineal múltiple un error de predicción del 9.03%, lo que proporciona una precisión estimada de 91.7% en el aumento de las enfermedades de riesgo de tipo bronquial, el valor fue alto por utilizar una técnica de agrupación adicional como la k-means que reforzó su modelo para optimizar la precisión y el otro resultado fue un error del 7.6%, lo que le dio una precisión del 92.4% debido a valores tan altos porque su pronóstico se preparó para la planificación de los servicios de salud que se reguló por diferentes modelos de la serie Arima y que aumentó su precisión, respectivamente. Los resultados del presente estudio fueron inferiores a los descritos porque tomamos 29 variables con precisión en el momento del incremento de la atención de urgencias y para mejorar el resultado es necesario tener en nuestros datos estas variables género, actividad laboral, turno, clasificación de Riesgo, los datos también se agruparon según el libro de la CIA-10 sobre las enfermedades que producían. en diferentes comportamientos y especificar qué variable generará más impacto y que provocó el reajuste del nivel de precisión.

Los resultados del estudio en la regresión logística ordinal fueron de una predicción estimada del 63,6% con datos de la región de Antofagasta con 29 variables para lograr la mejor precisión en el estudio, que fueron superiores a los resultados de los estudios de Innala et al. (2016) y fueron inferiores a los resultados de Gholipour et al. (2016) que encontraron una precisión en su

modelo de regresión logística ordinal de 52.3% en su estudio sobre el tabaquismo porque sólo trabajaron con un modelo de regresión logística ordinal asociado a otros modelos como el Ex-RA relacionado con el tabaquismo para obtener el mejor resultado posible es el mencionado anteriormente y en el otro modelo de regresión logística del 85,4%, el valor fue alto, ya que trabaja con datos factoriales en todas sus variables, que es la interpretación de los datos más adecuados para esta investigación para predecir el número de casos de tabaquismo, respectivamente. Los resultados del presente estudio permitieron encontrar un nuevo modelo de regresión logística ordinal para predecir la cantidad de atención de emergencias en el registro, para lo cual su precisión no aumentó mucho ya que en los datos que tiene no se encontraron estas variables género, actividad laboral, turno, clasificación de riesgo teniendo relación con la base de datos de atenciones de urgencias encontrada y no se encontró todas las variables de tipo factorial lo cual aumentaría el resultado de predicción obtenido. Sin embargo, los resultados del estudio son mayores que los resultados de Innala et al. (2016) e inferiores que los estudios de Gholipour et al. (2016) que encontraron mejores modelos predictivos para el paciente fumador, respectivamente.

Los resultados del estudio para la predicción y el número de urgencias para el modelo de Holt-Winter's fueron de una predicción estimada del 97,3% con la cantidad de atención de emergencia en la región de Antofagasta, que fueron superiores a los resultados de Zor y Cebi (2018) e inferiores a los resultados de Hernández et al. (2019) quienes encontraron en su modelo una precisión de 92.7% para pronósticos aceptables porque hicieron un estudio con la misma base de datos de entrenamiento para la predicción y el otro estudio reflejó un porcentaje de error para la atención de pacientes con cesárea de 99.881% para mejorar la atención para generar sostenibilidad porque los resultados fueron tan altos debido a que la interpretación de los datos de la atención es correlativa porque al generar pocos por día, el entrenamiento mejora el resultado, respectivamente. Los resultados del presente estudio nos permitieron encontrar un nuevo modelo de Holt-Winter's que fue diferente a los mencionados porque nuestra proyección fue de 2 años, lo que nos mostró una precisión en la serie temporal con todos los datos de atención de emergencias desde el 2014 al 2018

para generar puntos medios de atención para los años 2019 y 2020 en la región de Antofagasta de cada mes.

Los resultados del estudio para la predicción y el número de urgencias para el modelo del método Naive fueron de una predicción estimada del 96,02% con la cantidad de atención de urgencia en la región de Antofagasta, que fue menor que los resultados del estudio de Spreco (2017) y mayor que el estudio de Chand, Rao, Tekabu y Khan (2018) que encontraron una precisión en su modelo Naive de 99.93% para las predicciones, ya que hicieron un estudio con la misma interpretación de los datos de la CIA-10 de la formación para la predicción y el otro estudio nos muestra que con los datos de entrada los modelos Naive son aceptables si tienen al menos un error del 23.83%, lo que también se verifica con nuestros resultados. Los resultados del presente estudio nos permitieron encontrar un nuevo modelo de Naives que fue superior e inferior a los estudios mencionados, ya que uno refleja un mayor nivel de precisión, ya que sólo un capítulo de las enfermedades respiratorias sólo de la CIA-10 y en el otro estudio se registra es menor al estudio realizado supera el porcentaje del 23,83% como error porcentual de su modelo Naive.

Los resultados del estudio para la predicción y el número de atenciones de urgencias para el modelo exponencial de tendencia suavizada de urgencias fueron de una predicción estimada del 97,03% con la cantidad de atención de emergencia en la región de Antofagasta, que fueron superiores a los resultados de los estudios. de Zang et al. (2020) y Ke et al. (2016), que encontraron una precisión en su modelo de 95. 0% de precisión, toman como base esta tendencia, ya que al manejar el modelo les da un intervalo de predicción en el que conocerán la cantidad que puede ser incrementada ya que tienen un punto medio al cual referenciar y en otro estudio que refleja el crecimiento exponencial, pero de acuerdo a la cantidad obtenida en un estudio anterior que da la base de un cierto nivel de cantidad con una precisión del 51.23%, respectivamente. Los resultados del presente estudio nos permitieron encontrar un nuevo modelo exponencial de tendencia suavización que era diferente de los mencionados porque nuestra proyección era de 2 años, lo que nos mostró una precisión en la serie temporal con todos los datos de atenciones de urgencias desde 2014 hasta

2018 generando cantidad de atención para los años 2019 y 2020 en la región de Antofagasta cada mes porque nuestros resultados son mejores porque la herramienta utilizada interpreta mejor el resultado ya que se definió la tendencia y puedo agregar estabilidad pero en los otros casos sólo manejan la tendencia y sólo estiman el resultado final, dan un margen de precisión que también es válido, pero que depende de la estacionalidad que al no tenerla hace que la precisión sea demasiado baja.



## **VI. CONCLUSIONES**

En este capítulo se darán las conclusiones que se pudieron definir en nuestra investigación, todo lo que se logró con nuestra metodología aplicada y con nuestras técnicas de predicción para crear nuevos modelos de atención de emergencias en la región de Antofagasta. Las conclusiones de la investigación fueron las siguientes:

1. Según los resultados de la regresión lineal múltiple del estudio, el rango de 5 años y 14 años de edad es el que menos impacto tiene en la atención, ya que dará la menor cantidad de atención de emergencia, el rango que tendrá el mayor impacto para aumentar la cantidad de atención de emergencia son los pacientes menores de 11 meses de edad.
2. Según los resultados de la regresión lineal múltiple, las enfermedades de gran impacto son las que pertenecen al grupo B99, que son el grupo de enfermedades infecciosas y no especificadas, y la enfermedad de menor impacto, T66\_78, que pertenece al síndrome de irradiación, que son raras por el aumento de la atención de emergencia.
3. De acuerdo con los resultados, Holt-Winter's nos muestra que la predicción del número de cuidados de emergencia en 2020 me muestra que el mes que tendrá el mayor impacto de la cantidad de atenciones es octubre y el mes con menor impacto debido a la menor cantidad de atenciones en enero. del año anteriormente mencionado.
4. De acuerdo con los resultados, el Método Ingenuo nos muestra que la predicción del número de atenciones de emergencia en 2020 me muestra que el mes que tendrá mayor impacto de la atención es julio y el mes con menor impacto debido a la menor cantidad de atención en febrero del año mencionado anteriormente.
5. Según los resultados de la Exponencial de la tendencia suavizada, nos muestra que la predicción del número de atenciones de emergencia en 2020 me muestra que el mes que tendrá mayor impacto de la cantidad de atenciones es julio y el mes con menor impacto debido a la menor cantidad de atenciones el mes de febrero del año antes mencionado.

6. Según los resultados de la regresión logística ordinal, nos muestra que el rango de edad con mayor impacto en la generación de atenciones de emergencia es el comprendido entre los 15 y los 64 años y el rango de edad con menor impacto en las atenciones de emergencia es el de los pacientes menores de 11 meses. El resultado también nos muestra que las condiciones con mayor probabilidad de ocurrir son las que pertenecen al grupo J09\_11 que pertenecen al grupo de gripe y neumonía, y las menos probables son las que pertenecen al grupo T66\_78, que corresponde al síndrome de irradiación.
7. La precisión de la regresión lineal múltiple, si se definieran mejores variables con correlación o agrupaciones de datos, el R ajustado obtendría mejores resultados. Por lo tanto, en el presente proyecto de investigación se definieron 29 variables. Obteniendo un porcentaje de precisión del 66,9% en el número de servicios de emergencia en la región de Antofagasta, es decir, la precisión estimada para el comportamiento de cada variable para ver el crecimiento y la caída del número de atención de urgencia.
8. La precisión de la regresión logística ordinal, si define mejores variables con correlación o agrupación de datos además de mejorar la conversión de las variables a tipo factorial, el Psuedo R2 obtendría mejores resultados. Por lo tanto, en el presente proyecto de investigación se definieron 29 variables. Obteniendo un porcentaje de precisión del 63,6% en el número de servicios de emergencia en la región de Antofagasta, como resultado podemos ver el comportamiento de crecimiento y caída de la atención urgente de la atención urgente
9. La precisión de los holtwinter's obtenidos fue del 97,3%, lo que es una predicción muy alta. Esta precisión fue determinada y el modelo pudo adaptar la cantidad de atención y previsión a un nivel detallado, ya que nos dio un margen de error del 2,7 % la razón de este error mínimo el modelo refleja la proyección de 2 años que se ajusta a su funcionalidad ya que esta técnica determina la estacionalidad y la tendencia de aumento

evaluando las diferentes cantidades de los años anteriores para hacer la previsión.

10. La precisión del Método Naives obtuvo una precisión del 96,02%. El resultado de esta predicción estimada se debió a que el modelo se adaptó a las funciones de la herramienta Rstudio, ya que una vez definida la función debemos dar la estacionalidad y la tendencia se adquiere a partir de la cantidad de datos cargados y esa precisión mejora porque definimos un punto de partida para predecir y donde queremos terminar la estimación de la cantidad de atención.
11. El método exponencial de tendencia suavizada. Es un método que proviene del exponencial suavizado que sólo muestra constantes estacionales, pero su algoritmo de tendencia se utilizó para establecer una tendencia y mejorar los resultados, que tenían un nivel de precisión. 97,03% porque tenía una cantidad similar a la de Holtwinters porque los exponenciales suavizados son técnicas muy similares a las de Holt para las que se asemejan a su precisión ya que fueron evaluadas con la misma cantidad de atención de emergencia durante meses a lo largo de 5 años, pero definimos la tendencia ya que toma como tendencia el aumento medio de los datos por año. La tendencia se interpreta entonces para indicar la precisión estimada del aumento de nuestro pronóstico de atención de emergencia en la región de Antofagasta.
12. El uso de la metodología CRISP-DM es adaptable para utilizarla en la comparación de modelos ya que en todas sus fases previas nos ayuda a recopilar los datos, limpiar, interpretar y comprender para llevar a cabo la carga final, que son los datos que nos ayudarán a hacer nuestra nueva previsión de atención de emergencias con diferentes modelos para encontrar la mejor.
13. La presentación de un sistema de SIG que nos muestra la demografía para mostrar la cantidad de atención en los hospitales de la región de Antofagasta, puede adaptarse para que nuestra carga de datos muestre cuál es la cantidad en los diferentes hospitales, el aumento de la atención

por condición, la distribución del número de atención por rango de edad y preparar un pronóstico de la atención de emergencia para que se pueda reflejar la estimación de las cantidades esperadas por mes.

## **VII. RECOMENDACIONES**

Las recomendaciones para futuras investigaciones son las siguientes:

1. Desarrollar la investigación científica a nivel nacional con todos los hospitales de Chile. Por razones de recursos tecnológicos, sólo fue posible cargar datos de la región de Antofagasta. Se recomienda aplicar a todas las demás regiones o a una región específica debido a que el modelo realizado puede ser ejecutado con otras regiones y así poder evaluar cuál tendrá la mayor cantidad de atención por región o ver la cantidad a lo largo de los años.
2. Se recomienda realizar el mismo estudio para el Perú solicitando permiso al MINSA para poder obtener los datos e interpolarlos con los modelos propuestos para hacer un pronóstico de la atención de emergencias y urgencias a nivel nacional o replicarlo para ver su efectividad en una sola región.
3. Para futuros estudios que tengan en cuenta la pandemia que sufre la sociedad y que impacta sobre COVID-19 en todos los países, ésta podría considerarse una nueva enfermedad ya que está relacionada con diversas enfermedades como la fiebre, la tos, el dolor muscular, etc. Se recomienda realizar el mismo estudio para ver el cambio que se produjo en la atención a lo largo del año 2020.
4. Se recomienda realizar el mismo estudio para otros países como México y los Estados Unidos en América del Norte, en Europa los países España e Italia, lo que se pudo comprobar que su base de datos de atención de emergencia es pública, ya que facilita la recopilación de datos para este nuevo estudio. Se recomienda la aplicación de otros tipos de técnicas de previsión de extracción de datos para mejorar los resultados de la cantidad de atención o proponer una muestra más amplia para los nuevos resultados. Para poder obtener un resultado ya no a nivel regional sino para saber a nivel nacional y hacer predicciones por regiones.
5. Utilizar otras técnicas de agrupación o clasificación estadística que puedan conducir a una mejor interpretación de los datos procesados en futuras investigaciones. Esto sería realizar un estudio de impacto con el

GIS y poder conocer la región que ha tenido más atención de emergencias a lo largo del tiempo y continuar con los mismos indicadores o agregar otros adicionales para saber cuál es el punto principal que genera más atención de emergencias por región en todo el país de Chile.

6. Se recomienda mejorar la recopilación de datos para mejorar los procesos de investigación y no perder el tiempo con datos incompletos, ya que se recomienda disponer de una base de datos de registros de 5 años antes de la investigación para tener una predicción precisa. Y esto mejoraría la precisión ya que los registros completos al realizar la depuración no perderán datos para futuras investigaciones.
7. Recomendamos poner más énfasis en la parte de herramientas de minería de datos ya que en esta investigación se utiliza la herramienta para poder hacer los nuevos modelos se recomendaría utilizar 2 o más herramientas para validar que el mismo modelo creado puede ser ejecutado en diferentes tipos de herramientas de minería de datos y los resultados seguirán siendo los mismos variarán en décimas de la precisión.
8. Para el desarrollo la investigación ampliar la cantidad de variables es posible considerar una dimensión de impacto que se guiará de la variable enfermedad o padecimiento que permitiría conocer cuál es el padecimiento más recurrente o más común en las distintas regiones de Chile y también poder ampliar no sólo a nivel regional sino de provincias y comunas que permitan tener un mejor detalle en el sistema.
9. Se recomienda ampliar los criterios de ubicación del Sistema, para no ser una muestra de una región sino agregar establecimientos por provincias y comunas que permitan tener una detallada sobre las dolencias más comunes y conocer las pocas detectadas esto con el fin de que al finalizar la investigación se puedan recomendar campañas para prevenir esta dolencia en la provincia o comuna específica.
10. Para futuras investigaciones, recopilar artículos o libros en idioma mandarín o ruso donde se han observado información de la aplicación de



las técnicas en el area de medicina, debido a esta limitación de idioma, sólo se tomó el español y el inglés. Por lo tanto, se recomienda investigar artículos en esos idiomas. Ampliar la idea del proyecto de investigación ya que la atención de emergencia va de la mano de la atención urgencias y así poder tener un módulo de impacto de geolocalización por cantidad atenciones de emergencia y urgencia, además conocer cuál es el tipo de establecimientos que atiende al mayor número de pacientes que solicitan atención inmediata.

## **REFERENCIAS**

- 1&1 IONOS España S.L.U. (2019). *Software de data mining: las mejores herramientas*. Retrieved from 11 octubre, 2019, de [https://www.ionos.es/digitalguide/online-marketing/analisis-web/software-de-data-mining-las-mejores-herramientas/?fbclid=IwAR3roXHQLmLvzorlAwoe5daGskl-\\_BZWjInj4w7D8OGpy1Vwq5S634vM7MA](https://www.ionos.es/digitalguide/online-marketing/analisis-web/software-de-data-mining-las-mejores-herramientas/?fbclid=IwAR3roXHQLmLvzorlAwoe5daGskl-_BZWjInj4w7D8OGpy1Vwq5S634vM7MA)
- Adeyemi, G. A., Markus, M., Gbolahan, O. G., & Edeki, S. O. (2018). *Landfill Site Selection in ADO-ODO Ota LGA Using Geographic Information Systems [Selección de vertederos en ADO-ODO Ota LGA usando Sistemas de Información Geográfica]*. *International Journal of Civil Engineering and Technology (IJCIET)*, 9(8), 1434-1445.
- Akter, F., Hossin, M. A., Daiyan, G. M., & Hossain, M. M. (2018). *Classification of Hematological Data Using Data Mining Technique to Predict Diseases [Clasificación de los datos hematológicos mediante la técnica de extracción de datos para predecir enfermedades]*. *Journal of Computer and Communications*, 6(04), 76.
- Ali, M., Lee, S., & Kang, B. H. (2018). *KEM-DT: A Knowledge Engineering Methodology to Produce an Integrated Rules Set Using Decision Tree Classifiers [KEM-DT: Una metodología de ingeniería del conocimiento para producir un conjunto integrado de reglas usando clasificadores de árbol de decisión]*. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, 1-5.
- Amirabadizadeh, A., Nezami, H., Vaughn, M. G., Nakhaee, S., & Mehrpour, O. (2018). *Identifying Risk Factors for Drug Use in an Iranian Treatment Sample: A Prediction Approach Using Decision Trees [Identificación de los factores de riesgo del consumo de drogas en una muestra de tratamiento iraní: Un enfoque de predicción utilizando árboles de decisión]*. *Substance use & misuse*, 53(6), 1030-1040.
- Arancibia, J. N., Sánchez, F. J. M., del Rey Mejías, Á., del Castillo, J. G., Cháfer, J., Briñon, M. G., & Aguilar, G. S. (2019). *Evaluation of a Diagnostic Decision Support System for the Triage of Patients in a Hospital Emergency Department [Evaluación de un sistema de apoyo a las*

- decisiones diagnósticas para la selección de pacientes en el servicio de emergencia de un hospital]. IJIMAI, 5(4), 60-67.*
- Bansal, A., & Srivastava, S. (2018). *Tools Used in Data Analysis: A Comparative Study [Herramientas utilizadas en el análisis de datos: Un estudio comparativo].*
- Berente, N., Seidel, S., & Safadi, H. (2018). *Research Commentary—Data-Driven Computationally Intensive Theory Development [Comentario de Investigación-Desarrollo de Teoría Intensiva de Computación Dirigida por Datos]. Information Systems Research, 30(1), 50-64*
- Betancourt, D. F. (2016). *IT Businesses: New Revolution [Negocios de IT: Nueva Revolución]. Published December 15, 2015.* Retrieved from <https://ingenioempresa.com/suavizacion-exponencial-doble/>
- Betancourt, D. F. (February 22, 2016). *The trend in double exponential smoothing or holt model [La tendencia en el doble alisado exponencial o modelo Holt].* Retrieved on July 8, 2020, from Ingenio Empresa: [www.ingenioempresa.com/suavizacion-exponencial-doble](http://www.ingenioempresa.com/suavizacion-exponencial-doble).
- Bibri, S. E., & Krogstie, J. (2018). *The Big Data Deluge for Transforming the Knowledge of Smart Sustainable Cities: A Data Mining Framework for Urban Analytics [El gran diluvio de datos para transformar el conocimiento de las ciudades sostenibles inteligentes: Un marco de extracción de datos para el análisis urbano]. In Proceedings of the 3rd International Conference on Smart City Applications.*
- Cassiani, S. H. B. (2014). *Strategy for Universal Access to Health and Universal Health Coverage and the Contribution of the International Nursing Networks [Estrategia para el acceso universal a la salud y la cobertura sanitaria universal y la contribución de las redes internacionales de enfermería]. Revista Latino-Americana de Enfermagem, 22(6).*
- Catal, C., Kaan, E. C. E., Arslan, B., & Akbulut, A. (2019). *Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting [Benchmarking de algoritmos de regresión y técnicas de análisis de series temporales para la previsión de ventas]. Balkan Journal of Electrical and Computer Engineering, 7(1), 20-26.*
- Chand, R., Rao, D. K., Tekabu, T. B., & Khan, M. G. M. (2018). *Modeling Breast Cancer Cases in Fiji [Modelización de casos de cáncer de mama en Fiji].*

2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE).

