

**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE
MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y
COMPUTACIÓN**



**APLICACIÓN DE LA TÉCNICA DE CLASIFICACIÓN
DE MINERÍA DE DATOS PARA MEJORAR LOS
PROCESOS DE ATENCIÓN DE CITAS EN EL ÁREA
DE CONSULTORIOS EXTERNOS DE UN HOSPITAL
DEL DEPARTAMENTO DE LAMBAYEQUE**

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

JOSÉ ALONSO POLO CAPUÑAY

Chiclayo 10 de setiembre de 2018

**APLICACIÓN DE LA TÉCNICA DE CLASIFICACIÓN
DE MINERÍA DE DATOS PARA MEJORAR LOS
PROCESOS DE ATENCIÓN DE CITAS EN EL ÁREA
DE CONSULTORIOS EXTERNOS DE UN HOSPITAL
DEL DEPARTAMENTO DE LAMBAYEQUE**

POR:

JOSÉ ALONSO POLO CAPUÑAY

**Presentada a la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

APROBADA POR EL JURADO INTEGRADO POR

**Mgtr. María Ysabel Arangurí García
PRESIDENTE**

**Ing. Héctor Miguel Zelada Valdiviezo
SECRETARIO**

**Ing. Segundo José Castillo Zumarán
ASESOR**

DEDICATORIA

A Dios por haberme regalado la vida, por estar conmigo en todo momento e iluminarme en cada paso que doy en este mundo.

A mis padres que siempre me brindaron su apoyo incondicional, no sólo para desarrollarme profesionalmente sino por preocuparse en que sea mejor persona cada día.

A mis familiares y a las personas que, de alguna u otra forma, influyeron en mí para culminar esta etapa de mi vida e incentivarne a seguir adelante.

AGRADECIMIENTOS

A las autoridades del hospital en estudio y a sus trabajadores, quienes me brindaron su apoyo en todo momento, y que sin ellos no habría sido posible el desarrollo de mi investigación.

A mi asesor, el Ing. Segundo José Castillo Zumarán, por compartir sus conocimientos que facilitaron el desarrollo de la Tesis, así como también a la Ing. María Ysabel Arangurí García, por brindarme las pautas para el desarrollo del trabajo y por su inagotable paciencia.

ÍNDICE

I.	INTRODUCCIÓN	12
II.	MARCO TEÓRICO	15
2.1.	Antecedentes.....	15
2.2.	Bases Teórico Científicas	16
2.2.1.	Bases tecnológica	17
2.2.1.1.	Business Intelligence (Inteligencia de Negocios).....	17
2.2.1.2.	Datamart	18
2.2.1.3.	Extract, Transform and Load – ETL	18
2.2.1.4.	Minería de Datos (Data Mining)	19
2.2.1.5.	Metodologías para proyectos de Data Mining.....	20
2.2.1.6.	Tareas de Minería de Datos	26
2.2.1.7.	Técnicas de Minería de Datos	26
2.2.1.8.	Herramientas de Inteligencia de Negocios	27
III.	MATERIALES Y MÉTODOS	28
3.1.	Diseño de investigación.....	28
3.2.	Población, muestra y muestreo.....	28
3.3.	Hipótesis	28
3.4.	Variables.....	28
3.5.	Indicadores	29
3.6.	Metodología.....	29
IV.	RESULTADOS	30
4.1.	COMPRENSIÓN DEL NEGOCIO.....	30
4.1.1.	Determinar los objetivos del Negocio (Consejo Directivo, 2012)	30
4.1.2.	Evaluación del Negocio.....	30
4.1.3.	Determinar los objetivos de la Minería de Datos	33
4.1.4.	Realizar el Plan del Proyecto.....	33
4.2.	COMPRENSIÓN DE LOS DATOS	34
4.2.1.	Recolectar Datos Iniciales	34
4.2.2.	Descripción de los Datos	35
4.2.3.	Exploración de los Datos.....	38
4.2.4.	Verificar la calidad de los Datos.....	49
4.3.	PREPARACIÓN DE LOS DATOS	52
4.3.1.	Seleccionar los Datos	52
4.3.2.	Limpiar los Datos	53
4.3.3.	Estructurar los Datos	57
4.3.4.	Integrar los Datos	59
4.3.5.	Formateo de los Datos	61
4.4.	MODELADO	63
4.4.1.	Seleccionar técnica de modelado.....	63
4.4.2.	Generar el plan de prueba.....	65
4.4.3.	Construir el modelo	66
4.4.4.	Evaluar el modelo.....	83
4.5.	EVALUACIÓN	83
4.5.1.	Evaluar los resultados.....	84
4.5.2.	Proceso de Revisión	85
4.5.3.	Determinación de futuras fases.....	85

4.6. IMPLEMENTACIÓN	85
4.6.1. Plan de implementación	85
4.6.2. Plan de monitoreo y mantención	86
4.6.3. Informe final	86
4.6.4. Revisión del proyecto	87
V. DISCUSIÓN	88
VI. CONCLUSIONES	90
Recomendaciones	91
VII. REFERENCIAS BIBLIOGRÁFICAS	92
VIII. ANEXOS	95

ÍNDICE DE FIGURAS

Figura 1: Herramientas de BI.....	18
Figura 2: El Proceso de Data Mining.....	19
Figura 3: Comparación de fases de las metodologías SEMMA y CRISP-DM.....	20
Figura 4: Fases de la Metodología CRISP-DM.....	21
Figura 5: Técnicas de Minería de Datos.....	26
Figura 6: Herramientas para Inteligencia de Negocios.....	27
Figura 7: Interfaz del Sistema de Gestión Hospitalaria.....	31
Figura 8: Interfaz del Sistema de Referencia y Contrarreferencia.....	32
Figura 9: Interfaz del Sistema de Acreditación.....	32
Figura 10: Total de atenciones de los diez servicios con más cantidad de citas.....	39
Figura 11: Total de citas anuales por indicador de atención.....	40
Figura 12: Total de asegurados por tipo de asegurado y sexo.....	44
Figura 13: Cantidad de asegurados por tipo de asegurado y policlínico.....	45
Figura 14: Cantidad de asegurados por tipo de asegurado y rango de edades.....	46
Figura 15: Total de asegurados por tipo de asegurado y ubigeo.....	47
Figura 16: Selección de datos de la tabla Adcitas.....	52
Figura 17: Estableciendo la conexión ODBC.....	53
Figura 18: Diseño de DatamartDeserción.....	54
Figura 19: Origen ADO Net – Consulta Citados.....	55
Figura 20: División Condicional – CMP diferentes a “0”.....	56
Figura 21: Columna derivada de tabla Adcitas.....	57
Figura 22: Conversión de tipos de datos de campos de la tabla adcitas.....	58
Figura 23: Columnas de entrada y salida del componente de script.....	58
Figura 24: Sentencias para reemplazar valores.....	59
Figura 25: Generación de la columna llamada Edad.....	60
Figura 26: Combinación de datos de la tabla de citas y de dimensión médico.....	60
Figura 27: Ordenar datos por CMP.....	61
Figura 28: Eliminando caracteres especiales.....	62
Figura 29: Flujo de datos – HechoAdmisión.....	62
Figura 30: Flujo de control del ETL – Poblando Hechos.....	63
Figura 31: Gráfico de Precisión de Minería de Datos - Estructura_Con.....	64
Figura 32: Matriz de Clasificación – Estructura_Con.....	64
Figura 33: Gráfico de Precisión de Minería de Datos - Estructura_Proc.....	64
Figura 34: Matriz de Clasificación – Estructura_Proc.....	64
Figura 35: Dashboard de Deserción.....	65
Figura 36: Vista de Origen de Datos.....	68
Figura 37: Estructura de Minería de Datos – Estructura_Con.....	69
Figura 38: Estructura de Minería de Datos – Estructura_Proc.....	69
Figura 39: Ramificaciones del árbol de decisión – Actividad Consulta Ambulatoria.....	70
Figura 40: Variable más influyente – Actividad Consulta Ambulatoria.....	70
Figura 41: Detalle del nodo más cercano o influyente – Actividad Consulta Ambulatoria.....	71
Figura 42: Ramificaciones del árbol de decisión – Actividad Procedimientos.....	71
Figura 43: Variable más influyente – Actividad Procedimientos.....	72
Figura 44: Detalle del nodo más cercano o influyente – Actividad Procedimientos.....	72
Figura 45: Elaboración de la consulta DMX – Actividad Consulta Ambulatoria.....	73
Figura 46: Relación de pacientes con probabilidad de deserción – Actividad Consulta Ambulatoria.....	73
Figura 47: Elaboración de la consulta DMX – Actividad Procedimientos.....	74
Figura 48: Relación de pacientes con probabilidad de deserción – Actividad Procedimientos.....	74
Figura 49: Flujo de Control - Tareas de Consulta de minería de datos.....	75
Figura 50: Selección del modelo de minería de datos – Actividad Procedimientos.....	75
Figura 51: Consulta DMX – Actividad Procedimientos.....	77
Figura 52: Salida del resultado de la consulta DMX – Actividad Procedimientos.....	77

Figura 53: Selección del modelo de minería de datos – Actividad Consulta Ambulatoria	78
Figura 54: Consulta DMX – Actividad Consulta Ambulatoria	80
Figura 55: Salida del resultado de la consulta DMX – Actividad Consulta Ambulatoria.....	80
Figura 56: Flujo de datos “Resultado Final”	82
Figura 57: Flujo de Control Final.....	83
Figura 58: Dashboard Presupuestal.....	84
Figura 59: Hoja de cálculo – Pacientes con probabilidad de deserción Consulta Ambulatoria.....	86
Figura 60: Hoja de cálculo – Pacientes con probabilidad de deserción Procedimientos	86
Figura 61: Porcentaje de desertores a la consulta externa	95
Figura 62: Porcentaje de desertores por sexo.....	96
Figura 63: Porcentaje de desertores por rango de edades.	97
Figura 64: Porcentaje de desertores por tipo de asegurado.....	98
Figura 65: Porcentaje de desertores por procedencia	99
Figura 66: Porcentaje de motivos de deserción	100

ÍNDICE DE TABLAS

Tabla 1: Cálculo de la Tasa de Deserción a Consulta Médica	16
Tabla 2: Tareas de Data Mining	26
Tabla 3: Tabla de Indicadores	29
Tabla 4: Descripción de la tabla DBF Adctivi	35
Tabla 5: Descripción de la tabla DBF Adcitas	35
Tabla 6: Descripción de la tabla DBF Admedico	35
Tabla 7: Descripción de la tabla DBF Adservic	36
Tabla 8: Descripción de la tabla DBF Adtaseg	36
Tabla 9: Descripción de la tabla DBF Asegur	36
Tabla 10: Descripción de la tabla DBF Adturno	37
Tabla 11: Descripción de la tabla DBF Policlin	37
Tabla 12: Descripción del archivo Ubigeo.xls	37
Tabla 13: Lista de actividades	38
Tabla 14: Diez primeros registros de pacientes citados	38
Tabla 15: Presupuesto anual para citas	40
Tabla 16: Lista del personal asistencial con sus colegiaturas	41
Tabla 17: Lista de servicios	42
Tabla 18: Lista de tipos de asegurado	43
Tabla 19: Lista de turnos	48
Tabla 20: Lista de Policlínicos	48
Tabla 21: Lista de Ubigeo	49
Tabla 22: Códigos de servicios erróneos o nulos	49
Tabla 23: Códigos de turnos erróneos o nulos	49
Tabla 24: Fechas fuera de rango o nulos	50
Tabla 25: CMP y nombre del personal asistencial erróneos, nulos o repetidos	50
Tabla 26: Actos médicos duplicado y nulos	51
Tabla 27: Datos nulos en estado de cita	51
Tabla 28: Datos nulos o erróneos en precio	51
Tabla 29: Datos nulos o erróneos en el campo código de actividad	52
Tabla 30: Descripción de Dimensiones y Hechos	53
Tabla 31: Índice de deserción antes de minería de datos	88
Tabla 32: Índice de deserción después de aplicar minería de datos	88
Tabla 33: Presupuesto desaprovechado para la Consulta Externa antes de la minería de datos	89
Tabla 34: Total de presupuesto desaprovechado	101

RESUMEN

La creciente demanda de pacientes que acuden al hospital en estudio da origen a que el diferimiento de citas sea más prolongado, lo que dificulta cumplir con uno de los objetivos institucionales, el cual es brindar una atención oportuna a los pacientes. A su vez y en consecuencia, no se aprovecha el presupuesto asignado; esto deriva en que la tasa de deserción a la consulta supere el estándar que es 5%.

Ante esta situación, se realiza el presente trabajo para demostrar que, aplicando técnicas de minería de datos se puede determinar la probabilidad de deserción a la consulta médica, lo que repercutirá en una mejor gestión de asignación y/o aprovechamiento en el otorgamiento de citas; siendo justificado tecnológicamente, pues se haría uso de algoritmos de minería de datos para demostrar la probabilidad de deserción; económicamente, se demostraría que existe una mejora en el aprovechamiento del presupuesto asignado; socialmente, pues estas mejoras repercutirían positivamente en los pacientes; científico, pues esta investigación quedará como base para futuros estudios.

Para lograr el objetivo propuesto, se debe disminuir el índice de citas no atendidas, disminuir el presupuesto desaprovechado destinado para las atenciones en consultorios externos, y aumentar el índice de pacientes citados producto de la deserción.

La presente investigación es de tipo Tecnológica Aplicada, optando como diseño de contrastación de hipótesis, la de sucesión o llamada también pre test / post test o en Línea, en la que se consideró, como población, la data histórica con la que cuenta la institución desde el año 2010 al año 2016, de la cual se tomará como muestra los pacientes atendidos durante el año 2017; y se hizo uso de encuestas y entrevistas para la recolección de datos.

Como resultado del trabajo se pudo concluir que durante el periodo de prueba, se pudo disminuir 1.00% el índice de deserción en consulta ambulatoria, y un 29.79% en procedimientos. En lo que respecta al presupuesto, se aprovecharía S/.110,670.00 en consulta ambulatoria y S/.1,247,400.00 en procedimientos; asimismo, se determinó que, mediante el uso de la minería de datos, se hubiera aprovechado 2,635 citas en consulta ambulatoria y 29,700 en procedimientos, destinados a pacientes en lista de espera.

PALABRAS CLAVE: diferimiento de cita, consulta externa, deserción, presupuesto, minería de datos.

Abstract

The ever-increasing demand of patients that the hospital receives gives rise to an even more spaced deferral of appointments, which, at the same time, goes against the compliance of one of the institutional objectives, which is to provide due care to patients. As a consequence, the allotted budget is not taken advantage of; this means that the rate of desertion to appointments surpasses the figure of 5%, which is the standard.

In view of this situation, the present study is carried out to demonstrate that, applying data mining techniques, the probability of desertion to the medical consultation can be determined, which will result in a better management of allocation and/or use in the granting of appointments, being technologically justified, since data mining algorithms would be used to demonstrate the probability of desertion; economically, it'd be demonstrated that there is an improvement in the use of the allocated budget; socially, because these improvements would have a positive effect on patients; scientific, because this research will be the basis for future studies

In order to achieve this objective, the rate of non-attended appointments has to decrease; the budget not being fully used that are thought for the external offices also has to decrease; and the rate of non-attending patients, that are product of the desertion, has to increase.

The herein study is of the Applied Technology type, which opts for the succession as the contrast hypothesis design, or so called, pre test/post test, or in line, in which, the population considered for this study lies within the historical data owned by the Institution from the year 2010 until 2016, out of which the sample that was taken is of all patients under medical care during the year 2017. Also, to gather data, polls and interviews were used.

As a result, we conclude that during the trial period, the rate of desertion in ambulatory cares diminished 1.00% and 29.79% in procedures. As for the budget, S/.110,670.00 would be taken advantage of in ambulatory care and S/.1,247,400.00 in procedures. At the same time, it was determined that through data mining 2,635 appointments could have been effectuated and 29,700 in procedures, for patients on the waiting list.

Keywords: appointment deferral, outpatient, desertion, budget, data mining.

I. INTRODUCCIÓN

El derecho a la salud incluye el acceso a una atención sanitaria oportuna, aceptable, asequible y de calidad satisfactoria.¹

El artículo 12 del Pacto Internacional de Derechos Económicos, Sociales y Culturales (1966) establece que, entre las medidas que se deberán adoptar a fin de asegurar la plena efectividad del derecho a la salud, figurarán las necesarias para la creación de condiciones que aseguren el acceso de todos a la atención de la salud. Para ello el Comité de Derechos Económicos, Sociales y Culturales de las Naciones Unidas, que supervisa el cumplimiento del Pacto Internacional de Derechos Económicos, Sociales y Culturales, adoptó en 2000 una Observación general sobre el derecho a la salud; en la que se afirma que el derecho a la salud no sólo abarca la atención de salud oportuna, sino también los factores determinantes de la salud, tales como el acceso a condiciones sanitarias adecuadas.

En definitiva, el derecho a la salud significa que los gobiernos deben crear las condiciones que permitan a todas las personas vivir lo más saludablemente posible y esas condiciones incluyen la disponibilidad garantizada de servicios de salud (OMS, 2013).

La igualdad en la accesibilidad a los servicios de salud establece teóricamente las mismas oportunidades de utilización a sujetos con el mismo nivel de necesidad. En el acceso se incluyen variables como el coste de oportunidad, que implica acudir a los servicios sanitarios, los tiempos y listas de espera, el medio de transporte y la distancia a los centros más cercanos, etc. Esta noción de equidad respeta las diferencias interpersonales de partida y dispone un marco idóneo para la igualdad de oportunidades. (Jabalera, 2015)

Cuando hay problemas de acceso a servicios de salud, las listas de espera para las citas se hacen más largas y esto a su vez conduce a una mayor falta de asistencia (George, 2003)

La inasistencia de los usuarios a las consultas médicas programadas sin realizar la cancelación de las mismas es un problema que afecta de forma directa la calidad en la prestación de los servicios de salud, aumentando los costos sociales y económicos de la enfermedad. La no asistencia a las consultas influye negativamente en la adherencia de los pacientes a los tratamientos médicos indicados, contribuyendo al empeoramiento de su estado de salud y a la cronificación de sus enfermedades de base, esta situación conlleva al colapso de los servicios de urgencias, de los centros de referencia de segundo y tercer nivel, redundando en el deterioro de la salud del paciente y en el incremento de los costos operacionales. (Pérez y Rendón, 2016)

Esta problemática existe en diversos lugares del mundo, según Pereira-Victorio (2016) se determinó que existió un 14.6% de deserción a la consulta médica en un hospital de tercer nivel en España, donde se tuvo como población 1,885 citas nuevas registradas

¹ Organización Mundial de la Salud. 2013.

entre el 01 al 05 de agosto de 2011, siendo las justificaciones más empleadas: 16,1% desconocimiento, 15,6% tener doble cita o acudir en otra fecha, 15,1% cancelar su cita previamente, 12,9% olvidar la cita y 12,4% recuperarse de la enfermedad antes.

Asimismo en Colombia, la IPS Integrados determinó que durante el periodo comprendido entre diciembre del 2013 y diciembre del 2014 estuvo entre el 4%-6% para la consulta especializada básica y entre un 6% y 8% para la consulta médica general, la meta esperada es del 4% y 6% respectivamente. (Pérez y Rendón, 2016)

Mientras que en Chile se registró que el 19% de pacientes no asisten a su consulta médica programada para el año 2015, según el Departamento de Estadística e Informaciones del Ministerio de Salud. (La Tercera, 2017).

En nuestro entorno, las instituciones de la salud, se encuentran comprometidas en desarrollar acciones para lograr una buena atención a los pacientes. El hospital en estudio, no es la excepción; teniendo como uno de los indicadores corporativos la tasa de deserción a consulta médica, cuyo estándar debe mantenerse en $\leq 5\%$. Para el año 2016 este indicador superaba el 30%.

Entre los motivos² de deserción a la consulta médica más frecuente tenemos: el personal de módulo o referencias no hacen conocer o recordar oportunamente a los pacientes su fecha de cita, el desconocimiento de los canales o medios para anular o postergar su cita, fallecimiento del paciente producto del prolongado diferimiento de cita, olvido del paciente, tener dos citas para diferentes especialidades en la misma fecha; problemas como la no atención de pacientes por ubicar tardíamente la historia clínica así como sus análisis y/o placas radiográficas y el no registro de las atenciones por parte del médico en el Sistema de Gestión Hospitalaria, es considerado también como deserción; entre otros. Teniendo como referencia dicho indicador y sabiendo del diferimiento muy amplio que existe de cita en cita, lo que obliga a los pacientes a acudir al Servicio de Emergencia para obtener la atención que necesitan, inclusive, en algunos casos, produciéndose su hospitalización por haberse agravado su salud, nace la necesidad de adoptar medidas preventivas y aprovechar los cupos disponibles que se generarían por tal deserción. El que un paciente deserte a su cita genera un tiempo ocioso promedio de 15 minutos, significando que no se aproveche al máximo la productividad del médico.

Considerando la problemática descrita, nació la interrogante de ¿cómo mejorar los procesos de atención como el registro, reprogramación o eliminación de citas en el área de consultorios externos?, la cual obtuvo su respuesta a través de la aplicación de minería de datos utilizando la técnica de clasificación.

El presente trabajo tuvo como objetivo principal mejorar los procesos de atención en el área de consultorios externos (registro, reprogramación o eliminación de citas), el cual se logrará a través de la disminución del índice de citas no atendidas, la disminución del presupuesto desaprovechado destinado para las atenciones en consultorios externos, y aumentando el índice de pacientes citados producto de la deserción. En torno al objetivo

² Encuesta “Motivos de Deserción a la Consulta Externa”, 2012.

principal, hay bibliografía que detalla la implementación de sistemas hospitalarios, que ayudan a gestionar las atenciones de los pacientes; implementación de sistemas expertos para la detección temprana de cáncer de mama; también aplicaciones de minería de datos para determinar el perfil de pacientes con ciertos diagnósticos y su relación del presupuesto necesario para su atención.

Asimismo, el presente trabajo se encuentra justificado tecnológicamente, pues hace uso de la técnica de clasificación de minería de datos para determinar los probables pacientes desertores; desde el punto económico, permitirá disminuir el desaprovechamiento de dinero presupuestado para pacientes citados por consultorios externos y evitar también el gasto que se genera al atender pacientes por otros servicios, como Emergencia y Hospitalización; socialmente, recae en el interés continuo que prestaría la institución al estar pendiente del perfil de los pacientes que no acuden a sus citas y tomar las medidas correspondientes para mejorar la calidad de atención, con la finalidad de detectar falencias en el personal de la institución de ser el caso, y así evaluar sus procedimientos y calidad de atención, aumentando así su satisfacción por contribuir a que los pacientes mejoren y lleven una mejor calidad de vida y de salud; científicamente, pues el desarrollo del presente trabajo servirá de guía o base para futuros proyectos de investigación con la finalidad de seguir mejorando la atención de los pacientes.

II. MARCO TEÓRICO

En el presente capítulo se exponen los antecedentes considerados para la investigación; así como las bases teóricas científicas.

2.1. Antecedentes

Las técnicas de minería de datos en el sector salud no ha sido muy aplicada; sin embargo si existen trabajos sobre implementaciones de sistemas transaccionales para la mejora de procesos en diversas instituciones de salud.

Teniendo en cuenta lo antes mencionado, se consideraron los siguientes antecedentes; los cuales se acercan a lo propuesto en el presente trabajo.

Candela (2015), en su trabajo plantea como objetivo mostrar el proceso automatizado de cada etapa del proceso de descubrimiento con el fin de predecir el abandono en los tratamientos de cáncer de una entidad de salud pública con una precisión eficiente basándose en características o factores determinados en la etapa de análisis junto con los miembros de la institución. El trabajo es un claro ejemplo de la implementación de un Datamart y aplicación de técnicas de minería de datos para predecir el abandono en los tratamientos de cáncer, por lo que aportó mucho con respecto a la manera en cómo desarrollar o considerar en el presente trabajo.

Sevilla (2014), realizó un trabajo para mejorar el itinerario de atención a pacientes con problemas de salud mental, así como determinar el coste en que se incurriría en la atención de los mismos; clasificándolos por perfil que tenían como variables sociodemográficas, relacionadas con trastornos psiquiátricos, variables relacionadas con los servicios sanitarios, etc. Para la aplicación de minería de datos utilizaron la metodología SEMMA, donde se determinó por ejemplo el prototipo de paciente atendido, trastornos más comunes, se tuvo como resultado también que las variables sociodemográficas no se tuvieron en cuenta para clasificar a los pacientes en los diferentes niveles de coste y solamente son relevantes las variables clínicas; entre otros resultados. Este antecedente ayudó a conocer o identificar las diferentes metodologías que existen para desarrollar proyecto de minería de datos.