- Chen, J., Li, K., Rong, H., Bilal, K., Yang, N., & Li, K. (2018). *A Disease Diagnosis and Treatment Recommendation System Based On Big Data Mining and Cloud Computing [Un sistema de diagnóstico y recomendación de enfermedades basado en la gran minería de datos y la computación en nube]*. *Information Sciences*, 435, 124-149.
- Chen, P. (2019). *Reasonable Evaluation Model of Pediatric Disease Based on pharmacology and Data Mining [Modelo de evaluación razonable de enfermedades pediátricas basado en la farmacología y la extracción de datos]*. *Investigación Clínica*, 60(5).
- Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). *Data Cleaning [Limpieza de datos]*. *Proceedings of the 2016 International Conference on Management of Data - SIGMOD'16*.
- Cruz, C., Olivares, S. y González, M. (2014). *Metodología de la investigación*. México D.F.: Grupo Editorial Patria.
- Departamento de Estadísticas e información de salud (DEIS). (s.f.). *Preguntas Frecuentes - DEIS*. Recuperado 23 octubre, 2019, de <http://www.deis.cl/preguntas-frecuentes/>
- Departamento de Estadísticas e Información de Salud. (s. f.). DEIS. Recuperado 1 de marzo de 2020, de <https://deis.minsal.cl/#datosabiertos>
- Domadiya, N., & Rao, U. P. (2018). *Privacy-Preserving Association Rule Mining for Horizontally Partitioned Healthcare Data: A Case Study on the Heart Diseases [La Asociación para la Preservación de la Privacidad de la Minería de Reglas para los datos de atención médica divididos horizontalmente: Un estudio de caso sobre las enfermedades cardíacas]*. *Sādhanā*, 43(8), 127.
- El Jerjawi, N. S., & Abu-Naser, S. S. (2018). *Diabetes Prediction Using Artificial Neural Network [Predicción de la diabetes usando la red neuronal artificial]*.
- Farooqui, N. A., & Mehra, R. (2018). *Design of A Data Warehouse for Medical Information System Using Data Mining Techniques [Diseño de un almacén de datos para un sistema de información médica utilizando técnicas de*

- minería de datos*. 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), 199-203.
- Field, A. (2009). *Discovering Statistics Using SPSS [Descubriendo las estadísticas usando SPSS]*. SAGE Publications.
- Ford, E. W., Hesse, B. W., & Huerta, T. R. (2016). *Personal health record use in the United States: forecasting future adoption levels [Uso de registros sanitarios personales en los Estados Unidos: previsión de futuros niveles de adopción]*. *Journal of medical Internet research*, 18(3), e73.
- Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., & Martínez-Álvarez, F. (2019). *Multi-Step Forecasting For Big Data Time Series Based On Ensemble Learning [Pronóstico de múltiples pasos para grandes series temporales de datos basado en el aprendizaje en conjunto]*. *Knowledge-Based Systems*, 163, 830-841.
- Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). *Big Data Techniques in Auditing Research and Practice: Current Trends and Future Opportunities [Técnicas de grandes datos en la investigación y la práctica de la auditoría: Tendencias actuales y oportunidades futuras]*. *Journal of Accounting Literature*, 40, 102-115.
- Ghaibeh, A. A., Kasem, A., Ng, X. J., Nair, H. L. K., Hirose, J., & Thiruchelvam, V. (2018). *Gaining Insights on Nasopharyngeal Carcinoma Treatment Outcome Using Clinical Data Mining Techniques [Obtención de conocimientos sobre los resultados del tratamiento del carcinoma nasofaríngeo mediante técnicas de extracción de datos clínicos]*. *MIE*, 386-390.
- Gholipour K; Asghari-Jafarabadi M; Iezadi S; Jannati A; Keshavarz S. (2018) *Modelling the prevalence of diabetes mellitus risk factors based on artificial neural network and multiple regression [Modelización de la prevalencia de los factores de riesgo de la diabetes mellitus basada en la red neural artificial y la regresión múltiple]*. *East Mediterr Health J.*;24(8):770–777.
- GmbH, R. (s. f.). 404 - RapidMiner Documentation [Documentación sobre RapidMiner]. RapidMiner. Recuperado 2 de mayo de 2020, de <https://docs.rapidminer.com/9.5/studio/installation/system-requirements.html>)

- González R., M. J. (1997). *Metodología de la investigación social. Técnicas de recolección de datos*. Aguacalera: Madrid (España).
- Grecol, M. L. (2012). *Microsoft SSIS and Pentaho Kettle: A Comparative Study for Three-Tier Data Warehouses [Microsoft SSIS y Pentaho Kettle: Un estudio comparativo para almacenes de datos de tres niveles]*.
- Hernández, R. (2014). *Metodología de la investigación*. Edit. McGrawHill. México.
- Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación (6ª ed.)*. México D. F.: McGraw Hill.
- Hsieh, P.-H., Yang, S.-J., & Yang, D.-L. (2015). *Decision support for unrelated parallel machine scheduling with discrete controllable processing times [Soporte de decisiones para la programación de máquinas paralelas no relacionadas con tiempos de procesamiento discretos y controlables]*. *Applied Soft Computing*, 30, 475–483.
- Huang, Y., Zhao, Q., Zhou, Q., & Jiang, W. (2018). *Air Quality Forecast Monitoring and Its Impact on Brain Health Based on Big Data and the Internet of Things [El monitoreo del pronóstico de la calidad del aire y su impacto en la salud del cerebro basado en grandes datos y el Internet de las cosas]*. *IEEE Access*, 6, 78678-78688.
- Huang, Z., Lin, S., Long, L., Cao, J., Luo, F., Qin, W., ... Gregersen, H. (2020). *Predicting the Morbidity of Chronic Obstructive Pulmonary Disease Based on Multiple Locally Weighted Linear Regression Model with K-means Clustering [Predicción de la morbilidad de la enfermedad pulmonar obstructiva crónica basada en un modelo de regresión lineal múltiple ponderada localmente con agrupación de K-means]*. *International Journal of Medical Informatics*, 104141
- Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). *DMME: Data Mining Methodology for Engineering Applications—A Holistic Extension to the CRISP-DM Model [DMME: Metodología de Minería de Datos para Aplicaciones de Ingeniería - Una Extensión Holística al Modelo CRISP-DM]*. *Procedia CIRP*, 79, 403-408.
- IBM Corporation. (1994-2012). *Manual CRISP-DM de IBM SPSS Model [Manual CRISP-DM de IBM SPSS Model]*. *IBM SPSS Modeler 15*. Recuperado de <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>

- Innala, L., Sjöberg, C., Möller, B., Ljung, L., Smedby, T., Södergren, A., ... Wållberg-Jonsson, S. (2016). *Co-morbidity in patients with early rheumatoid arthritis - inflammation matters [La comorbilidad en pacientes con artritis reumatoide temprana - la inflamación]*. *Arthritis Research & Therapy*, 18(1).
- Ionita, I., & Ionita, L. (2016). *Applying Data Mining Techniques in Healthcare [Aplicación de técnicas de minería de datos en la atención de la salud]*. *Studies in Informatics and Control*, 25(3), 385-394.
- Ionita, I., & Ionita, L. (2018). *Classification Algorithms of Data Mining Applied for Demographic Processes [Algoritmos de clasificación de la minería de datos aplicados a los procesos demográficos]*. *BRAIN Broad Research in Artificial Intelligence and Neuroscience*, 9(1), 94-100.
- Islam, M. M., Wu, C. C., Poly, T. N., Yang, H. C., & Yu-Chuan (Jack) Li. (2018). *Applications of Machine Learning in Fatty Liver Disease Prediction [Aplicaciones del aprendizaje automático en la predicción de las enfermedades de los gordos vivos]*. *MIE*, 166-170.
- Islam, M. S., Hasan, M. M., Wang, X., & Germack, H. D. (2018). *A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining [Una revisión sistemática de los análisis de la atención sanitaria: Aplicación y perspectiva teórica de la minería de datos]*. *Healthcare*, 6(2).
- Jain, G., & Mallick, B. (2017). *A study of time series models ARIMA and ETS [Un estudio de los modelos de series temporales ARIMA y ETS]*. Available at SSRN 2898968.
- Jun-Ya Zeng, Jian-Bang Lin & Tian Wang (2017). *A New Competing Risks Model for Predicting Prepayment and Default Using Data Mining [Un nuevo modelo de riesgos en competencia para predecir el pago anticipado y el incumplimiento mediante la minería de datos]*. *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, 985-989.
- Kaur, P., & Sharma, M. (2018). *Analysis of Data Mining and Soft Computing Techniques in Prospecting Diabetes Disorder in Human Beings: A Review [Análisis de las técnicas de minería de datos y de computación suave en la prospección del trastorno de la diabetes en los seres humanos: Reseña]*. *Int. J. Pharm. Sci. Res*, 9, 2700-2719.



- Ke, G., Hu, Y., Huang, X., Peng, X., Lei, M., Huang, C., et al. (2016). *Epidemiological analysis of hemorrhagic fever with renal syndrome in China with the seasonal-trend decomposition método and the exponential smoothing model [Análisis epidemiológico de la fiebre hemorrágica con síndrome renal en China con el método de descomposición de tendencia estacional y el modelo de alisamiento exponencial]*. *Scientific Reports*, 6(1).
- Kefelegn, S., & Kamat, P. (2018). *Prediction and Analysis of Liver Disorder Diseases by using Data Mining Technique: Survey [Predicción y análisis de enfermedades de trastornos hepáticos mediante la técnica de la minería de datos: Encuesta]*. *International Journal of Pure and Applied Mathematics*, 118(9), 765-770.
- Ku, C. C., & Dodd, P. J. (2019). *Forecasting the impact of population ageing on tuberculosis incidence [Previsión de los efectos del envejecimiento de la población en la incidencia de la tuberculosis]*. *PloS one*, 14(9), e0222937.
- Kumar, B. R. (2018). *A Survey on Patient Flow Prediction via Mutually-Correcting Process [Una encuesta sobre la predicción del flujo de pacientes mediante un proceso de corrección mutua]*. *Journal of Network Communications and Emerging Technologies (JNCET)*, 8(8).
- Li, P., Yates, S. N., Lovely, J. K., & Larson, D. W. (2015). *Patient-like-mine: A Real Time, Visual Analytics Tool for Clinical Decision Support [Paciente como una mina: Una herramienta analítica visual en tiempo real para el apoyo de decisiones clínicas]*. *2015 IEEE International Conference on Big Data (Big Data)*, 2865-2867
- Livieris, I., Kotsilieris, T., Dimopoulos, I., & Pintelas, P. (2018). *Decision Support Software for Forecasting Patient's Length of Stay [Software de apoyo a la decisión para pronosticar la duración de la estancia del paciente]*. *Algorithms*, 11(12), 199.
- Lopes, M. L. M., Minussi, C. R., & Lotufo, A. D. P. (2005). *Electric load forecasting using a fuzzy ART & ARTMAP neural network [Pronóstico de carga eléctrica usando una red neural borrosa de ART & ARTMAP]*. *Applied Soft Computing*, 5(2), 235–244.
- Lousa, A., Pedrosa, I., & Bernardino, J. (2019). *Evaluation and Analysis of Business Intelligence Data Visualization Tools [Evaluación y análisis de*

- los instrumentos de visualización de datos de inteligencia commercial]. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI).*
- Luna, R. Jr., Sapio, A., Sawyer, K. R. Bloechl. M. O. (2019). U.S. Patent No 10,234,536. Annapolis, MD:US, Patent and Trademark Office
- Manikandan, V., Porkodi, V., Mohammed, A. S., & Sivaram, M. (2018). *Privacy Preserving Data Mining Using Threshold Based Fuzzy Cmeans Clustering [La preservación de la privacidad de la minería de datos mediante la agrupación de medios difusos basados en umbrales]. ICTACT Journal on Soft Computing, 9(1).*
- Massaro, A., Maritati, V., & Galiano, A. (2018). *Data Mining Model Performance of Sales Predictive Algorithms Based On Rapidminer Workflows [Modelo de minería de datos Rendimiento de los algoritmos de predicción de ventas basado en los flujos de trabajo de Rapidminer]. Int. J. Comp. Sci. Inf. Technol, 10, 39-56.*
- Messias, V. R., Estrella, J. C., Ehlers, R., Santana, M. J., Santana, R. C., & Reiff-Marganiec, S. (2016). *Combining time series prediction models using genetic algorithm to autoscaling web applications hosted in the cloud infrastructure [Combinación de modelos de predicción de series de tiempo utilizando un algoritmo genético para la auto-escalada de aplicaciones web alojadas en la infraestructura de la nube]. Neural Computing and Applications, 27(8), 2383-2406.*
- Mirza, S., Mittal, S., & Zaman, M. (2018). *Decision Support Predictive Model for Prognosis of Diabetes Using SMOTE and Decision Tree [Modelo predictivo de apoyo a la decisión para el pronóstico de la diabetes utilizando SMOTE y el árbol de decisiones]. International Journal of Applied Engineering Research, 13(11), 9277-9282.*
- Mishra, N., Soni, H. K., Sharma, S., & Upadhyay, A. K. (2018). *Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data [Desarrollo y análisis de modelos de redes neuronales artificiales para la predicción de lluvias mediante el uso de datos de series temporalis]. International Journal of Intelligent Systems and Applications, 11(1), 16.*

- Modhukur, V., Iljasenko, T., Metsalu, T., Lokk, K., Laisk-Podar, T., & Vilo, J. (2018). *Methsurv: A Web Tool to Perform Multivariable Survival Analysis Using DNA Methylation Data [Methsurv: Una herramienta web para realizar análisis multivariados de supervivencia utilizando datos de metilación de AND]*. *Epigenomics*, 10(3), 277-288.
- Montalvo-Garcia, J., Quintero, J. B., & Manrique-Losada, B. (2019). *CRISP-DM/SMEs: A Data Analytics Methodology for Non-profit SMEs [CRISP-DM/PYME: Una metodología de análisis de datos para las PYMES sin ánimo de lucro]*. *Advances in Intelligent Systems and Computing*, 449-457.
- Mukhina, K. D., Visheratin, A. A., Mbogo, G. K., & Nasonov, D. (2018). *Forecasting of the Urban Area State Using Convolutional Neural Networks [Previsión del estado de la zona urbana utilizando redes neuronales convolucionales]*. *2018 23rd Conference of Open Innovations Association (FRUCT)*, 268-275.
- Mustafee, N., Powell, J. H., & Harper, A. (2018). *RH-RT: A Data Analytics Framework for Reducing Wait Time at Emergency Departments and Centres for Urgent Care [RH-RT: Un marco de análisis de datos para reducir el tiempo de espera en los departamentos de emergencia y centros de atención urgente]*. *2018 Winter Simulation Conference (WSC)*, 100-110.
- Nahar, N., & Ara, F. (2018). *Liver Disease Prediction by Using Different Decision Tree Techniques [Predicción de enfermedades hepáticas mediante el uso de diferentes técnicas de árbol de decisión]*. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 8(1).
- Nichols, J., England, R., Holliday, S., & Newton, J. L. (2019). *Clinical Care Pharmacists in Urgent Care in North East England: A Qualitative Study of Experiences after Implementation [Farmacéuticos de atención clínica en atención de urgencia en el noreste de Inglaterra: Un estudio cualitativo de las experiencias después de la aplicación]*. *Pharmacy*, 7(3), 114.
- NicoleLeaper, V. A. P. B. (2019, 9 febrero). *A visual guide to CRISP-DM methodology [Una guía visual de la metodología CRISP-DM]*. EXDE. <https://exde.wordpress.com/2009/03/13/a-visual-guide-to-crisp-dm-methodology/>

- Ñaupas, H., Mejía, E., Novoa, E. y Villagómez, A. (2014). *Metodología de la investigación cuantitativa - cualitativa y redacción de la tesis*. Bogotá, Colombia: Ediciones de la U.
- Obermeyer, Z., Clarke, A. C., Makar, M., Schuur, J. D., & Cutler, D. M. (2016). *Emergency Care Use and the Medicare Hospice Benefit for Individuals with Cancer with a Poor Prognosis [El uso de cuidados de emergencia y el beneficio de hospicio de Medicare para personas con cáncer de mal pronóstico]*. *Journal of the American Geriatrics Society*, 64(2), 323–329.
- Orenga-Roglá, S., & Chalmeta, R. (2017). *Methodology for the Implementation of Knowledge Management Systems 2.0 [Metodología para la aplicación de sistemas de gestión del conocimiento 2.0]*. *Business & Information Systems Engineering*, 61(2).
- Prayudani, S., Hizriadi, A., Lase, Y. Y., & Fatmi, Y. (2019, November). *Analysis Accuracy of Forecasting Measurement Technique on Random K-Nearest Neighbor (RKNN) Using MAPE and MSE [Análisis de la Precisión de la Técnica de Medición de Previsión sobre el Vecino Más Cercano al Azar (RKNN) usando MAPE y MSE]*. In *Journal of Physics: Conference Series* (Vol. 1361, No. 1, p. 012089). IOP Publishing.
- Prerana, Sehgal, P., & Taneja, K. (2015). *Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network [Minería de datos predictiva para el diagnóstico de enfermedades de la tiroides mediante la red neuronal]*. *International Journal of Research in Management, Science & Technology*, 3(2), 75-80.
- Pulido P., M. (2015). *Ceremonial y protocolo: métodos y técnicas de investigación científica*. *Universidad de Cádiz: España*, 31(1), 1137-1156.
- Pyle, D. (2003). *Business Modeling and Data Mining [Modelado de negocios y minería de datos]*. San Francisco, USA: Morgan Kaufmann Publishers.
- Rasitha Banu, G. (2016). *Predicting Thyroid Disease using Linear Discriminant Analysis (LDA) Data Mining Technique [Predicción de la enfermedad de la tiroides mediante la técnica de minería de datos del Análisis Discriminante Lineal (LDA)]*. *Communications on Applied Electronics*, 4(12), 4-6.
- Rodríguez, M., León, Á., Alvin, H., Mora, M. E., Díaz Rodríguez, M., & Díaz Mora, M. E. (2016). *Introducción al análisis estadístico multivariado aplicado*.

- Experiencia y casos en el Caribe colombiano*. Universidad del Norte.  
[https://books.google.com.pe/books?id=x3daDwAAQBAJ&dq=libro+estad%C3%ADstico+multivariado+pdf&lr=&hl=es&source=gbs\\_navlinks\\_s](https://books.google.com.pe/books?id=x3daDwAAQBAJ&dq=libro+estad%C3%ADstico+multivariado+pdf&lr=&hl=es&source=gbs_navlinks_s)
- Romero Sevilla, R., et al. (2016). *A Stroke Care Management System Prevents Outcome Differences Related to Time of Stroke Unit Admission [Un sistema de gestión del cuidado de los pacientes con apoplejía evita las diferencias de resultados relacionadas con el tiempo de admisión en la unidad de apoplejía]*. *Neurología*, 31(3), 149—156
- Rosas-Hernandez, L., Tlapa, D., Báez-López, Y.A., Limon-Romero, J., & Pérez-Sánchez, A. (2019). *Use of Lean Healthcare and DMAIC to Reduce Waste in a Public Hospital [El uso de Lean Healthcare y DMAIC para reducir los residuos en un hospital público]*.
- Roy, A. K. (2016). *Impact of Big Data Analytics on Healthcare and Society [Impacto de los grandes análisis de datos en la salud y la sociedad]*. *Journal of Biometrics & Biostatistics*, 7(3).
- Rózańska, A., Dymerski, T., & Namieśnik, J. (2018). *Novel Analytical Method for Detection of Orange Juice Adulteration Based On Ultra-Fast Gas Chromatography [Novedoso método analítico para la detección de la adulteración del zumo de naranja basado en la cromatografía de gases ultrarrápida]*. *Monatshefte für Chemie-Chemical Monthly*, 149(9), 1615-1621.
- RStudio. (s. f.). *RStudio Documentation [Documentación de RStudio]*. RStudio Administration Guide. <https://docs.rstudio.com/rsp/documentation>.
- Saggi, M. K., & Jain, S. (2018). *A Survey towards an Integration of Big Data Analytics to Big Insights for Value-Creation [Una encuesta para la integración de grandes análisis de datos con grandes conocimientos para la creación de valor]*. *Information Processing & Management*, 54(5), 758-790.
- Shahzadeh, A., Khosravi, A., & Nahavandi, S. (2015). *Improving load forecast accuracy by clustering consumers using smart meter data [Mejorar la precisión de la previsión de carga agrupando a los consumidores con datos de medidores inteligentes]*. *2015 International Joint Conference on Neural Networks (IJCNN)*.