Villanueva (2008), en su proyecto implementa un DataWarehouse donde se aplica sólo las nociones básicas de construcción: Diseño de BD, Diseño de Repositorio de metadata, Diseño ETL, Desarrollo ETL, Desarrollo de la aplicación e implementación. Se implementó una plataforma de BI para explotar el DataWarehouse en múltiples cubos para mostrar información importante a las autoridades pertinentes a través de informes finales. Se consideró este trabajo como antecedente por amplia terminología y el detalle de los procesos ETL y de la construcción del Datamart, el cual será la base para aplicar las técnicas de minería de datos y obtener el resultado esperado..

2.2. Bases Teórico Científicas

En el presente apartado, se considera las bases teóricas científicas que sustentan el presente trabajo; por un lado de la institución y por el otro de la tecnología o herramienta utilizada.

Para entender el contexto de la problemática abordada es necesario saber que la consulta médica es la actividad prestacional que incluye evaluación integral general, diagnóstico, prescripción, tratamiento, monitoreo, seguimiento, orientación de pacientes con condiciones clínicas y/o exposición a riesgos, que se brinda en el centro asistencial, la cual tiene un tiempo aproximada de duración de 15 minutos.³

Asimismo existen procesos de atención que permiten gestionar la consulta médica, los cuales son un conjunto de actividades relacionadas entre sí, cuya finalidad es brindar una atención adecuada al paciente para solicitar, reprogramar o anular una consulta médica, en adelante cita.

- Solicitar cita: proceso en el que el paciente solicita una consulta médica, a través de los diferentes medios que la institución brinda.
- Reprogramación de cita: proceso en el cual, el paciente, solicita que se re programe su consulta médica ya programadas, por distintos motivos.
- Anulación de citas: cuando el paciente determina que le será imposible asistir a su consulta médica, debe solicitar la anulación de su cita oportunamente, para así ofrecer su cupo a un paciente que se encuentra en espera⁴

Los pacientes al no asistir a su cita generan o incrementan la tasa de deserción, la cual permite evaluar la deserción de los pacientes en los servicios y/o especialidades con la finalidad de adoptar las medidas preventivas o correctivas. (Gerencia Central de Salud – EsSalud, 2002)

El método de cálculo de la tasa de deserción es el siguiente:

Tabla 1: Cálculo de la Tasa de Deserción a Consulta Médica.

<i>FORMULA DE CALCULO</i>	<i>STD</i>
$\frac{\text{Número de pacientes que no asistieron a Consulta médica en un determinado período de tiempo} \times 100}{\text{Total de pacientes citados a consulta médica para en el mismo período}}$	5%

Fuente: (Gerencia Central de Salud – EsSalud, 2002)

³ Seguro Social de Salud. 2013. “Glosario de Términos de la Seguridad Social”: 36, (último acceso 01 de junio de 2018).

⁴ Gerencia Central de Planificación y Desarrollo - EsSalud. 2012. “Glosario de términos normativos institucional”: 37, (último acceso 01 de abril de 2018).

Donde:

- Pacientes que no asistieron a consulta médica, es el número de pacientes citados para consulta médica que no concurrieron a su cita (deserción).
- Pacientes citados, se refiere al número de pacientes que recibieron una boleta de cita para determinada consulta médica. Se incluye las citas telefónicas.
- Período de Medición: Mensual

2.2.1. Bases tecnológica

Puesto que se trata de un trabajo de Inteligencia de Negocios (Business Intelligence – BI), es necesario comprender lo que es propiamente BI y todo lo que ésta implica, puesto van de la mano con el presente trabajo.

2.2.1.1. Business Intelligence (Inteligencia de Negocios)

Vieira (2009), nos dice que BI es un conjunto de conceptos y metodologías que, haciendo uso de acontecimientos (hechos) y sistemas sustentados en los mismos, apoya la toma de decisiones en los negocios. Para esto ser posible es necesario adquirir los datos, por ejemplo, por medio de un sistema de procesamiento on-line de transacciones (OLTP), almacenarlos en un sistema de base de datos, como un Data Warehouse del cual se puede generar aún un subconjunto más específico de datos, Datamart y, finalmente procesar estos datos con una herramienta de análisis que puede ser: una herramienta de procesamiento analítico on-line (OLAP), un sistema de informaciones para ejecutivos (EIS); un sistema de apoyo a la decisión (DSS); o aun, un sistema de descubierta y predicción (Data mining).

Para Curto y Conesa (2010) , Business Intelligence es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización.

Stair y Reynolds (2000) manifiestan que la inteligencia de negocios es un proceso que implica recopilar información correcta, suficiente, oportuna y utilizable, y analizarla para que tenga efecto positivo en la estrategia, tácticas y operaciones del negocio.

Considerando las definiciones descritas anteriormente, se puede concluir que Business Intelligence permite transformar:

- Convertir datos en información.
- Tomar mejores decisiones rápidamente que ayuden a mejorar el rendimiento de una compañía o institución y que además que mejore su ventaja competitiva (objetivo primario del BI)

En la siguiente figura se puede observar, según Viera et al (2009), las herramientas de Business Intelligence.

Tipo de Herramienta	Cuestión básica	Ejemplo de respuesta
Data Mining	¿Qué es interesante? ¿Qué puede suceder?	Tipos de clientes Predicción de ventas
OLAP	¿Qué sucedió y por qué?	Ventas mensuales versus variaciones de precios de los competidores
EIS/DSS	¿Qué necesito saber ahora ?	Cotizaciones diversas
Estudios y informes	¿Qué sucedió ?	Ventas del último mes

Figura 1: Herramientas de BI
Fuente: (Vieira et al, 2009)

2.2.1.2. Datamart

Según Date (2001) un Datamart es un almacén de datos especializado, orientado a un tema, integrado, volátil y variante en el tiempo para apoyar un subconjunto específico de decisiones de administración, asimismo, el Datamart es especializado y volátil. Por especializado queremos decir que contiene datos para dar apoyo (solamente) a un área específica de análisis de negocios; por volátil queremos decir que los usuarios pueden actualizar los datos e incluso, posiblemente, crear nuevos datos (es decir, nuevas tablas) para algún propósito.

Stair y Reynolds (2000) nos dicen que los Datamarts contienen un subgrupo de datos relacionados con un solo aspecto de los negocios de una compañía, por ejemplo finanzas, inventario o personal son de mucha utilidad para grupos pequeños que deseen acceder a datos detallados.

Considerando que se abordará la problemática relacionada a la gestión de citas médicas, se ha construido un Datamart que permita determinar el perfil de los pacientes desertores utilizando técnicas de minería de datos.

2.2.1.3. Extract, Transform and Load – ETL

ETL es una tecnología que permite extraer datos del entorno de origen, transformarlos según nuestras necesidades de negocio para integración de datos y cargar estos datos en los entornos destino. Los entornos de origen y destino son usualmente bases de datos, ficheros, u otras fuentes estructuradas, semi estructurada o no estructurada. (Conesa y Curto 2010)

Este proceso es necesario para realizar la integración de los datos de las diferentes fuentes existentes en el hospital en estudio. Cabe recalcar que estos datos deben pasar un proceso de transformación que garantice su calidad y con ello obtener un buen perfilamiento de los pacientes desertores.

2.2.1.4. Minería de Datos (Data Mining)

Según Rodríguez et al (2003), los orígenes de Data Mining se pueden establecer a finales de los años 80, cuando la administración de hacienda estadounidense, desarrolló un programa de investigación para detectar fraudes en las declaraciones de impuestos. Sin embargo, la gran expansión de Data Mining no se produce hasta la década de 1990 originada principalmente por tres factores:

- Incremento de la potencia de los computadores.
- Incremento del ritmo de adquisición de datos, debido al abaratamiento del hardware y a la automatización de los procesos de recolección de datos.
- Aparición de nuevas técnicas de aprendizaje y almacenamiento de datos.

Ferrer et al (2000) hace mención a la definición de minería de datos que hace (Piatetsky-Shapiro et al, 1992); donde la define como un conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar conocimiento implícito (previamente desconocido), potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir de forma automatizada tendencias y comportamientos y/o describir de forma automatizada modelos previamente desconocidos.

La minería de datos se sitúa en los campos de la estadística, la tecnología de base de datos, patrón de reconocimiento, visualización de datos, entre otros.

El objetivo de la minería de datos en el presenta trabajo radica en determinar el perfil de los pacientes desertores a la consulta médica.

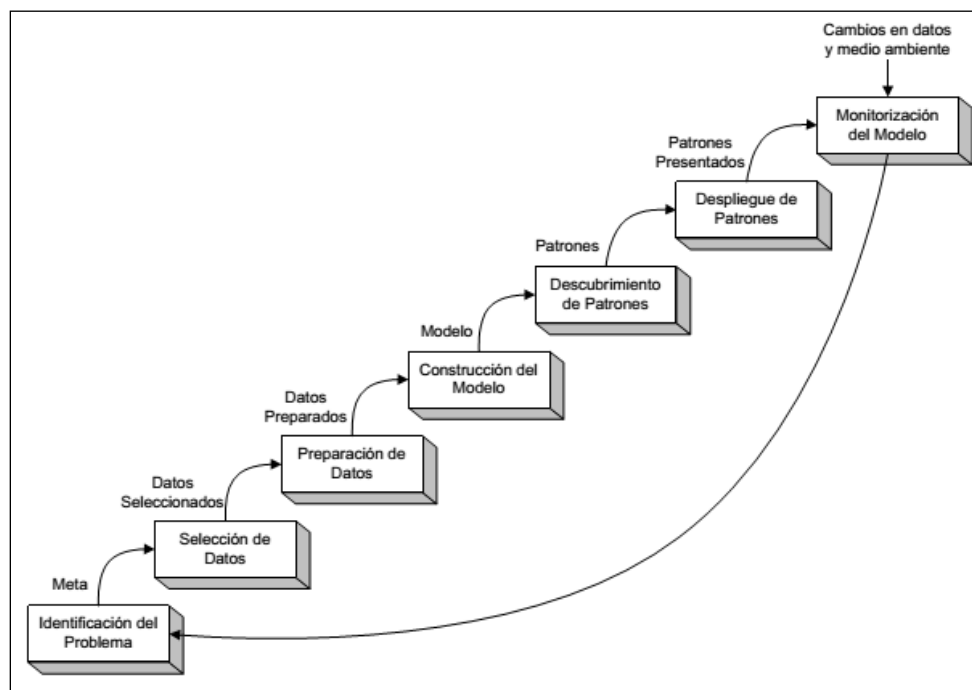


Figura 2: El Proceso de Data Mining

Fuente: (Caniupan, 2000)

Mediante la Figura 2, podemos observar el proceso de minería de datos, como un proceso independiente de extracción y exploración de información oculta.

Sin embargo existen diversas metodologías para los proyectos de minería de datos.

En el siguiente apartado se mencionará algunas metodologías existentes para proyectos de minería de datos.

2.2.1.5. Metodologías para proyectos de Data Mining

Son diversos los modelos de proceso que han sido propuestos para el desarrollo de proyectos de Data Mining tales como:

- SEMMA (Sample, Explore, Modify, Model, Assess) (SAS, 2003)
- DMAMC (Definir, Medir, Analizar, Mejorar, Controlar) (Isixsigma, 2005)
- CRISP-DM (Cross Industry Standard Process for Data Mining) (Chapman et al., 2000)

Básicamente las metodologías SEMMA y CRISP-DM comparten la misma naturaleza y el mismo objetivo ya que se busca solucionar un problema que permanece en la empresa y que los resultados están ocultos dentro toda la información; sin embargo se decidió por utilizar la metodología CRISP-DM, pues adicionalmente incorpora un análisis previo del problema existe en el negocio, tal como se muestra en la siguiente figura.

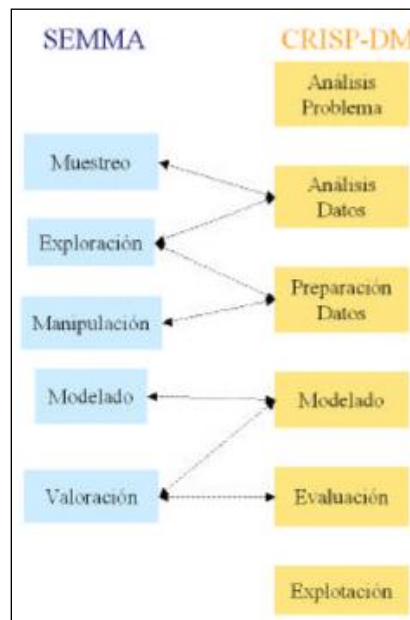


Figura 3: Comparación de fases de las metodologías SEMMA y CRISP-DM

Fuente: (Rodríguez et al., 2003)

- **Metodología CRISP-DM**

Los orígenes de CRISP-DM se remontan hacia el año 1999, cuando un importante consorcio conformado por empresas europeas tales como NCR (Dinamarca), AG (Alemania), SPSS (Inglaterra), OHRA (Holanda), Teradata, SPSS, y Daimler-Chrysler, proponen a partir de diferentes versiones de KDD (Knowledge Discovery in Databases) [Reinartz, 1995], [Adraans, 1996], [Brachman, 1996], [Fayyad, 1996], el desarrollo de una guía de referencia de libre distribución denominada CRISP-DM (Cross Industry Standard Process for Data Mining). (Gallardo, 2009)

Por ser una metodología de libre distribución puede trabajar con cualquier herramienta para desarrollar el proyecto que esté disponible en el mercado aplicando así una característica adicional que es el de ser una metodología equitativa.

Así se distingue esta metodología como un modelo que está basado en situaciones reales que ocurren en la empresa, organizando sus fases principales y estableciendo tareas secundarias que especifican con más detalle las obligaciones de cada una de las seis tareas.

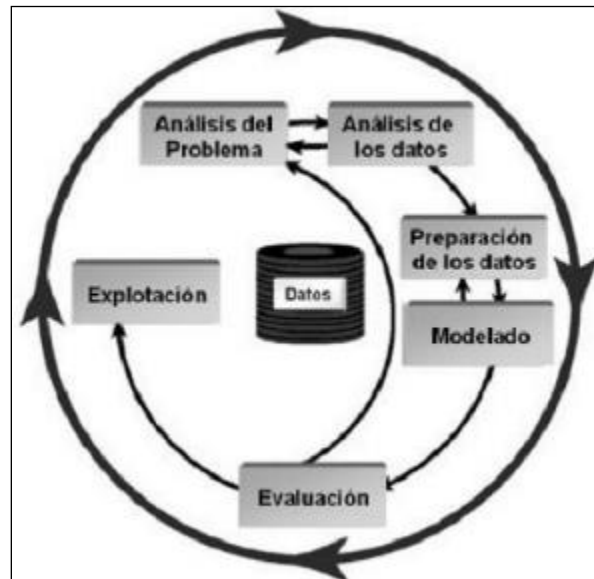


Figura 4: Fases de la Metodología CRISP-DM
Fuente: (Chapman et al., 2000)

A continuación se detalla las fases de la metodología CRISP-DM, según Gallardo (2009):

Análisis del Problema o Comprensión del negocio

Esta fase se enfoca en la comprensión de los objetivos del proyecto y los requerimientos desde la perspectiva del negocio. Esta información se convierte en conocimiento para la definición del problema y el diseño del plan para alcanzar los objetivos.

Está compuesta por las siguientes tareas o actividades:

- Determinar los objetivos del negocio: tiene como metas, determinar cuál es el problema que se desea resolver, por qué la necesidad de utilizar Data Mining y definir los criterios de éxito
- Evaluación o valoración de la situación: se definen los requisitos del problema, tanto en términos de negocio como en términos de Data Mining.
- Determinar los objetivos de minería de datos: tiene como objetivo representar los objetivos del negocio en términos de las metas del proyecto de DM, como por ejemplo, si el objetivo del negocio es el desarrollo de una campaña publicitaria para incrementar la asignación de créditos hipotecarios, la meta de DM será por ejemplo, determinar el perfil de los clientes respecto de su capacidad de endeudamiento.
- Realizar el plan del proyecto: tiene como meta desarrollar un plan para el proyecto, que describa los pasos a seguir y las técnicas a emplear en cada paso.

Análisis de los Datos o Comprensión de los datos

Esta fase comienza con una colección de datos y sobre la cual se realizan actividades para familiarizarse con ellos, para identificar los problemas de calidad y detectar subconjuntos interesantes para formar hipótesis sobre la información escondida.

Está compuesta por las siguientes actividades:

- Recolectar datos iniciales: tiene como objetivo, elaborar informes con una lista de los datos adquiridos, su localización, las técnicas utilizadas en su recolección y los problemas y soluciones inherentes a este proceso.
- Descripción de los datos: después de adquiridos los datos iniciales, estos deben ser descritos. Este proceso involucra establecer volúmenes de datos (número de registros y campos por registro), su identificación, el significado de cada campo y la descripción del formato inicial.
- Exploración de los datos: el fin es encontrar una estructura general para los datos. Esto involucra la aplicación de pruebas estadísticas básicas, que revelen propiedades en los datos recién adquiridos, se

crean tablas de frecuencia y se construyen gráficos de distribución.

- Verificar la calidad de los datos: se efectúan verificaciones sobre los datos, para determinar la consistencia de los valores individuales de los campos, la cantidad y distribución de los valores nulos, y para encontrar valores fuera de rango, los cuales pueden constituirse en ruido para el proceso. La idea en este punto, es asegurar la completitud y corrección de los datos.

Preparación de los datos

Esta fase involucra todas las actividades para construir el conjunto final de datos a partir del conjunto inicial. Estas actividades se llevan a cabo varias veces y no en un orden predefinido. Entre estas tareas se encuentran la selección de atributos así como la transformación y limpieza de los datos para las herramientas del modelado.

Para cumplir con esta fase, se realiza las siguientes actividades:

- Seleccionar los datos: se selecciona un subconjunto de los datos adquiridos en la fase anterior, apoyándose en criterios previamente establecidos en las fases anteriores.
- Limpiar los datos: Esta tarea complementa a la anterior, y es una de las que más tiempo y esfuerzo consume, debido a la diversidad de técnicas que pueden aplicarse para optimizar la calidad de los datos a objeto de prepararlos para la fase de modelación.
- Estructurar los datos: esta tarea incluye las operaciones de preparación de los datos tales como la generación de nuevos atributos a partir de atributos ya existentes, integración de nuevos registros o transformación de valores para atributos existentes.
- Integrar los datos: involucra la creación de nuevas estructuras, a partir de los datos seleccionados, por ejemplo, generación de nuevos campos a partir de otros existentes, creación de nuevos registros, fusión de tablas que contengan atributos diferentes para un mismo objeto, agregación de nuevos campos o nuevas tablas donde se resumen características de múltiples registros o de otros campos en nuevas tablas de resumen.
- Formateo de los datos: consiste principalmente, en la realización de transformaciones sintácticas de los datos sin modificar su significado, esto, con la idea de permitir o facilitar el empleo de alguna técnica de DM en particular, como por ejemplo la reordenación de los campos, etc.

Modelado

En esta fase se seleccionan y aplican varias técnicas de modelado o algoritmos y sus parámetros son calibrados para obtener los mejores resultados. Algunos algoritmos tienen requerimientos específicos para el formato de los datos, en cuyo caso se debe regresar a la fase de preparación de datos y realizar las tareas necesarias para obtener dicho formato, las veces que sean necesarias.

Las tareas que se deben realizar para cumplir con esta fase son:

- **Seleccionar la técnica de modelado:** consiste en la selección de la técnica de DM más apropiada al tipo de problema a resolver. Para esta selección, se debe considerar el objetivo principal del proyecto y la relación con las herramientas de DM existentes. Por ejemplo, si el problema es de clasificación, se podrá elegir de entre árboles de decisión, k-nearest neighbour o razonamiento basado en casos (CBR); si el problema es de predicción, análisis de regresión, redes neuronales; o si el problema es de segmentación, redes neuronales, técnicas de visualización, etc.
- **Generar el plan de prueba:** se debe generar un procedimiento destinado a probar la calidad y validez del mismo. Por ejemplo, en una tarea supervisada de DM como la clasificación, es común usar la razón de error como medida de la calidad. Entonces, típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba.
- **Construir el modelo:** se ejecuta sobre los datos previamente preparados para generar uno o más modelos.
- **Evaluar el modelo:** se interpretan los modelos de acuerdo al conocimiento preexistente del dominio y los criterios de éxito preestablecidos.

Evaluación

En esta fase se revisan los pasos ejecutados para construirlo y estar seguros de que se han alcanzado los objetivos iniciales. También es importante verificar si existe algún punto del negocio que no se haya considerado antes.

Las tareas a realizar son:

- **Evaluar los resultados:** involucra la evaluación del modelo en relación a los objetivos del negocio y busca determinar si hay alguna razón de negocio para la cual, el modelo sea deficiente, o si es aconsejable probar el modelo, en un problema real si el tiempo y restricciones lo permiten.
- **Proceso de revisión:** se refiere a calificar al proceso entero de DM, a objeto de identificar elementos que pudieran ser mejorados.

- Determinación de futuras fases: si se ha determinado que las fases hasta este momento han generado resultados satisfactorios, podría pasarse a la fase siguiente, en caso contrario podría decidirse por otra iteración desde la fase de preparación de datos o de modelación con otros parámetros.

Explotación o Implementación

En esta no se acaba el proyecto, puesto que, se deben hacer pruebas de su implementación así como un seguimiento periódico.

Los resultados que presentan deben ser claros de tal manera que se puedan comprender fácilmente por personas que no tengan mucha intervención en el área de informática, sino también de profesionales o no que pertenezcan a otras áreas, presentando resultados de una manera visual usando cuadros estadísticos que representen las variables usadas y su aporte para la empresa.

- Plan de implementación: esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación.
- Plan de monitoreo y mantenimiento: es aconsejable preparar estrategias de monitorización y mantenimiento para ser aplicadas sobre los modelos.
- Informe final: es la conclusión del proyecto de DM realizado.
- Revisión del proyecto: se evalúa qué fue lo correcto y qué lo incorrecto, qué es lo que se hizo bien y qué es lo que se requiere mejorar.

2.2.1.6. Tareas de Minería de Datos

Una tarea es un tipo de problema de data mining. Por ejemplo: clasificar piezas en defectuosas, no defectuosas, defectuosas reparables y defectuosas no reparables es una tarea. Esta tarea se podría resolver mediante árboles de decisión o redes neuronales, entre otros métodos, éstos, son métodos o técnicas que permiten resolver las tareas (Hernández, 2004).

Existen dos grandes grupos en los que se pueden dividir las tareas, tal como podemos observar en la siguiente figura:

Tabla 2: Tareas de Data Mining

Tipo de Tareas	Descripción	Ejemplos
Predictivas	Se trata de problemas en los que hay que predecir uno o más valores para uno o más ejemplos	Clasificación.
		Clasificación suave.
		Estimación de probabilidad de clasificación.
		Categorización.
		Preferencias o priorización.
Descriptivas	Los ejemplos se presentan como un conjunto de datos sin ordenar ni etiquetar de ninguna manera. Por lo tanto, el objetivo, no es predecir nuevos datos sino describir los existentes	Regresión
		Clustering.
		Correlación y factorizaciones
		Reglas de asociación.
		Dependencias funcionales.
		Detección de valores e instancias anómalas.

Fuente: (González, 2007)

2.2.1.7. Técnicas de Minería de Datos

Las técnicas de la minería de datos provienen de la Inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener resultados. (Belinchón, 2010).

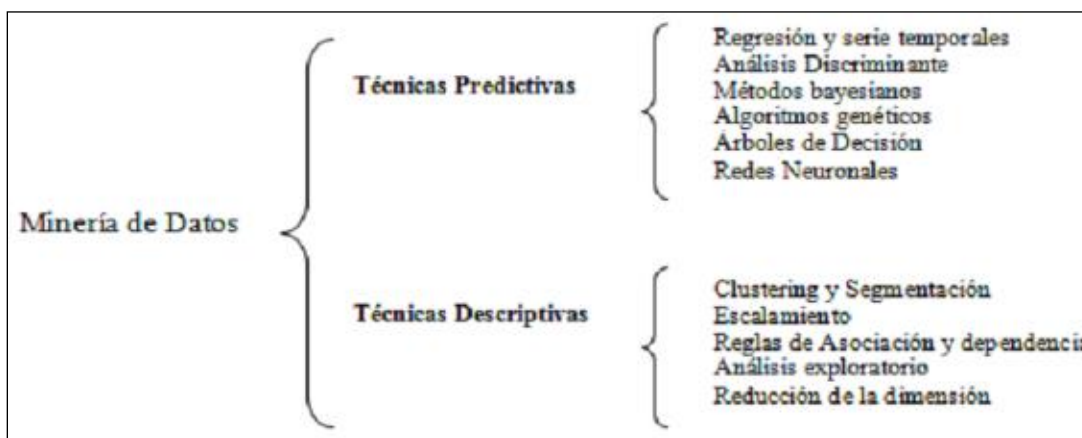


Figura 5: Técnicas de Minería de Datos

Fuente: (Curso, Sin fecha)

2.2.1.8. Herramientas de Inteligencia de Negocios

Para el desarrollo de aplicación de inteligencia de negocios, en las que incluyen el proceso de minería de datos, se cuenta con una gran diversidad de herramientas, tal como se puede visualizar en la siguiente figura.



Figura 6: Herramientas para Inteligencia de Negocios
Fuente: (González, 2007, capítulo 3: 16)

Debido a que la institución en estudio cuenta con licencia de Microsoft SQL Sever y a que se cuenta con profesionales capacitados en el uso de ese gestor, se decidió utilizarla, la cual brinda una plataforma completa para el análisis de datos e inteligencia empresarial, tales como Microsoft SSIS (SQL Server Integration Services), que permite integrar datos de diferentes orígenes, construye administra y despliega soluciones de integración escalables y construye o alimenta un datawarehouse.

Asimismo se hace uso del servicio Microsoft SSAS (SQL Server Analysis Services), que provee una vista unificada e integrada para todos los datos de la empresa, reportes, análisis OLAP, marcadores de indicadores de desempeño y data mining.

III. MATERIALES Y MÉTODOS

En este capítulo se desarrolla el diseño de investigación, se detalla la población, muestra y muestreo, se plantea la hipótesis, variables e indicadores, y por último la se menciona la metodología elegida para el desarrollo del proyecto.

3.1 Diseño de investigación

Tipo de estudio

La presente investigación es de tipo Tecnológica Aplicada, puesto que, se ha generado conocimiento relevante a raíz de la aplicación de técnicas de minería de datos, cuyo diseño de contrastación es diseño de sucesión o en línea, donde se considerará los reportes y encuestas emitidas por la Oficina de Gestión de Calidad y Control Interno.

Diseño de contrastación de hipótesis

El diseño de contrastación de hipótesis es de sucesión o llamada también pre test / post test o en Línea, donde para determinar el perfil de los pacientes desertores, se consideró aquellas citas comprendidas entre los años 2010 y 2016. Una vez establecido el perfil de deserción, éste se contrastará con las características de las citas del año 2017 que será el grupo experimental.

3.2. Población, muestra y muestreo

La población que se consideró fue data histórica con la que cuenta la institución desde el año 2010 al año 2016, y como muestra, se consideraron los pacientes atendidos durante el año 2017.

Población	=	Registros de citas del año 2010 al 2016
Grupo experimental	=	Registros de citas del año 2017

3.3. Hipótesis

Aplicando la técnica de clasificación de minería de datos, se mejora los procesos de atención de citas en el área de consultorios externos.

3.4. Variables

- Variable Independiente
Aplicación de la técnica de clasificación de minería de datos.
- Variable Dependiente
Procesos de atención de citas en el área de consultorios externos.

3.5. Indicadores

En este apartado se detallan los indicadores establecidos para el proyecto, según la Tabla 3.

Tabla 3: Tabla de Indicadores

Indicador	Descripción	Unidad de Medida
Índice de deserción	Evalúa la variación del número de citas no atendidas, para adoptar medidas de corrección.	Entero: Cantidad
Porcentaje de presupuesto desaprovechado para la consulta externa	Define el índice del presupuesto desaprovechado por citas no atendidas	Decimal: Cantidad
Índice de citas otorgadas por la deserción	Número de citas aprovechadas producto de la deserción	Entero: Cantidad

3.6. Metodología

La metodología que se ha elegido para el desarrollo de la tesis, ha sido CRISP-DM, fundamentalmente porque incorpora como una de sus fases la comprensión del negocio, a comparación de las demás metodologías mencionadas en el capítulo anterior, que no lo hacen.

IV. RESULTADOS

A continuación se detallan los resultados del proyecto, los cuales se desarrollan en función a la metodología CRISP-DM descrita en el capítulo anterior; la cual cuenta con las fases comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación.

4.1. COMPRENSIÓN DEL NEGOCIO

Esta primera fase cuenta con las siguientes tareas: determinar los objetivos del negocio, evaluación del negocio, determinar los objetivos de la minería de datos y realizar el plan del proyecto.

4.1.1. Determinar los objetivos del Negocio (Consejo Directivo, 2012)

- Extender la cobertura de la seguridad social, incluyendo a los trabajadores independientes e informales.
- Brindar atención integral a los asegurados, con los más altos estándares de calidad, en el marco de un fuerte compromiso del Estado con el bienestar de los asegurados; mejorar el trato a los asegurados, cambiar el modelo de atención por uno basado en la atención primaria y actuar sobre los determinantes sociales de la salud, con énfasis en los aspectos preventivo–promocionales, contando para ello con el apoyo técnico de la OPS/OMS⁵.
- Garantizar la sostenibilidad financiera de la seguridad social en salud.
- Implementar una gestión transparente basada en el mérito y la capacidad, con personal calificado y comprometido.
- Brindar una atención oportuna para mejorar la calidad de vida de nuestros asegurados.

4.1.2. Evaluación del Negocio

- El indicador de deserción que maneja la institución en estudio, es de $\leq 5\%$. Este indicador al 2016 fue de 37.59%, correspondiente a 137,371 citas perdidas.
- Al superar el indicador, se genera un desaprovechamiento de presupuesto aproximado de S/.10,358.04, sólo en la muestra (209 pacientes) obtenida la encuesta “Motivos de Deserción a la Consulta Externa”. (Oficina de Gestión de la Calidad y Control Interno, 2012)
- Al no encontrar citas oportunamente los pacientes acuden al Servicio de Emergencia y Hospitalización, generando un gasto adicional a lo previsto de promediando los S/.3,116.77.
- La propuesta a desarrollar, ayudará a aprovechar de mejor manera las citas que quedarían disponibles por la deserción.
- El Hospital en estudio, cuenta con un Sistema de Gestión Hospitalaria, desarrollado en la plataforma FoxPro, teniendo como base de datos tablas DBF.

⁵ OPS/OMS: Organización Panamericana de la Salud / Organización Mundial de la Salud.

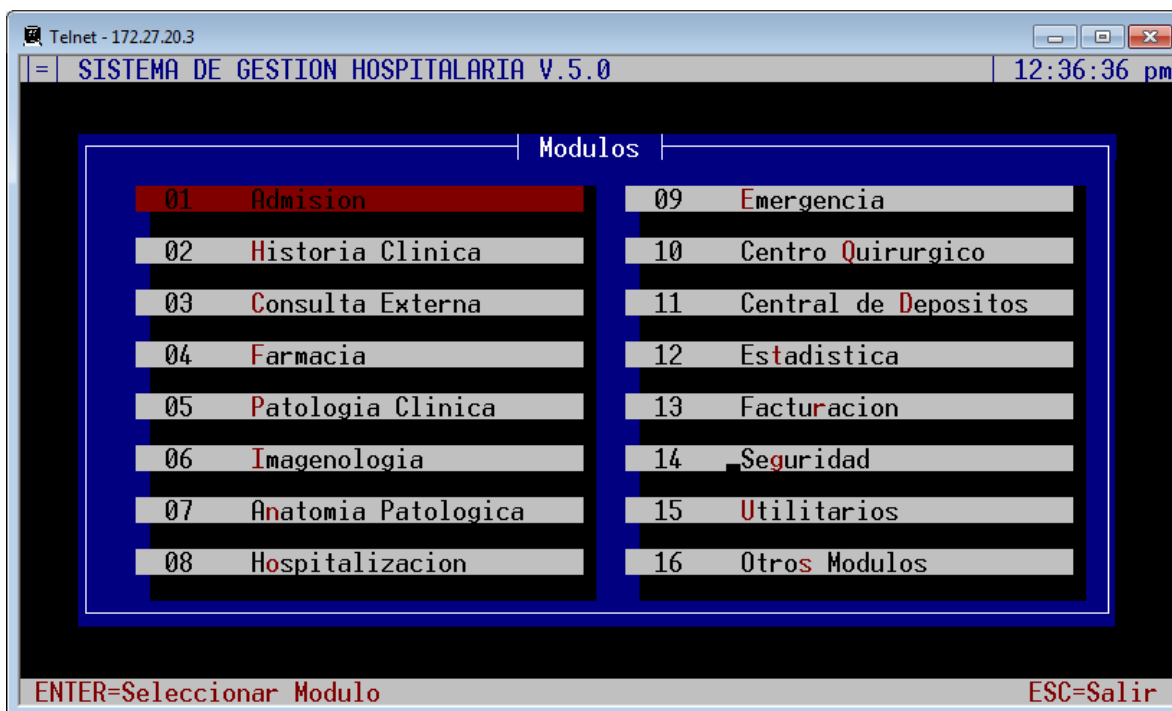


Figura 7: Interfaz del Sistema de Gestión Hospitalaria
Fuente: Sistema de Gestión Hospitalaria.

Para el caso en estudio, es el personal de Admisión de Citas y de Módulos, quienes utilizan el Sistema de Gestión Hospitalaria – SGH, haciendo uso del módulo de Admisión, módulo resaltado en la Figura 7; con la finalidad de realizar el proceso de otorgamiento de citas, realizar reprogramaciones de citas, eliminaciones, consultas de cupos, consultas de programación médica, pacientes citados, etc.

El personal del área de Referencias y Contrarreferencias, también tiene acceso al SGH, para consultar disponibilidad de cupos y otorgar citas por primera vez.

- Asimismo, cuentan con un Sistema Web de Referencias y Contrarreferencias, para gestionar las mismas, que serán la base de la atención de los pacientes asegurados; con interfaz desarrollada en Java, y la base de datos bajo la plataforma Oracle.



Figura 8: Interfaz del Sistema de Referencia y Contrarreferencia
Fuente: Sistema de Referencias y Contrarreferencias.

Es el personal del área de Referencias y Contrarreferencias, el encargado de utilizar el sistema mostrado en la Figura 8 para recepcionar, visar, citar, según la disponibilidad de cupos, y contrarreferir las referencias una vez expiradas. El personal del área de Admisión de Citas y Módulos cuenta con un usuario para éste sistema, con la finalidad de realizar sólo consultas.

- Para consultas y/o apoyo, cuentan con el Sistema de Acreditación, también bajo la interfaz Java, y con la base de datos en Oracle.



Figura 9: Interfaz del Sistema de Acreditación
Fuente: Sistema de Acreditación.

El Sistema de Acreditación, como ya se mencionó, es un sistema de consulta o apoyo, y es utilizado tanto por el personal de Admisión de Citas y de Módulos; cuando hay problemas con la acreditación de los pacientes.

- Se ha utilizado la Suite Business Intelligence de Microsoft SQL Server 2012 y posteriormente para mostrar el resultado de minería, se utilizará los complementos Power BI y Power View de la herramienta de Excel 2013.

4.1.3. Determinar los objetivos de la Minería de Datos

Utilizar la técnica árboles de decisión para determinar el perfil de los pacientes desertores respecto a su cita programada. Los pacientes citados serán evaluados según el perfil obtenido y el resultado permitirá predecir si acudirán o no a su consulta médica, basado en la clasificación que realizan los árboles de decisión según las siguientes variables:

- Edad.
- Sexo.
- Servicio.
- Médico.
- Dirección del paciente (según ubigeo del domicilio), entre otras.

4.1.4. Realizar el Plan del Proyecto

Se ha seleccionado la técnica de árboles de decisión, pues es necesario clasificar a los pacientes desertores de citas, según sus variables (edad, sexo, servicio, etc.) y obtener el perfil de éstos, el cual se ha confrontado con los datos pacientes citados a futuro, para predecir si asistirán o no a su cita.

Para realizar lo antes mencionado, primero hemos identificado las tablas DBF donde están almacenados los datos necesarios para el estudio; posteriormente, mediante la herramienta Microsoft SQL Server Integration Services, se ha realizado el proceso ETL necesario para preparar los datos utilizados en el modelo de minería de datos.

Los datos obtenidos del punto último, se han utilizados para el modelado de minería de datos mediante la técnica de árbol de decisión, de donde se ha obtenido el perfil deseado, el cual ha sido confrontado con los datos de los pacientes citados, obteniendo como resultado los posibles pacientes desertores, a quienes se procede a llamarlos, dos días antes de su cita, para confirmar su asistencia; de ser negativa su respuesta, se ofrecerá su cupo (cita) a un paciente que se encuentre en lista de espera.

4.2. COMPRENSIÓN DE LOS DATOS

Esta fase está comprendida por las tareas de recolección de datos iniciales, descripción de los datos, exploración de los datos y verificación de la calidad de los datos.

4.2.1. Recolectar Datos Iniciales

De las 876 tablas que contiene la base de datos del hospital en estudio, se trabajarán con las siguientes:

- **Adactivi** : tabla donde se almacena las actividades de atención con las que cuenta el hospital, como son: consulta ambulatoria, procedimientos, etc.
- **Adcitas** : tabla en la que se registran los pacientes citados.
- **Admedico** : tabla donde se registra el personal asistencial (médicos, enfermeras, tecnólogo médicos, obstetricas, etc.) que brindarán atención a los pacientes en consulta externa.
- **Adservic** : tabla donde se listan todos los servicios que brinda el hospital.
- **Adtaseg** : lista los tipos de asegurados.
- **Adtelcom** : registro de teléfonos y correo electrónico de pacientes
- **Asegur** : tabla donde se almacena toda la información del asegurado.
- **Adturno** : tabla donde se almacena los turnos de atención
- **Policlin** : lista todos los centros asistenciales a nivel nacional con los que cuenta la institución.

Asimismo se utilizará un archivo excel llamado Ubigeo.

4.2.2. Descripción de los Datos

- **Adactivi**

Tabla 4: Descripción de la tabla DBF Adactivi

Adactivi		Lista todas las actividades				
Total registros		14 - 2.09 KB				
Nombre del Campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumerico	Descripción campo
AC_CACTV	varchar	2	0	SI	NO	Código de actividad
AC_DACTV	varchar	40	0	SI	NO	Nombre de actividad

- **Adcitas**

Tabla 5: Descripción de la tabla DBF Adcitas

Adcitas		Pacientes citados				
Total registros		3823008 - 1 Gb				
Nombre del Campo	Tipo	Longitud	Decimales	Permite Nulls	Es Autonumerico	Descripción campo
CE_CSER	varchar	3	0	SI	NO	Código del servicio
CE_CTUR	varchar	2	0	SI	NO	Código del turno de atención
CE_FECHC	date	8	0	SI	NO	Fecha de cita
CE_CASE	varchar	15	0	SI	NO	Autogenerado del asegurado
CE_HATENC	float	8	0	SI	NO	Hora de atención
CE_IATE	varchar	1	0	SI	NO	Indicador de atención
CE_NCMP	varchar	5	0	SI	NO	Colegiatura del médico
CE_NACTM	varchar	7	0	SI	NO	Acto médico
CE_NOMBRE	varchar	40	0	SI	NO	Nombre de paciente
CE_CPOLICL	varchar	3	0	SI	NO	Policlinico de origen
CE_DIA	date	8	0	SI	NO	Fecha de otorgamiento de cita
CE_HORA	varchar	8	0	SI	NO	Hora de otorgamiento de cita
CE_PRECIO	float	8	2	SI	NO	Precio de consulta
CE_EDAD	float	8	0	SI	NO	Edad de paciente
CE_SEXO	varchar	1	0	SI	NO	Sexo de paciente
CE_TASEH	varchar	2	0	SI	NO	Tipo de asegurado titular
CE_NDOC	varchar	15	0	SI	NO	Número de documento de identidad de titular
CE_CACT	varchar	2	0	SI	NO	Código de actividad

- **Admedico**

Tabla 6: Descripción de la tabla DBF Admedico

Admedico		Lista de personal médico				
Total registros		3281 - 372 kb				
Nombre del Campo	Tipo	Longitud	Decimales	Permite Nulls	Es Autonumerico	Descripción campo
ME_NCMP	varchar	5	0	SI	NO	Colegiatura del médico
ME_DMED	varchar	40	0	SI	NO	Nombre del médico
ME_CSER	varchar	3	0	SI	NO	Identificador de servicio
ME_NTELF	float	8	0	SI	NO	Número de teléfono
ME_SERVIC	varchar	3	0	SI	NO	Código del Servicio
ME_DOCID	varchar	10	0	SI	NO	Número de documento de identidad

- Adservic

Tabla 7: Descripción de la tabla DBF Adservic

Adservic	Listado de especialidades y/o servicios					
Total registros	199 - 28,4 kb					
Nombre del Campo	Tipo	Longitud	Decimales	Permite Nulls	Es Autonumerico	Descripción campo
SE_CSER	varchar	3	0	SI	NO	Código del servicio
SE_DSER	varchar	40	0	SI	NO	Nombre del servicio

- Adtaseg

Tabla 8: Descripción de la tabla DBF Adtaseg

Adtaseg	Tipo de asegurado				
Total registros	59 - 4,93 kb				
Nombre del Campo	Tipo	Longitud	Permite Nulls	Es Autonumerico	Descripción campo
TA_CTASE	varchar	2	SI	NO	Código de tipo de asegurado
TA_DTASE	varchar	30	SI	NO	Tipo de asegurado

- Asegur

Tabla 9: Descripción de la tabla DBF Asegur

Asegur	Detalles del paciente					
Total registros	433561 - 140 MB					
Nombre del Campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumerico	
AS_NUMRUC	varchar	15	0	SI	NO	Número de RUC del Empleador del Titular
AS_TIPUSR	varchar	2	0	SI	NO	Tipo de asegurado
AS_AUTTIT	varchar	15	0	SI	NO	Autogenerado del titular
AS_AUTASE	varchar	15	0	SI	NO	Autogenerado del Paciente asegurado
AS_NOMBRES	varchar	59	0	SI	NO	Nombre del paciente
AS_SEXO	varchar	1	0	SI	NO	Sexo del paciente
AS_FECNAC	date	8	0	SI	NO	Fecha de nacimiento del paciente
AS_CPOLIC	varchar	3	0	SI	NO	Policlínico de origen
AS_DOCIDE	varchar	1	0	SI	NO	Tipo de documento de identidad
AS_NUMIDE	varchar	10	0	SI	NO	Número de documento de indentidad
AS_ESTCIV	varchar	1	0	SI	NO	Tipo de estado civil
AS_UBIDOM	varchar	6	0	SI	NO	Ubigeo del paciente
AS_NROTEL	varchar	9	0	SI	NO	Numero de telefono
AS_DIA	date	8	0	SI	NO	Día de otorgamiento de cita
AS_HORA	varchar	8	0	SI	NO	Hora de otorgamiento de cita
AS_DIRECC	varchar	40	0	SI	NO	Dirección de asegurado

- **Adturno**

Tabla 10: Descripción de la tabla DBF Adturno

Adturno		Lista de turnos de programación				
Total registros		154 - 6 KB				
Nombre del Campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumerico	Descripción campo
TU_CTUR	varchar	2	0	SI	NO	Código de turno
TU_DTUR	varchar	15	0	SI	NO	Nombre de turno
TU_HINI	float	8	0	SI	NO	Hora de inicio de turno
TU_HFIN	float	8	0	SI	NO	Hora de fin de turno
TU_NHORAS	float	8	0	SI	NO	Total de horas por turno
TU_HORARIO	varchar	1	0	SI	NO	Abreviatura de turno

- **Policlin**

Tabla 11: Descripción de la tabla DBF Policlin

Policlin		Lista de policlínicos u hospitales				
Total registros		508 - 59,6 MB				
Nombre del Campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumerico	Descripción campo
PO_CPOLIC	varchar	3	0	SI	NO	Código de policlínico u hospital
PO_NPOLIC	varchar	30	0	SI	NO	Nombre de policlínico u hospital
PO_UBIGEO	varchar	6	0	SI	NO	Ubigeo del policlínico u hospital

- **Ubigeo**

Tabla 12: Descripción del archivo Ubigeo.xls

Ubigeo.xls		Lista todas la ubicaciones geográficas				
Total registros		1833 - 329 KB				
Nombre del Campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumerico	Descripción campo
CÓDIGO DEPARTAMENTO	nvarchar	510	NO	-	NO	Código del departamento
CÓDIGO PROVINCIA	nvarchar	510	NO	-	NO	Código del provincia
CÓDIGO DISTRITO	nvarchar	510	NO	-	NO	Código del distrito
Codigo Ubigeo	nvarchar	510	NO	-	NO	Código Ubigeo
Departamento	nvarchar	510	NO	-	NO	Nombre del departamento
Provincia	nvarchar	510	NO	-	NO	Nombre del provincia
Distrito	nvarchar	510	NO	-	NO	Nombre del distrito

4.2.3. Exploración de los Datos

- **Adactivi**

Tabla 13: Lista de actividades

AC_CACTV	AC_DACTV
91	ATENCION MEDICA AMBULATORIA
92	ATENCION HOSPITALIZACION GENERAL
93	ATENCION EMERGENCIA
94	ATENCION QUIRURGICA
95	APOYO AL DIAGNOSTICO Y TRATAMIENTO
96	ATENCION PROCEDIMIENTOS
97	ATENCION PAAD (PROG.ATEN.AMBUL.DESCENTR)
98	ATENCION PADOMI (PROG.ATEN.DOMICILIARIA)
90	ACTIVIDADES ADMINISTRATIVAS
A0	ATENCION CRIPC (CENT.REH.INTEG.PAC.CRON)
99	OTRAS ACTIVIDADES EN HORAS SANITARIAS
A1	ATENCION EXTRAMUROS
A4	ATENCION UNIDAD DE CUIDADOS INTERMEDIOS
A3	ATENCION UNIDAD DE CUIDADOS INTENSIVOS

De la Tabla 13, podemos ver que el campo de **AC_CACTV** (código de actividades) deberá ser de tipo de dato CHAR(2) y el campo **AC_DACTV** (nombre de actividades) mantendrá el tipo de dato VARCHAR(100).

- **Adcitas**

Tabla 14: Diez primeros registros de pacientes citados

CE_CSER	CE_CTUR	CE_FECHC	CE_CASE	CE_NCMP	CE_NACTM	CE_IATE	CE_PRECIO	CE_CACT
A51	4	04/01/2005 00:00	8211081SDCHA016	14874	4297181	1	74.56	91
A51	4	04/01/2005 00:00	6703081GCMLJ003	14874	4297199	1	74.56	91
A51	4	04/01/2005 00:00	5502121QSVAJ006	14874	4297563	1	74.56	91
A51	4	04/01/2005 00:00	1707271PEOON009	14874	4297723	1	74.56	91
A51	4	04/01/2005 00:00	7902161MCAEC009	14874	4298006	1	74.56	91
A51	18	04/01/2005 00:00	5302151ZACZS000	23911	4280133	1	74.56	91
A51	18	04/01/2005 00:00	5510091AAPRE007	23911	4280749	1	74.56	91
A51	18	04/01/2005 00:00	4210011SSSTJ010	23911	4288058	1	74.56	91
A51	18	04/01/2005 00:00	2904220SCMCA008	23911	4285807	0	74.56	91
A51	18	04/01/2005 00:00	5804090AENEA009	23911	4288349	1	74.56	91

De la Tabla 14, se puede apreciar los campos, de la tabla adcitas, que se ha necesitado para realizar el presente trabajo; para cada campo se determinó los siguientes tipos de datos:

- CE_CSER : CHAR(3)
- CE_CTUR : CHAR(2)
- CE_FECHC : DATE
- CE_CASE : VARCHAR(15)
- CE_NCMP : INT
- CE_NACTM : INT

- CE_IATE : INT
- CE_PRECIO : FLOAT
- CE_CACT : CHAR(2)

Todos estos campos no deben ser nulos.