- Sidiq, U. & Ahmad Khan, R. (2017). *Comparative Study of Existing Techniques for Diagnosing Various Thyroid Ailments [Estudio comparativo de las técnicas existentes para el diagnóstico de diversas enfermedades de la tiroides]*. *International Research Journal of Engineering and Technology*, 4(9).
- Silva, H. C., Pessoa, R. L., & Menezes, R. M. P. (2016). *Trauma in Elderly People: Access to the Health System through Pre-Hospital Care [Trauma en personas mayores: Acceso al sistema de salud a través de la atención prehospitalaria]*. *Revista Latino-Americana de Enfermagem*, 24.
- Spreco, A., Eriksson, O., Dahlström, Ö., & Timpka, T. (2017). *Influenza detection and prediction algorithms: comparative accuracy trial in Östergötland county, Sweden, 2008–2012 [Algoritmos de detección y predicción de la gripe: ensayo de precisión comparativa en el condado de Östergötland, Suecia, 2008-2012]*. *Epidemiology and Infection*, 145(10), 2166–2175.
- Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, et al. (2013) *Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research [Estrategia de Investigación de Pronóstico (PROGRESS) 3: Investigación de Modelo de Pronóstico]*. *PLoS Med* 10(2): e1001381.
- Tan, Y., & Shi, Y. (2016). *Data Mining and Big Data: First International Conference [Minería de datos y Big Data: Primera Conferencia Internacional]*, *DMBD 2016*, Bali, Indonesia: Springer International Publishing.
- Tanyavutti, A., & Tanlamai, U. (2018). *ARIMAX versus Holt-Winter's Methods: The Case of Blood Demand Prediction in Thailand [ARIMAX Vs. los métodos de Holt-Winter: El caso de la predicción de la demanda de sangre en Tailandia]*. *International Journal of Environmental & Science Education*, 13(6), 519-525.
- Tariq, H. I., Sohail, A., Aslam, U., & Batcha, N. K. (2019). *Loan Default Prediction Model Using Sample, Explore, Modify, Model, and Assess (SEMMA) [Modelo de Predicción de Préstamo por Defecto Usando la Muestra, Exploración, Modificación, Modelo y Evaluación (SEMMA)]*. *Journal of Computational and Theoretical Nanoscience*, 16(8), 3489–3503.
- Teixeira, S. (2018). *Qualitative Geographic Information Systems (GIS): An Untapped Research Approach for Social Work [Sistemas de información*

- geográfica cualitativa (SIG): Un enfoque de investigación sin explotar para el trabajo social]. Qualitative Social Work, 17(1), 9-23.*
- Tobback, E., Naudts, H., Daelemans, W., de Fortuny, E. J., & Martens, D. (2018). *Belgian Economic Policy Uncertainty Index: Improvement through Text Mining [Índice de incertidumbre de la política económica belga: Mejora a través de la minería de textos]. International journal of forecasting, 34(2), 355-365.*
- Vargas C., Z. R. (2009). *La investigación aplicada: una forma de conocer las realidades con evidencia científica. Revista Educación, 33 (1),155-165.*
- Vásquez, R., Muñoz, C., and Ramos, C. (2019). *El triage hospitalario en los servicios de emergencia. [Artículo de Revisión].19 (1):90-100.*
- Villani, M., Earnest, A., Nanayakkara, N., Smith, K., de Courten, B., & Zoungas, S. (2017). *Time series modelling to forecast prehospital EMS demand for diabetic emergencies [Modelización de series temporales para pronosticar la demanda de servicios médicos de urgencia prehospitalarios para emergencias diabéticas.]. BMC Health Services Research, 17(1).*
- Villena, J. (2012, 2 agosto). *CRISP-DM: La metodología para poner orden en los proyectos. Sngular. Extraído de <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>*
- Wang, W. J., Su, M. C., Lee, S. H., Hung, C. S., & Chen, C. C. (2018). *A Guideline to Determine the Training Sample Size When Applying Big Data Mining Methods in Clinical Decision Making [Una guía para determinar el tamaño de la muestra de entrenamiento cuando se aplican grandes métodos de minería de datos en la toma de decisiones clínicas]. 2018 IEEE International Conference on Applied System Invention (ICASI), 678-681*
- Wang, X., Lee, W.-J., Huang, H., Szabados, R. L., Wang, D. Y., & Van Olinda, P. (2016). *Factors that Impact the Accuracy of Clustering-Based Load Forecasting [Factores que influyen en la precisión de la previsión de carga basada en la agrupación]. IEEE Transactions on Industry Applications, 52(5), 3625–3630.*
- Wang, Y., Kung, L., & Byrd, T. A. (2018). *Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations [Gran análisis de datos: Comprensión de sus capacidades y beneficios*

- potenciales para las organizaciones de atención de la salud]. *Technological Forecasting and Social Change*, 126, 3-13.
- World Health Organization & World Health Organization. (2015). *International Statistical Classification of Diseases and Related Health Problems [Clasificación Estadística Internacional de Enfermedades y Problemas de Salud Relacionados]*. World Health Organization.
- Wu, H., Xu, Z., Ren, P., & Liao, H. (2018). *Hesitant Fuzzy Linguistic Projection Model to Multi-Criteria Decision Making for Hospital Decision Support Systems [Modelo de proyección lingüística difusa para la toma de decisiones multicriterio para los sistemas de apoyo a las decisiones de los hospitales]*. *Computers & Industrial Engineering*, 115, 449-458.
- Xia, T., Jin, X., Xi, L., Zhang, Y., & Ni, J. (2013). *Operating load based real-time rolling grey forecasting for machine health prognosis in dynamic maintenance schedule [La carga operativa basada en la predicción en tiempo real de los grises rodantes para el pronóstico de la salud de la máquina en el programa de mantenimiento dinámico]*. *Journal of Intelligent Manufacturing*, 26(2), 269–280.
- Xu, G., Gao, S., Daneshmand, M., Wang, C., & Liu, Y. (2017). *A Survey for Mobility Big Data Analytics for Geolocation Prediction [Un estudio para la movilidad Gran análisis de datos para la predicción de la geolocalización]*. *IEEE Wireless Communications*, 24(1), 111–119.
- Yamunadevi, K., & Nagaraj, R. (2018). *An Optimized Classification of Human Cancer Disease for Gene Expression Data [Una clasificación optimizada de la enfermedad del cáncer humano para los datos de la expresión genética]*. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(2), 8-15.
- Yang, J. K., & Lee, D. H. (2018). *Optimization of Mean and Standard Deviation of Multiple Responses Using Patient Rule Induction Method [Optimización de la media y la desviación estándar de las respuestas múltiples usando el método de inducción de la regla del paciente]*. *International Journal of Data Warehousing and Mining (IJDWM)*, 14(1), 60-74.
- Yang, S., Santillana, M., Brownstein, J. S., Gray, J., Richardson, S., & Kou, S. C. (2017). *Using electronic health records and Internet search information for accurate influenza forecasting [Utilización de los registros sanitarios*



*electrónicos y la información de búsqueda en Internet para el pronóstico preciso de la gripe]. BMC Infectious Diseases, 17(1).*

Zhang, Yanqiu & Li, Xinxu & Li, Weibin & Jiang, Jianguo & Zhang, Guolong & Zhuang, Yan & Xu, Jiying & Shi, Jie & Sun, Dingyong. (2020). *Tuberculosis registration rates analysis and prediction based on exponential smoothing model: could tuberculosis control milestones be achieved in 2020, 2025 in Henan province, China? [Análisis y predicción de las tasas de registro de la tuberculosis basados en un modelo de suavizado exponencial: ¿podría alcanzarse un hito en el control de la tuberculosis en 2020, 2025 en la provincia de Henan (China)?].*

Zor, C., & Çebi, F. (2018). *Demand prediction in health sector using fuzzy grey forecasting [Predicción de la demanda en el sector de la salud mediante el pronóstico gris difuso]. Journal of Enterprise Information Management.*

Zulkurnain, N. F., & Shah, A. (2017). *HYBRID: An Efficient Unifying Process to Mine Frequent Itemsets. [HYBRID: Un eficiente proceso de unificación para la extracción de artículos frecuentes] 2017 IEEE 3rd International Conference on Engineering Technologies and Social Sciences (ICETSS), 1-5*

## Anexo 1: Matriz de operacionalización variable

Tabla 27

*Matriz de operacionalización variable*

Variable	Definición conceptual	Definición operativa	Dimensiones	Indicadores	Escala de medida
Impacto o efecto del uso de la previsión	Las previsiones de incidencia específicas de la edad, junto con las previsiones demográficas, pueden (Ku and Dodd, 2019, p.1)	Utilizamos datos secundarios de la Encuesta Nacional de Tendencias de Información sanitaria (HINTS) de ciudadanos estadounidenses (Ford, Hesse and & Huerta, 2016, p.1)	Precisión del pronóstico (Shahzadeh, Khosravi, & Nahavandi, 2015, p. 6)	Incremento de la precisión (Wang, Lee, Huang, Szabados, Wang, & Van Olinda, 2016, p. 5).	intervalo
				Reducción de error (Yang, Santillana, Brownstein, Gray, Richardson, & Kou, 2017, p. 4)	intervalo

## Anexo 2: Matriz de consistencia

Tabla 28

### Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE	DIMENSIONES	INDICATORS
General	General	General			
<p>No se han encontrado estudios que hagan que el pronóstico de la atención de urgencia en los hospitales aplique técnicas predictivas junto con la metodología CRISP-DM con la que se puede desarrollar un sistema de geolocalización para medir el impacto de la atención hospitalaria. En consecuencia, no hay resultados que muestren cambios en el aumento o disminución del número de atención urgentes en los hospitales.</p>	<p>Determinar a través de un nuevo modelo predictivo con la ayuda de la minería de datos y la técnica de predicción de minería de datos de regresión lineal, a través de un sistema de geolocalización para medir los impactos y cambios de atención de urgencia en los hospitales.</p>	<p>La implementación de metodología aplicando combinaciones de técnicas de minería de datos para la previsión de la atención de urgencia en los hospitales.</p>	-	-	-
ESPECÍFICO	ESPECÍFICO	ESPECÍFICO			INDICATORS
	<p>Determinar un mejor cálculo del pronóstico donde podamos identificar el impacto, la precisión de la atención urgente en los hospitales</p>	<p>La implementación de la minería de datos con un sistema de geolocalización nos ayudará a predecir el aumento o la disminución de la atención de urgencia en los hospitales.</p>	<p>Impacto o efecto del uso del sistema de previsión (Xu, G., Gao, S., Daneshmand, M., Wang, C., &amp; Liu, Y. (2017, p.9)</p>	<p>Precisión del pronóstico (Shahzadeh, Khosravi, &amp; Nahavandi, 2015, p. 6).</p>	<p>Incremento de la precisión (Wang, Lee, Huang, Szabados, Wang, &amp; Van Olinda, 2016, p. 5).</p>
	<p>Determinar demográficamente qué hospital recibe la atención más urgente renovando los datos de entrada anualmente para identificar con precisión los que reciben más atención que otros.</p>	<p>El desarrollo de la fusión de técnicas de minería de datos garantizará mejores resultados de previsión con los que generaremos una precisión del 0,90%</p>			<p>Reducción de error (Yang, Santillana, Brownstein, Gray, Richardson, &amp; Kou, 2017, p. 4)</p>

### **Anexo 3: Antecedentes adicionales a la investigación**

Según Prerana, Sehgal y Taneja (2015) indicaron: "En las bases de datos de los sistemas de información de los hospitales se almacena una gran cantidad de datos históricos" (Prerana, 2015, p. 76). Además, "Este sistema de información puede utilizarse para la toma de decisiones; pero la mayoría de los sistemas están diseñados para apoyar el inventario, la facturación a los pacientes y la generación de estadísticas sencillas" (Prerana, 2015, p. 76). Concluyeron que, "los métodos existentes para la previsión de la geolocalización basados en grandes datos móviles, incluida la región de geolocalización minera popular y las trayectorias mineras personales" (Prerana, 2015, p. 76).

Xu, Gao, Daneshmand, Wang y Liu (2017) explicaron: "Los métodos existentes para la predicción de la geolocalización (GP) basados en los grandes datos móviles (MBD), incluida la región de geolocalización popular minera y las trayectorias personales mineras". (p. 114). También Xu et al. (2017) propusieron: Aunque esos métodos han resuelto ciertos problemas, todavía hay cuestiones de investigación pendientes que requieren un estudio más a fondo, la precisión de la predicción de la geolocalización disminuirá cuando los algoritmos existentes predigan la geolocalización basada en regiones de geolocalización discretas en la situación de que falten algunos datos intermedios. (p. 113).

Además del desarrollo del software, se realiza un estudio de minería de datos para ello "se utiliza el modelo OLAP, que consiste en filtrar la información más relevante y necesaria para obtener los mejores resultados. Para estudiar las variables de información de la base de datos de emergencias y así predecir la información futura" (Xu et al., 2017, p. 113). Por lo tanto, Xu et al. (2017): indicaron que "equilibrar la relación entre la complejidad de los algoritmos y la precisión de los algoritmos de predicción es un gran desafío" (p. 113), los resultados de estos estudios y análisis permiten al cuerpo médico tomar decisiones oportunas y eficaces.

A través de este estudio realizado con la minería de datos y la adaptación del sistema de atención de urgencia. Para Xu et al. (2017) se preguntaron "cómo

determinar el alcance de la región y cómo mejorar la precisión de la minería de las regiones de geolocalización es otro problema que debe ser resuelto" (p. 114). Como conclusión permite al personal médico estar informado sobre las características de los pacientes que atiende, lo que le da una visión más clara de cómo tratar a cada tipo de paciente y saber qué personal se necesita en el servicio de urgencias para satisfacer la demanda de los usuarios de los Hospitales.

Como proponen Li, Yates, Lovely y Larson (2015): Un enfoque de grandes datos crea un entorno transparente e interactivo que permite a un proveedor formular un plan más específico para un paciente determinado utilizando la evidencia del mundo real que se encuentra en el EMR institucional de la INAN. Mediante el uso de una interfaz de usuario muy flexible, el proveedor o el equipo también pueden explorar escenarios "qué pasaría si" que anteriormente habrían requerido un equipo estadístico/base de datos y un tiempo considerable para su desarrollo (p. 56).

Farooqui y Mehra (2018) propusieron el diseño de un almacén de datos para un sistema de información médica, que se basa en las nuevas metodologías. Por lo tanto, para construir un modelo que implemente el almacén de datos clínicos. Este modelo comprende las técnicas de minería de datos que se promueven en la construcción de un almacén de datos clínicos. (p. 202). Por lo tanto, Farooqui y Mehra (2018) establecieron un "modelo que gestiona los datos clínicos y ayuda al analista de datos y a los gestores clínicos a realizar la minería de datos y el análisis de datos sobre los datos que se almacenan en el almacén" (p. 202). Además, Xu et al. (2017) afirman que "la mayoría de los métodos existentes están diseñados para un escenario específico, y es difícil encontrar un método que pueda funcionar para una amplia gama de escenarios" (p. 115).

Disponer de un marco general sobre los servicios hospitalarios basado en la extracción de datos para un análisis de los datos extraídos de las actividades clínicas en el hospital, mostrando a través de una descripción cronológica de las actividades del hospital: el comportamiento periódico de la cantidad de atención de emergencia. Puede considerarse como un "ciclo de vida" de un hospital, un

patrón de seguimiento a largo plazo de los pacientes con respecto al número de atención se extrajo mediante el registro de los sistemas de atención (Li et al., 2015). Sin embargo, Li et al. (2015): "El número de pacientes "similares" en un determinado depósito puede ser limitado, lo que reduce la solidez de este paradigma basado en el recuerdo. Construyendo este sistema sobre una arquitectura basada en la nube, puede resolver este problema simplemente incluyendo más pacientes de otras instituciones" (p. 56).

Rasitha Banu (2016) explicó: "La minería de datos está desempeñando un papel importante en la predicción de muchas enfermedades. La clasificación es una de las técnicas más significativas en la minería de datos" (Rasitha Banu, 2016, p. 4). Por lo tanto, "Es un aprendizaje supervisado. Se utiliza para clasificar conjuntos de datos predefinidos. Los datos de atención médica están teniendo un crecimiento exponencial en volumen y complejidad" (Rasitha Banu, 2016, p. 4). Concluye que "La técnica de minería de datos se utiliza principalmente en las organizaciones de salud para la toma de decisiones, el diagnóstico de enfermedades y para dar un mejor tratamiento a los pacientes" (Rasitha Banu, 2016, p. 4).

La gran arquitectura de datos propuesta para aplicar la minería de datos en colaboración parece prometedora y puede ser muy útil para extraer conjuntos de datos específicos de estudios más pequeños dentro de un gran almacén de datos clínicos. El proceso de extracción de conocimientos a partir de los datos existentes y se utiliza como instrumento bancario para encontrar información útil en datos históricos y operacionales que puedan ayudar en la adopción de decisiones Xu et al. (2017) concluyeron:

"Cuando un modelo de predicción de geolocalización se basa en los puntos de geolocalización continuos, el alcance de los algoritmos típicos es demasiado pequeño y sólo puede aplicarse en la trayectoria local" (p. 452), además, los autores indicaron que "debido a los efectos de las diferentes velocidades y direcciones de movimiento, la mayoría de los métodos existentes están diseñados para un escenario específico, y es difícil encontrar un método que pueda funcionar para una amplia gama de escenarios" (Xu et al., 2017, p. 452).

Además, Roy (2016), "el poder de las percepciones que proporcionan los grandes datos creados por el resultado oportuno y significativo de los grandes conjuntos de datos suele ayudar en la toma de decisiones empresariales eficaces que maximizan el impacto comercial" (p. 3). Complementando con Ionita y Ionita (2016) dijeron que: "la industria de la salud genera enormes cantidades de datos complejos sobre los pacientes, los recursos de los hospitales, el diagnóstico y el tratamiento de enfermedades, los registros electrónicos de los pacientes, los dispositivos médicos y así sucesivamente" (p. 393). Además, Ionita y Ionita (2016) explicaron que: "Todos estos datos almacenados en bases de datos, comúnmente llamados almacenes de datos, representan la materia prima que debe ser procesada y analizada para extraer el conocimiento que permita apoyar el ahorro de costos y el proceso de toma de decisiones" (p. 393).

Domadiya y Rao (2018) Señalaban: "mejorar la precisión de la investigación médica mediante la colaboración de múltiples sistemas de EHR con datos divididos horizontalmente. Se discutieron cuestiones de privacidad en el caso de la colaboración de múltiples sistemas EHR"(p. 8). Así mismo, determinaron Domadiya y Rao (2018) concluyeron: "Para resolver esto, propusimos un algoritmo eficiente y seguro para el PPDARM sobre datos particionados horizontalmente. El análisis experimental del algoritmo propuesto con datos de enfermedades cardíacas demostró que la precisión de la predicción de las enfermedades cardíacas ha mejorado y se preserva la privacidad" (p. 8).

Yamunadevi y Nagaraj (2018) mencionaron: "La clasificación del cáncer basada en datos de expresión genética es un área de investigación alentadora en el campo de la minería de datos. El algoritmo sugerido tendía a que el tema del diagnóstico precoz del cáncer fuera uno de los problemas de salud más genuinos del mundo" (p. 8). Además, Yamunadevi y Nagaraj (2018) definieron: "se presenta una nueva metodología para clasificar las enfermedades de cánceres humanos basadas en los perfiles de expresión genética. En este esquema propuesto, inicialmente el IG se utiliza para la selección de características y el CS para la reducción de características" (p. 8). Por lo tanto, Yamunadevi y Nagaraj (2018) indicaron que "el enfoque CS se utiliza para optimizar el valor umbral de los rasgos determinados por la IG. El resultado

experimental muestra que el enfoque IG/CS propuesto supera al enfoque existente IG/GP en cuanto a la precisión de la clasificación, la especificidad y la sensibilidad" (p. 8).

Se utilizaron técnicas de aprendizaje automático para predecir la enfermedad del hígado graso, y la regresión logística y el resultado de la predicción tiene el potencial de ayudar a los clínicos a tomar decisiones más precisas y significativas (Islam, Wu, Poly, Yang y Yu-Chuan, 2018). Islam, Wu, Poly, Yang y Yu-Chuan (2018) Concluyeron: "El resultado de esta predicción tiene el potencial de ayudar a los clínicos a tomar decisiones más precisas y significativas sobre el diagnóstico y el tratamiento de la enfermedad del hígado graso." (p.5).

Así pues, el enfoque de clasificación precisa para identificar y examinar los factores de riesgo en las personas que consumen opiáceos naturales recomienda que los investigadores y los médicos aprovechen este modelo Amirabadizadeh et al. (2018) explicaron: "Han desarrollado en este estudio un enfoque de clasificación preciso para identificar y examinar los factores de riesgo en los individuos que consumen opiáceos naturales y narcóticos basados en la heroína" (p. 10). Además, afirmaron en su estudio "Dada su interpretación simple e inteligible y su gran precisión en comparación con los métodos ordinarios, recomendamos que los investigadores y los médicos aprovechen este modelo" (Amirabadizadeh et al., 2018, p. 10).

Por otro lado, el menor tamaño de la ventana llevará a errores desencadenados por el evento que tuvo lugar en las horas previas. Se eligieron ventanas de cinco horas debido a la gran necesidad de equilibrio entre el ciclo de actividad diaria y los eventos masivos. Según Mukhina, Visheratin, Mbogo y Nasonov (2018) explicaron:

Debe tenerse en cuenta que durante este trabajo asumimos que cinco horas previas son suficientes para la predicción del siguiente estado. Pero el tamaño de una ventana retrospectiva debería estudiarse más, ya que el uso de la ventana de cinco horas puede conducir a un ruido causado por los patrones diarios. Por otro lado, el tamaño más corto de la ventana



llevará a errores provocados por eventos que tuvieron lugar en horas anteriores. Se eligieron las ventanas de cinco horas debido a la gran necesidad de equilibrio entre el ciclo de actividad diaria y los eventos masivos. Por lo tanto, la ventana elegida parece óptima para disminuir la influencia del ruido causado por diferentes razones (p.6).