– Cantidad de citas anuales por servicio

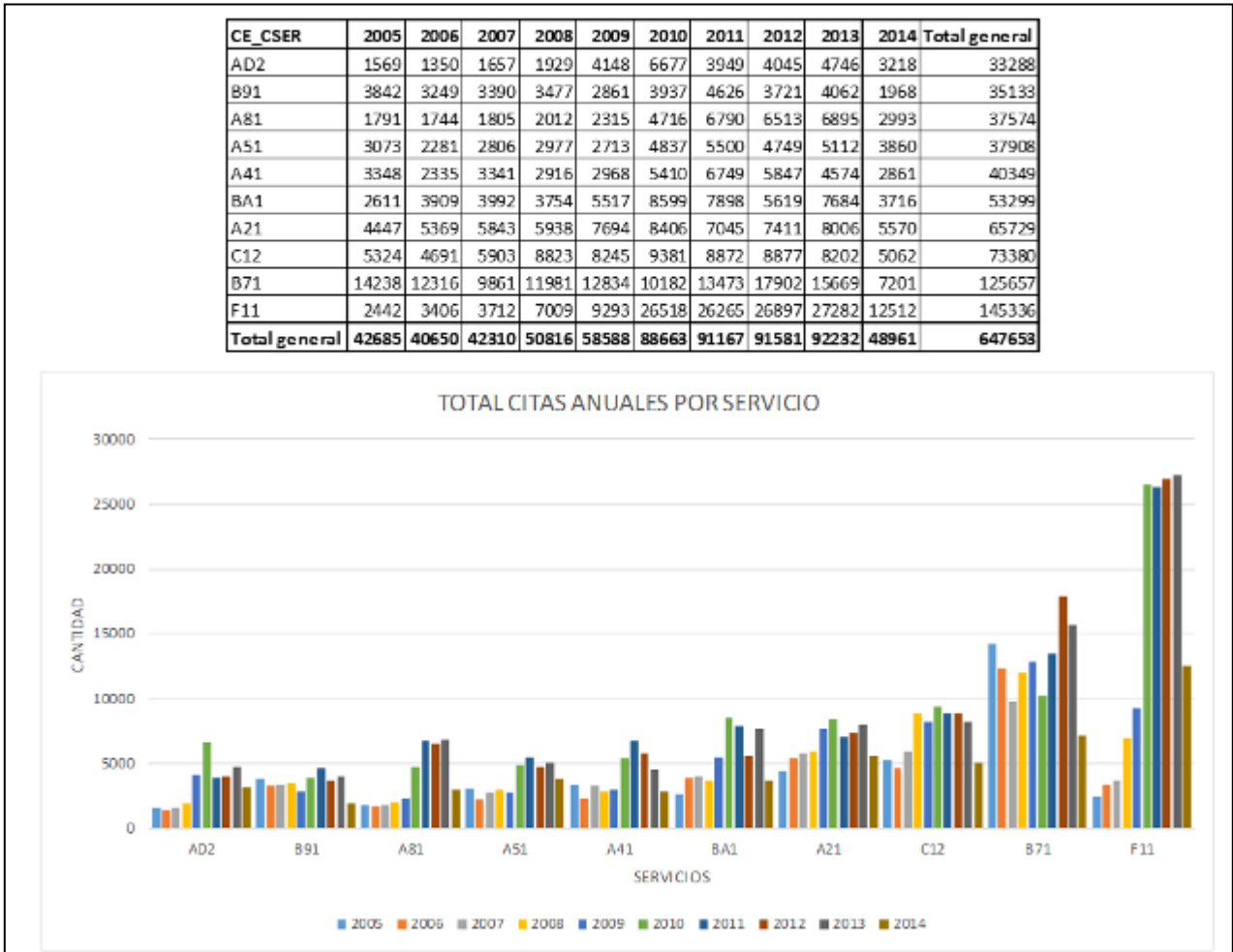


Figura 10: Total de atenciones de los diez servicios con más cantidad de citas
Fuente: Sistema de Gestión Hospitalaria. Tabla Adcitas.

En la Figura 10, podemos ver la cantidad de citas anuales, que han sido registradas a través del tiempo.

Han sido filtrados los 10 servicios que tienen mayor demanda.

- Total de presupuesto anual por actividades.

Tabla 15: Presupuesto anual para citas.

Cod Actividad	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2020	NULL	Total general
16			S/. 37										S/. 37
17		S/. 75											S/. 75
18	S/. 5,816	S/. 4,921	S/. 4,123										S/. 14,859
29	S/. 746												S/. 746
30	S/. 23,412	S/. 447	S/. 185										S/. 24,044
32	S/. 75	S/. 373											S/. 447
70	S/. 75												S/. 75
91	S/. 4,033,621	S/. 3,664,773	S/. 3,581,946	S/. 3,309,489	S/. 2,376,560	S/. 3,136,440	S/. 5,573,912	S/. 3,335,220	S/. 3,422,916	S/. 2,092,146	S/. 40		S/. 34,527,064
92			S/. 518	S/. 5,309	S/. 7,400	S/. 5,320	S/. 3,994	S/. 2,646	S/. 3,654	S/. 1,554			S/. 30,395
94	S/. 123,546	S/. 122,651	S/. 74,893	S/. 65,238									S/. 386,328
95							S/. 3,276	S/. 42					S/. 3,318
96	S/. 653,966	S/. 577,989	S/. 588,164	S/. 796,067	S/. 799,120	S/. 1,357,240	S/. 1,723,084	S/. 1,057,434	S/. 994,686	S/. 704,802			S/. 9,252,552
98				S/. 5,066	S/. 3,560								S/. 8,626
99					S/. 240	S/. 1,200		S/. 1,932	S/. 1,932	S/. 1,638			S/. 6,942
A1				S/. 560									S/. 560
NULL												S/. 0	S/. 0
Total general	S/. 4,841,255	S/. 4,371,229	S/. 4,249,866	S/. 4,181,729	S/. 3,186,880	S/. 4,500,200	S/. 7,304,266	S/. 4,397,274	S/. 4,423,188	S/. 2,800,140	S/. 40	S/. 0	S/. 44,256,068

En la Tabla 15, podemos observar el presupuesto que se ha necesitado para cubrir la demanda de pacientes, según las actividades de atención, a través de los años; siendo la actividad con código 91 (Atención médica ambulatoria), la más demandada.

- Total de cantidad de citas por indicador de atención.

Ind. Atención	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2020	NULL	Total general
0	78731	58577	72773	71266	60916	95993	93091	108468	99652	64115	1		803583
1	141143	145393	143137	178396	207498	247992	253670	229038	226071	70303		1	1842642
NULL												203	203
Total general	219874	203970	215910	249662	268414	343985	346761	337506	325723	134418	1	204	2646428

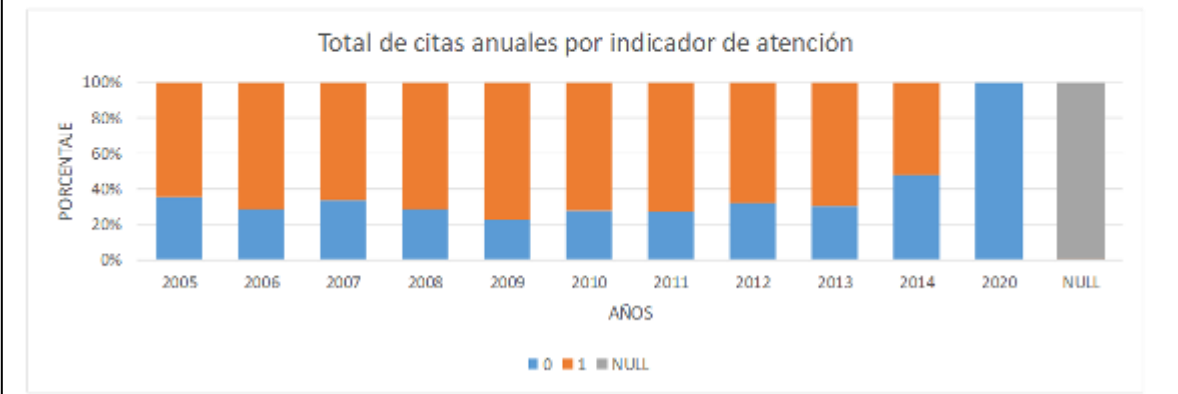


Figura 11: Total de citas anuales por indicador de atención
Fuente: Sistema de Gestión Hospitalaria. Tabla Adcitas.

En la Figura 11, ya podemos observar el porcentaje de deserción anual, pues el indicador “0” es una cita no atendida.

- Admedico

Tabla 16: Lista del personal asistencial con sus colegiaturas.

CMP	Médico	Total
99999	HEREDIA FERNANDEZ ROSA	1
	MANAY	1
	MEDICO DE EMERGENCIA	1
	MERA CABALLERO BRITANIA	1
	PRUEBA	1
	VARGAS TOURS LUIS	1
	Total 99999	6
29506	GUTIERREZ CELESTINO SEGURA WALTER L	1
	GUTIERREZ CELESTINO SEGURA WALTER LUIS	1
	GUTIERREZ CELESTINO WALTER	1
	Total 29506	3
33198	RIVAS SIMBANA FRANKLIN	1
	RIVAS SIMBANA FRANKLIN HENRRY	1
	RIVAS SIMBAÑA FRANKLIN HENRRY	1
	Total 33198	3
1	CORNEJO CARRANZA JESSICA	1
	MEDICO	1
	VISA TALAVERA	1
	Total 1	3
2163	BARBA NURENA CARMEN	1
	BARBA NURENA CARMEN AURORA	1
	BARBA NUREÑA CARMEN	1
	Total 2163	3
34680	TAFUR ROJAS ROXANA	1
	TAFUR ROJAS ROXANA HERMELINDA	1
	Total 34680	2
28408	GIL YUPANQUI MARCO	1
	GIL YUPANQUI MARCO ANTONIO	1
	Total 28408	2
2	MEDICO RESIDENTE	1
	MOZO GARCIA NANCY	1
	Total 2	2
	Total	24

En la Tabla 16, se puede apreciar la lista de colegiaturas con los nombres del personal asistencial.

El campo CMP (colegiatura) deberá tener un tipo de dato INTEGER, mientras que el campo del nombre del médico, será VARCHAR(100).

- Adservic

Tabla 17: Lista de servicios

Cod_Ser	Nombre de Servicio	Cantidad
NULL	NULL	4
19	SALA OBSERV. GINECO-	1
A11	ALERGIA	1
A12	ALERGIA PEDIATRICA	1
A21	CARDIOLOGIA	1
A22	CARDIOLOGIA INVASIVA	1
A23	CARDIOLOGIA NO INVASIVA	1
A24	CARDIOLOGIA PEDIATRICA	1
A31	DERMATOLOGIA	1
A32	DERMATOLOGIA PEDIATRICA	1
A41	ENDOCRINOLOGIA	1
A42	ENDOCRINOLOGIA PEDIATRICA	1
A51	GASTROENTEROLOGIA	1
A52	GASTROENTEROLOGIA PEDIATRICA	1
A61	GENETICA	1
A71	GERIATRIA	1
A81	HEMATOLOGIA CLINICA	1
A82	HEMATOLOGIA PEDIATRICA	1
A83	HEMATOLOGIA ESPECIAL	1
A91	MEDICINA COMPLEMENTARIA	1
AA1	MEDICINA FISICA Y REHABILITACION	1
AA2	REHABILITACION ADULTOS	1
AA3	REHABILITACION PEDIATRICA	1
AA4	REHABILITACION SISTEMA LOCOMOTOR	1
AA5	REHABILITACION TRAUMATOLOGICA Y OCUPACIO	1
AA6	OTRAS MINUSVALIAS Y DISCAPACIDADES	1
AA7	PATOLOGIA DEL DESARROLLO	1
AB1	MEDICINA GENERAL	1
AC1	MEDICINA INTERNA	1
AC2	MEDICINA INTERNA 1	1
AC3	MEDICINA INTERNA 2	1
AC4	MEDICINA INTERNA 3	1
AC5	MEDICINA INTERNA 4	1

En la tabla anterior, se lista algunos Servicios con los que cuenta el Hospital Nacional Almanzor Aguinaga Asenjo.

El campo Cod_Ser (código del servicio) tendrá un tipo de dato CHAR(03), mientras que el nombre será VARCHAR(100).

- **Adtaseg**

Tabla 18: Lista de tipos de asegurado

Cod_T.Aseg	Descripción Tipo de Asegurado	Cantidad
NULL	NULL	6
1	OBLIGATORIO DEPEND.	1
2	FACULT. INDEPENDITE REG.ESP	1
3	FACULT. CONTINUADOR REG.ESP	1
4	S.REGULAR - TRAB.DEL HOGAR	1
5	AMA DE CASA REG.ESP	1
6	AGRARIO INDEPENDIENTE	1
7	PENSIONISTA PENSIO.	1
8	CONYUGE DER.HAB	1
9	HIJO DER.HAB	1
10	HIJO INCAPACITADO DER.HAB	1
11	FOPASEF TERCERO	1
12	TERCERO TERCERO	1
13	ORFANDAD PENSIO.	1
14	INVALIDEZ PENSIO.	1
15	CHOFER PROFESIONAL REG.ESP	1
16	VIUDEZ PENSIO.	1
17	ACCION CIVICA OTROS	1
18	LEY EMERGENCIA OTROS	1
19	CONSTRUCCION CIVIL DEPEND.	1
20	PRO-JOVEN DEPEND.	1
21	MAGISTERIO DEPEND.	1
22	TRABAJADOR ESSALUD DEPEND.	1
23	CONCUBINO DER.HAB	1
24	GESTANTE DER.HAB	1
25	BOMBEROS OTROS	1
26	HEROES - EX COMBATIENTES 1941	1

Cod_T.Aseg	Descripción Tipo de Asegurado	Cantidad
27	VARIOS OTROS	1
29	SECTOR PUBLICO DEPEND.	1
30	SEGURO ESCOLAR OTROS	1
31	SIS INTERCAMBIO PRESTACIONAL	1
35	SEGURO UNIVERSITARIO OTROS	1
36	AGRARIO DEPENDIENTE	1
37	PESCADOR ARTESANAL	1
40	ESSALUD INDEPENDIENTE	1
41	ESSALUD PERSONAL	1
42	ESSALUD FAMILIAR	1
43	CONVENIO PERU - ESPANA	1
44	A TRABAJAR URBANO (SCTR)	1
45	A TRABAJAR RURAL (SCTR)	1
46	HEROES DEL CENEP - CIVIL	1
47	HEROES DEL CENEP - MILITAR	1
48	HEROES DEL CENEP - SOBREVIV.	1
49	POTESTATIVO UNICO	1
50	POTESTATIVO - PLAN TOTAL/VITAL	1
60	CONVENIO FOSPOLI	1
67	PENSIONISTA - LEY 28320	1
68	TRABAJAD. PESQUERO - LEY 28320	1
69	SOLICITANTE DE PENSION	1
70	VOLEYBOLISTAS	1
73	SEG. SALUD TRABAJADORES INDEP.	1
74	SEGURO DEL EJERCITO	2
75	S.REGULAR - D.LEG.1057 (CAS)	1
TOTAL		59

Según los datos de la tabla anterior, el campo Cod_T.Aseg (código del tipo de asegurado) será de tipo INTEGER, y la descripción del tipo de asegurado VARCHAR(100).

- Asegur

- Total de asegurados por tipo de asegurado y sexo

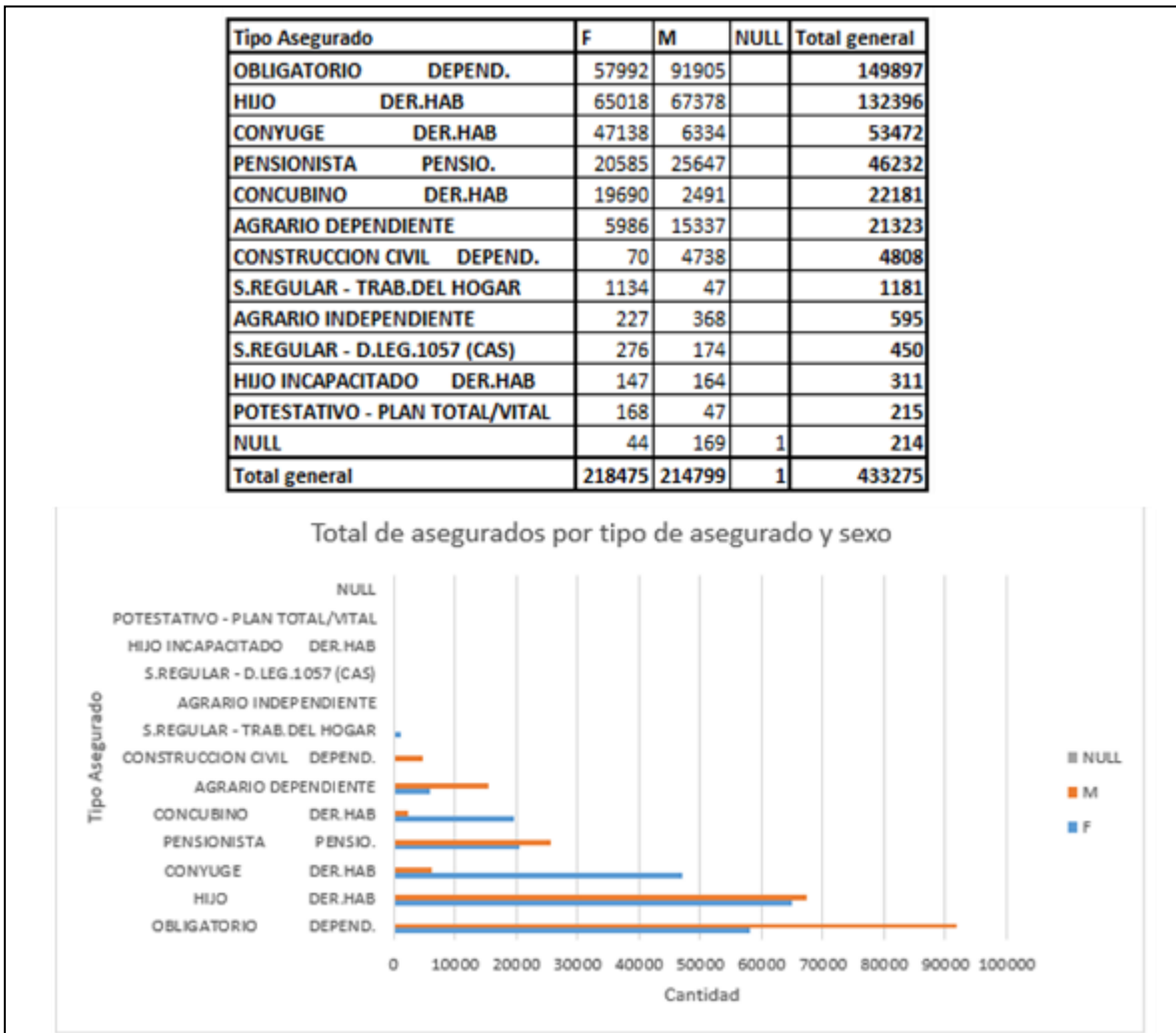


Figura 12: Total de asegurados por tipo de asegurado y sexo
Fuente: Sistema de Gestión Hospitalaria. Tabla Asegur – Essalud.

En la Figura 12 se muestra la cantidad de asegurado por tipo de asegurado y por sexo, donde el campo Sexo ha de ser de tipo CHAR(1) y el código del tipo de asegurado, según la tabla DBF ADTASEG, antes explorada, será de tipos INTEGER.

- Total de asegurados por tipo de asegurado y policlínico.

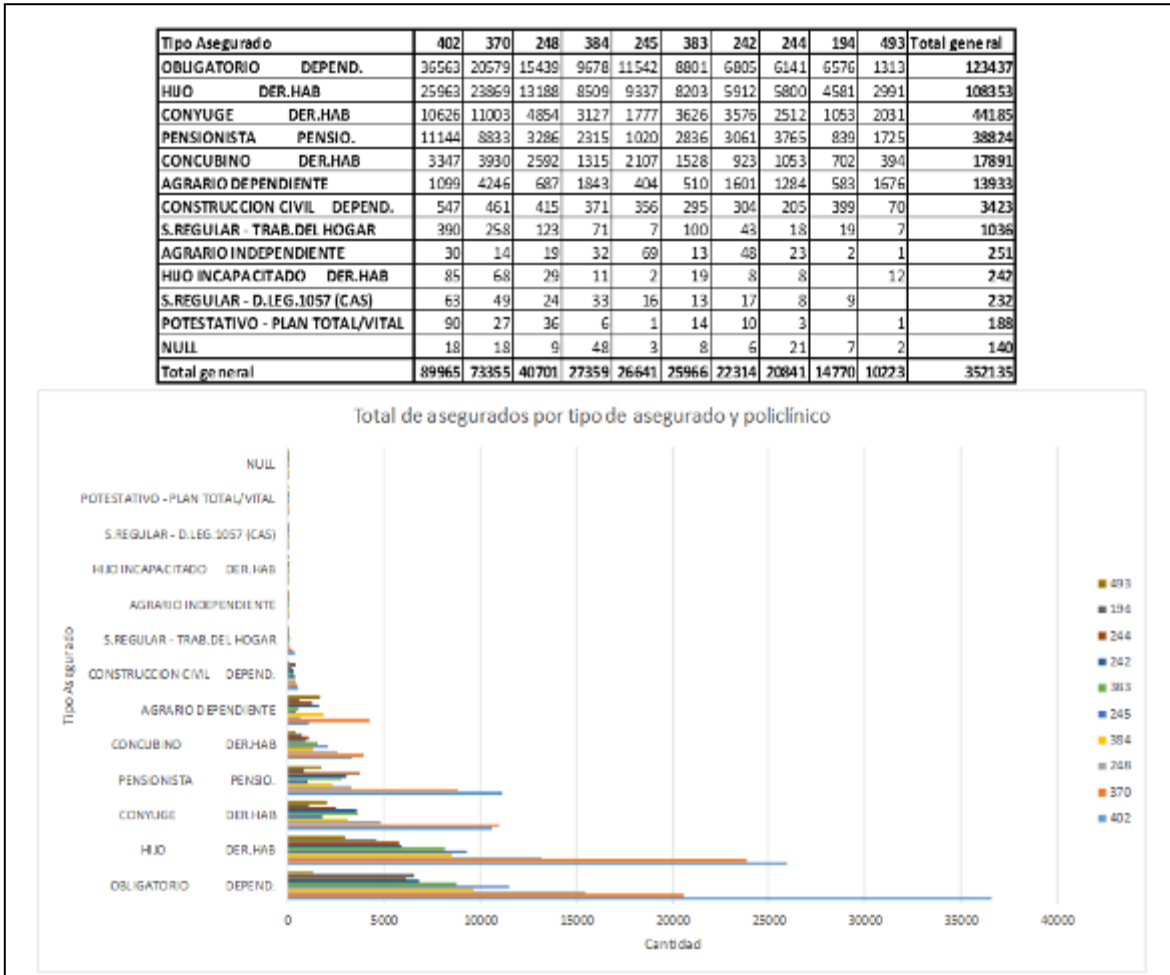


Figura 13: Cantidad de asegurados por tipo de asegurado y policlínico
Fuente: Sistema de Gestión Hospitalaria. Tabla Asegur.

Mediante la figura mostrada anteriormente, se estableció que el tipo de dato del campo código de policlínico sería INTEGER.

– Total de asegurados por tipo de asegurado y rango de edades



Figura 14: Cantidad de asegurados por tipo de asegurado y rango de edades
Fuente: Sistema de Gestión Hospitalaria. Tabla Asegur.

Mediante la Figura 14, se determinó que el tipo de dato para el campo edad sería INTEGER; asimismo no permitió identificar que hay mayor cantidad de asegurados registrados que son pensionistas.

- Total de tipo de asegurados por ubigeo.

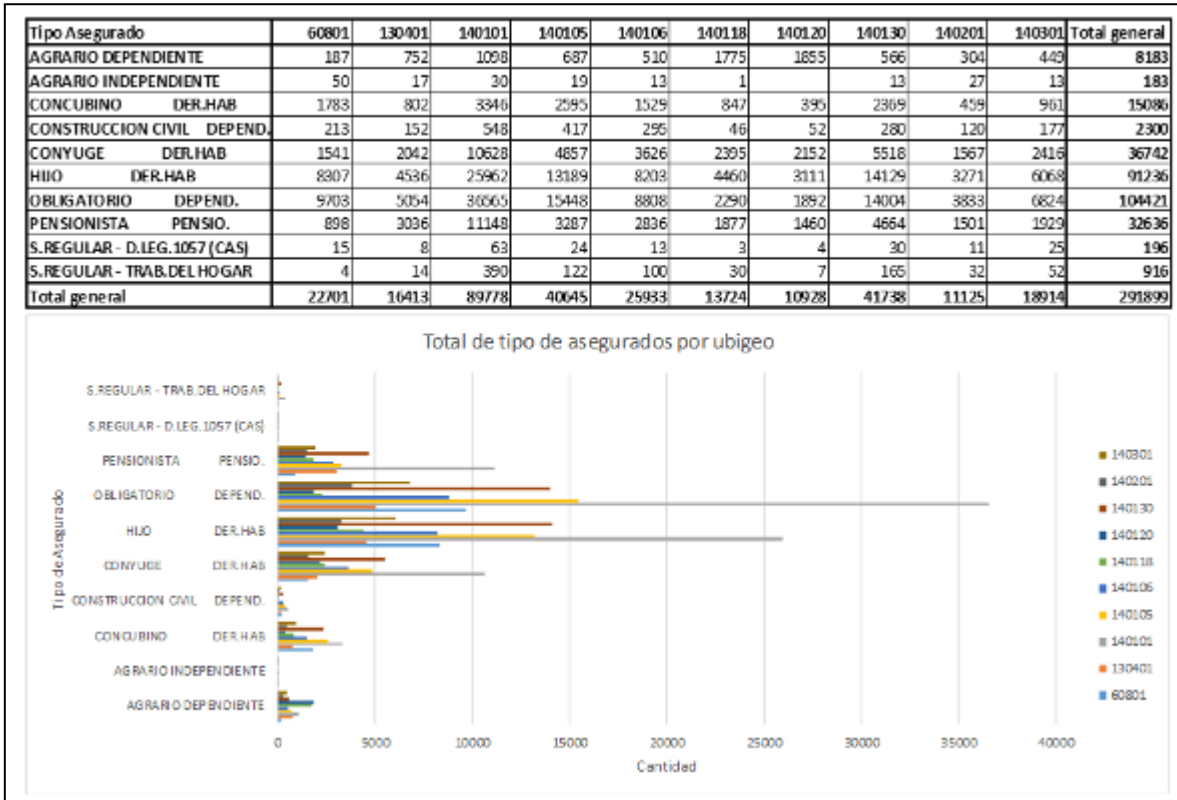


Figura 15: Total de asegurados por tipo de asegurado y ubigeo
Fuente: Sistema de Gestión Hospitalaria. Tabla Asegur.

Mediante la Figura 15, se determinó que el tipo de dato del campo ubigeo, ha de ser CHAR(06), además que la mayor cantidad de asegurado provienen o pertenecen de la ciudad con código de ubigeo 140101 (Chiclayo) y tienen tipo de seguro Obligatorio.

- **Adturno**

En esta tabla se puede observar los 150 tipos de turnos que maneja el hospital para la programación del personal.

Tabla 19: Lista de turnos.

Codigo	Turno	H. Inicio	H. Fin	Horas	Abrev.	Codigo	Turno	H. Inicio	H. Fin	Horas	Abrev.	Codigo	Turno	H. Inicio	H. Fin	Horas	Abrev.
77	TURNO MAÑANA	8.3	10.3	2	M	A4	TURNO TARDE	13.3	14.3	1	T	S3	TURNO MAÑANA	7.45	10	2.15	M
78	TURNO TARDE	12.3	18.3	6	T	A5	TURNO MAÑANA	7.3	13.5	6.2	M	S4	TURNO MAÑANA	7.45	14	6.15	M
79	TURNO MAÑANA	8	16	8	M	A6	TURNO MAÑANA	11.3	12	0.3	M	S5	TURNO MAÑANA	7.45	14	6.15	M
80	TURNO TARDE	15	17	2	T	A7	TURNO MAÑANA	11	11.3	0.3	M	S6	TURNO MAÑANA	7.45	10	2.15	M
81	TURNO MAÑANA	11.3	13.3	2	M	A8	TURNO MAÑANA	12	12.3	0.3	M	S7	TURNO MAÑANA	10	12.45	2.45	M
82	TURNO MAÑANA	8.3	13.3	5	M	B1	MAÑANA	6	11.3	5.3	M	S8	TURNO TARDE	14	18.45	4.45	T
83	TURNO MAÑANA	7.3	11.3	4	M	E1	TURNO MAÑANA	7	7.2	0.2	M	S9	TURNO TARDE	14.3	19.3	5	T
84	TURNO MAÑANA	7.3	8.3	1	M	E2	TURNO TARDE	13.3	13.5	0.2	T	T1	TURNO TARDE	11.3	15.3	4	T
85	TURNO MAÑANA	7	8	1	M	E3	TURNO NOCHE	19.3	19.5	0.2	N	T2	TURNO TARDE	13	14	1	T
86	TURNO MAÑANA	11	13.3	2.3	M	E4	TURNO TARDE	13	13.2	0.2	T	T3	TURNO TARDE	17	20	3	T
87	TURNO MAÑANA	10	18	8	M	E5	TURNO NOCHE	19	19.2	0.2	N	T4	TURNO TARDE	16	20	4	T
88	TURNO TARDE	15	19	4	T	E6	TURNO TARDE	13.3	17.3	4	T	T5	TURNO NOCHE	19	7	12	N
89	TURNO MAÑANA	7	12	5	M	E7	TURNO MANANA	7.3	9.3	2	M	T6	TURNO MAÑANA	7	8	1	M
90	TURNO MAÑANA	8.3	11.3	3	M	E8	TURNO TARDE	13.3	16.3	3	T	T7	TURNO NOCHE	19	20	1	N
91	TURNO MAÑANA	6	10	4	M	G1	GUARDIA DIURNA	8	14	6	M	T8	TURNO NOCHE	18	22	4	N
92	TURNO TARDE	13	15	2	T	G2	GUARDIA DIURNA	14	20	6	T	T9	TURNO MANANA	7	11	4	M
93	TURNO TARDE	14	21	7	T	G3	GUARDIA DIURNA	8	20	12	M	X3	NOCHE(08PM-02AM)	20	2	6	N
94	TURNO TARDE	14	19	5	T	G4	GUARDIA NOCTURN	20	8	12	N	Z1	TURNO TARDE	11	16	5	T
95	TURNO MAÑANA	7	10	3	M	H1	TURNO TARDE	17	19	2	T	Z2	TURNO TARDE	15	20	5	T
96	TURNO MAÑANA	7.3	11	3.3	M	N1	TURNO NOCHE	20	24	4	N	Z3	TURNO NOCHE	20	2	6	N
97	TURNO MAÑANA	9.3	11.3	2	M	N2	TURNO NOCHE	0	8	8	N	Z4	TURNO TARDE	17	21	4	T
98	TURNO TARDE	14.3	18.3	4	T	N3	TURNO NOCHE	20	8	12	N	Z5	TURNO MAÑANA	7	14	7	M
99	TURNO MAÑANA	6	12	6	M	N4	TURNO NOCHE	20	22	2	N	Z6	TURNO MANANA	8	14.3	6.3	M
A1	TURNO MAÑANA	7	9	2	M	N5	TURNO NOCHE	18	21	3	N	Z7	TARDE 13.30-15.	13.3	15.3	2	T
A2	TURNO MAÑANA	9	14	5	M	S1	TURNO MAÑANA	8	12.45	4.45	M						
A3	TURNO NOCHE	19.3	7.3	12	N	S2	TURNO MAÑANA	7.45	12	4.15	M						

A través de la Tabla 19, se puede determinar que el tipo de dato del campo código de turno será INTEGER.

- **Policlin**

La siguiente tabla almacena los 503 policlínicos con los que cuenta EsSalud, de los cuales tenemos como ejemplo los siguientes:

Tabla 20: Lista de Policlínicos.

Cod_Poli	Nombre Policlínico	Ubigeo	Cantidad
NULL	NULL	NULL	4
t	NULL	NULL	1
1	H. N. EDGARDO REBAGLIATI M	150133	2
2	H. N. GUILLERMO ALMENARA I	150108	1
3	H. N. CARLOS SEGUIN ESCOBEDO	40101	1
4	H. N. ALMANZOR AGUINAGA	140101	1
5	H. N. ALBERTO SABOGAL	70101	1
6	HOSP. II ANGAMOS	150114	1
7	HOSP. III JUAN A. PHILIPPS VELI	150101	1
8	HOSP. II VITARTE	150103	1
9	POLI. CANTO GRANDE (AHORA 406)	150137	1
10	POLI. CHINCHA	150101	1

Además se pudo determinar que el tipo de dato del campo Cod_Poli (código de policlínico) sería INTEGER.

- **Ubigeo**

Esta tabla es una fuente externa, que fue obtenida de la página web del Instituto Nacional de Estadística e Informática – INEI.

Tabla 21: Lista de Ubigeo.

CÓDIGO DEPARTAMENTO	CÓDIGO PROVINCIA	CÓDIGO DISTRITO	Código Ubigeo	Departamento	Provincia	Distrito
01	01	01	010101	AMAZONAS	CHACHAPOYAS	CHACHAPOYAS
01	01	02	010102	AMAZONAS	CHACHAPOYAS	ASUNCION
01	01	03	010103	AMAZONAS	CHACHAPOYAS	BALSAS
01	01	04	010104	AMAZONAS	CHACHAPOYAS	CHETO
01	01	05	010105	AMAZONAS	CHACHAPOYAS	CHILQUIN
01	01	06	010106	AMAZONAS	CHACHAPOYAS	CHUQUIBAMBA
01	01	07	010107	AMAZONAS	CHACHAPOYAS	GRANADA

Mediante esta tabla se determinó que el tipo de dato del campo código ubigeo sería CHAR(06).

4.2.4. Verificar la calidad de los Datos

- **Adcitas**

– Verificamos si el servicio existe o si existen datos nulos, donde el resultado fue el siguiente:

Tabla 22: Códigos de servicios erróneos o nulos.

Cod_Serv	Servicio	Cantidad
NULL	NULL	221
AAT	NULL	1
HEM	NULL	270
MEN	NULL	6
OFT	NULL	3742
P01	NULL	240
P03	NULL	311
P11	NULL	101
SA2	NULL	1
SIC	NULL	1
TOTAL		4894

– Verificamos si el turno existe o si existen datos nulos:
Dando el siguiente resultado:

Tabla 23: Códigos de turnos erróneos o nulos.

CE_CTUR	TU_DTUR	Cantidad
NULL	NULL	221
P5	NULL	17
TOTAL		238

– Verificamos la fecha de la cita:

Resultado:

Tabla 24: Fechas fuera de rango o nulos.

CE_FECHC	Cantidad
NULL	221
14/04/2020	1
11/06/2020	1
03/11/2020	1
Total	224

– Valores nulos o la no existencia de médicos:

Resultado:

Tabla 25: CMP y nombre del personal asistencial erróneos, nulos o repetidos.

CE_NCMP	ME_DMED	Cantidad	CE_NCMP	ME_DMED	Cantidad
0	NULL	2	91006	RAMIREZ MEDINA DEIDAMIA	10
2212	NULL	6	91006	RAMIREZ MEDINA DEIDAMIA	10
3667	NULL	2	91007	RAMIREZ CASTRO ELISA	55
13789	NULL	739	91008	CUMPA SUYON MERCEDES	530
17063	NULL	13	91012	CABREJO VIDARTE JUANA REQUILDA	47524
23489	NULL	1	91012	CABREJO VIDARTE JUANA REQUILDA	47524
38103	NULL	44	91013	TAPIA BAZAN ALICIA	659
43692	NULL	48	91014	ALFARO LOSSIO EDITH	2581
NULL	NULL	491	91015	DIAZ ZEVALLOS MARIA AMABLE	14034
60001	HOSPITAL PERU OFERTA FLEXIBLE	130	91015	DIAZ ZEVALLOS MARIA AMABLE	14034
60001	HOSPITAL PERU OFERTA FLEXIBLE	130	91016	SAMILLAN PANTA ANTONIA	21744
60002	HOSPITAL PERU OFERTA FLEXIBLE	327	91017	GRANDEZ FERNANDEZ ISABEL	5343
60002	HOSPITAL PERU OFERTA FLEXIBLE	327	91018	CUZQUEN SAAVEDRA MANUEL	15406
60003	HOSPITAL PERU OFERTA FLEXIBLE	341	91021	ERRADO	9600
60003	HOSPITAL PERU OFERTA FLEXIBLE	341	91021	ERRADO	9600
60004	HOSPITAL PERU OFERTA FLEXIBLE	254	94002	DIAZ CELIS EVILEIDA	31996
60004	HOSPITAL PERU OFERTA FLEXIBLE	254	94004	PICON PESANTES BERTHA	3698
60005	HOSPITAL PERU OFERTA FLEXIBLE	194	94005	GOMEZ GALLARDO JULIA ROSA	37595
60005	HOSPITAL PERU OFERTA FLEXIBLE	194	94010	BARRETO MONTENEGRO ROSALINA	5328
60006	HOSPITAL PERU OFERTA FLEXIBLE	227	94011	VALLEJOS CASTAÑEDA ZOILA	702
60006	HOSPITAL PERU OFERTA FLEXIBLE	227	99001	VILCHEZ MENDOZA LUIS MARTIN	9070
60007	HOSPITAL PERU OFERTA FLEXIBLE	94	99003	CASTRO REVILLA FLOR	1
60007	HOSPITAL PERU OFERTA FLEXIBLE	94	99004	CASTRO RODRIGUEZ FLOR	110
60008	HOSPITAL PERU OFERTA FLEXIBLE	31	99007	MORALES CASTAÑEDA PILAR	4216
60008	HOSPITAL PERU OFERTA FLEXIBLE	31	99009	PAZ RAMOS MARITZA	5
60009	HOSPITAL PERU OFERTA FLEXIBLE	103	99012	CABREJO VIDARTE JUANA REQUILDA	3
60009	HOSPITAL PERU OFERTA FLEXIBLE	103	99012	CABREJO VIDARTE JUANA REQUILDA	3
60010	HOSPITAL PERU OFERTA FLEXIBLE	174	99012	LOPEZ PARDO MARIA DOLORES	3
60010	HOSPITAL PERU OFERTA FLEXIBLE	174	99013	ALVARADO NIQUEN DE FLORES ANA	1
60011	HOSPITAL PERU OFERTA FLEXIBLE	244	99014	LARREA ROJAS VALERIANO	6090
60011	HOSPITAL PERU OFERTA FLEXIBLE	244	99999	HEREDIA FERNANDEZ ROSA	296
60012	HOSPITAL PERU OFERTA FLEXIBLE	61	99999	MANAY	148
60012	HOSPITAL PERU OFERTA FLEXIBLE	61	99999	MEDICO DE EMERGENCIA	148
90490	CASTILLO PEREZ MARIA ANTONIETA	9786	99999	MERA CABALLERO BRITANIA	148
91001	RIOJAS REYES ARACELI	414	99999	PRUEBA	148
91001	RIOJAS REYES ARACELI VICTORIA	4554	99999	VARGAS TOURS LUIS	148
91002	MEJIA DELGADO MAGNA	240009		TOTAL	548980

- Datos nulos en el campo de acto médico:
Resultado:

Tabla 26: Actos médicos duplicado y nulos.

Acto Médico	Cantidad	Acto Médico	Cantidad	Acto Médico	Cantidad	Acto Médico	Cantidad
NULL	221	5446727	2	5457145	2	6034274	2
4271607	2	5447677	2	5457229	2	6035102	2
4347864	2	5449236	2	5457249	2	6039989	2
4466078	2	5449603	2	5457271	2	6083566	2
4540186	2	5449691	2	5457328	2	6114820	2
4616001	2	5450101	2	5457336	2	6114846	2
4656223	2	5450604	2	5457371	2	6221781	2
4734008	2	5451076	2	5457413	2	6247891	2
4808640	2	5451478	2	5457441	2	6287359	2
4875352	2	5453192	2	5457445	2	6292644	2
4936937	2	5453381	2	5457561	2	6297177	2
5007292	2	5453632	2	5457572	2	6439599	2
5070819	2	5453661	2	5457580	2	6526405	2
5141631	2	5454026	2	5457613	2	6591013	2
5253556	2	5454067	2	5457655	2	6657940	2
5335586	2	5454191	2	5457682	2	6817860	2
5412928	2	5455199	2	5457796	2	7005532	2
5414368	2	5455448	2	5483686	2	7055951	2
5415421	2	5455768	2	5506007	2	7057803	2
5421697	2	5455963	2	5593575	2	7061526	2
5421989	2	5456219	2	5612184	2	7072651	2
5422793	2	5456448	2	5643792	2	7137680	2
5422844	2	5456527	2	5691066	2	7181355	2
5423067	2	5456613	2	5770202	2	7235051	2
5424383	2	5456616	2	5883163	2	8738109	2
5428608	2	5456620	2	5991212	2	8740258	2
5429013	2	5456634	2	5996641	2	8740726	2
5432184	2	5456684	2	6010897	2	8741169	2
5433269	2	5456687	2	6013014	2	8745371	2
5434045	2	5456741	2	6015148	2	8752593	2
5437386	2	5456838	2	6016390	2	8755766	2
5437411	2	5456899	2	6016657	2		
5441846	2	5456922	2	6021102	2		
5446034	2	5457024	2	6033009	2		
						Total	485

- Datos nulos en el campo estado de atención:

Tabla 27: Datos nulos en estado de cita.

CE_IATE	Cantidad
NULL	203
0	1270559
1	2522089
TOTAL	3792851

- Datos nulos o inconsistentes en el campo precio.

Tabla 28: Datos nulos o erróneos en precio.

CE_PRECIO	Cantidad
NULL	1148
0	2234
37	514765
40	1157091
42	1419622
74.5	1
74.56	697990
Total	3792851

- Datos nulos o inconsistentes en el campo código de actividad.

Tabla 29: Datos nulos o erróneos en el campo código de actividad.

CE_CACT	AC_CACTV	Cantidad
NULL	NULL	204
16	NULL	2
17	NULL	1
18	NULL	3741
29	NULL	270
30	NULL	652
32	NULL	6
70	NULL	1

Según lo verificado en las tablas descritas, debe asegurarse que los campos de esta tabla, así como los campos de las demás tablas que se relacionarán con la tabla adckitas, se encuentren sin datos erróneos y/o nulos.

4.3. PREPARACIÓN DE LOS DATOS

Las tareas de selección, limpieza, estructuración, integración y formateo de datos, componen esta fase.

4.3.1. Seleccionar los Datos

En la tarea Descripción de los Datos de la fase Comprensión de los Datos, se ha especificado los datos que se utilizarán para el modelado del proyecto. Adicionalmente se utilizará datos del Ubigeo de una fuente externa (archivo Excel).

- **Adckitas**

Adckitas	Pacientes citados
Nombre del Campo	Descripcion campo
CE_CSER	Código del servicio
CE_CTUR	Código del turno de atención
CE_FECHC	Fecha de cita
CE_CASE	Autogenerado del asegurado
CE_HATENC	Hora de atención
CE_IATE	Indicador de atención
CE_NCMP	Colegiatura del médico
CE_NACTM	Acto médico
CE_NOMBRE	Nombre de paciente
CE_CPOLICL	Policlínico de origen
CE_DIA	Fecha de otorgamiento de cita
CE_HORA	Hora de otorgamiento de cita
CE_PRECIO	Precio de consulta
CE_EDAD	Edad de paciente
CE_SEXO	Sexo de paciente
CE_TASEH	Tipo de asegurado titular
CE_NDOC	Número de documento de identidad de titular
CE_CACT	Código de actividad



Adckitas	Pacientes citados
Nombre del Campo	Descripcion campo
CE_CSER	Código del servicio
CE_CTUR	Código del turno de atención
CE_FECHC	Fecha de cita
CE_CASE	Autogenerado del asegurado
CE_IATE	Indicador de atención
CE_NCMP	Colegiatura del médico
CE_NACTM	Acto médico
CE_PRECIO	Precio de consulta
CE_CACT	Código de actividad

Figura 16: Selección de datos de la tabla Adckitas

En relación a los campos de la tabla adckitas, se deberán seleccionar los campos de las demás tablas, según las relaciones y necesidades para el estudio

4.3.2. Limpiar los Datos

Para realizar la limpieza de los datos se creó una conexión ODBC, con la finalidad de conectarnos con las tablas DBF a utilizar; dicha conexión tiene por nombre: ConexionHospital, y ha servido para crear nuestro Datamart.

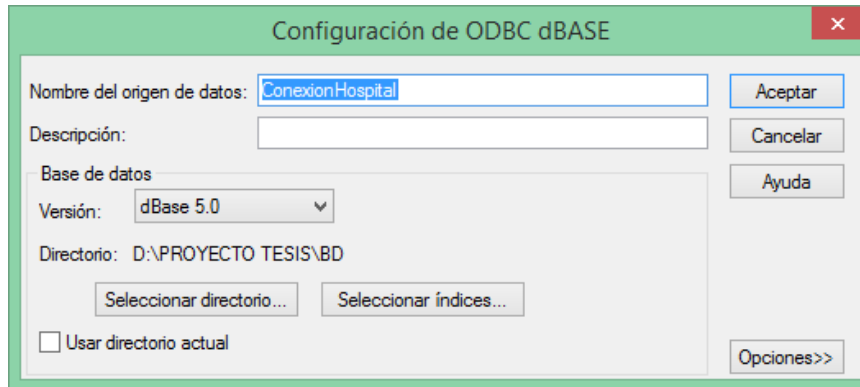


Figura 17: Estableciendo la conexión ODBC

Esta tarea, se realizó conjuntamente con la tarea de Estructurar, Integrar y Formatear los Datos, mediante un proceso de ETL, utilizando la plataforma de SQL Server Integration Services, que es parte de la herramienta de SQL Server Business Intelligence Development Studio.

Se trabajó con 10 *Tareas Flujo de Datos*, mediante las cuales hemos poblado nuestras *Dimensiones* y *Hechos* de nuestro *DatamartDeserción*:

Tabla 30: Descripción de Dimensiones y Hechos

	Tablas DBF	Archivo XLS	Dimensiones	Procedimiento Almacenado	Descripción
DimensiónActividad	Adactivi	-	-	-	Registrar las actividades asistenciales
DimensiónAsegurado	Asegur, Adtaseg y Adtelcom	-	DimensionPoliclinico y DimensionUbigeo	-	Registrar datos de los asegurados
DimensiónMedico	Admedico	-	-	-	Registrar el personal asistencial
DimensiónPoliclinico	Policlin	-	-	-	Registrar los policlínicos
DimensiónServicio	Adservic	-	-	-	Registrar los servicios
DimensiónTiempo	-	-	-	X	Poblar las fechas a explorar
DimensiónTurno	Adturno	-	-	-	Registrar los turnos
DimensiónUbigeo	-	Ubigeo	-	-	Registrar el ubigeo
HechoAdmision	-	-	Todas las dimensiones	-	Registrar histórico de citas
HechoCitas	-	-	Todas las dimensiones	-	Registrar citas a futuro

En primera instancia, se registra en el HechoAdmision el histórico de las citas hasta la fecha de implementación de la aplicación; posterior a ésta, se estableció esperar 2 días para continuar con el registro; con esto se tendrá un histórico con 2 días de retraso y así garantizamos que todos los datos de las citas se encuentren registradas, pues de éstos se realizará el entrenamiento de minería de datos.

Por otro lado, en el HechoCitas, se registra citas con dos días posteriores a la fecha actual, pues estos serán los pacientes a quienes se les determinará la probabilidad de deserción.

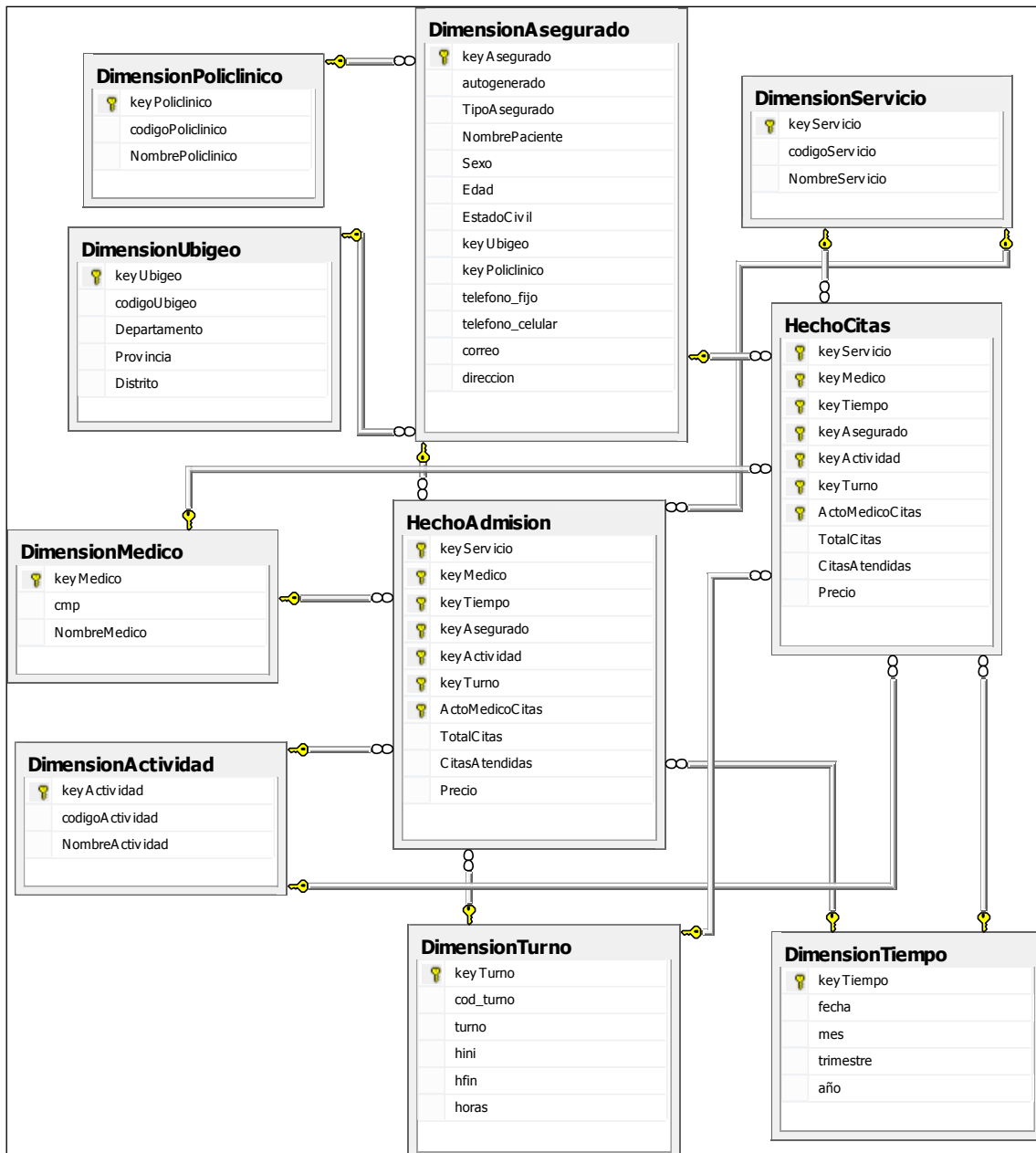


Figura 18: Diseño de DatamartDeserción

En la herramienta SQL Server Business Intelligence Development Studio, se creó el Proyecto de Integration Services llamado ETL y seguidamente un Origen de Datos llamado “*Conexión Hospital*”.