Mediktor en este segmento de pacientes muestra que un mayor nivel de gravedad corresponde a un mayor número de ingresos hospitalarios, readmisiones hospitalarias y regresos a días de urgencia. Según Arancibia, Sánchez, Cháfer, Briñon y Aguilar (2019) indicaron:

Los niveles de Mediktor, en este grupo de pacientes urgentes de baja complejidad, no coinciden con los asignados por el sistema de Manchester, lo que podría indicar un modelo de clasificación diferente. La clasificación de Mediktor en este segmento de pacientes muestra que un mayor nivel de gravedad corresponde a un mayor número de ingresos hospitalarios, readmisiones hospitalarias y visitas a urgencias a los 30 días, aunque sin significación estadística (p.6).

El usuario desempeña un papel importante en la creación de la cadena de suministro para el tratamiento de la atención de urgencia. El despliegue de la analítica de datos tiene por objeto apoyar este sistema poniendo a disposición de los pacientes análisis en tiempo real. Según Mustafee, Powell y Harper (2018) indicaron:

En última instancia, la decisión de acudir a un servicio de atención de urgencias en particular recae en el paciente, y factores como las percepciones de su condición clínica, la tolerancia al tiempo de espera, la distancia de viaje implicada, las experiencias pasadas y las percepciones de la atención disponible contribuirán a la decisión final (esto incluye la predicción de los estados futuros y el análisis de los escenarios hipotéticos), permitiendo a los pacientes tomar una decisión informada sobre las opciones de tratamiento y cambiar potencialmente el comportamiento de asistencia al SUH (p.9).

Una inducción formal y estandarizada a 111 servicios, que un defensor de este innovador programa me dentro de la atención de urgencias y emergencias sería una gran ventaja para los farmacéuticos y la integración departamental (Nichols, Inglaterra, Holliday y Newton, 2019, p. 9). Además, “Se han destacado los beneficios indicados de la integración de los farmacéuticos clínicos de atención de urgencia en los servicios de atención de urgencia, y para progresar en estos beneficios” (Nichols, Inglaterra, Holliday y Newton, 2019, p. 9), se respaldan lo siguiente: “Se introduzca una inducción formal y estandarizada a los servicios 111, el desarrollo de procedimientos operativos estándar y directrices para todos los farmacéuticos dentro de los servicios 111” (Nichols, Inglaterra, Holliday y Newton, 2019, p. 9).

También es imperativo que cada Trust desarrolle una conciencia de múltiples niveles sobre el papel del farmacéutico para prevenir la falta de comunicación, y finalmente que un defensor de este innovador programa me dentro de la atención de emergencia y urgencia sería una gran ventaja para los farmacéuticos y la integración de los departamentos (Nichols, Inglaterra, Holliday y Newton, 2019, p. 9).

## **Anexo 4: Teorías relacionadas adicionales a la investigación**

### **Metodología CATALYST**

Durante el proceso de modelación y explotación minera, cualquier modelador y minero tiene percepciones, que son parte del propósito de la modelación y la explotación minera después de todo. Sea lo que sea y como sea que se manifiesten, cuando significa cambios en la forma en que se preparan, modelan o presentan los datos, significa, tal vez, preparar los datos, crear un nuevo modelo, o de otra manera dejar una etapa e ir a otra parte de la metodología.

Según Pyle (2003) propuso: "La metodología no es un documento secuencial. Cualquier intento de leerlo secuencialmente va a resultar en una confusión total. No hay un solo camino a través de las cajas". (p. 379) Además de eso "La esencia de estas metodologías es que un modelador o minero tome acciones, evalúe los resultados y decida, basado en el resultado, qué recuadro leer a continuación". (Pyle, 2003, p. 379). Esta metodología tiene las siguientes fases: Modelización de negocios, Preparación de datos, Selección de herramientas y modelización inicial, Refinamiento del modelo, Comunicación (Pyle, 2003, pp. 376-367)

### **Metodología SEMMA**

SEMMA y el resto de los dos métodos que son CRISP-DM y KDD es el requisito y la comprensión de los requisitos previos de la comprensión empresarial y las bases de datos. Además, incluso después de la fase de aplicación del modelo, ambas metodologías KDD y CRISP-DM requieren la estimación del modelo para su evaluación del rendimiento, lo que tampoco se requiere en SEMMA.

Según Tariq et al. (2019) añadieron: "Sobre la base de estos parámetros, se seleccionó, explicó y sugirió el mejor método por sus características significativas en cuanto a la predicción de los impagos de préstamos en el sector financiero" (p. 3503). Para Tariq et al. (2019) propuesto:

En la fase de desarrollo del modelo, se elaboraron los cinco pasos clave de SEMMA antes del despliegue del modelo. En la fase de experimentación se emplearon tres técnicas diferentes de minería de

datos para el desarrollo del modelo propuesto y se evaluaron sus resultados en diversos parámetros (p. 3503).

Además, agregaron, “Esta metodología tiene las siguientes fases: Muestrear, Explorar, Modificar, Modelar, Evaluar” (Tariq et al., 2019, pp.3492-3500)

### **Metodología KEM-DT**

En el KEM-DT, cinco pasos clave se refieren a la construcción de una base de conocimientos integrada basada en reglas, después de extraer las reglas de producción de un conjunto de datos, como se muestra. Según Ali et al. (2018) indicaron:

El objetivo de todo sistema basado en el conocimiento es resolver los problemas del mundo real y ayudar en la toma de decisiones; por lo tanto, el conocimiento creíble y enorme puede desempeñar un papel importante en el ámbito de la atención de la salud (p. 5).

Además, complementaron "para lograr este objetivo, proponemos una metodología de ingeniería del conocimiento (KEM-DT) que construye una base de conocimiento integrada basada en reglas" (Ali et al., 2018, p. 5).

Esta metodología tiene las siguientes fases: “Aprendizaje del modelo, pre procesamiento del modelo, transformación del modelo, análisis sintáctico del modelo, verificación y validación de las reglas” (Ali et al., 2018, p. 5). Por lo tanto, al desarrollar esta metodología está basado en arboles de decisiones, según Ali et al. expusieron: “Para construir una base de conocimiento creíble, este estudio extrae las reglas de producción del árbol de decisiones múltiples. Para realizar la metodología KEM-DT, se desarrolla una herramienta de adquisición de conocimiento basada en datos (DDKAT)” (2018, p. 5).

### **Metodología W2KM**

Es una metodología capaz de ofrecer una guía a lo largo del proceso de desarrollo e implementación de un KMS usando herramientas de Web y Big Data". Según Orenga-Roglá y Chalmeta (2017) explicado: "La metodología de

gestión del conocimiento de la Web 2.0 (W2KM) consiste en fases que se pueden descomponer en actividades, que a su vez se componen de tareas" (p. 6). Además, Orenga-Roglá y Chalmeta (2017) complementaron:

El W2KM utiliza las fases tradicionales de un proyecto de sistema de información. Puede ayudar a recopilar, generar, gestionar y aplicar el conocimiento generado tanto dentro de la organización como de las relaciones externas de la misma, para luego transferirlo a las personas adecuadas de manera fácil y rápida (p. 6).

Esta metodología tiene las siguientes fases: "Borrador, Planificación, Análisis, Diseño, Desarrollo, Implementación, Control" (Orenga-Roglá y Chalmeta, 2017, pp. 6-10). Por lo tanto, Orenga-Roglá y Chalmeta (2017) ejemplificaron que "Una empresa puede decidir seguir un modelo iterativo, y llevar a cabo una serie de mini-cascadas desde la fase de análisis" (p. 6). Por lo que en "todas las fases de la cascada se completan para un bloque de conocimiento, ya que dentro del desarrollo de un bloque de conocimiento cada fase depende de los resultados obtenidos en las fases anteriores (Orenga-Roglá y Chalmeta, 2017, p. 6).

### **Redes neuronales**

En cuanto al rendimiento del modelo y las predicciones resultantes, la selección de los elementos de procesamiento se redujo cuidadosamente. Este modelo de RNA genera una predicción muy precisa para el rango de 5 a 40 neuronas. Este modelo de RNA consiste en una capa oculta y un número variable de elementos de procesamiento o neuronas.

Es el modelo más óptimo obtenido mediante el método de ensayo y error. Según Mishra, Soni, Sharma y Upadhyay (2018) explicaron: "El modelo de Red Neural Artificial (RNA) se desarrolla usando MATLAB. En este modelo se ha utilizado Feed Forward Neural Network (FFNN) usando el algoritmo Back Propagation y la función de entrenamiento de Levenberg- Marquardt" (p. 15). Además, Mishra et al. (2018) comentaron: "Las pruebas se realizan primero seleccionando al azar el número de elementos de procesamiento al final del

entrenamiento, sobre la base de un error cuadrado medio, la precisión será predicha" (p. 15).

### **Arboles de decisiones**

Es un modelo de predicción utilizado para representar visualmente las diferentes posibilidades un escenario, como aspecto de la vida cotidiana, decisiones difíciles en el entorno familiar, hasta aplicaciones tecnológicas de los negocios y en la inteligencia artificial. Según Nahar y Ara (2018) emplearon:

Algoritmo de árbol de decisión como J48, LMT, Random Forest, Random Tree, REPTree, Decisión Stump y Hoeffding Tree para predecir la enfermedad hepática en una etapa más temprana. Este algoritmo da varios resultados basados en la Precisión, el Error Medio Absoluto, la Precisión, el Recall, las estadísticas Kappa y el Tiempo de Ejecución (p. 6).

### **Agrupación**

Consiste en la organización de los datos dentro de clases. Según Manikandan, Porkodi, Mohammed y Sivaram (2018) indicaron: "La fase inicial de la minería de datos es agrupar los datos como se indica en alguna norma (p. 13).

Sin embargo, los cálculos de investigación de los conglomerados pueden delegarse ampliamente los cálculos progresivos, es la diferencia en esta técnica, que se desconocen las etiquetas de las clases, por lo que se centra en la agrupación difusa de medios que es un cálculo de partición encargándose de descubrir clases aceptables. Según Manikandan et al.,2018 explicaron:

La investigación de los conglomerados es una tarea general que debe resolverse y no un cálculo particular. Los cálculos de la investigación de conglomerados pueden delegarse ampliamente en los cálculos progresivos y de partición. Los cálculos de partición agrupan los datos según lo indicado por alguna norma y los evalúan. El trabajo se centra en la agrupación de c-means difusas, que es un cálculo de partición (p. 14).

## **Regla de inducción**

Si el investigador trata de encontrar el ajuste óptimo para las variables de entrada que maximice o minimice la respuesta, la hipótesis de la varianza constante podría no ser válida en la práctica. En tales casos, no sólo la respuesta media, sino también la desviación estándar de la respuesta. Según Yang y Lee (2018) explicaron:

La metodología de la superficie de respuesta (RSM) consiste en un grupo de técnicas utilizadas en el estudio empírico entre una respuesta y una serie de variables de entrada. El investigador intenta encontrar el ajuste óptimo para las variables de entrada que maximice o minimice la respuesta (p. 15).

En esos casos, no sólo la respuesta media, sino también las desviaciones estándar de la respuesta deben considerarse al determinar las condiciones óptimas de las variables de entrada. Por lo tanto, Yang y Lee (2018) concluyen: “La RSM supone que la varianza de la respuesta es constante y se centra en la optimización de la media de la respuesta. Sin embargo, la suposición de la varianza constante podría no ser válida en la práctica” (p. 15).

## **RAPIDMINER**

RapidMiner está compuesto dentro de la programación de Java. Se utiliza para la minería de contenidos, la minería de marcos visuales y sonoros, incluida la construcción, la minería de flujo de datos y el seguimiento del pensamiento flotante, el progreso de los procedimientos de la empresa y la minería de datos transmitidos. Según Massaro, Maritati y Galiano, (2018) definieron:

RapidMiner, una vez en el pasado YALE (Yet Another Learning Environment), es una situación para dar minería de datos y sistemas de aprendizaje automático que incluye: apilamiento y cambio de datos (ETL), pre procesamiento de datos e imagen visual, visualización, evaluación y disposición (p. 9).

RapidMiner se utilizará para la minería de contenidos, la minería de marcos visuales y sonoros, incluye la construcción, la minería de flujos de datos

y el seguimiento de pensamientos flotantes, el avance de los procedimientos de la compañía y la minería de datos transmitidos. Por lo tanto, la herramienta para Massaro et al. (2018) indicaron que: “los procedimientos de minería de datos estarán formados por administradores de mesa de casa arbitrarios, hablados en documentos XML y hechos en la interfaz gráfica de cliente (GUI) de RapidMiner” (p. 9).

## **WEKA**

Contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo complementado con una interfaz gráfica de usuario de fácil entendimiento a sus funcionalidades para modelar algoritmos.

Según Akter, Hossin, Daiyan y Hossain (2018) explicaron: “Se ha utilizado una herramienta de minería de datos abierta a todos, WEKA (versión 3.8.0). Se han utilizado dos conjuntos de datos diferentes y se ha examinado el rendimiento de los algoritmos de clasificación (clasificadores)” (p. 6). Por lo tanto, en su trabajo de Akter et al. (2018) indicaron:

El análisis ha sido realizado por SONY VIAO Windows versión 8 con Intel® Core™ i3 Unidad Central de Procesamiento, Procesador de 1.70 Gigahertz y 4 Gigabytes de RAM. Los conjuntos de datos se han seleccionado de manera que varían en tamaño, predominantemente con el número de atributos (p. 6).

## **ORANGE**

Es un programa para realizar minería de datos y análisis predictivo. se caracteriza en la media y la escala por la desviación estándar Además de operaciones de pre procesamiento y representación gráfica de datos. Las áreas de los picos cromatográficos correspondientes a los compuestos químicos seleccionados se utilizaron como datos de entrada para el análisis de conglomerados jerárquicos.

Rózańska, Dymerski y Namieśnik (2018) mencionados: "El análisis quimiométrico se realizó utilizando el programa informático Orange Canvas Data Mining v. 3.3.9 (Laboratorio de Bioinformática, Universidad de Ljubljana,



Eslovenia)". (p. 6) También se mencionó Róžańska et al. (2018): "Para normalizar las características, se centraron por la media y se escalaron por la desviación estándar. Luego se seleccionaron diez compuestos químicos basados en el análisis unidireccional de la varianza (ANOVA)". (p. 6).

### **SAS (Stands for Statistical Analysis System),**

Es una herramienta de Business Intelligence para ayudar a los clientes de todo tipo de industrias, desde las empresas farmacéuticas y los bancos hasta las entidades académicas y gubernamentales. Señalando que SAS - tanto el software como la compañía - prosperó a lo largo de las próximas décadas, el paquete más despojado incluirá el siguiente SAS de base, SAS/STAT, SAS Graph y SAS acceso a ODBC

Bansal y Srivastava (2018) definieron: "SAS" significa Sistema de Análisis Estadístico. Los creadores del SAS son Jim Goodnight y Jim Barr. Comenzó en la Universidad Estatal de Carolina del Norte como un proyecto para analizar la investigación agrícola" (p. 12). Además, para Bansal y Srivastava (2018) concluyeron que: "La demanda de estas capacidades de software comenzó a crecer, y SAS fue fundada en 1976 para ayudar a los clientes de todo tipo de industrias, desde compañías farmacéuticas y bancos hasta entidades académicas y gubernamentales". (p. 12).

### **SISENSE**

Sisense soporta datos de varias fuentes, lo que permite estructurar e importar fuentes de datos, así como relacionarlas y prepararlas para su análisis y visualización mediante la metodología ETL. Según Lousa et al. (2019) indicaron:

Las fuentes de datos soportadas pueden ser archivos o servidores de datos ubicados en servidores locales o servicios web como Salesforce y Google AdWords. Tiene conectores de datos nativos, ODBC y REST (Representational State Transfer). La metodología ETL (Extracción, Transformación y Carga). Las fuentes de datos soportadas pueden ser archivos o servidores de datos (SQL Server, MySQL, Snowflake, etc.) ubicados en servidores locales o servicios web. (p. 3)

## **Microsoft Power BI**

Consiste en la aplicación Power BI Desktop que permite la transformación de datos para la visualización de los mismos. Según Lousa, et al. (2019) indicaron: "Power BI es un conjunto de herramientas, conectores (personalizados y certificados) y servicios que permiten la visualización y el análisis de datos importados de muchas fuentes desde archivos, servidores de datos, así como servicios en línea y Azure" (p. 3).

## **QLIKVIEW**

Es una herramienta de Business Intelligence que permite recolectar datos desde diferentes orígenes, bases de datos SQL, datos de Excel, etc., para modelarlos a nuestra necesidad para facilitar su manejo y presentación visual.

Según Lousa, et al. (2019) indicaron: "En el Front End es donde los usuarios pueden interactuar y ver los documentos y datos creados por la aplicación QV Developer que se conecta a diferentes fuentes de datos (Data Warehouses, MS Excel, SAP, Salesforce, Oracle, etc.)" (p. 3)

## **Información geográfica**

Un Sistema de Información Geográfica (SIG) es un conjunto de componentes específicos permitiendo a los usuarios crear consultas, analizar y representar cualquier información geográfica referenciada de un territorio de forma eficaz. Según Teixeira (2018) explico: "El GIS cualitativo es prometedor para los trabajadores sociales porque tiene el potencial de permitirnos cuestionar las explicaciones existentes de los problemas sociales" (p. 45).

Además, Teixeira (2018) agrego: "examinar las cuestiones desde múltiples puntos de vista, comprometiéndonos con múltiples significados y realidades que comprenden los fenómenos de nivel macro". Por lo tanto, Teixeira (2018) concluyo: "Aunque existen barreras para la adopción de un GIS cualitativo más ampliamente dentro de la investigación del trabajo social, es un área rica para la exploración futura." (p. 45).

## **Clasificación internacional de enfermedades (CIE-10)**

Entre ellos figuran el análisis de la situación sanitaria general de grupos de población y la vigilancia de la incidencia y prevalencia de enfermedades y otros problemas de salud en relación con otras variables, como las características y circunstancias de las personas afectadas.

Según Organización Mundial de la Salud, 2016 definieron: “El CIE se ha convertido en la clasificación diagnóstica estándar internacional para todos los fines epidemiológicos generales y muchos de los de gestión de la salud” (p. 3).

Además, el CIE es un instrumento de la organización de enfermedades o problemas, “no tiene por objeto ni es adecuado para la indización de entidades clínicas distintas. También existen algunas limitaciones en el uso de la CIE para estudios de aspectos financieros, como la facturación o la asignación de recursos” (Organización Mundial de la Salud, 2016, p. 3)

### **Pronostico de atenciones de urgencia**

Se desconoce la pauta de ingresos hospitalarios resultantes de las visitas al servicio de urgencias de las personas que se encuentran en un hospicio y de las que no lo están. Como resultado, el efecto de los hospicios en el uso de los servicios de urgencias y, más ampliamente, la relación entre los cuidados de emergencia y los hospicios siguen siendo poco conocidos. Según Obermeyer, Clarke, Makar, Schuur y Cutler (2016) explicaron:

El 51% de los adultos mayores visitan el departamento de emergencias (ED) en el último mes de vida. Aunque la inscripción en el hospicio de los ancianos beneficiarios de Medicare se asocia con menos visitas al ED en el último mes de vida (p. 325).

### **Geolocalización**

La geolocalización es un término genérico que denota las constelaciones de satélites que proporcionan señales desde el espacio que transmiten datos de posicionamiento y cronometraje que permiten a los usuarios determinar su posición tridimensional con cobertura mundial. Según Zhu (2018) indicó: “El Sistema de Posicionamiento Global (GPS) de los Estados Unidos ha estado en pleno funcionamiento desde 1995” (p. 14). Por ejemplo:

El sistema chino de navegación por satélite BeiDou (BDS) y el sistema europeo GALILEO están actualmente en desarrollo, aunque el BDS inició un servicio operacional inicial (Fase II) a finales de diciembre de 2011 y Galileo se ha declarado operacional en diciembre de 2016 para los servicios iniciales (Zhu, 2018, p. 14).