Luego se ha creado un Paquete de SQL Server Integration Services (SSIS), llamado “*ETL*”, donde se realizará el proceso, valga la redundancia, de ETL (Extracción, transformación y carga). Es aquí donde se realiza el poblado de las Dimensiones y los Hechos, cumpliendo con las tareas de esta fase.

Sólo la DimensionTiempo fue poblada por un procedimiento almacenado, el resto de dimensiones e incluso los hechos fueron poblados mediante las herramientas que nos brinda el SSIS.

- **DimensionTiempo**

Esta dimensión se pobló a través de un procedimiento almacenado, el cual registró fechas desde el 01 de enero de 2005 al 31 de diciembre de 2018:

```
USE [DataMartCitas]
GO
/***** Object: StoredProcedure [dbo].[pa_generaTiempo]  Script Date:
06/25/2014 10:34:41 *****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE proc [dbo].[pa_generaTiempo]
as
    declare @fi smalldatetime=cast('01/01/2005' as smalldatetime)
    while @fi<=cast('31/12/2018' as smalldatetime)
        begin
            insert into DimensionTiempo
values(@fi,MONTH(@fi),DATEPART(QQ,@fi),YEAR(@fi))
            set @fi=@fi+1
        end
end
```

Para el desarrollo de estas tareas se tomó como ejemplo el poblado del HechoAdmision, y del HechoCitas; y en donde se utilizaron herramientas similares para el poblado de las demás dimensiones.

Cabe señalar que para el poblado de las tablas hechos, antes se deben poblar las dimensiones.

- **HechoAdmision y Hecho Citas**

Para iniciar con la tarea de limpieza de datos, se procedió a eliminar los datos nulos del campo CE_CASE (autogenerado de paciente) de la tabla Adcitas.

Mediante la herramienta Origen ADO Net, que permite extraer datos de una base de datos relacional mediante un proveedor OLE DB, se realizó la conexión con la tabla DBF antes mencionada y se procedió a eliminar datos nulos y limitar los datos que se necesitan, mediante una consulta sql, según sea el caso:

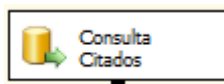


Figura 19: Origen ADO Net – Consulta Citados

HechoAdmision

- Consulta para registrar el histórico de citas:
select ce_cser,ce_ctur, ce_fecha, ce_case, ce_ncmp, ce_nactm,
ce_cact, ce_precio,ce_iate, 1 from adcitas where ce_case is not null
and ce_fecha<=dateadd('d', -1, now())

- Consulta para registrar progresivamente con dos día de retraso:
`select ce_cser,ce_ctur, ce_feche, ce_case, ce_ncmp, ce_nactm, ce_cact, ce_precio,ce_iate, 1 from adcitas where ce_case is not null and ce_feche<= dateadd('d',-2,getdate())`

HechoCitas

- Consulta para registrar citas con días posteriores a la fecha actual:
`select ce_cser,ce_ctur, ce_feche, ce_case, ce_ncmp, ce_nactm, ce_cact, ce_precio, ce_iate, 1 from adcitas where ce_case is not null and ce_feche between dateadd('d', +1, now()) and dateadd('d', +2, now())`

Otra manera de realizar limpieza de datos es utilizando la herramienta de división condicional, la cual evalúa y dirige las filas de un conjunto de datos; como ejemplo se tomó a la DimensionMedico:

- **DimensionMedico**

En este ejemplo se evaluó el campo de colegiatura (CMP), en donde sólo seguirían el proceso, aquellos datos o registros que tuvieran colegiaturas diferentes a “0”.

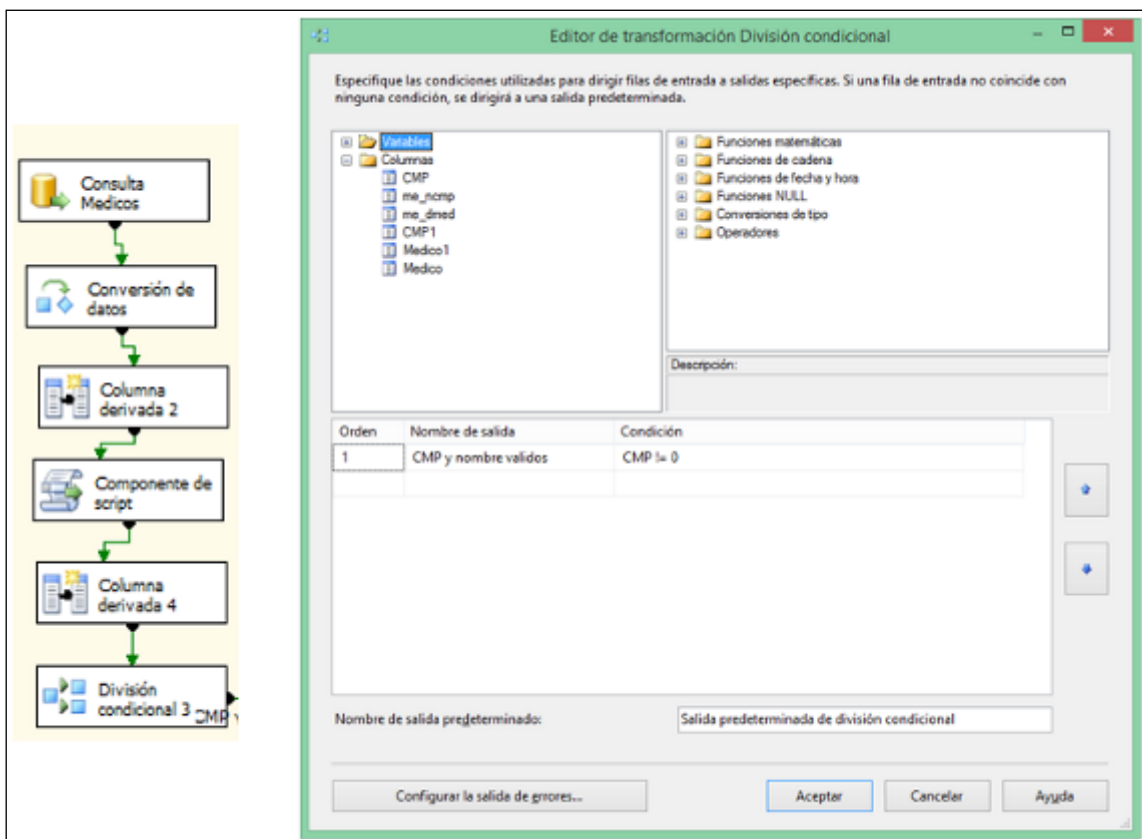


Figura 20: División Condicional – CMP diferentes a “0”

4.3.3. Estructurar los Datos

Para el desarrollo de esta tarea, se tomó como ejemplo el poblado del HechoAdmision, cuyo proceso es igual al poblado del HechoCitas y se utilizó herramientas comunes para el poblado de las dimensiones:

- **HechoAdmision y HechoCitas**

- Columna Derivada; actualiza valores de una columna mediante expresiones.

En este caso se utilizó para eliminar espacios en blanco de los campos de la tabla Adcitas y para reemplazar valores de la columna precio a 74.50, siempre y cuando los valores sean nulos o igual a "0" o 74.56; de lo contrario que mantenga el mismo valor.

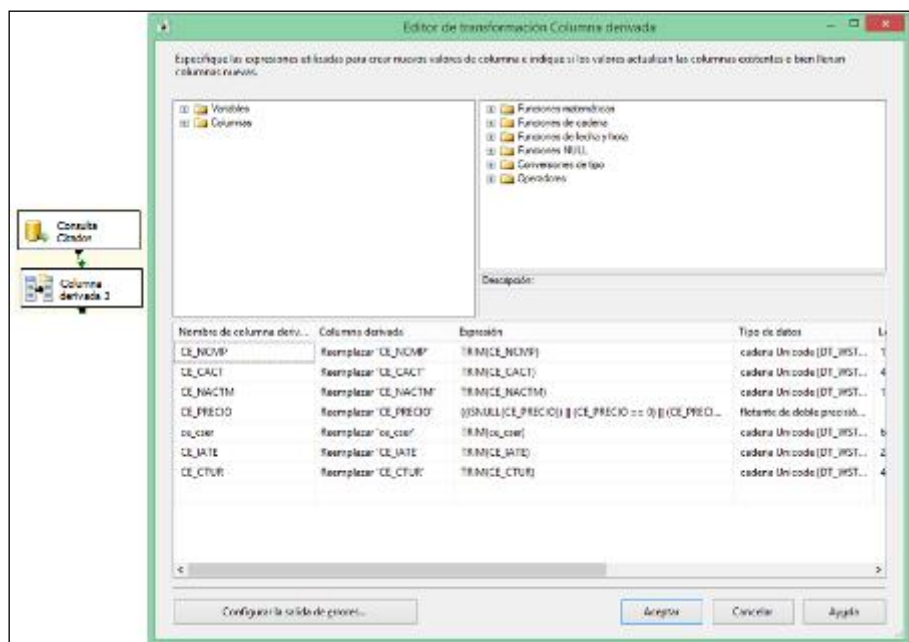


Figura 21: Columna derivada de tabla Adcitas

- Conversión de datos; convierte columnas a distintos tipos de datos y agrega las columnas al conjunto de datos.

Como se puede observar en la siguiente figura, se utilizó esta herramienta, para convertir el tipo de datos de las columnas y genera nuevas columnas con los datos convertidos.

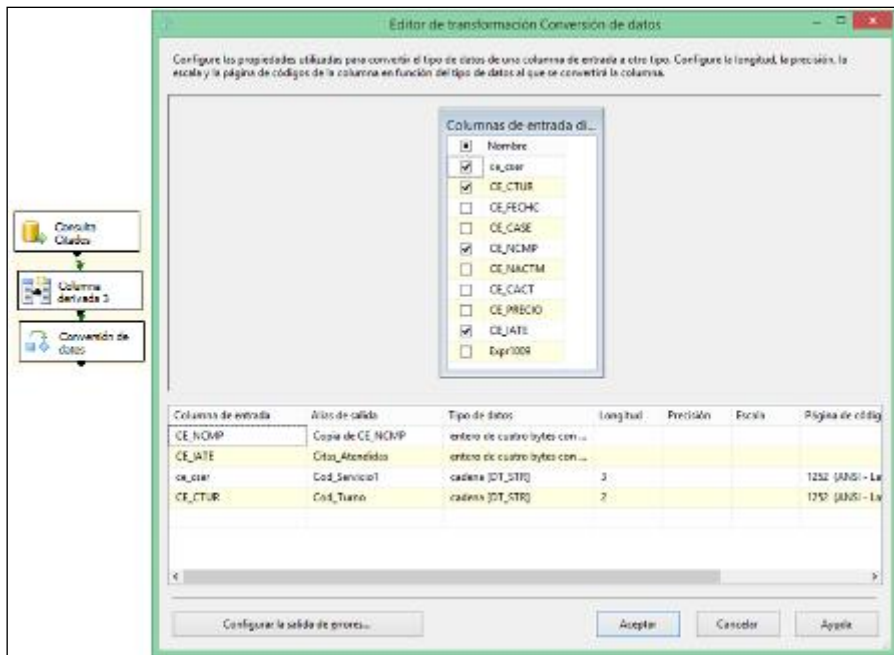


Figura 22: Conversión de tipos de datos de campos de la tabla adctitas

- Componente de script; actualiza valores de una columna mediante la ejecución de un script personalizado. Mediante esta herramienta, se reemplazó códigos de servicio y colegiaturas; esto se pudo realizar con la herramienta columna derivada, pero como eran muchos los valores a reemplazar se optó por utilizar el componente de script; en el cual se necesita especificar las columnas de entrada y las de salida.

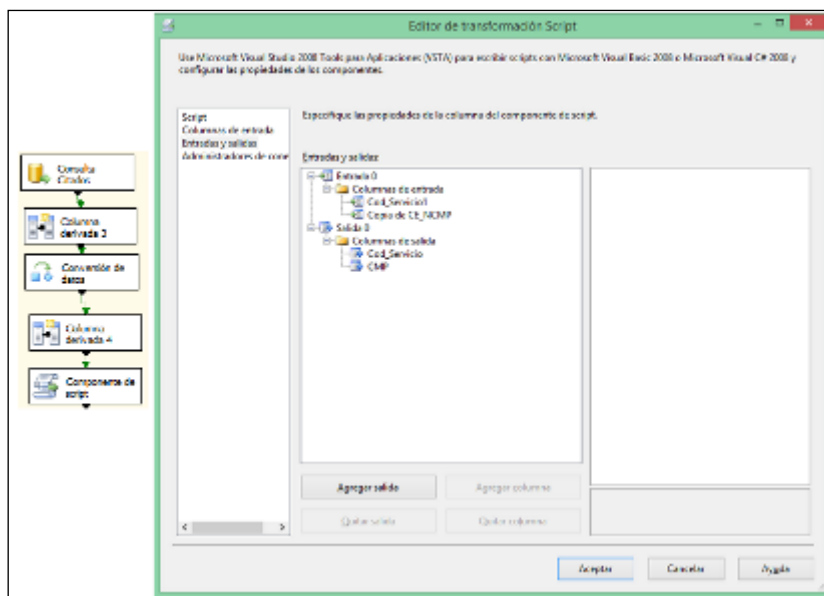


Figura 23: Columnas de entrada y salida del componente de script

En la siguiente figura se puede observar algunas de las sentencias ejecutadas en el componente de script.

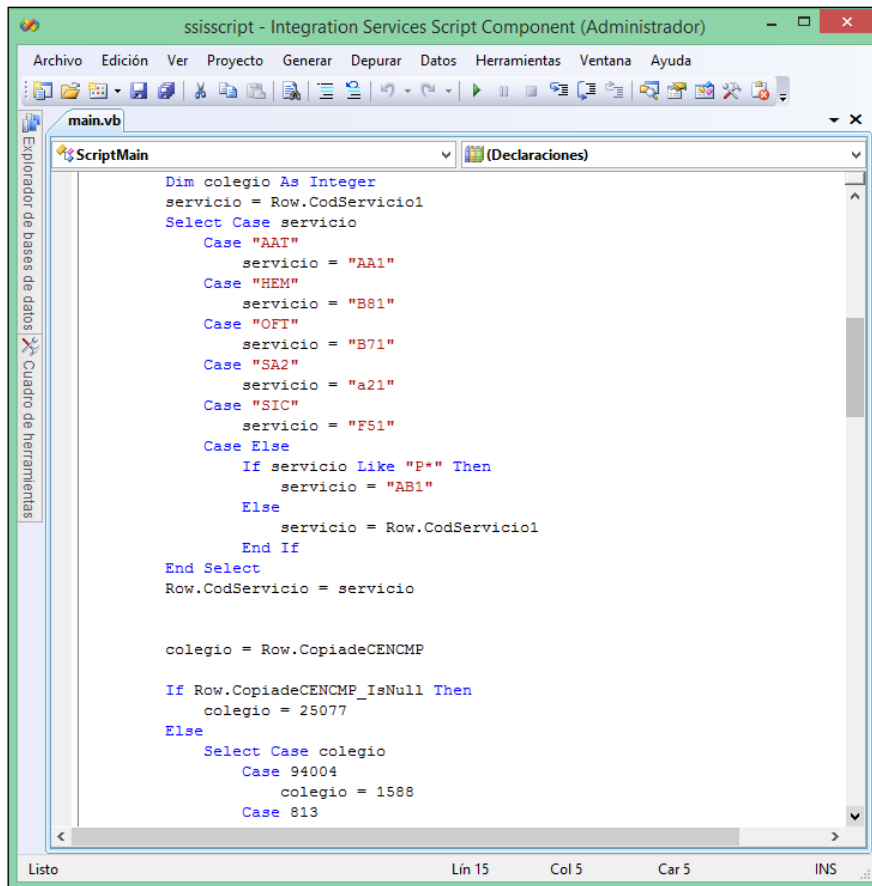


Figura 24: Sentencias para reemplazar valores

4.3.4. Integrar los Datos

Una de las finalidades de esta tarea es generar nuevas columnas a partir de datos de otras columnas; esto se puede realizar con la herramienta columna derivada, ya antes vista y como ejemplo se tomó la DimensionAsegurado.

- **DimensionAsegurado**

En esta dimensión se utilizó una columna derivada para generar la columna que calcule la edad del paciente o asegurado, la cual fue llamada “Edad”; para realizar esto se consideró la fecha de nacimiento y el autogenerated del paciente y se utilizó la siguiente expresión:

```

YEAR(GETDATE()) - (ISNULL(as_fecnac) ?
(((DT_I4)SUBSTRING(Autogen_Paciente,1,2) > 14 || Tipo_Asegur ==
73) ? (DT_I4)("19" + SUBSTRING(Autogen_Paciente,1,2)) :
(DT_I4)("20" + RIGHT("0" + SUBSTRING(Autogen_Paciente,1,2),2)))
: YEAR(as_fecnac))

```

Asimismo se reemplazó valores en los campos tipo de asegurado y código de policlínico, cuyo proceso se describió en la tarea anterior.

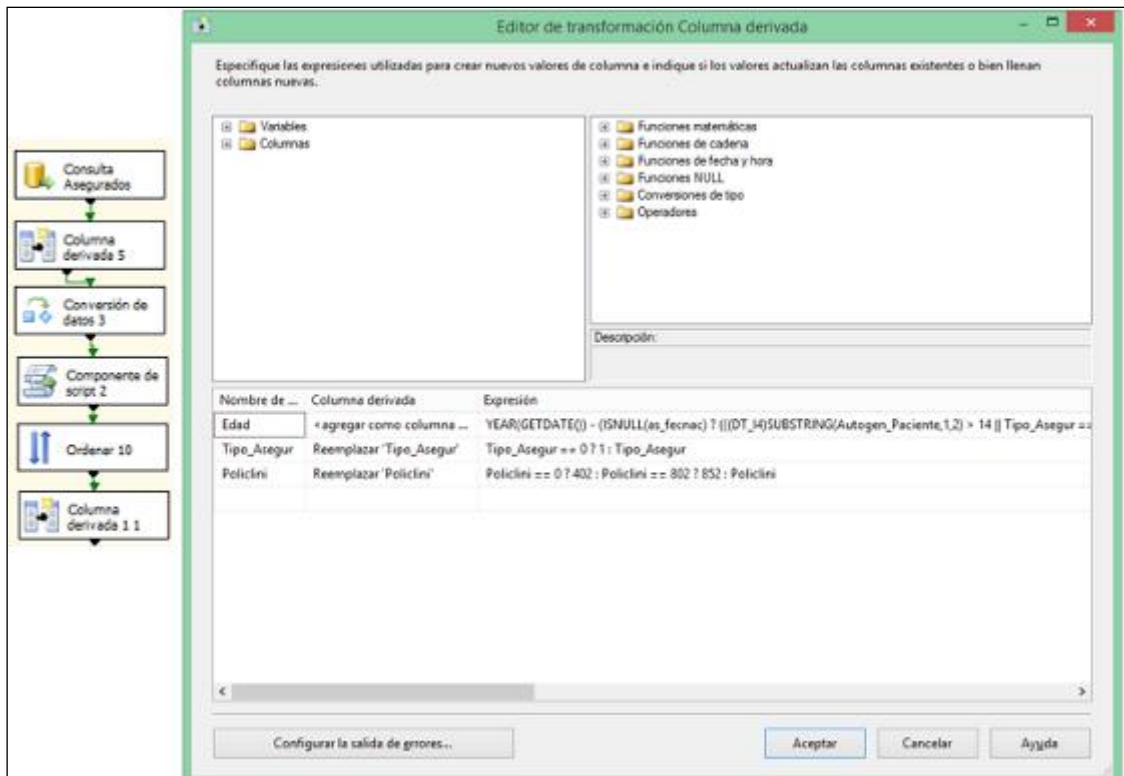


Figura 25: Generación de la columna llamada Edad

Asimismo en esta tarea se utilizó la herramienta combinación de mezcla; la cual mezcla datos de dos conjuntos de datos mediante una combinación. Veamos un ejemplo con las tablas hechas:

- **HechoAdmision y HechoCitas**

En la siguiente figura, se puede apreciar la mezcla de datos del conjunto de datos proveniente de la del origen de datos “Consulta Citados” con el conjunto de datos provenientes de la “Dimensión Médico”, en la cual se seleccionan los datos que se requieren para continuar el proceso.

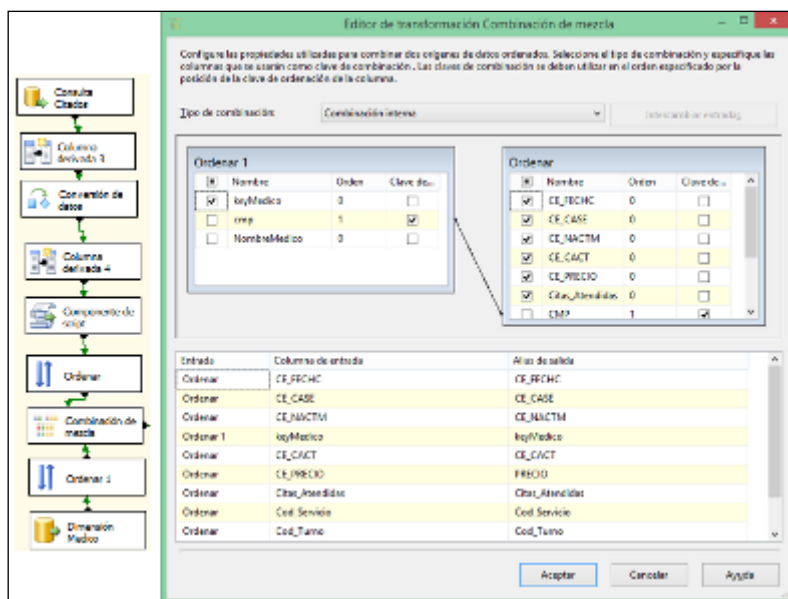


Figura 26: Combinación de datos de la tabla de citas y de dimensión médico

4.3.5. Formateo de los Datos

Una de las finalidades de esta tarea es la reordenación de los campos; lo cual se puede realizar a través de la herramienta llamada Ordenar.

En la siguiente figura se puede apreciar que los datos seleccionados en la opción “Paso a través”, han sido ordenados ascendentemente por la columna CMP.

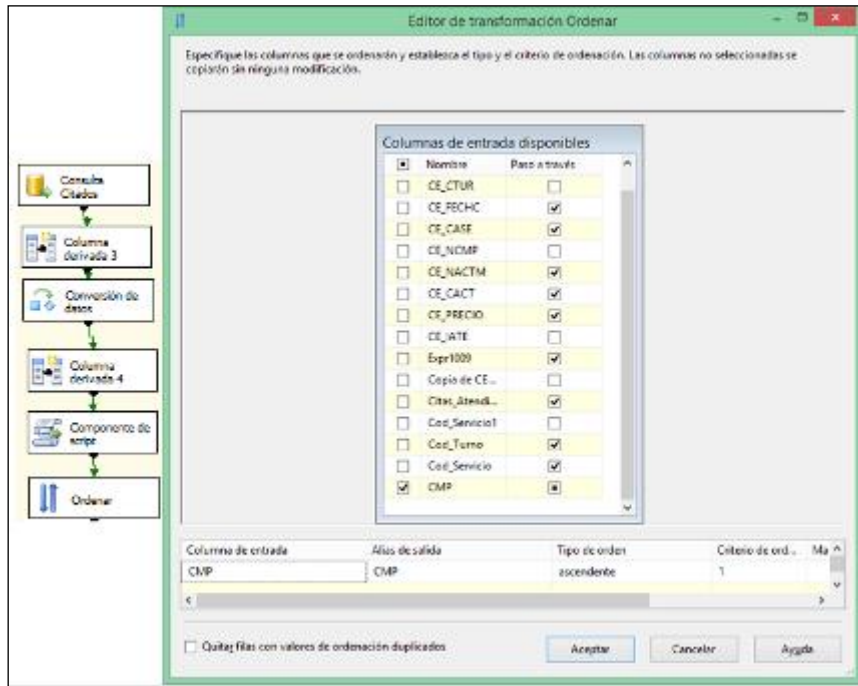


Figura 27: Ordenar datos por CMP

Asimismo mediante esta tarea, se busca eliminar comas, caracteres especiales, u otros datos que impidan un buen proceso ETL. Esto se puede realizar con varias herramientas; como por ejemplo la columna derivada, componente de script, pero también se puede realizar a partir de la consulta sql del origen de datos:

La siguiente figura representa el proceso inicial de la DimensionPoliclinico; donde, mediante una consulta sql se está especificando que se seleccionen los campos código y nombre de policlínico, pero con la condición que en los registros de la columna nombre de policlínico no contenga el carácter especial “.”.

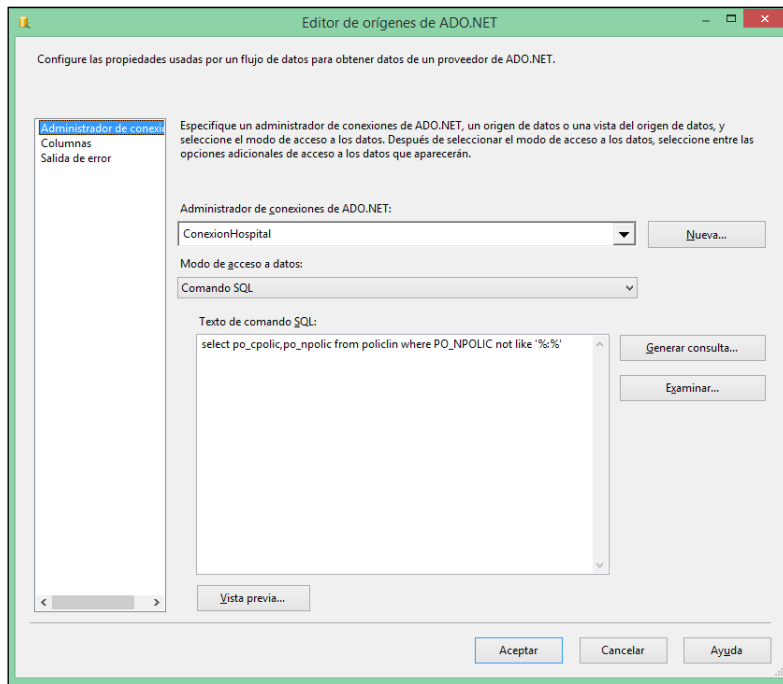


Figura 28: Eliminando caracteres especiales

Mediante la siguiente figura se muestra las herramientas utilizadas para el poblado del HechoAdmisión:

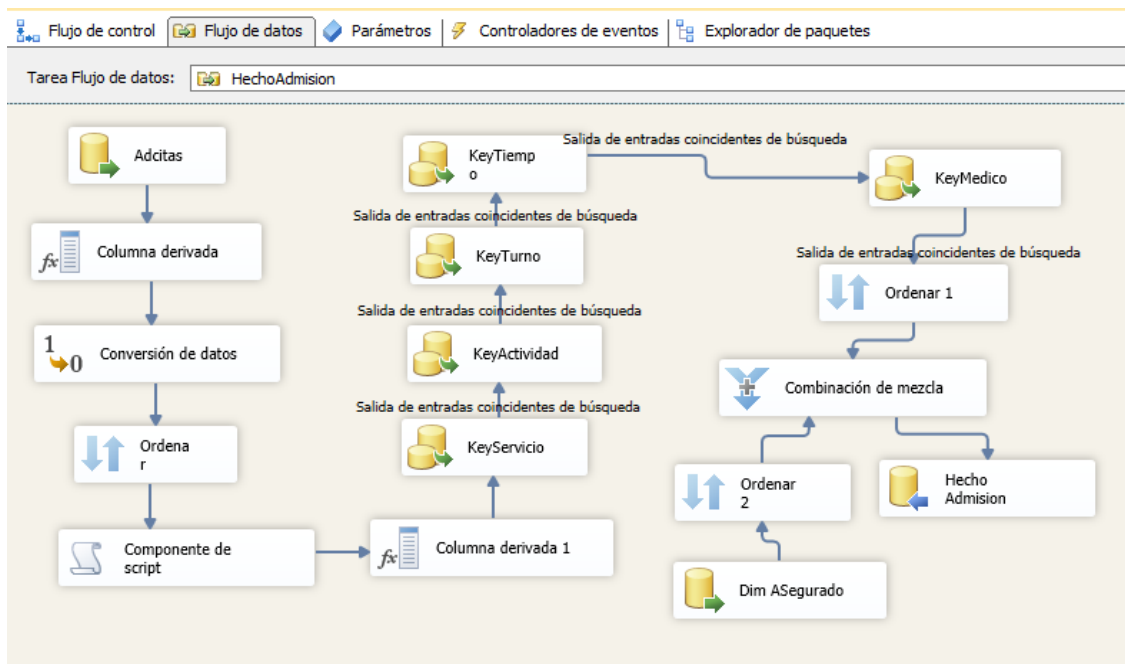


Figura 29: Flujo de datos – HechoAdmisión

Asimismo la figura siguiente muestra el flujo de control que contiene las relaciones que existe entre las dimensiones y los hechos.

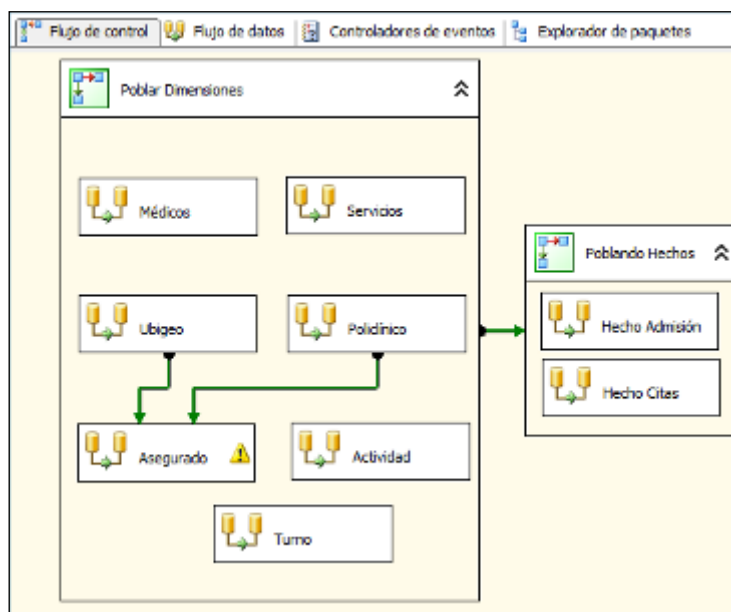


Figura 30: Flujo de control del ETL – Poblando Hechos

4.4. MODELADO

Los resultados de esta fase se detallan según las tareas de selección de la técnica de modelado, generación del plan de prueba, construcción del modelo y la evaluación del modelo.

4.4.1. Seleccionar técnica de modelado

Debido a que el problema es crear el perfil de los pacientes desertores para determinar la probabilidad de deserción de los pacientes futuros, se consideró 2 técnicas de modelado:

- Árboles de Decisión
- Clustering

Al final se tomó la decisión de elegir la técnica de Árboles de Decisión, pues es la que más se acerca al modelo ideal con un 83.26% de probabilidad de predicción frente a un 81.05% del modelo de clustering en la estructura de la Actividad de Consulta Ambulatoria (Estructura_Con) y un 96.52% de probabilidad de predicción del modelo de árboles de decisión frente a un 79.36% del modelo de clustering en la estructura de la Actividad de Procedimientos (Estructura_Proc).

Se consideró sólo las citas registradas de las actividades Consulta Ambulatoria y de Procedimientos, pues son las actividades con mayor demanda y actividades que se realizan por el área de Consultorios Externos.

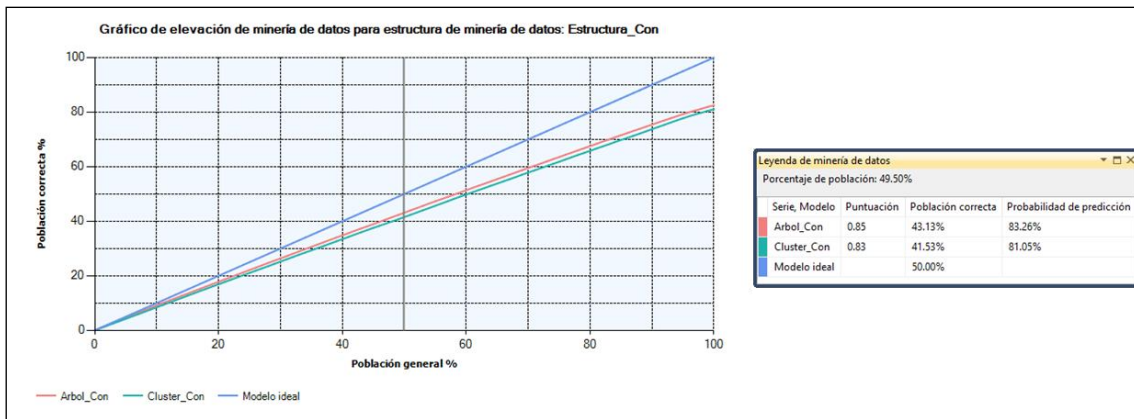


Figura 31: Gráfico de Precisión de Minería de Datos - Estructura_Con

Recuentos para Arbol_Con en Citas Atendidas		
Previsto	0 (Real)	1 (Real)
0	10740	2523
1	94674	448362

Recuentos para Cluster_Con en Citas Atendidas		
Previsto	0 (Real)	1 (Real)
0	0	0
1	105414	450885

Figura 32: Matriz de Clasificación – Estructura_Con

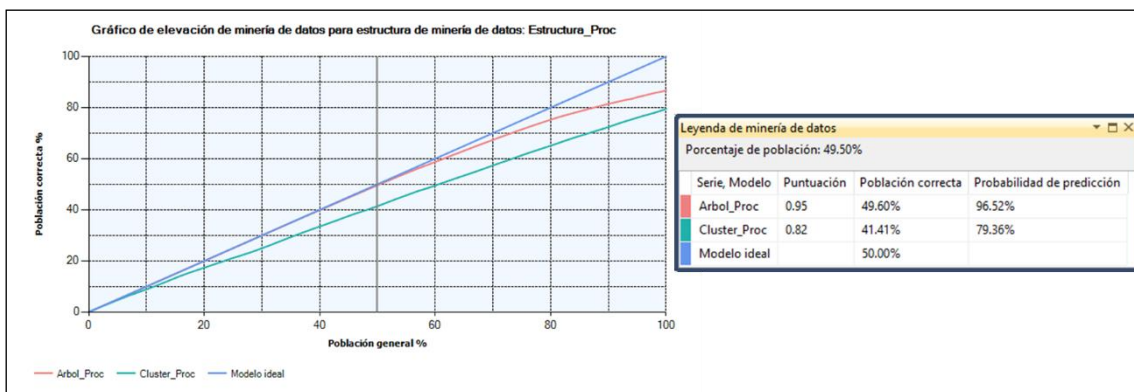


Figura 33: Gráfico de Precisión de Minería de Datos - Estructura_Proc

Recuentos para Arbol_Proc en Citas Atendidas		
Previsto	0 (Real)	1 (Real)
0	118274	15462
1	5175	16337

Recuentos para Cluster_Proc en Citas Atendidas		
Previsto	0 (Real)	1 (Real)
0	123449	31799
1	0	0

Figura 34: Matriz de Clasificación – Estructura_Proc

4.4.2. Generar el plan de prueba

La data histórica se remonta desde el año 2005 hasta el mes de enero de 2018, donde para realizar el plan de prueba se dividió los datos en dos grupos:

- Grupo de entrenamiento : 2010 – 2016
- Grupo de prueba : 2017

El resultado de este plan de prueba fue el siguiente:

- Actividad Consulta Ambulatoria
De un total de 264,213 (100%) citas registradas en el periodo del grupo de prueba, 105,393 (39.89%) fueron deserciones, mientras que 158,820 (60.11%) fueron citas atendidas.
El resultado del modelo de minería de datos determinó un total de 6,715 posibles desertores; de los cuales 2,635 fueron realmente pacientes que desertaron a su cita, teniendo un 39.24% de certeza en el pronóstico.
- Actividad Procedimientos
De un total de 99,704 (100%) citas registradas en el periodo del grupo de prueba, 39,909 (40.02%) fueron deserciones, mientras que 59,795 (59.98%) fueron citas atendidas.
El resultado del modelo de minería de datos determinó un total de 74,314 posibles desertores; de los cuales 29,700 fueron realmente pacientes que desertaron a su cita, teniendo un 39.97% de certeza en el pronóstico.

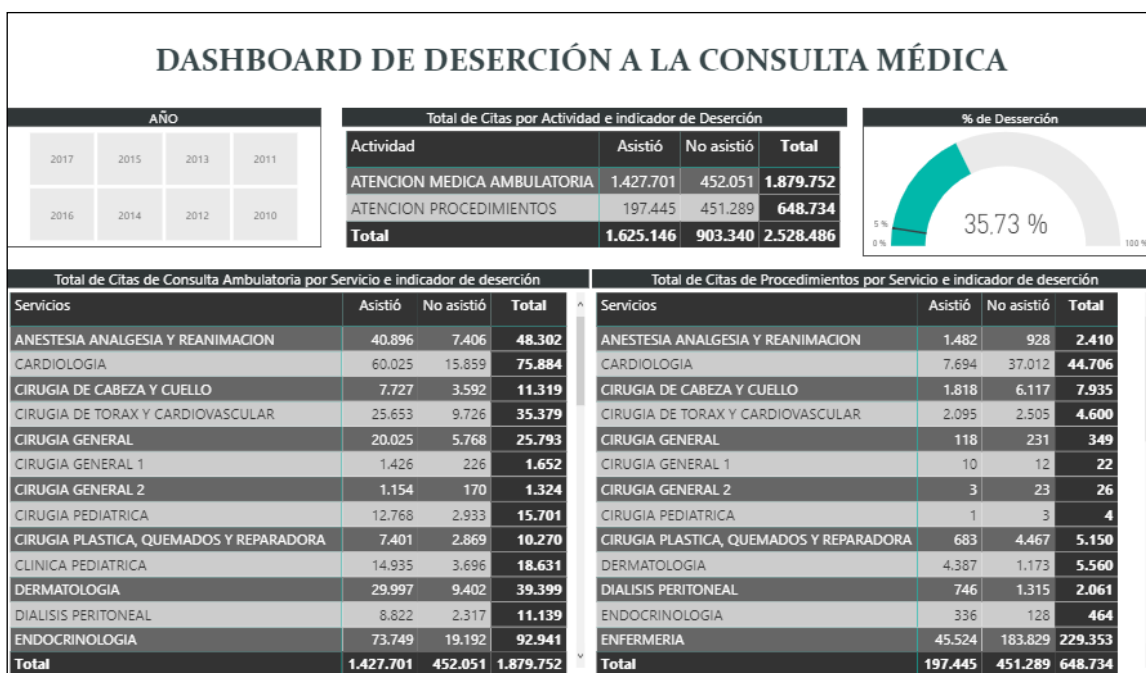


Figura 35: Dashboard de Deserción

4.4.3. Construir el modelo

Para construir el modelo de minería de datos, se realizaron las siguientes acciones:

1. En la herramienta SQL Server Management Studio, se creó, dentro del DatamartDesercion, 4 vistas llamadas:

- Vista_Mineria_Con

Almacena el histórico de citas de la Actividad Médica Ambulatoria; utilizando el asistente para la creación de la vista, se generó la siguiente consulta SQL:

```
SELECT A.TipoAsegurado, A.Sexo, A.Edad, A.EstadoCivil,
       M.NombreMedico, P.NombrePoliclinico, U.Departamento,
       U.Provincia, U.Distrito, H.ActoMedicoCitas, H.CitasAtendidas,
       TU.turno, S.NombreServicio
```

```
FROM DimensionPoliclinico P INNER JOIN DimensionAsegurado
A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN
DimensionUbigeo U ON A.keyUbigeo = U.keyUbigeo INNER
JOIN HechoAdmision H ON A.keyAsegurado =
H.keyAsegurado INNER JOIN DimensionActividad AC ON
H.keyActividad = AC.keyActividad INNER JOIN
DimensionServicio S ON H.keyServicio = S.keyServicio INNER
JOIN DimensionTiempo T ON H.keyTiempo = T.keyTiempo
INNER JOIN DimensionTurno TU ON H.keyTurno =
TU.keyTurno INNER JOIN DimensionMedico M ON
H.keyMedico = M.keyMedico
```

```
WHERE (AC.codigoActividad = '91') AND (S.codigoServicio <>
'I21') AND (YEAR(T.fecha) < 2017)
```

- Vista_Mineria_Proc

Almacena el histórico de citas de la Actividad Procedimientos; utilizando el asistente para la creación de la vista, se generó la siguiente consulta SQL:

```
SELECT A.TipoAsegurado, A.Sexo, A.Edad, A.EstadoCivil,
       M.NombreMedico, P.NombrePoliclinico, U.Departamento,
       U.Provincia, U.Distrito, H.ActoMedicoCitas, H.CitasAtendidas,
       TU.turno, S.NombreServicio
```

```
FROM DimensionPoliclinico P INNER JOIN DimensionAsegurado
A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN
DimensionUbigeo U ON A.keyUbigeo = U.keyUbigeo INNER
JOIN HechoAdmision H ON A.keyAsegurado =
H.keyAsegurado INNER JOIN DimensionActividad AC ON
H.keyActividad = AC.keyActividad INNER JOIN
DimensionServicio S ON H.keyServicio = S.keyServicio INNER
JOIN DimensionTiempo T ON H.keyTiempo = T.keyTiempo
INNER JOIN DimensionTurno TU ON H.keyTurno =
TU.keyTurno INNER JOIN DimensionMedico M ON
H.keyMedico = M.keyMedico
```

```
WHERE (AC.codigoActividad = '96') AND (S.codigoServicio <>
'I21') AND (YEAR(T.fecha) < 2017)
```

Al momento de implementar la solución ambas consultas variarían con respecto a la condición de la fecha, siendo la siguiente:
WHERE (YEAR(T.fecha) < getdate()), el resto de la consulta seguirá igual; asimismo no se está considerando los registros de citas que sean del servicio “Técnico en enfermería” con código I21, pues no existe dicho servicio.

- Vista_Datos_Citas_Con

Almacena los registros, de Actividad Médica Ambulatoria, que serán evaluados para determinar su probabilidad de deserción.

```
SELECT      P.NombrePoliclinico,      A.TipoAsegurado,
            A.NombrePaciente,      A.Sexo,      A.Edad,      A.EstadoCivil,
            U.Departamento, U.Provincia, U.Distrito, H.ActoMedicoCitas,
            H.CitasAtendidas, H.Precio, T.fecha, AC.codigoActividad,
            S.NombreServicio,      M.NombreMedico,      TU.turno,
            A.autogenerado, M.cmp, S.codigoServicio, TU.codigoTurno
FROM DimensionPoliclinico P INNER JOIN DimensionAsegurado
A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN
DimensionUbigeo U ON A.keyUbigeo = U.keyUbigeo INNER
JOIN HechoAdmision H ON A.keyAsegurado =
H.keyAsegurado INNER JOIN DimensionActividad AC ON
H.keyActividad = AC.keyActividad INNER JOIN
DimensionServicio S ON H.keyServicio = S.keyServicio INNER
JOIN DimensionTiempo T ON H.keyTiempo = T.keyTiempo
INNER JOIN DimensionTurno TU ON H.keyTurno =
TU.keyTurno INNER JOIN DimensionMedico M ON
H.keyMedico = M.keyMedico
WHERE (YEAR(T.fecha) = 2017) AND (AC.codigoActividad = '91')
AND (S.codigoServicio <> 'I21')
```

- Vista_Datos_Citas_Proc

Almacena los registros, de Actividad Procedimientos, que serán evaluados para determinar su probabilidad de deserción.

```
SELECT      P.NombrePoliclinico,      A.TipoAsegurado,
            A.NombrePaciente,      A.Sexo,      A.Edad,      A.EstadoCivil,
            U.Departamento, U.Provincia, U.Distrito, H.ActoMedicoCitas,
            H.CitasAtendidas, H.Precio, T.fecha, AC.codigoActividad,
            S.NombreServicio,      M.NombreMedico,      TU.turno,
            A.autogenerado, M.cmp, S.codigoServicio, TU.codigoTurno
FROM DimensionPoliclinico P INNER JOIN DimensionAsegurado
A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN
DimensionUbigeo U ON A.keyUbigeo = U.keyUbigeo INNER
JOIN HechoAdmision H ON A.keyAsegurado =
H.keyAsegurado INNER JOIN DimensionActividad AC ON
H.keyActividad = AC.keyActividad INNER JOIN
DimensionServicio S ON H.keyServicio = S.keyServicio INNER
JOIN DimensionTiempo T ON H.keyTiempo = T.keyTiempo
INNER JOIN DimensionTurno TU ON H.keyTurno =
TU.keyTurno INNER JOIN DimensionMedico M ON
H.keyMedico = M.keyMedico
```

WHERE (YEAR(T.fecha) = 2017) AND (AC.codigoActividad = '96')
AND (S.codigoServicio <> 'I21')

Al momento de implementar la solución ambas consultas variaran con respecto al HECHO a consultar, en lugar de HechoAdmision, será HechoCitas, con respecto a la condición, variaran con respecto a la fecha, quedando de la siguiente manera:

WHERE (T.fecha BETWEEN DATEADD(day, 1, GETDATE())
AND DATEADD(day, 2, GETDATE())), el resto de la consulta seguirá igual.

- Mediante el uso de la plataforma de SQL Server Analysis Services; se agregó un proyecto de analysis services llamado “*OLAP*”, posteriormente se creó un Origen de Datos llamado “*Origen Minería*”.

Establecida la conexión y creadas las vistas, se procedió a crear una Vista de Origen de Datos llamada “*Vista Minería*”, la cual estará compuesta por las cuatro vistas creadas en el SQL Server Management Studio.

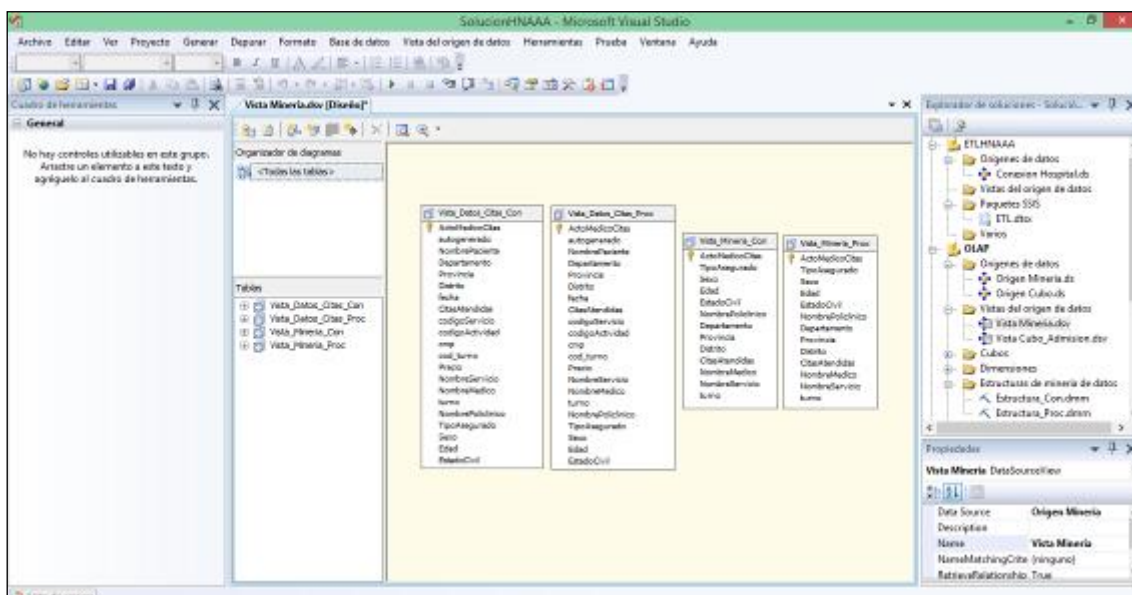


Figura 36: Vista de Origen de Datos

Luego de crear la vista, se creó las estructuras de minería de datos, una para citas de Actividad Médica Ambulatoria y la otra para la Actividad Procedimientos.