### Anexo 5: Tabla de puntuación de metodología

Tabla 29

*Tabla de puntuación de metodología*

	Funcionamiento		Fases		Modelado		Resultado
		15%		60%		35%	100%
<b>CATALYST</b>	1	0,3	1	0,5	2	0,4	1,2
<b>SEMMA</b>	1	0,3	3	1,5	2	0,4	2,2
<b>KEM-DT</b>	1	0,3	3	1,5	2	0,4	2,2
<b>CRISP-DM</b>	3	0,9	2	1	3	0,6	2,5
<b>W2KM</b>	2	0,6	2	1	3	0,6	2,2

En la tabla de puntuación se caracterizaron por 3 criterios (función, fases y modelado) que se distribuyó por medio de una puntuación de criterio basada en cada componente de las metodologías con las que se decidió elegir CRISP-DM como metodología de desarrollo.(Ver anexo 10)

### Anexo 6: Tabla de puntuación Herramientas de minería de datos

Tabla 30

*Tabla de puntuación Herramientas de minería de datos*

	Características		Lenguaje de programación		Sistema Operativo		Precio/Licencia		Resultado
		45%		30%		15%		10%	100%
<b>RapidMiner</b>	3	1,35	2	0,6	2	0,3	2	0,2	2,45
<b>WEKA</b>	3	1,35	2	0,6	1	0,15	1	0,1	2,2
<b>Orange</b>	1	0,45	3	0,9	2	0,3	2	0,2	1,85
<b>RStudio</b>	3	1,35	2	0,6	1	0,15	1	0,1	2,2
<b>SAS</b>	2	0,9	1	0,3	2	0,3	2	0,2	1,7
<b>Visual Studio (SISS)</b>	3	1,35	2	0,6	2	0,3	3	0,3	2,55

En el marcador se marcaron con 4 criterios (características, lenguaje de programación, sistema operativo y licencia) que se distribuyó a través de una puntuación de criterios basados en cada componente a juicio del uso de las herramientas y la fiabilidad de cada distribuidor con el que se decidió elegir RapidMiner para la aplicación de la minería de datos. (Ver anexo 11)

## Anexo 7: Puntuación de herramientas de visualización de datos

Tabla 31

*Tabla de puntuación de herramientas de visualización de datos*

	Nombre		Software		Estructura de visualización		Sistema operativo		Licencia		Escalabilidad		Extensibilidad		Resultado
	10%	25%	7%	3%	20%	25%	10%	100%							
<b>Sisense</b>	1	0,1	3	0,75	2	0,14	2	0,06	2	0,4	1	0,25	5	0,5	2,2
<b>QlikView</b>	3	0,3	3	0,75	2	0,28	2	0,6	3	0,015	2	0,08	3	0,3	2,325
<b>Tableau</b>	5	0,5	5	0,2	2	0,28	3	1	3	0,015	2	0,08	4	0,4	2,475
<b>Power BI</b>	2	0,2	4	0,16	3	0,42	3	1	3	0,016	1	0,25	4	0,4	2,446

En la tabla de puntuación se marcaron con 7 criterios (Software, Visualización, Estructura, Sistema Operativo, Licencia, Escalabilidad, Extensibilidad y Última Versión y Fecha) que se distribuyeron por medio de una puntuación de criterios basada en cada componente a juicio del uso de las herramientas y la fiabilidad de cada distribuidor con el que se decidió elegir TABLEAU para la demostración ya que su facilidad y comprensión era la mejor. (Ver anexo 12)

## Anexo 8: Clasificación de metodologías de minería de datos

Tabla 32

### Clasificación de metodologías de minería de datos

Metodología	Funcionamiento	Fases	Modelado
SEMMA	En la fase de desarrollo del modelo, se elaboraron todos los pasos clave de SEMMA antes de la implementación del modelo. Mientras que, en la fase de experimentación, se emplearon tres técnicas de minería de datos diferentes para el desarrollo del modelo propuesto y sus resultados se evaluaron en varios parámetros. Sobre la base de estos parámetros, se seleccionó, explicó y sugirió el mejor método debido a sus características significativas con respecto a la predicción de los impagos de préstamos en el sector financiero. (Tariq, Sohail, Aslam & Batcha, 2019)	<ul style="list-style-type: none"> <li>● Muestra</li> <li>● Explorar</li> <li>● Modificar</li> <li>● Modelo</li> <li>● Evaluación</li> </ul> <p>(Tariq, Sohail, Aslam &amp; Batcha, 2019)</p>	SEMMA y el resto de los dos métodos que son CRISP-DM y KDD es el requisito y la comprensión de los requisitos previos de la comprensión del negocio y las bases de datos. Además, incluso después de la fase de implementación del modelo, ambas metodologías KDD y CRISP-DM requieren la estimación del modelo para su evaluación del desempeño, que tampoco es necesaria en SEMMA (Tariq, Sohail, Aslam & Batcha, 2019)
CATALYST	La metodología no es un documento secuencial. Cualquier intento de leerlo secuencialmente va a dar lugar a una confusión total. No hay un solo camino a través de las cajas. La esencia de estas metodologías es que un modelador o minero tome medidas, evalúe los resultados y decida, en función del resultado, qué caja leer a continuación. (Pyle, 2003 ,p. 379)	<ul style="list-style-type: none"> <li>● Modelado de negocios</li> <li>● Preparación de datos</li> <li>● Selección de herramientas y modelado inicial</li> <li>● Refinamiento de modelos</li> <li>● Comunicación</li> </ul> <p>(Pyle, 2003 ,p. 376-367)</p>	Durante el proceso de modelado y minería, cualquier modelador y minero tiene información, que son parte del propósito para el modelado y la minería después de todo. Pero los conocimientos llegarán impredeciblemente, tal vez como destellos de inspiración o simplemente como un ligero indicio. cuando significa cambios en la forma en que los datos se preparan, modelan o presentan, significa, tal vez, preparar los datos, crear un nuevo modelo, o dejar una etapa e ir a otra parte de la metodología. (Pyle, 2003, p. 367)



Metodología	Funcionamiento	Fases	Modelado
KEM-DT	En KEM-DT, hay cinco pasos clave implicados en la construcción de una base de conocimiento integrada basada en reglas después de extraer reglas de producción de un conjunto de datos como se muestra (Ali, Lee and Kang, 2018).	<ul style="list-style-type: none"> <li>● Aprendizaje de modelos</li> <li>● Pre procesamiento de modelos</li> <li>● Transformación del modelo</li> <li>● Análisis de modelos</li> <li>● Verificación y validación de las reglas (Ali, Lee and Kang, 2018).</li> </ul>	Para lograr este objetivo, proponemos una metodología de ingeniería del conocimiento (KEM-DT) que construye una base de conocimiento basada en reglas integrada. Para construir una base de conocimiento creíble, este estudio extrae reglas de producción de varios árboles de decisión. Por último, para realizar la metodología KEM-DT, se desarrolla una Herramienta de Adquisición de Conocimientos Basados en Datos (DDKAT) (Ali, Lee and Kang, 2018).
.W2KM	La metodología de gestión del conocimiento Web 2.0 (W2KM) consta de fases que se pueden desglosar en actividades, que a su vez se componen de tareas. El W2KM utiliza las fases tradicionales de un proyecto de sistema de información. Puede ayudar a recopilar, generar, administrar y aplicar el conocimiento generado tanto dentro de la organización como desde las relaciones externas de la organización, y luego transferirlo a las personas adecuadas de forma fácil y rápida (Orenga and Chalmeta, 2019).	<ul style="list-style-type: none"> <li>● Borrador</li> <li>● Planificación</li> <li>● Análisis</li> <li>● Diseño</li> <li>● Desarrollo</li> <li>● Implementación</li> <li>● Control</li> </ul> (Orenga and Chalmeta, 2019).	La empresa puede decidir seguir un modelo iterativo y llevar a cabo una serie de mini-cascadas desde la fase de Análisis, entonces, todas las fases de la cascada se completan para un bloque de conocimiento, ya que dentro del desarrollo de un bloque de conocimiento en cada fase depende de los resultados obtenidos en fases anteriores (Orenga and Chalmeta, 2019)

Metodología	Funcionamiento	Fases	Modelado
CRISP-DM	<p>Un subproyecto se dedicó al desarrollo de modelado impulsado por datos y simulación de cadenas de proceso para un ajuste específico de características predefinidas, así como asegurar una fabricación reproducible de componentes compuestos textiles termoplásticos. El enfoque estaba en la producción de una torreta de suspensión de componentes compuestos termoplásticos de fibra de vidrio con un módulo electrónico integrado. En el proyecto, se utilizó el IoT-Dashboard Detact®, que proporciona las funciones necesarias desde la conexión de los orígenes de datos y el análisis de los datos hasta la visualización de los resultados (Huber, Wiemer, Schneider and Ihlenfeldt, (Huber, Wiemer, Schneider y Ihlenfeldt 2019).</p>	<ul style="list-style-type: none"> <li>● Entendimiento de negocios</li> <li>● Comprensión y preparación de datos</li> <li>● Modelado</li> <li>● Evaluación y Despliegue (Huber, Wiemer, Schneider y Ihlenfeldt, 2019).</li> </ul>	<p>Para el modelado, probamos diferentes algoritmos de clasificación en MATLAB Statistics and Machine Learning Toolbox. En el resultado, elegimos el algoritmo de clasificación KNN (k-Nearest-Neighbor) ya que podría manejar muy bien el pequeño tamaño del conjunto de datos de entrenamiento. Además, es capaz de agregar vectores de entrenamiento y clases adicionales durante el tiempo de ejecución. A continuación, llevamos a cabo una validación de cruce de Leave-one-out (LOOCV) en el conjunto de datos de entrenamiento. Los resultados oscilaron entre el 90 % y el 100 % de precisión de clasificación para los diferentes rangos de frecuencia (Huber, Wiemer, Schneider y Ihlenfeldt, 2019).</p>

## Anexo 9: Clasificación de herramientas de minería de datos

Tabla 33

*Clasificación de herramientas de minería de datos*

	Características	Lenguaje de programación	Operating system	Precio/Licencia
<b>RapidMiner</b>	Adecuado para todos los procesos. Sobresaliente en el análisis predictivo	Java	Windows, MacOS, Linux	Freeware, diferentes versiones de pago
<b>WEKA</b>	Muchos métodos de clasificación	Java	Windows, MacOS, Linux	Free Software (GPL)
<b>Orange</b>	Crea una atractiva visualización de datos sin requerir mucho conocimiento previo para hacerlo	Software core: C++, extension and input language: Python	Windows, MacOS, Linux	Free Software (GPL)
<b>RStudio</b>	Software de minería de datos de código abierto que ha democratizado el acceso al análisis predictivo	Java	Windows, MacOS, Linux	Free Software (GPL)
<b>SAS</b>	Caro pero poderoso para las grandes empresas	Language SAS	Windows, MacOS, Linux	El software gratuito se limita a las instituciones públicas, el precio se fija a petición, se dispone de diferentes modelos
<b>Visual Studio (SSIS)</b>	Microsoft proporciona una herramienta ETL patentada llamada SQL Server Integración Services (SSIS).	Visual Basic o Visual C#	Windows, MacOS, Linux	Desarrollador de la versión de software libre

*Nota: Adaptado de* "Software de data mining: las mejores herramientas." por IONOS España S.L.U., 2019.

## Anexo 10: Clasificación de herramientas de visualización de datos

Tabla 34

### Clasificación de herramientas de visualización de datos

Nombre	Software	Estructura de visualización	Sistema operativo	Licencia	Escalabilidad	Extensibilidad	Última versión y fecha
Tableau	Aplicación de escritorio	Gráficos, mapas, tablas múltiples, Tableros, etc..	- Windows 7/ Windows Server 2008 R2 or later; -OSX 10.11 o más adelante; Dispositivos móviles: -iOS -Android	Comercial; Software-propiedad; Versión de prueba durante 14 días;	++ Apache Hive NoSQL,etc.	Tableau Developer Program; REST API; JavaScript API;	2019.1 14 Febrero, 2019
Power BI	Aplicación de escritorio;	Líneas de tendencia, gráficos, mapas, cuadros de mando, etc..	- Windows 7/Windows Server 2008 R2 o más adelante; Dispositivos móviles: -iOS -Android -Windows	Comercial; Versión gratuita	+++ Apache Hive, MongoDB, Teradata, etc.	REST API; JavaScript API;	2.66.5376.2521 Febrero 2019
Sisense	- Plataforma web; -Aplicación de escritorio Elasticube Manager;	Gráficos, tablas, Tableros, etc..	- Windows 7/Windows Server 2008 R2 o más adelante; Dispositivos móviles: -iOS -Android	Software propietario; versión de prueba durante 15 días;	++ Apache Hive, Teradata etc;	JavaScript API; REST API; Elasticube Plugins; SQL RUNNER.	7.3.0.11022 19 Febrero, 2019
Qlik-View	Aplicación de escritorio	Gráficos, tablas, Tableros, etc..	- Windows 7/Windows Server 2008 R2 o más adelante; Dispositivos móviles: -iOS	Comercial; Versión gratuita	++ MongoDB, Teradata, etc.	API JavaScript/COM/ QMS; QVX SDK;	12.30.20100.0 SR1 Noviembre 2018

Note: Adaptado de "Evaluation and Analysis of Business Intelligence Data Visualization Tools." By Lousa, A., Pedrosa, I., & Bernardino, J., 2019, 2019 14th Iberian Conference on Information Systems and Technologies (CISTI).

## Anexo 11: Metodología CRISP-DM

En la década de 1990, se desarrollaron una serie de metodologías para la minería de datos, en Inglés Cross Industry Standard Process for Data Mining (CRISP-DM), una de las metodologías más difundidas, es un modelo estándar abierto del proceso que describe los enfoques comunes utilizados por los expertos en minería de datos. Según Mayorga (2016) define:

Esta metodología proporciona algunos pasos que se requieren para un proyecto de minería de procesos, como la comprensión de los objetivos del negocio, la recopilación, la descripción, el análisis y la limpieza de datos. Sin embargo, en la etapa de modelado, los pasos son específicos y no los respaldan adecuadamente, lo que requiere diagnósticos y análisis de minería de procesos. (p. s.f.)

Además, según el Manual CRISP-DM de IBM SPSS Modmeler, CRISP-DM, "Es un método probado para guiar sus trabajos de minería de datos. Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas requeridas en cada etapa."(p. 2)

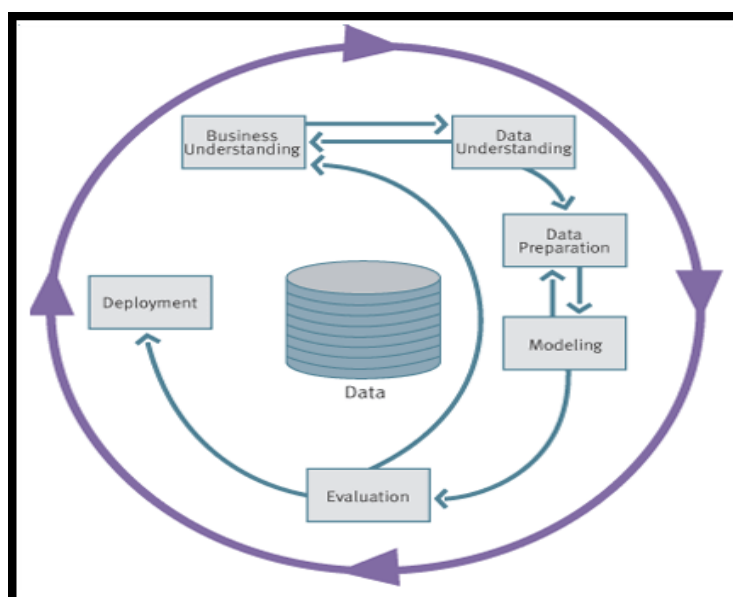


Figura 67 Etapas de metodología CRISP-DM

Nota: Etapas de metodología nítida. Esta imagen ha sido tomada de: Manual CRISP-DM de IBM SPSS Modeler 15 por IBM Corporation 1994-2012, recuperado de <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>

El ciclo de vida del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario. Según la autora describo cada una de las fases CRISP-DM de la siguiente manera:

Fase I. Comprensión del negocio. Definición de las necesidades del cliente (comprensión del negocio). Esta fase se centra en la comprensión de los objetivos del proyecto. Esta comprensión del contexto en el que nos encontramos nos permitirá generar un conocimiento de los datos para la definición del problema de aprendizaje automático y nos permitirá generar un modelo preliminar para alcanzar nuestros objetivos (Villena, 2016, prr. 8 - 14).

Fase II. Comprensión de datos. Estudio y comprensión de los datos. La fase de comprensión de datos comienza con la recopilación de datos inicial y continúa con actividades que le permiten familiarizarse con los datos, identificar problemas de calidad, descubrir conocimientos preliminares sobre los datos y/o descubrir subconjuntos interesantes para formar hipótesis con respecto a la información oculta (Villena, 2016, prr. 8 - 14).

Fase III. Preparación de datos. Análisis de datos y selección de características La fase de preparación de datos cubre todas las actividades necesarias para crear el conjunto final de datos (los datos que se utilizarán en las herramientas de modelado) a partir de los datos sin procesar iniciales. Estas tareas implican limpiar y transformar la información en instancias (filas) y atributos (columnas) para su posterior procesamiento mediante herramientas de modelado (Villena, 2016, prr. 8 - 14).

Fase IV. Modelado. El modelado es una fase en la que seleccionaremos las técnicas de modelado relevantes para el problema de predicción, y definiremos o puliremos los parámetros a valores válidos y óptimos. Algunas técnicas tienen requisitos específicos para la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto terminas volviendo a la fase de preparación de datos (Villena, 2016, prr. 8 - 14).

Fase V. Evaluación. (resultados) En esta etapa del proyecto, se han creado uno o más modelos que parecen lograr una calidad suficiente desde una perspectiva de análisis de datos. Antes de continuar con la implementación final del modelo, es importante evaluarlo a fondo y revisar los pasos ejecutados para crearlo, comparar el modelo obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay algún problema empresarial importante que no se haya considerado suficientemente. Al final de esta fase, se debe obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos (Villena, 2016, prr. 8 - 14).

Fase VI. despliegue. Implementación (producción) Por lo general, la creación del modelo no es el final del proyecto. Incluso si el objetivo del modelo es aumentar el conocimiento de los datos, el conocimiento adquirido tendrá que ser organizado y presentado antes de que el cliente pueda utilizarlos. Dependiendo de los requisitos, la fase de desarrollo puede ser tan simple como generar un informe o tan compleja como realizar un proceso de análisis de datos periódico y tal vez automatizado en la organización (Villena, 2016, prr. 8 - 14).

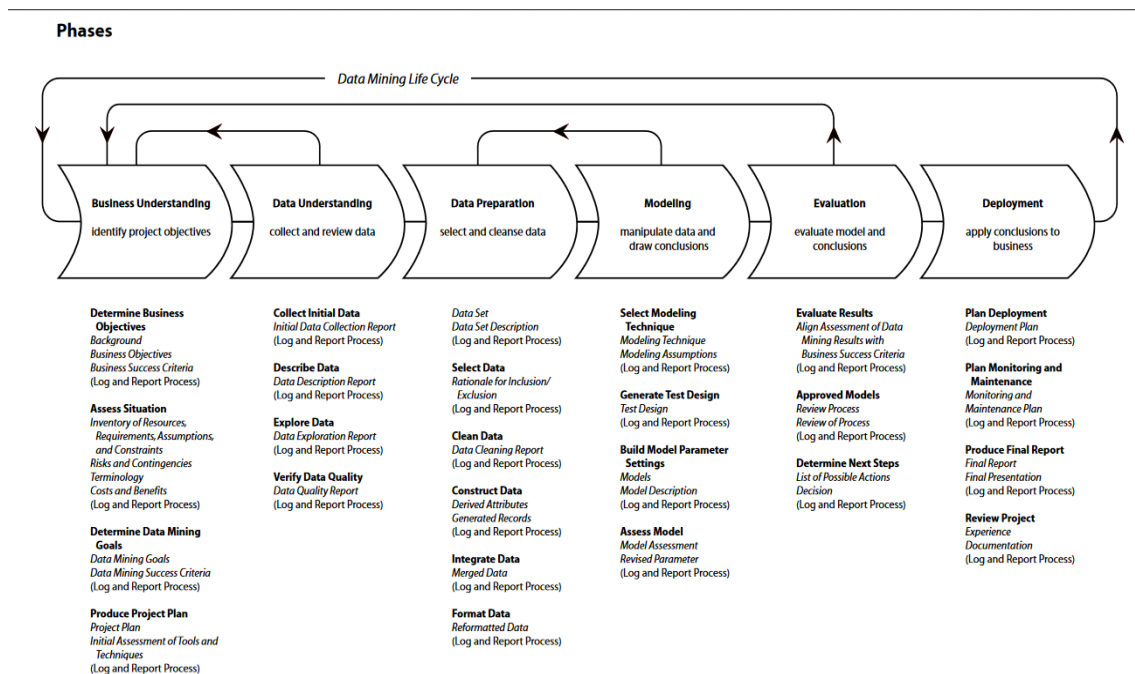


Figura 68 Fases y tareas del CRISP DM

Nota: Fases y tareas de la metodología CRISP DM Source: Esta imagen ha sido tomada de: una guía visual de la metodología CRISP-DM de NicoleLeaper 2019, EXDE, <https://exde.wordpress.com/2009/03/13/a-visual-guide-to-crisp-dm-methodology>

A partir de las descripciones anteriores podemos afirmar que la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) es una metodología completa porque tienen en cuenta el entorno empresarial y la parte de la secuencia de las fases no es rígida, y nos permitirá desarrollar este proyecto en cada una de sus fases descritas a continuación:

Para Montalvo-Garcia et al. (2019), clasifico:

#### Fase 1: Comprensión del negocio

- Comprender el contexto y determinar los objetivos Evaluación de la situación.
- Determinación de los objetivos de minería de datos.
- Desarrollo del plan del proyecto.

#### Fase 2: Comprender los datos iniciales de recopilación de datos.

- Descripción de los datos.
- Exploración de datos.
- Verificación de la calidad de los datos.

#### Fase 3: Preparación de los datos de selección de datos más relevantes.

- Limpieza de datos.
- Construcción de nuevos datos (atributos).
- Integración de datos.

#### Fase 4: Modelado

- Selección de técnicas de modelado.
- Generación de un diseño de verificación.
- Generación de modelos.
- Evaluación y verificación del modelo.

#### Fase 5: Evaluación

- Evaluación de los resultados.
- Proceso de revisión.
- Determine los siguientes pasos en función de los resultados.



## Fase 6: Despliegue

- Planificación de la distribución.
- Creación del informe final.
- Revisión final del proyecto. (p. 455)

## Anexo 12: Modelo dimensional con variables dummy

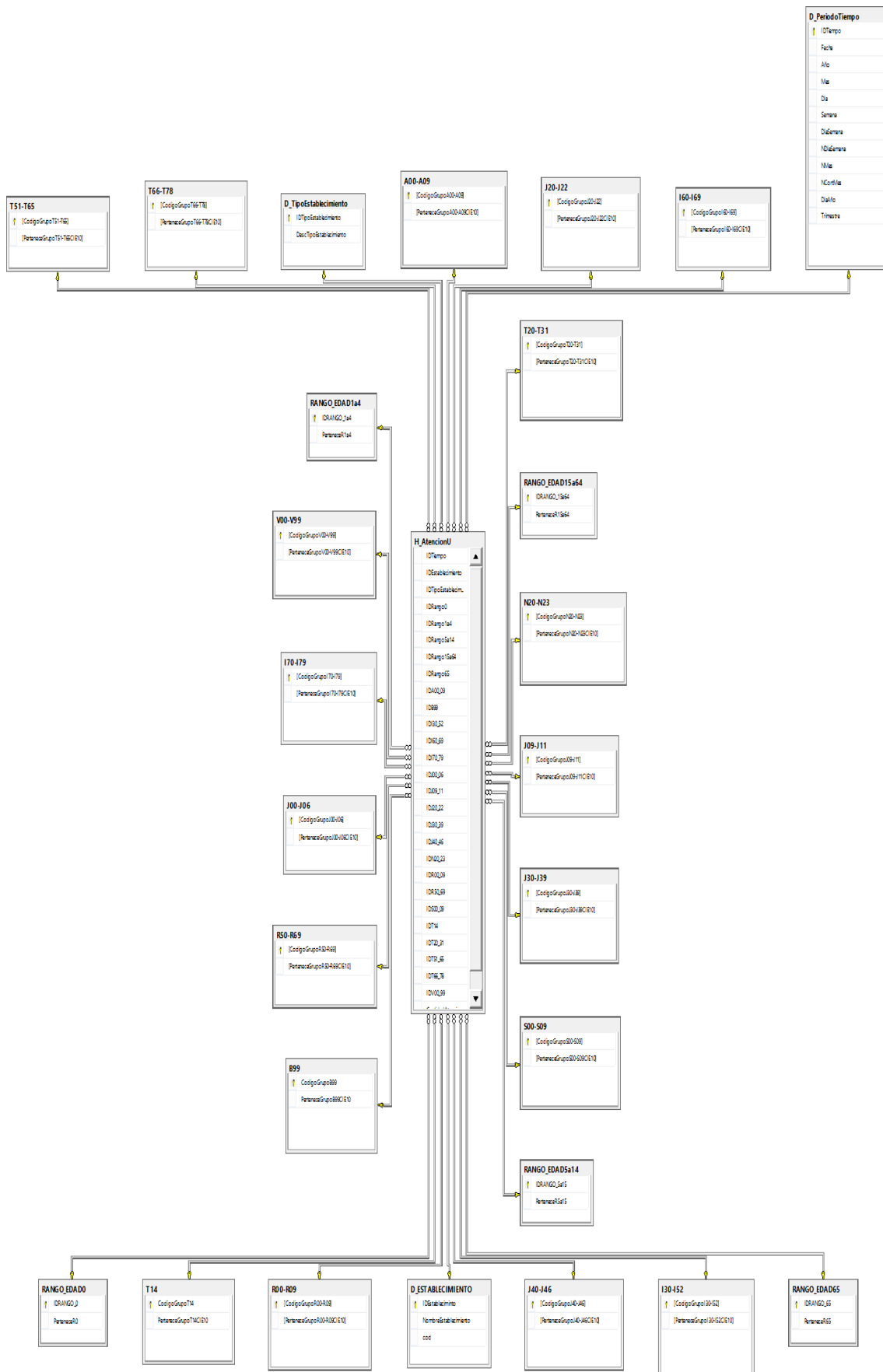


Figura 69 Modelo dimensional con variables dummy

### Anexo 13: Diccionario de datos del modelo dimensional

- Facts\_AttentionUrgencias, representa los eventos de los pacientes a lo largo del tiempo

Tabla 35

Dimensión H\_AtencionUrgencia

Columna	CP	CF	Tipo	Tamaño	Descripción
IDEstablecimiento		✓	N	10	Código Establecimiento
IDMedication		✓	N	10	Código Medicamento
IDPadecimiento		✓	N	10	Código Enfermedades
IDCampaña		✓	N	10	Código Campaña
IDTipoAtencion		✓	N	10	Código Tipo Atención
IDTiempo		✓	N	10	Código PeriodoTiempo
CAtecciones			N	10	Cantidad de atenciones
CAtecciones0			N	10	número de atenciones de edad 0
CAtecciones14			N	10	número de atenciones de edad entre 1 y 4
CAtecciones514			N	10	número de atenciones de edad entre 5 y 14
CAtecciones1564			N	10	número de atenciones de edad entre 15 y 64
CAtecciones65			N	10	número de atenciones de edad 65
CMedicamiento			N	10	cantidad de medicamentos utilizados

- D\_Establecimiento, contendrá información de los establecimientos

Tabla 36

Dimensión D\_Establecimiento

Columna	CP	CF	Tipo	Tamaño	Descripción
IDEstablecimiento	✓		N	10	Identificador de la tabla
NomEstablecimiento			A	50	Nombre del establecimiento
IDUbicacion		✓	N	10	Código de localización

- D\_TipoEstablecimiento, contendrá información sobre los tipos de establecimiento.

*Tabla 37*

*Dimensión D\_TipoEstablecimiento*

Columna	CP	CF	Tipo	Tamaño	Descripción
IDTipoEstablecimiento	✓		N	10	Identificador de la tabla
TipoEDescription			A	30	Descripción del tipo de centro de atención

- D\_Ubicacion, contendrá información sobre las ubicaciones de los hospitales

*Tabla 38*

*Dimensión D\_Ubicacion*

Columna	CP	CF	Tipo	Tamaño	Descripción
IDUbicacion	✓		N	10	Identificador de la tabla
UUbicacion			A	30	Nombre de la localidad, región, provincia

- D\_PeriodoTiempo, subdivisión de tiempo para consultas

*Tabla 39*

*Dimensión D\_PeriodoTiempo*

Columna	CP	CF	Tipo	Tamaño	Descripción
IDTiempo	✓		N	10	Identificador de la tabla
Fecha			F		Hora fecha de las atenciones
Año			N	10	Tiempo en años
Mes			N	10	Tiempo en el mes
Día			N	10	Tiempo en días
Semana			N	10	Tiempo en semanas
DiaSemana			N	10	el número del día de la semana
NDiaSemana			Ch	0	nombre del día de la semana
NMes			Ch	10	nombre del mes del año
NCortMes			Ch	10	abreviatura del mes del año
DiaAño			N	10	número que pertenece al día del año
Trimestre			N	10	Tiempo en el trimestre

- D\_TipoAtencion, contendrá información sobre los tipos de atenciones

*Tabla 40*

*Dimensión D\_TipoAtencion*

Columna	CP	CF	Tipo	Tamaño	Descripción
IDTipoAtencion	✓		N	10	Identificador de la tabla
TipoANombre			A	30	Descripción del tipo de atención

- D\_Campaña, contendrá información sobre las campañas llevadas a cabo en los hospitales.

*Tabla 41*

*Dimensión D\_Campaña*

Columna	CP	CF	Tipo	Tamaño	Descripción
IDCampaña	✓		N	10	Identificador de la tabla
Descripción			A	30	Descripción de la campaña

- D\_Padecimiento, contendrá información sobre enfermedades

*Tabla 42*

*Dimensión D\_Padecimiento*

Columna	CP	CF	Tipo	Tamaño	Descripción
IDPadecimiento	✓		N	10	Identificador de la tabla
EnfNombre			A	100	Nombre de la enfermedad
IDCEI10			N	10	Código CIE

- CEI10, contendrá los nombres de las causas de las cuales se realiza la atención urgente

*Tabla 43*

*Dimensión CEI10*

Columna	CP	CF	Tipo	Tamaño	Descripción
IDCEI10	✓		N	10	Identificador de la tabla
CodeCEI			A	30	Código del grupo de enfermedades
Descripción			A		Definición de grupos de enfermedades

- D\_ Medicamento, lista de medicamentos para enfermedades

*Tabla 44*

*Dimensión D\_Medicamento*

<b>Columna</b>	<b>CP</b>	<b>CF</b>	<b>Tipo</b>	<b>Tamaño</b>	<b>Descripción</b>
IDMedicamento	✓		N	10	Código de la causa de la atención
MedNombre			A	30	Nombre de la causa de la atención

## Anexo 14: Sentencias SQL para ETL

```

/*****Dimensión PeriodoTiempo*****/
SELECT DISTINCT
YEAR (CONVERT (date, fecha)) *10000+MONTH(CONVERT(date,
fecha))*100+DAY(CONVERT(date, fecha)) AS IDTIEMPO
, FORMAT(CONVERT(date, fecha) , 'yyyy-MM-dd') as fecha
, DATEPART(yy, CONVERT(date, fecha)) as Año
, DATEPART(m, CONVERT(date, fecha)) as mes
, RIGHT('0' + DATEPART(dd, CONVERT(date, fecha)),2) as dia
, DATEPART(wk, CONVERT(date, fecha)) as semana
, DATEPART(DW, CONVERT(date, fecha)) as Diasemana
, DATENAME(dw, CONVERT(date, fecha)) as Nomdiasemana
, DATENAME(mm, CONVERT(date, fecha)) as NomMes
, LEFT(CONVERT(date, fecha), 3) as NomCortMes
, DATEPART(dy, CONVERT(date, fecha)) as DíaAño
, DATEPART(qq, CONVERT(date, fecha)) as trimestre
FROM [2019].[dbo].[Atenciones Urgencias 2019] h

WHERE
NOT EXISTS (SELECT 1 FROM [dbo].D_PeriodoTiempo DT WHERE (YEAR(CONVERT(date,
h.fecha))*10000+MONTH(CONVERT(date, h.fecha))*100+DAY(CONVERT(date,
h.fecha))=DT.IDTiempo))
group by YEAR(fecha)*10000+MONTH(fecha)*100+DAY(fecha)
, FORMAT(fecha, 'yyyy-MM-dd')
, DATEPART(yy, fecha)
, DATEPART(m, fecha)
, RIGHT('0' + DATEPART(dd, fecha),2)
, DATEPART(wk, fecha)
, DATEPART(DW, fecha)
, DATENAME(dw, fecha)
, DATENAME(mm, fecha)
, LEFT(fecha, 3)
, DATEPART(dy, fecha)
, DATEPART(qq, fecha)

/*****Dimensión Padecimiento*****/

SELECT DISTINCT IdCausa,GlosaCausa
FROM [2015].[dbo].DatosLineal h
WHERE
NOT EXISTS (SELECT 1 FROM [dbo].D_Padecimineto DT WHERE
(h.IdCausa=DT.IDPadecimineto))
order by IdCausa asc

/*****Dimensión TipoEstablecimientos *****/
SELECT DISTINCT GLOSATIPOESTABLECIMIENTO
FROM [2012].[dbo].DatosLineal h
WHERE
NOT EXISTS (SELECT 1 FROM [dbo].D_TipoEstablecimientos DT WHERE
(h.GLOSATIPOESTABLECIMIENTO=DT.TEDescripcion))

/*****Dimensión Tipo Atención*****/
SELECT DISTINCT GLOSATIPOATENCION
FROM [2010].[dbo].[DatosLineal] h
WHERE
NOT EXISTS (SELECT 1 FROM [dbo].D_Padecimineto DT WHERE
(h.IdCausa=DT.IDPadecimineto))

/***** Dimensión Campaña *****/

```

```

SELECT DISTINCT GlosaTipoCampana
FROM [2012].[dbo].[DatosLineal] h
WHERE
    NOT EXISTS (SELECT 2 FROM [dbo].D_Campaña DT WHERE
(h.GlosaTipoCampana=DT.[CDescripcion]))

/***** Dimensión Ubicación *****/
SELECT DISTINCT Localidad
FROM [2012].[dbo].[DatosLineal] h
WHERE
    NOT EXISTS (SELECT 2 FROM [dbo].D_Ubicacion DT WHERE (h. Localidad
=DT.[ULocalidad]))

/***** Dimensión Medicamento *****/
SELECT DISTINCT Medicamento
FROM [2012].[dbo].[DatosLineal] h
WHERE
    NOT EXISTS (SELECT 2 FROM [dbo].D_Medicamiento DT WHERE
(h.Medicamento=DT.[MedNombre]))

/*****HECHOS AtencionUrgencia*****/

SELECT YEAR(fecha)*1000+MONTH(fecha)*100+DAY(fecha) as idtiempo,
IdEstablecimiento,
(select IDTipoEstablecimientos from D_TipoEstablecimientos where TEdescripcion =
[2019].[dbo].[Atenciones Urgencias 2019].GLOSATIPOESTABLECIMIENTOS ) as
IdtEstablecimiento,
(select IdCausa from D_Padecimineto where CANombre = [2019].[dbo].[Atenciones
Urgencias 2019].GlosaCausa ) as idcausa,
(select IDMedicamiento from D_Medicamiento where MedNombre =
[2019].[dbo].[Atenciones Urgencias 2019].Medicamento ) as idmedi,
(select IDTipoAtencion from D_TipoAtencion where TANombre =
[2019].[dbo].[Atenciones Urgencias 2019].GLOSATIPOATENCION ) as Idtatencion,
(select IDCampaña from D_Campaña where CDescripcion = [2019].[dbo].[Atenciones
Urgencias 2019].GlosaTipoCampana ) as idcampaña,
(select IDUbicacion from D_Ubicacion where ULocalidad = [2019].[dbo].[Atenciones
Urgencias 2019].Localidad ) as idlocalidad,
[Total],
[Menores de 1 año],
[1-4 años],
[5-14 años],
[15-64 años],
[65 años y más],
[TotalMedicina]
FROM [2019].[dbo].[Atenciones Urgencias 2019]

```



## Anexo 15: Código de RStudio

```
#Instalacion de librerias
install.packages("readr")
install.packages("ggplot2")
install.packages("forecast")
install.packages("fpp2")
install.packages("TTR")
install.packages("dplyr")
install.packages("lmtest")
install.packages("car")
install.packages("carData")
install.packages("tseries")
install.packages("lm.beta")
install.packages("reliability")
install.packages("psych")
install.packages("ordinal ")
install.packages("rms ")
install.packages("pscl ")
library(pscl)
library(rms)
library(ordinal)
library(readr)
library(ggplot2)
library(forecast)
library(fpp2)
library(TTR)
library(dplyr)
library(lmtest)
library(car)
library(carData)
library(tseries)
library(readr)
library(lm.beta)
library(Reliability)
library(psych)

#Carga de la data
DataFinal <- read_delim("C:/Users/Angello/Desktop/DataFinal.csv",
+   ";", escape_double = FALSE, col_types = cols(`Año` = col_number(),
+   Mes = col_number(), Days = col_number(),
+   IDEstablecimiento = col_number(),
+   IDTipoEstablecimiento = col_number(),
+   REdad0 = col_number(), REdad1a4 = col_number(),
+   REdad5a14 = col_number(), REdad15a64 = col_number(),
+   REdad65 = col_number(), IDA00_09 = col_number(),
+   IDB99 = col_number(), IDI30_52 = col_number(),
+   IDI60_69 = col_number(), IDI70_79 = col_number(),
+   IDJ00_06 = col_number(), IDJ09_11 = col_number(),
+   IDJ20_22 = col_number(), IDJ30_39 = col_number(),
```

```

+     IDJ40_46 = col_number(), IDN20_23 = col_number(),
+     IDR00_09 = col_number(), IDR50_69 = col_number(),
+     IDS00_09 = col_number(), IDT14 = col_number(),
+     IDT20_31 = col_number(), IDT51_65 = col_number(),
+     IDT66_78 = col_number(), IDV00_99 = col_number(),
+     RangAtenciones = col_number()), trim_ws = TRUE)
View(DataFinal)

```

### **#Regresion lineal multiple**

```

modeloRegresion <- lm(NAtenciones ~ Año + Mes + Days + IDEstablecimiento
+ IDTipoEstablecimiento + REdad0 + REdad1a4 + REdad5a14 + REdad15a64
+ REdad65 + IDA00_09 + IDB99 + IDI30_52 + IDI60_69 + IDI70_79 + IDJ00_06
+ IDJ09_11 + IDJ20_22+ IDJ30_39 + IDJ40_46 + IDN20_23 + IDR00_09 +
IDR50_69 + IDS00_09 + IDT14 + IDT20_31 + IDT51_65 + IDT66_78 , data =
DataFinal )
summary(modeloRegresion)

```

```

#Vif
vif(modeloRegresion)

```

```

#DARWIN Watson test
dwtest(modeloRegresion)

```

```

#BETA
lm.beta(modeloRegresion)

```

```

#ANOVA
anovamodelo = aov(modeloRegresion)
summary(anovamodelo)

```

```

#Descriptive statistics
sd(DataFinal$NAtenciones)
sd(DataFinal$Año)
sd(DataFinal$Mes)
sd(DataFinal$Days)
sd(DataFinal$IDEstablecimiento)
sd(DataFinal$IDTipoEstablecimiento)
sd(DataFinal$REdad0)
sd(DataFinal$REdad1a4)
sd(DataFinal$REdad5a14)
sd(DataFinal$REdad15a64)
sd(DataFinal$REdad65)
sd(DataFinal$IDA00_09)
sd(DataFinal$IDB99)
sd(DataFinal$IDI30_52)
sd(DataFinal$IDI60_69)
sd(DataFinal$IDI70_79)
sd(DataFinal$IDJ00_06)
sd(DataFinal$IDJ09_11)
sd(DataFinal$IDJ20_22)

```

```
sd(DataFinal$IDJ30_39)
sd(DataFinal$IDJ40_46)
sd(DataFinal$IDN20_23)
sd(DataFinal$IDR00_09)
sd(DataFinal$IDR50_69)
sd(DataFinal$IDS00_09)
sd(DataFinal$IDT14)
sd(DataFinal$IDT20_31)
sd(DataFinal$IDT51_65)
sd(DataFinal$IDT66_78)
```

#### #TIME SERIES

```
DA=ts(timeseries[,2], start = c(2014, 1), end = c(2018, 12), frequency = 12)
autoplot(DA) # Histograma
```

#### #Holt-Winter's

```
modeloholt = Holt-Winter's(DA,seasonal = "additive")
fitted(modeloholt)# Filter interpretation
#Xhat = Average forecast
#Level = Level of interaction
#Trend = Data training
#Season = Parking
plot(fitted(modeloholt))
autoplot(fitted(modeloholt))
```

#### #prediction

```
predholt=predict(modeloholt,12,prediction.interval = TRUE)
predholt
#Fit = estimated quantity
#Upr = maximum acceptance margin
#Lwr = minimum declination margin
plot(modeloholt,predholt)
summary(forecast(modeloholt))
autoplot(forecast(modeloholt))
```

#### #Naive método

```
modelonaive = snaive(DA, h =24)
plot(modelonaive)
prednaive=predict(modelonaive,12,prediction.interval = TRUE)
plot(modelonaive)
autoplot(forecast(modelonaive))
summary(forecast(modelonaive))
```

#### #Exponential Trend Smoothing

```
modeloets<- ets(DA)
prediccionets<- forecast(fit_ets_default, h = 24)
plot(prediccionets)
summary(forecast(prediccionets))
```

#### #MAPE test

```
mape<- function(Actualt, pred){
  mape<- mean(abs((Actualt - pred)/actual))*100
  return (mape)}
```

```
#Actualt = estimación de promedio
#pred = la estimación oscila entre hi 80 y hi 95 que va
# cambiar de acuerdo con el modelo utilizado
```

### **#Regresion Ordinal Logist**

```
MLogit<- lrm(formula = RangAtenciones ~ Año+Mes+Days + IDEstablecimiento
+ IDTipoEstablecimiento + REdad0 + REdad1a4 + REdad5a14 +
REdad15a64 + REdad65 + IDA00_09 + IDB99 + IDI30_52 + IDI60_69 +
IDI70_79 + IDJ00_06 + IDJ09_11 + IDJ20_22 + IDJ30_39 + IDJ40_46 +
IDN20_23 + IDR00_09 + IDR50_69 + IDS00_09 + IDT14 + IDT20_31 +
IDT51_65 + IDT66_78, data = DataLog,tol=1e-12)
```

```
# ordinal formula summary
print(MLogit)
```

```
# McFadden
pR2(MLogit)[["McFadden"]]
```

```
# Nagelkerke/Cragg & Uhler's
pR2(MLogit)[["r2CU"]]
```

```
# Cox & Snell's
pR2(MLogit)[["r2ML"]]
```

```
#Summary variable
anova(MLogit)
```

## Anexo 16: Manual de usuario

### INTRODUCCIÓN

La aplicación de atención de emergencia geolocalizada está diseñada para visualizar la atención histórica del centro de salud y también mostrar predicciones de atenciones.

### ENTRADA AL SISTEMA

Para utilizar el sistema, primero debe cargar la base de datos generada por el datamining para el uso correspondiente. Primero cargamos la base de datos AtencionUrgencia de SQL Server a Tableau muestra la herramienta de análisis de inteligencia empresarial geoespacial.



Figura 70 Ícono Tableau

Al iniciar la herramienta nos proporciona una diversidad de conexión de base de datos como: Excel, archivo de texto, Json, archivo de espacio, etc. Además de conectarse a un servidor SQL, Mysql, Oracle Etc.



Figura 71 Ventana principal del Tableau

Para la carga de ejemplo elegiremos SQL Server donde se encuentran nuestros datos.

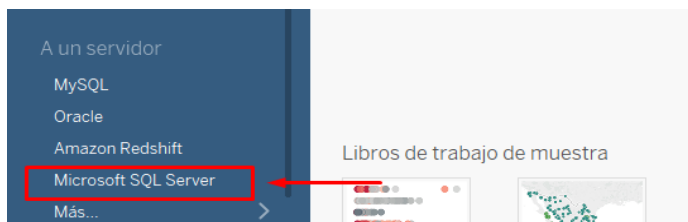


Figura 72 Selección del servidor de base de datos en Tableau

A continuación, ponemos el nombre del servidor y la base de datos, finalmente haga clic en el inicio de sesión para conectarse a SQL Server.

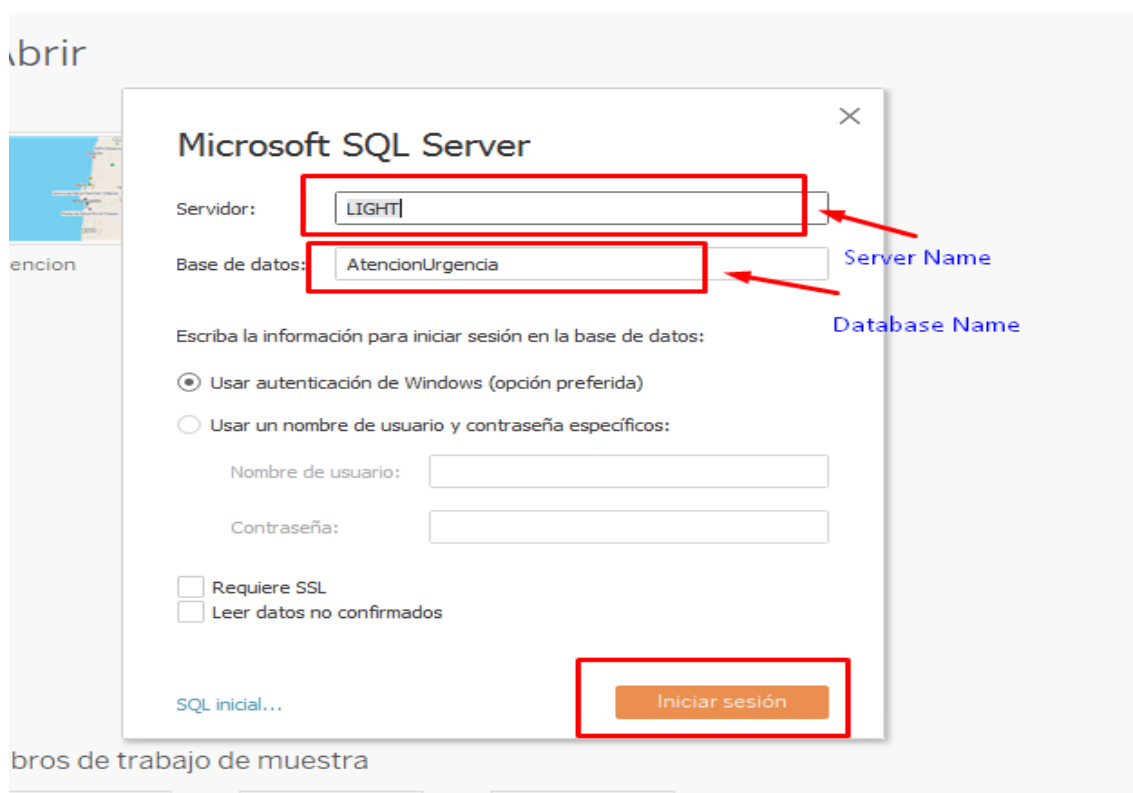


Figura 73 Entrada de datos para la conexión en Tableau

Esto podría tomar su tiempo de acuerdo con los requisitos del equipo, al terminar la espera nos muestra la siguiente ventana que contiene la siguiente:

- Conexión SQL

- Nombre de la base de datos
- Tablas y vistas
- Procedimientos Almacenados
- Vista para el manejo de tablas según sea necesario
- Vista de datos

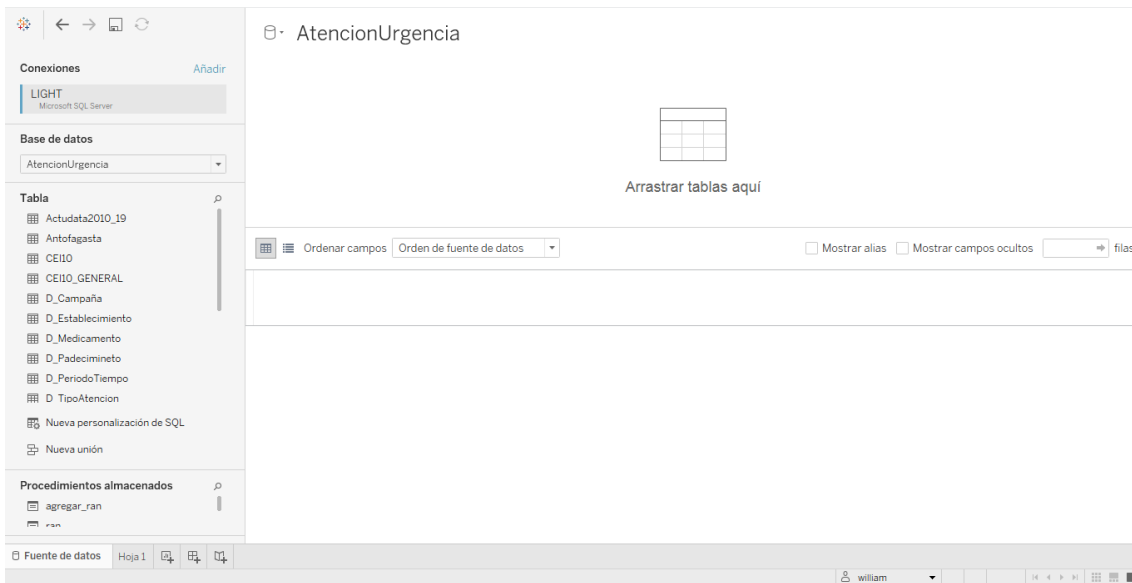


Figura 74 Tablas de ventanas de la base de datos en Tableau

Tomaremos una vista generada en SQL llamada Antofagasta que contiene los datos de 2010 a 2019 que se utilizara en la carga al Tableau.

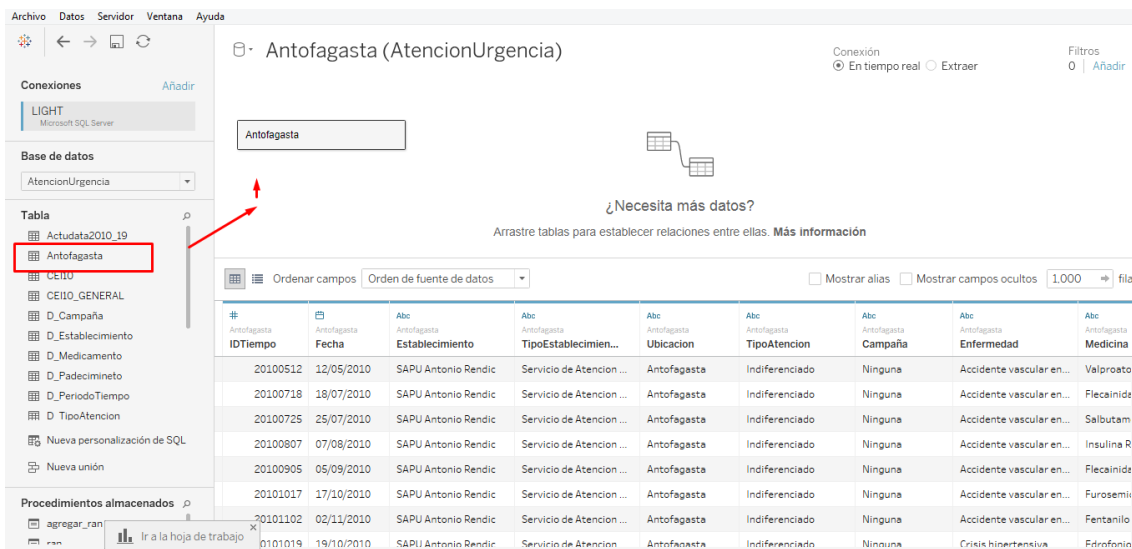


Figura 75 Selección de datos de análisis en Tableau

Cuando finaliza la carga de datos, abrimos el proyecto Tableau\_AtencionUrgencia y nos muestra la siguiente ventana.

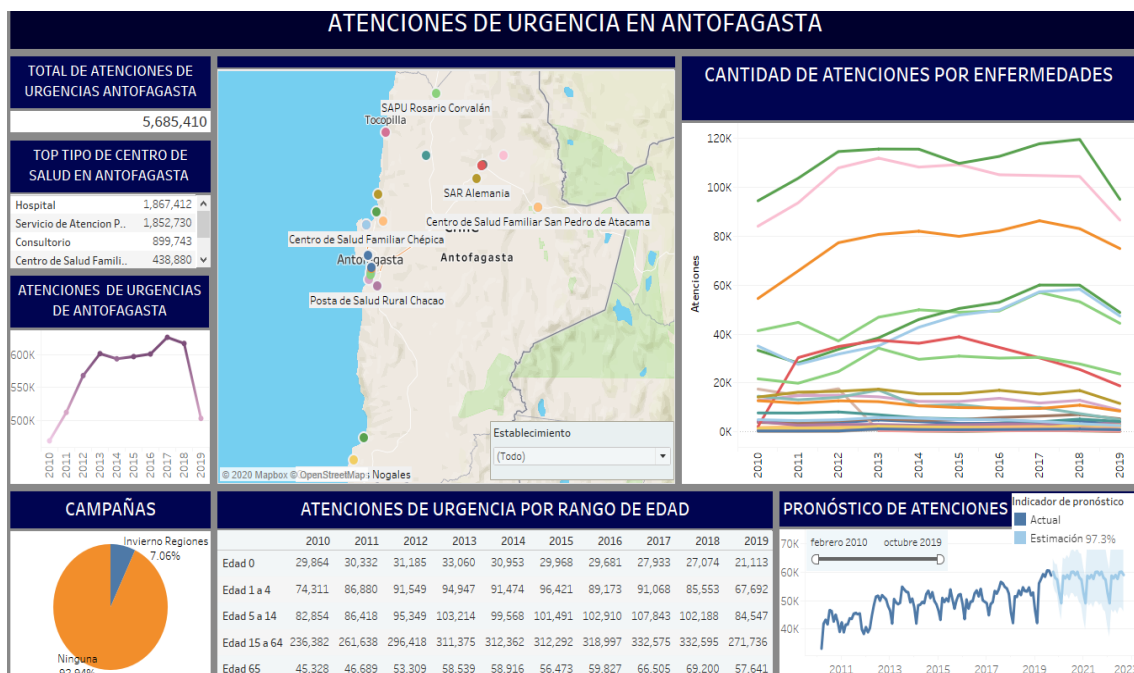


Figura 76 Ventana principal del sistema de predicción en Antofagasta

A continuación, vemos que el SIG de atención de emergencia de Antofagasta contiene los siguientes:

- Resultado total de atenciones de urgencia
- Tipo de instalación de atención médica superior
- Flujo de atenciones de urgencia
- Campañas
- Mapa de Antofagasta y sus establecimientos
- Cantidad de atención a enfermedades según el centro
- Atención de emergencia por rango de edad
- Pronóstico de atenciones de urgencias

Ahora, explicamos la funcionalidad del sistema SIG de Atención de Emergencia en Antofagasta.



## 1. RESULTADO TOTAL DE ATENCIONES DE URGENCIA

Aquí verá el resultado general de la atención de emergencia que tenemos en la base de datos en este momento.

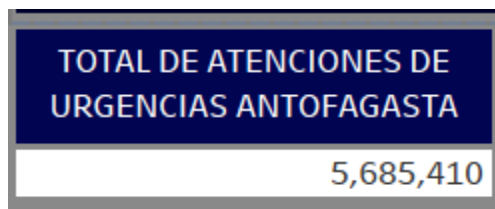


Figura 77 Vista de la atención de emergencia de resultados

## 2. TIPO DE INSTALACIÓN DE ATENCIÓN MÉDICA SUPERIOR

En vista, tenemos el tipo de establecimientos que tiene atención más urgente, también cuenta con una barra de desplazamiento para mostrar el resto.

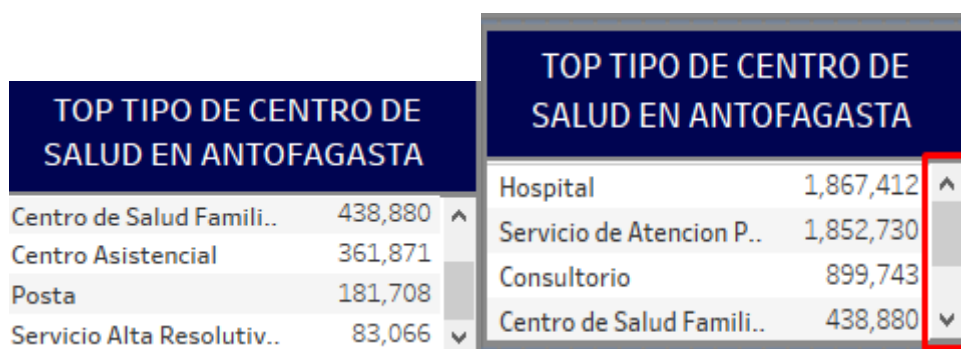


Figura 78 Vista superior del centro de salud con más atención

En algunos tipos de establecimiento, el nombre completo se muestra al señalar.

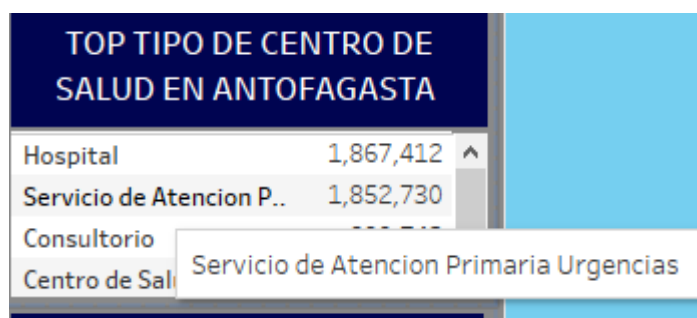


Figura 79 Visualización de etiquetas

### 3. FLUJO DE ATENCIONES DE URGENCIA

Esta sección muestra un gráfico de líneas con la cantidad de atención de emergencia en el período de tiempo. Cuando se señala cualquier punto de este gráfico, aparecerá una nota que muestra el año y la cantidad de atención en ese punto.

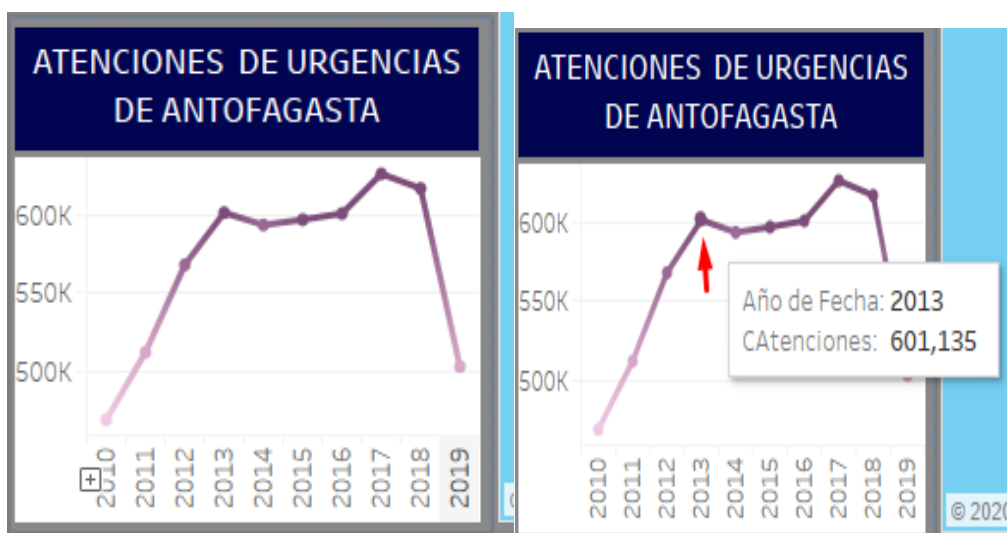


Figura 80 Gráfico de flujo de atención por año

### 4. CAMPAÑAS

En esta otra sección visualizaremos las campañas realizadas en la región de Antofagasta medidas en porcentaje. Además, para incluir el porcentaje se visualiza la cantidad de atención de urgencia al indicar.

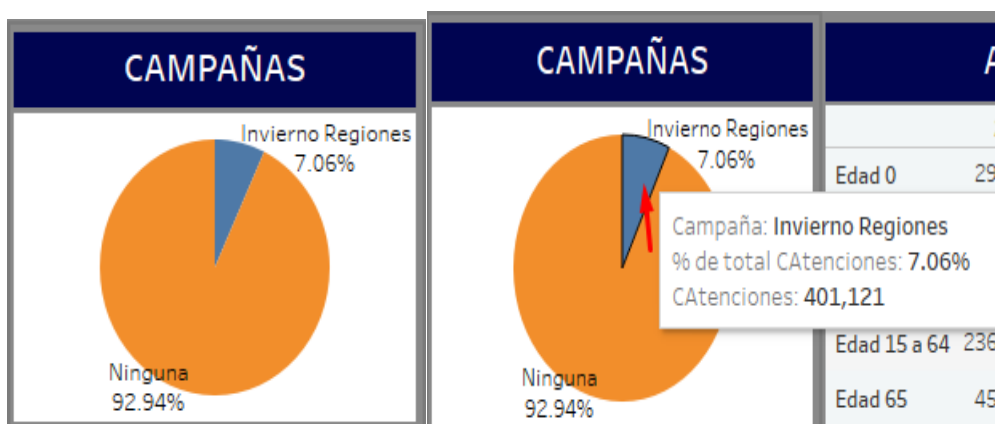


Figura 81 Gráfico de las campañas llevadas en el establecimiento

## 5. MAPA DE ANTOFAGASTA Y SUS ESTABLECIMIENTOS

Esta ventana muestra la región de Antofagasta-Chile con su respectivo establecimiento, contiene los siguientes:

- Cinta de herramientas de mapeo, para mover, acercar, buscar mapas y corregir mapas.
- Filtro de establecimientos.
- Punto de referencia de los establecimientos donde cada uno está representado con un color diferente.

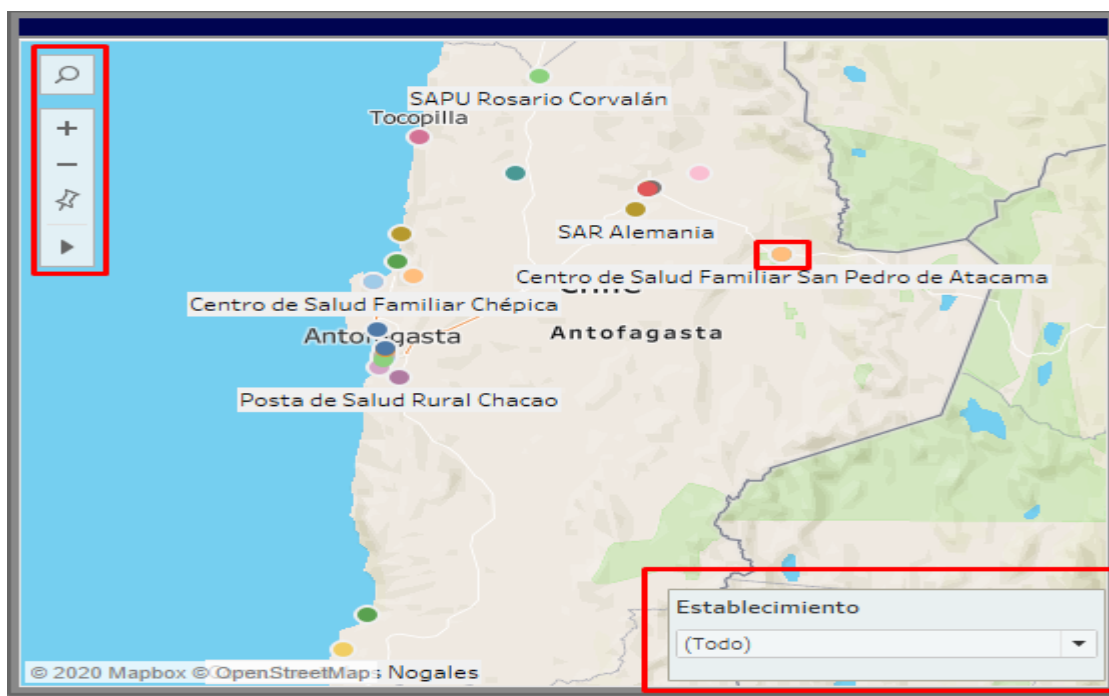


Figura 82 Mapa de Antofagasta y sus centros de salud

Cuando se resalta un punto, aparecerá una nota con la siguiente información: nombre del establecimiento, tipo de establecimientos y número de resultados de los servicios.

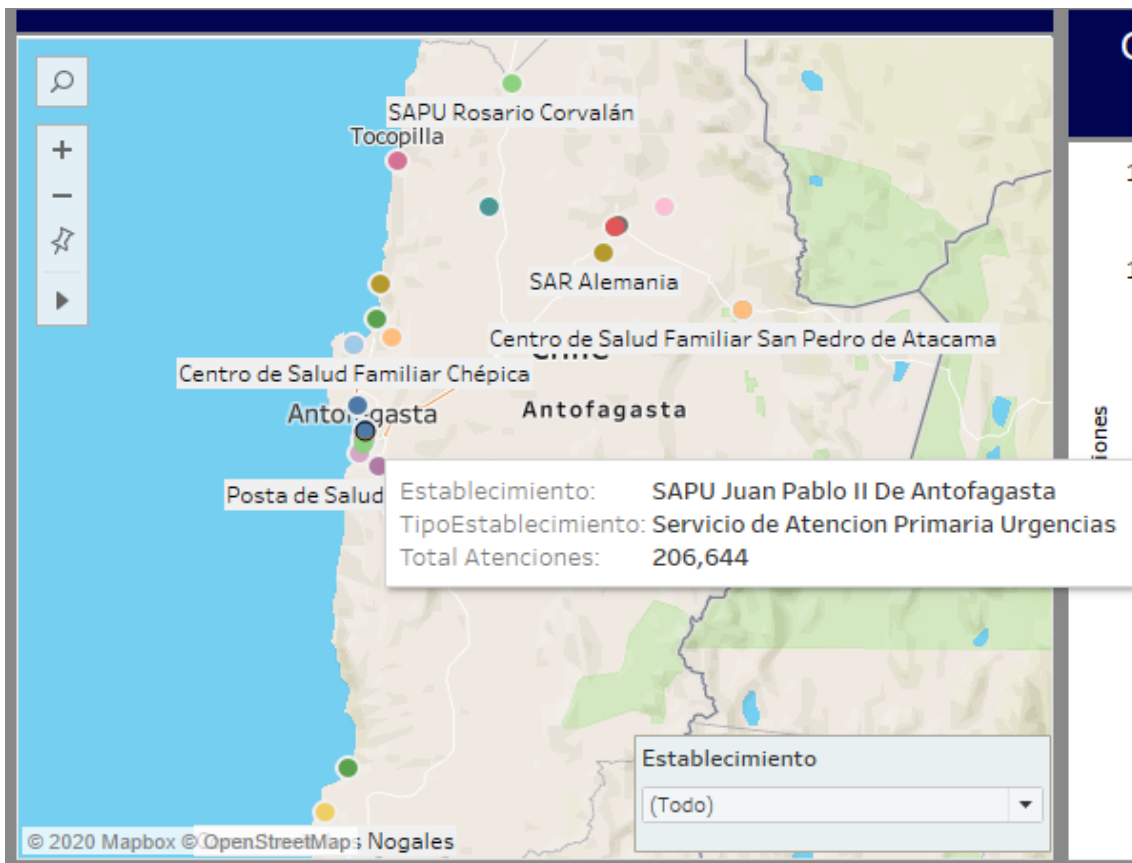


Figura 83 Información del centro de salud

Además, mediante el uso del filtro podemos localizar un establecimiento específico.

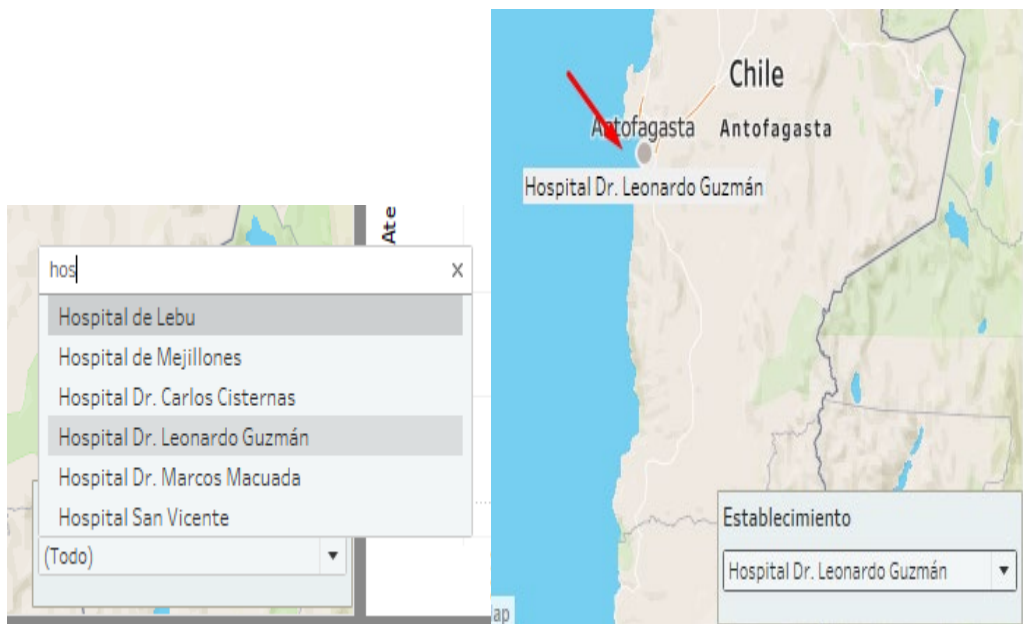


Figura 84 Filtro de establecimientos

Para eliminar el filtro hay un icono en forma de embudo, al que hacer clic se elimina el filtro.

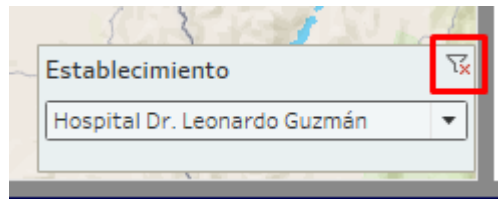


Figura 85 *Remove el filtro*

## 6. CANTIDAD DE ATENCIÓN A ENFERMEDADES SEGÚN EL CENTRO

Aquí en esta vista, se muestra el número de tratamientos de emergencia para enfermedades o causas. Además, al señalar un punto en la línea del gráfico donde cada color es una enfermedad diferente, se mostrarán los siguientes datos: enfermedad o causa, año, número de tratamientos.

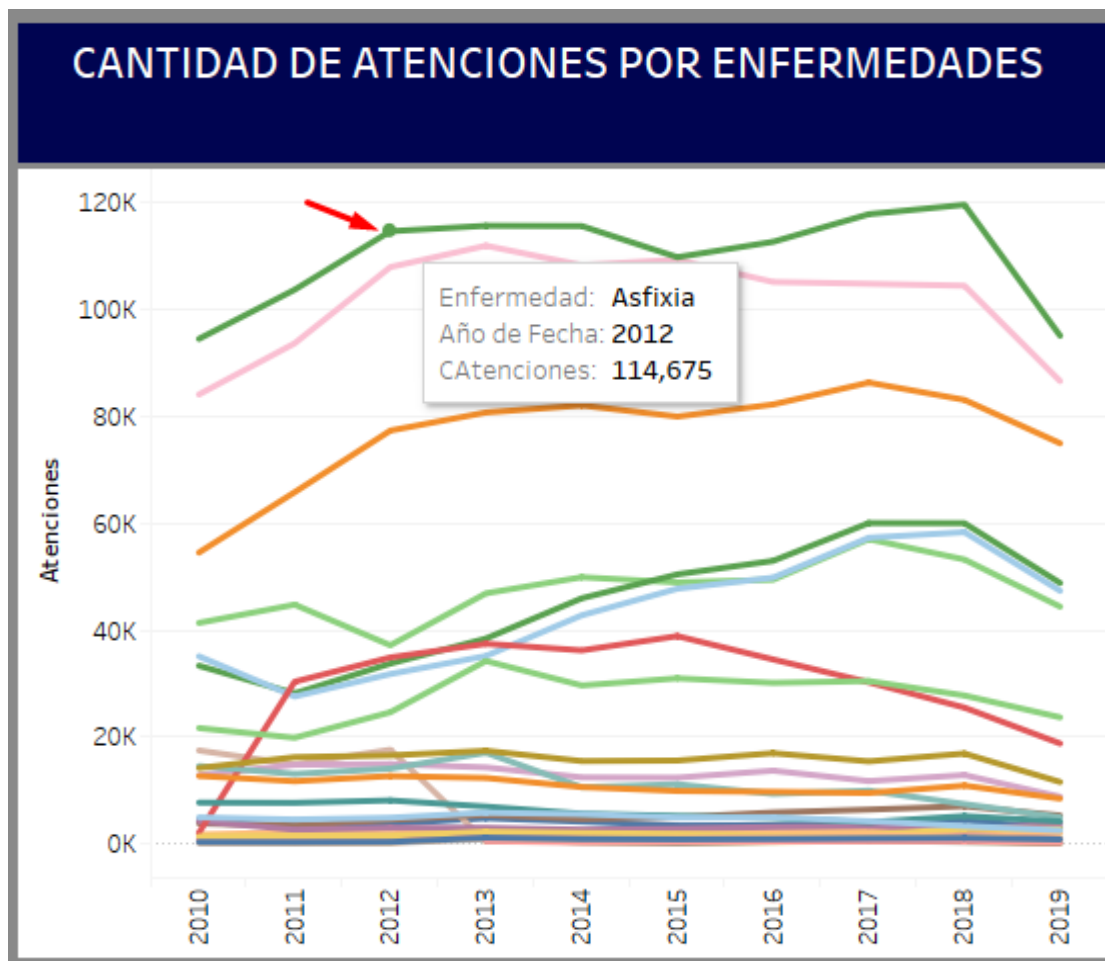


Figura 86 *Flujo de enfermedades al año*

## 7. ATENCIÓN DE EMERGENCIA POR RANGO DE EDAD

Esta tabla muestra el cuidado de resultados según el rango de edad con sus respectivos años.

ATENCIÓNES DE URGENCIA POR RANGO DE EDAD										
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Edad 0	29,864	30,332	31,185	33,060	30,953	29,968	29,681	27,933	27,074	21,113
Edad 1 a 4	74,311	86,880	91,549	94,947	91,474	96,421	89,173	91,068	85,553	67,692
Edad 5 a 14	82,854	86,418	95,349	103,214	99,568	101,491	102,910	107,843	102,188	84,547
Edad 15 a 64	236,382	261,638	296,418	311,375	312,362	312,292	318,997	332,575	332,595	271,736
Edad 65	45,328	46,689	53,309	58,539	58,916	56,473	59,827	66,505	69,200	57,641

Figura 87 Resultado de la edad de atención de emergencia por año

## 8. PRONÓSTICO DE ATENCIONES DE URGENCIAS

En este gráfico, nuestra predicción usando el Hola-Winter se muestra con una presión del 97,3%.



Figura 88 Pronostico de atenciones de urgencia

Al señalar la parte estimada vamos a visualizar los datos predichos.

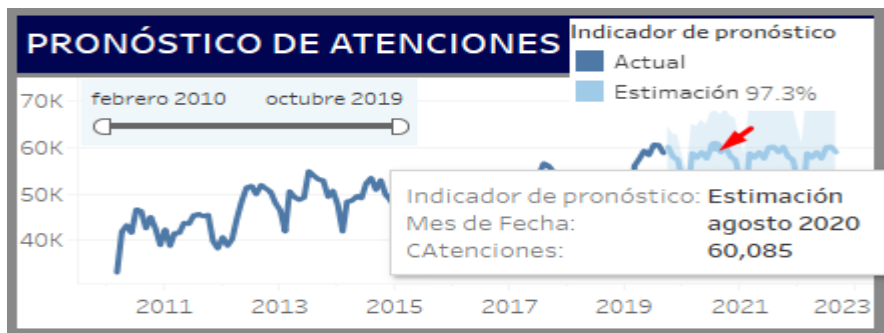


Figura 89 Estimación de atenciones de Urgencia 2020

También tiene un filtro de rango de fechas para la visualización.

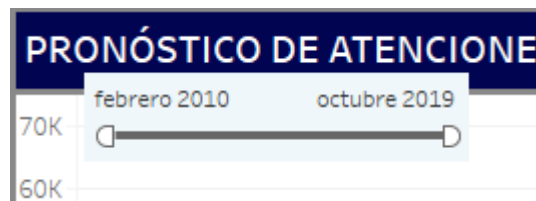


Figura 90 Filtrar la atención por fecha

Por lo tanto, cada viñeta representa los datos globales sólo de Antofagasta, por lo que ahora veremos cómo filtrar un establecimiento seleccionado para mostrarnos lo siguiente:

- Campañas
- Cantidad de atención a enfermedades según el centro
- Atención de emergencia por rango de edad
- Pronostico de Atenciones de Urgencia

En primer lugar, elegimos cualquier establecimiento, por ejemplo: Centro Familiar de Salud Chépica.

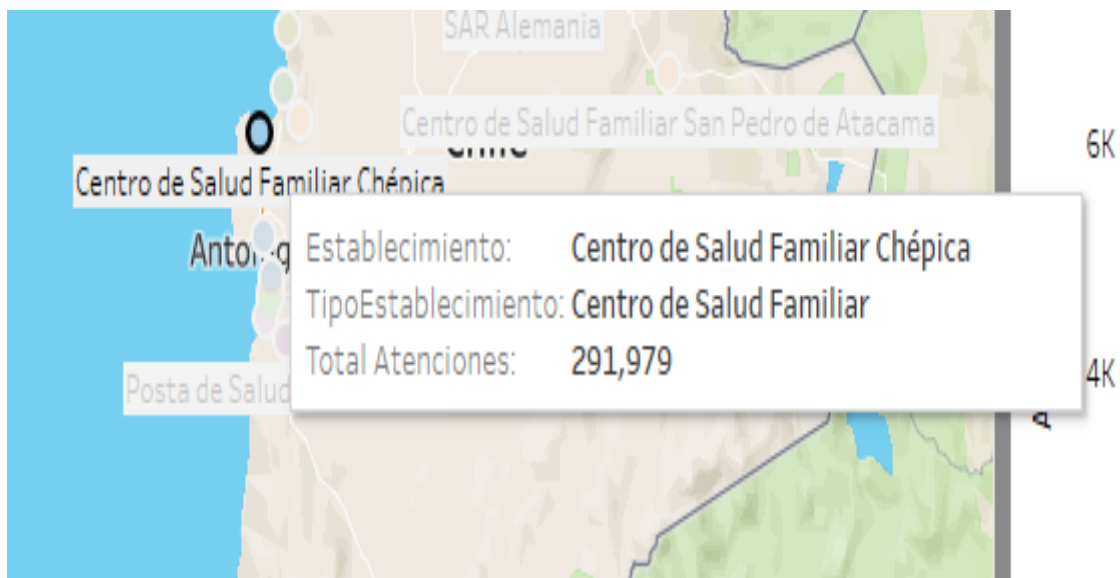


Figura 91 Selección de un establecimientos para realizar filtros

Al seleccionar el establecimiento, la viñeta mencionada mostrará solo información sobre esta selección.

- Campañas

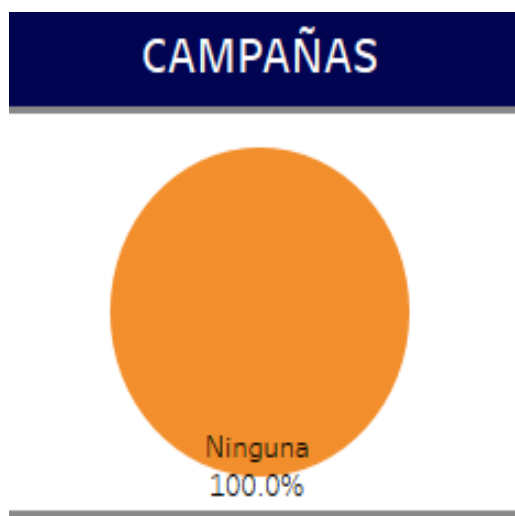


Figura 92 Filtro de campaña del Centro de Salud Familiar de Chépica

- Cantidad de atención a enfermedades según el centro



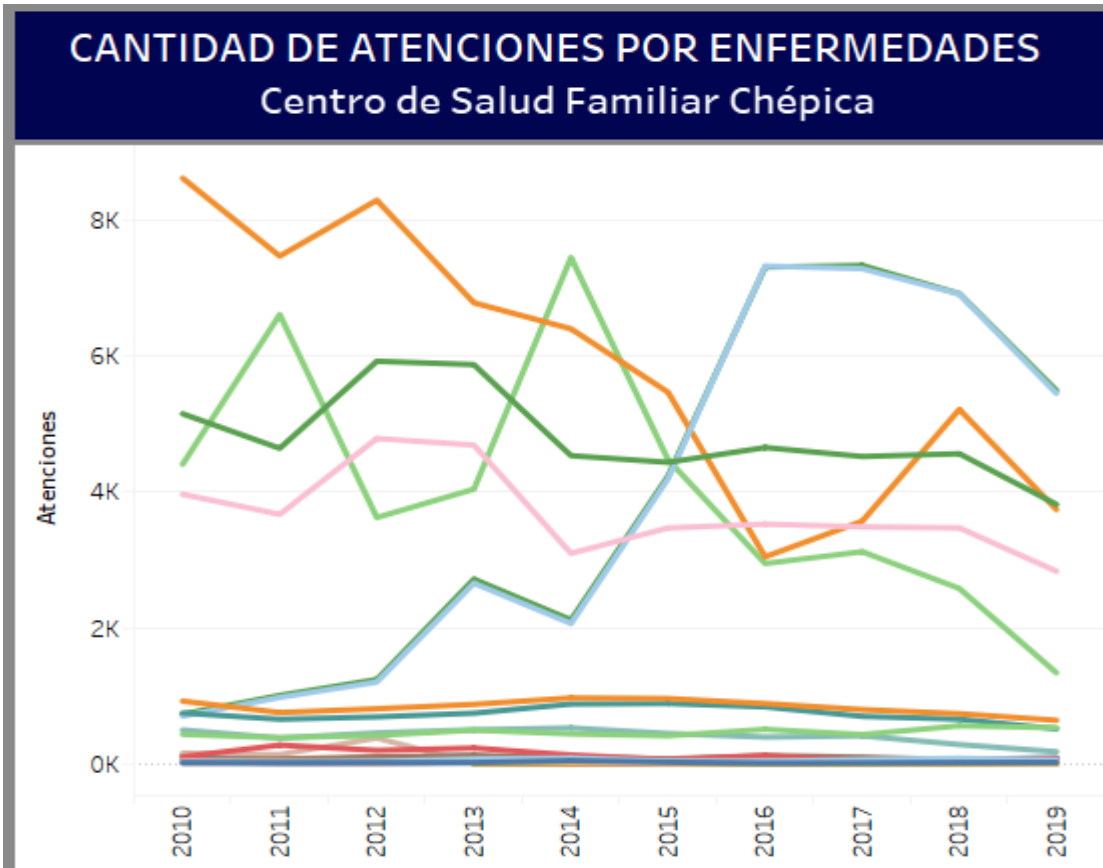


Figura 93 Filtro de atención de enfermedad del Centro de Salud Familiar de Chépica

- Atención de emergencia por rango de edad

### ATENCIONES DE URGENCIA POR RANGO DE EDAD

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Edad 0	800	1,096	841	1,139	818	753	679	630	759	490
Edad 1 a 4	3,421	3,150	3,764	3,831	3,362	2,850	3,170	3,038	2,744	2,432
Edad 5 a 14	4,281	4,483	5,273	5,452	4,640	4,532	4,784	4,856	4,763	4,105
Edad 15 a 64	15,728	15,883	15,689	16,529	16,719	16,914	18,670	18,980	19,733	14,366
Edad 65	2,519	2,560	2,725	3,142	3,588	4,306	4,534	4,534	4,299	3,625

Figura 94 El filtro de atención del Centro de Salud Familiar de Chépica por edad

- Pronóstico de Atenciones de Urgencia



Figura 95 Pronóstico de atención en el centro de salud familiar Chépica

Por lo tanto, nos muestra toda la información mencionada anteriormente, esto se puede hacer con todos los establecimientos, para volver a la información global simplemente desmarcarla.