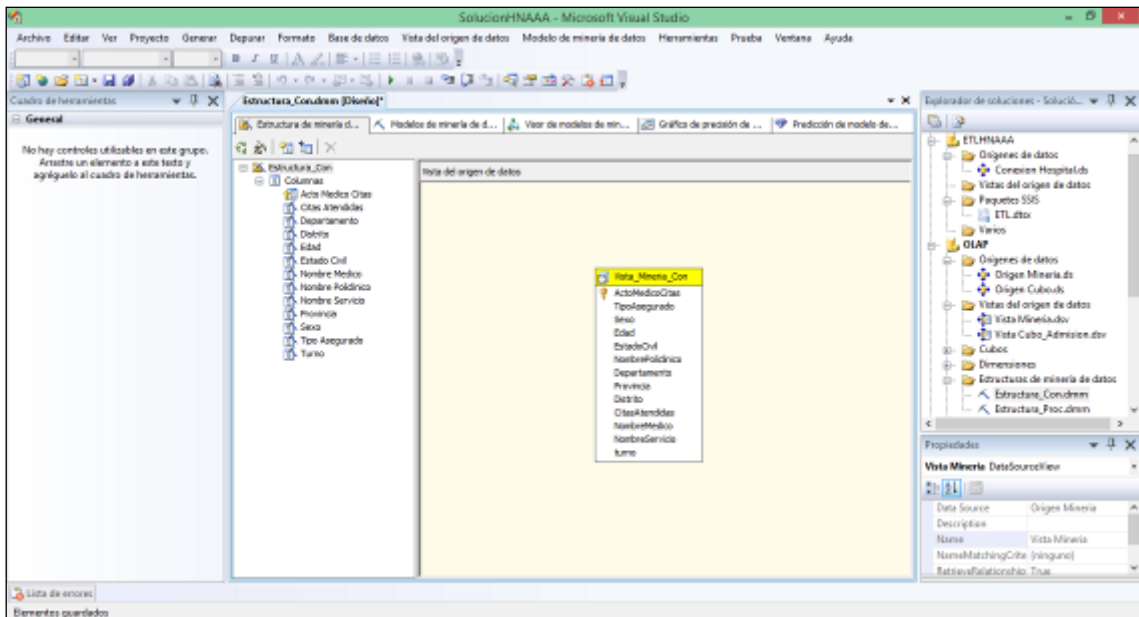


Figura 37: Estructura de Minería de Datos – Estructura_Con

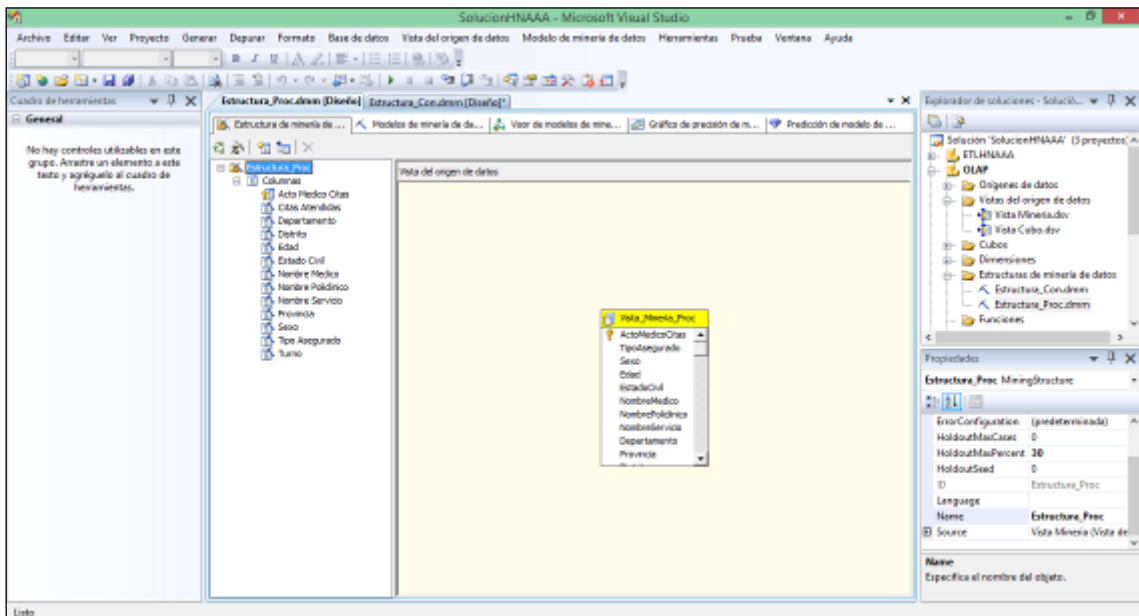


Figura 38: Estructura de Minería de Datos – Estructura_Proc

3. En la estructura de minería de datos, podemos visualizar el modelo de minería de datos a través del Visor de modelos de minería de datos, desde 2 tipos de visores.

- Estructura de Actividad Consulta Ambulatoria – Estructura_Con

- Visor de árboles de Microsoft

En este visor podemos encontrar 2 pestañas:

- Árbol de Decisión

En este tipo de vista, podemos visualizar la estructura del árbol, la cual tiene 32 niveles, nodos o ramificaciones.

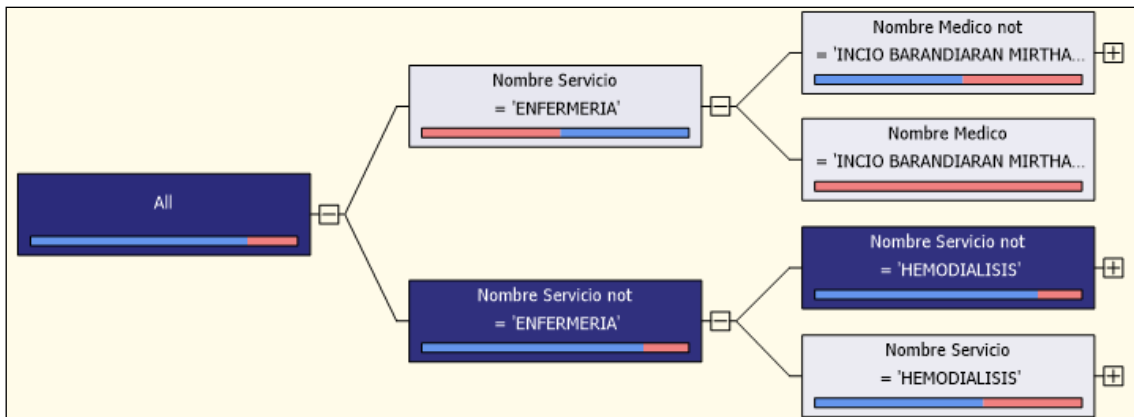


Figura 39: Ramificaciones del árbol de decisión – Actividad Consulta Ambulatoria

- Red de Dependencia
 En esta pestaña podemos ver la influencia de las variables en relación a la deserción; se puede apreciar que la variable más influyente es el Nombre de Servicio.

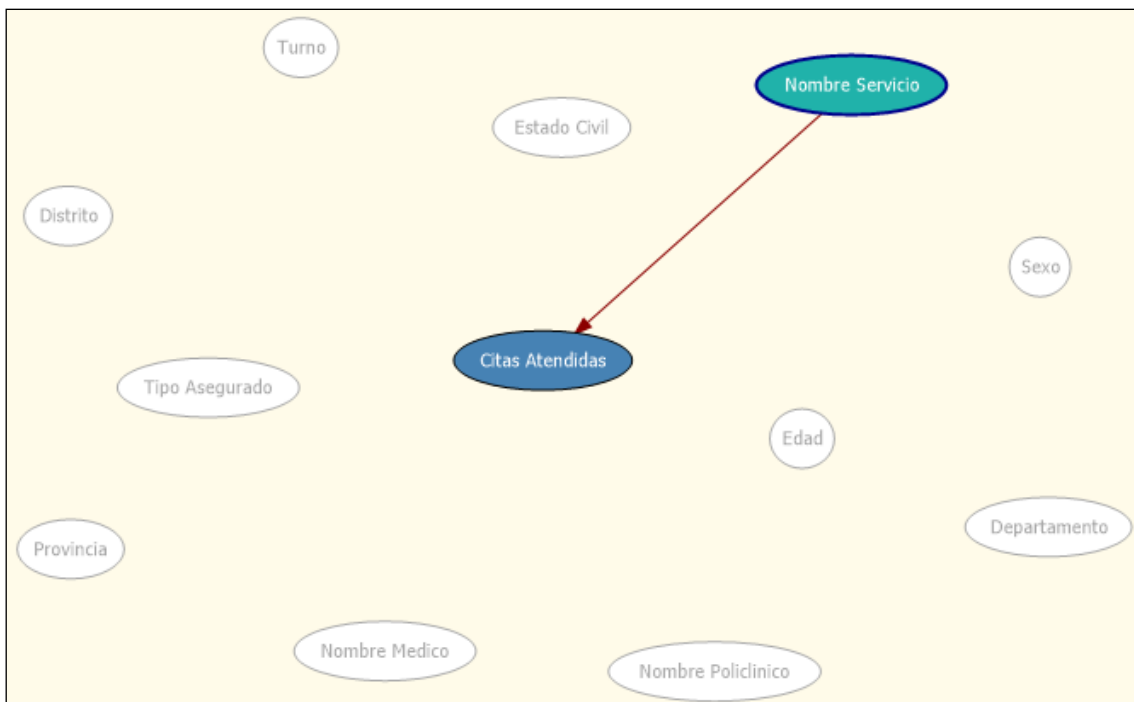


Figura 40: Variable más influyente – Actividad Consulta Ambulatoria

- Visor de árbol de contenido genérico de Microsoft
 En este tipo de vista podemos ver de manera diferente lo mostrado en las figuras anteriores, donde el Servicio más influyente en el modelo de minería es Enfermería.

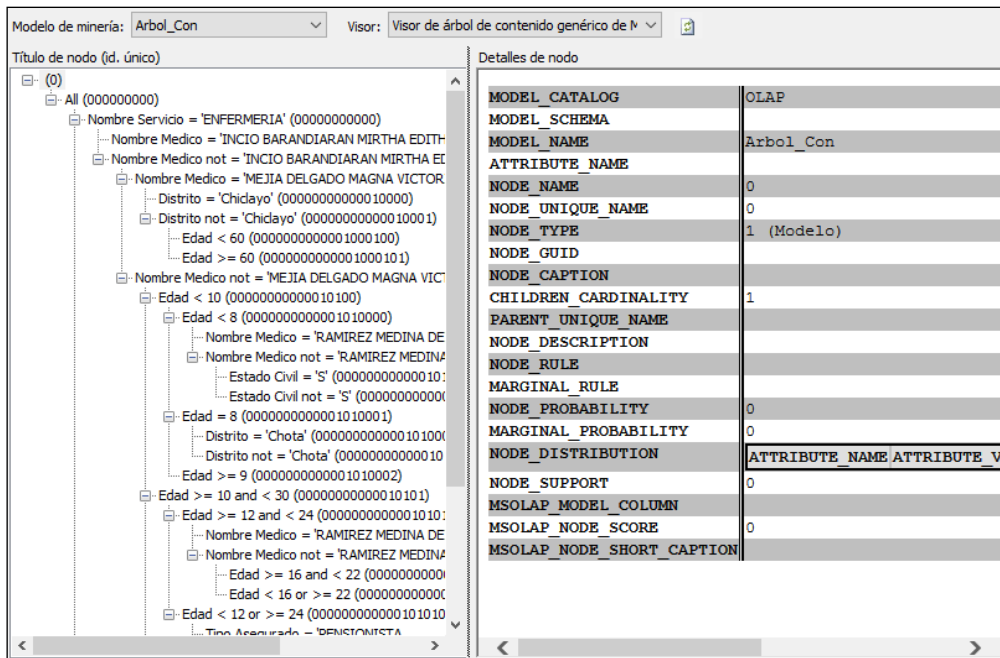


Figura 41: Detalle del nodo más cercano o influyente – Actividad Consulta Ambulatoria

- Estructura de Actividad Procedimientos – Estructura_Proc
 - Visor de árboles de Microsoft
 - En este visor podemos encontrar 2 pestañas:
 - **Árbol de Decisión**

En este tipo de vista, podemos visualizar la estructura del árbol, la cual tiene 45 niveles, nodos o ramificaciones.

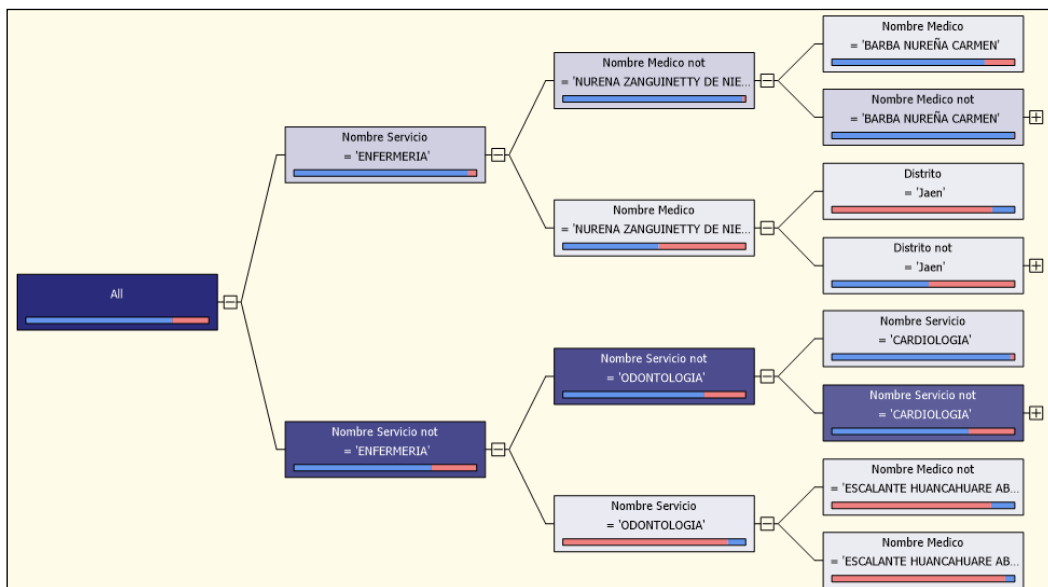


Figura 42: Ramificaciones del árbol de decisión – Actividad Procedimientos

- Red de Dependencia
 - En esta pestaña podemos ver la influencia de las variables en relación a la deserción; se puede apreciar que la variable más influyente es el Nombre de Servicio.

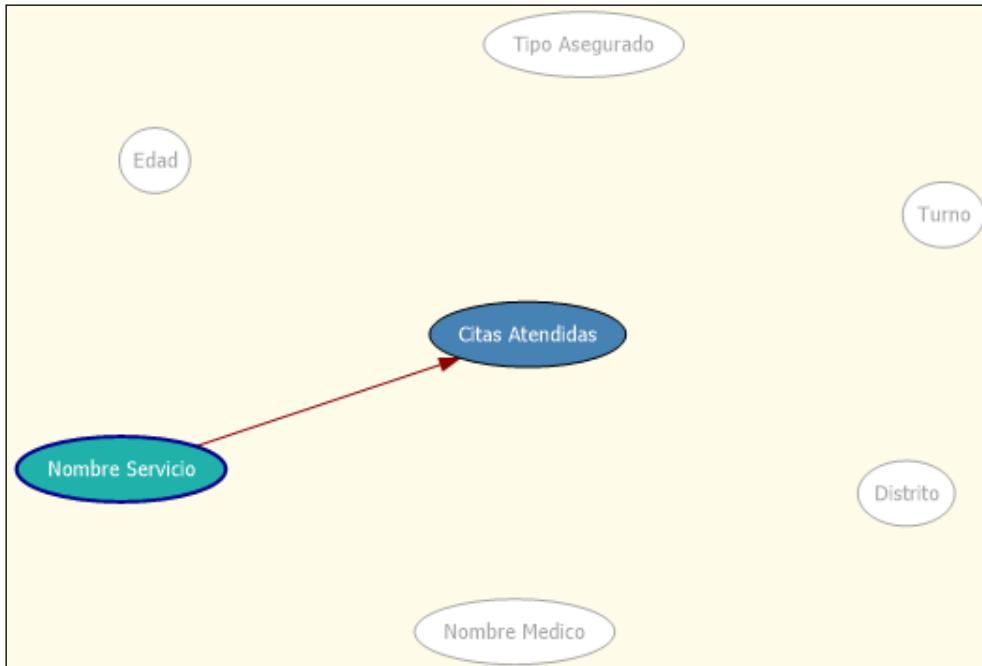


Figura 43: Variable más influyente – Actividad Procedimientos

- Visor de árbol de contenido genérico de Microsoft
En este tipo de vista podemos ver de manera diferente lo mostrado en las figuras anteriores.

Modelo de minería: Arbol_Proc Visor: Visor de árbol de contenido genérico de M

Título de nodo (id. único) Detalles de nodo

- (0)
- All (000000000)
- Nombre Servicio = 'ENFERMERIA' (0000000000)
- Nombre Medico = 'NURENA ZANGUINETTY DE NIET...
- Distrito = 'Jaen' (0000000000000000)
- Distrito not = 'Jaen' (00000000000000001)
- Distrito = 'Pacanga' (0000000000000000100)
- Distrito not = 'Pacanga' (000000000000000001)
- Nombre Medico not = 'NURENA ZANGUINETTY DE N...
- Nombre Medico = 'BARBA NUREÑA CARMEN' (0i...
- Nombre Medico not = 'BARBA NUREÑA CARMEN
- Nombre Medico = 'ROJAS DELGADO EMILIA
- Nombre Medico not = 'ROJAS DELGADO EM
- Nombre Medico = 'LUCERO VALERA ELV
- Nombre Medico not = 'LUCERO VALERA
- Nombre Servicio not = 'ENFERMERIA' (000000000001)
- Nombre Servicio = 'ODONTOLOGIA' (000000000001c
- Nombre Medico = 'ESCALANTE HUANCAHUARE
- Nombre Medico not = 'ESCALANTE HUANCAHUJ
- Nombre Servicio not = 'ODONTOLOGIA' (000000000
- Nombre Servicio = 'CARDIOLOGIA' (0000000000
- Nombre Servicio not = 'CARDIOLOGIA' (0000000
- Nombre Servicio = 'GASTROENTEROLOGIA'
- Nombre Servicio not = 'GASTROENTEROLOX
- Nombre Servicio = 'OBSTETRICIA' (000f
- Nombre Servicio not = 'OBSTETRICIA' (f
- Nombre Servicio = 'MEDICINA FISIC
- Nombre Servicio not = 'MEDICINA F
- Nombre Servicio = 'YTOODPTINI

MODEL_CATALOG	OLAP
MODEL_SCHEMA	
MODEL_NAME	Arbol_Proc
ATTRIBUTE_NAME	
NODE_NAME	0
NODE_UNIQUE_NAME	0
NODE_TYPE	1 (Modelo)
NODE_GUID	
NODE_CAPTION	
CHILDREN_CARDINALITY	1
PARENT_UNIQUE_NAME	
NODE_DESCRIPTION	
NODE_RULE	
MARGINAL_RULE	
NODE_PROBABILITY	0
MARGINAL_PROBABILITY	0
NODE_DISTRIBUTION	ATTRIBUTE_NAME ATTRIBUTE_VALU
NODE_SUPPORT	0
MSOLAP_MODEL_COLUMN	
MSOLAP_NODE_SCORE	0
MSOLAP_NODE_SHORT_CAPTION	

Figura 44: Detalle del nodo más cercano o influyente – Actividad Procedimientos

4. Creada las estructuras de minería de datos, se procede a realizar la consulta DMX (Referencia de Extensiones de minería de datos), en la pestaña Predicción de modelo de minería de datos, con la finalidad de obtener los pacientes con probabilidad de deserción.

Para esto se cuenta, a su vez, con 3 tipos de vista para elaborar la consulta DMX: diseño, consulta y resultado; una vez creado la consulta, ya sea en la vista de diseño o consulta, la relación de pacientes con probabilidad de deserción se muestra en la pestaña resultado.

- Atención de Consulta Ambulatoria
En el cuadro “Seleccionar tabla(s) de entrada”, seleccionamos la vista Vista_Datos_Citas_Con.
En este tipo de predicción se filtra o selecciona a aquellos pacientes que tenga una probabilidad de deserción mayor al 50%.

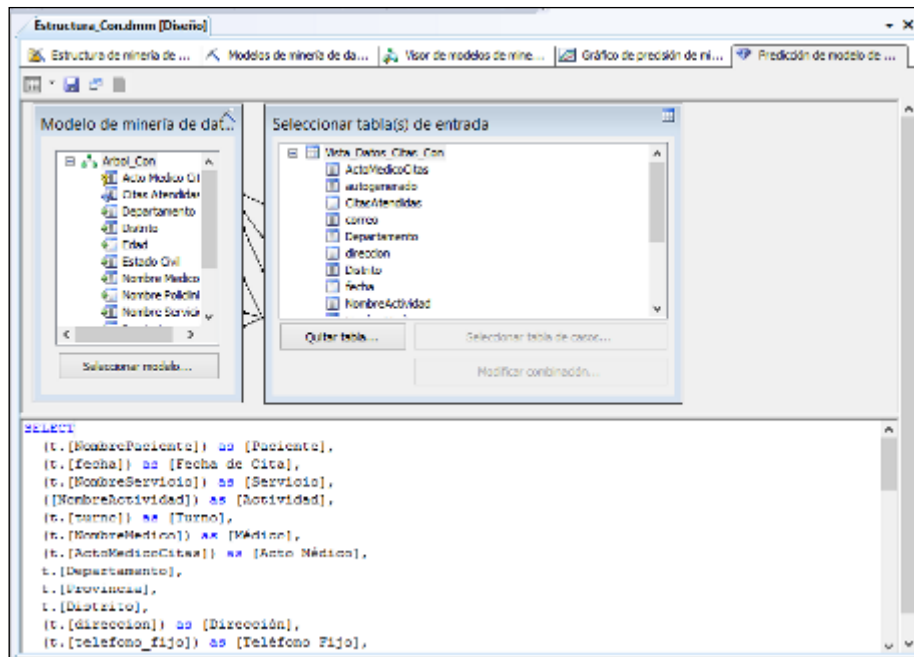


Figura 45: Elaboración de la consulta DMX – Actividad Consulta Ambulatoria

Creada la consulta, se visualiza el resultado.

Autogenerado	ActoMedicoCitas	Fecha	CitasAtendidas	codigoServicio	codigoActividad	ano	cod_turno	Precio	% Deserción
5211880286	9119826	07/05/2014 12...	0	F11	91	1318	78	42	59.99997538...
8395189386	9031887	13/01/2014 12...	0	F11	91	1348	46	40	59.99997538...
64051203ACT	9120171	07/03/2014 12...	0	F11	91	1348	46	40	59.99997538...
74051204CCO	9030382	06/01/2014 12...	0	F11	91	1318	82	42	59.99997538...
50083189SQ	9021727	15/01/2014 12...	0	F11	91	1318	78	42	59.99997538...
5815018495	9029688	06/01/2014 12...	0	F11	91	1348	46	40	59.99997538...
46102719GCT	9119130	13/03/2014 12...	0	F11	91	1348	46	40	59.99997538...
74010020GA	9030419	05/01/2014 12...	0	F11	91	1318	82	42	59.99997538...
02100189TU	9021380	14/01/2014 12...	0	F11	91	1318	78	42	59.99997538...
91100189AM	9029376	14/01/2014 12...	0	F11	91	1348	46	40	59.99997538...
20010497LE	9030060	20/01/2014 12...	0	F11	91	1348	46	40	59.99997538...
04100189TA	9037722	03/01/2014 12...	0	F11	91	1318	82	42	59.99997538...
07110000CC	9030026	28/01/2014 12...	0	F11	91	1318	78	42	59.99997538...
90030886CA	9039328	02/01/2014 12...	0	F11	91	1318	78	42	59.99997538...
441101031007	9121965	15/01/2014 12...	0	F11	91	1348	46	40	59.99997538...
40040189OG	9110859	13/03/2014 12...	0	F11	91	1348	46	40	59.99997538...
51020000DN	9110850	11/03/2014 12...	0	F11	91	1318	82	42	59.99997538...
54020189MO	9111028	07/05/2014 12...	0	F11	91	1318	78	42	59.99997538...
22100189PA	8996011	03/01/2014 12...	0	F11	91	1348	46	40	59.99997538...
56100189WA	9033256	03/01/2014 12...	0	F11	91	1348	46	40	59.99997538...
33061089SA	9017020	05/01/2014 12...	0	F11	91	1318	82	42	59.99997538...
3401070310D	9181023	20/05/2014 12...	0	F11	91	1318	78	42	59.99997538...
43040000PAP	9049126	15/01/2014 12...	0	F11	91	1348	46	40	59.99997538...

Figura 46: Relación de pacientes con probabilidad de deserción – Actividad Consulta Ambulatoria

- Atención de Procedimientos
 En el cuadro “Seleccionar tabla(s) de entrada”, seleccionamos la vista Vista_Datos_Citas_Proc.
 En este tipo de predicción se filtra o selecciona a aquellos pacientes que tenga una probabilidad de deserción mayor al 70%.

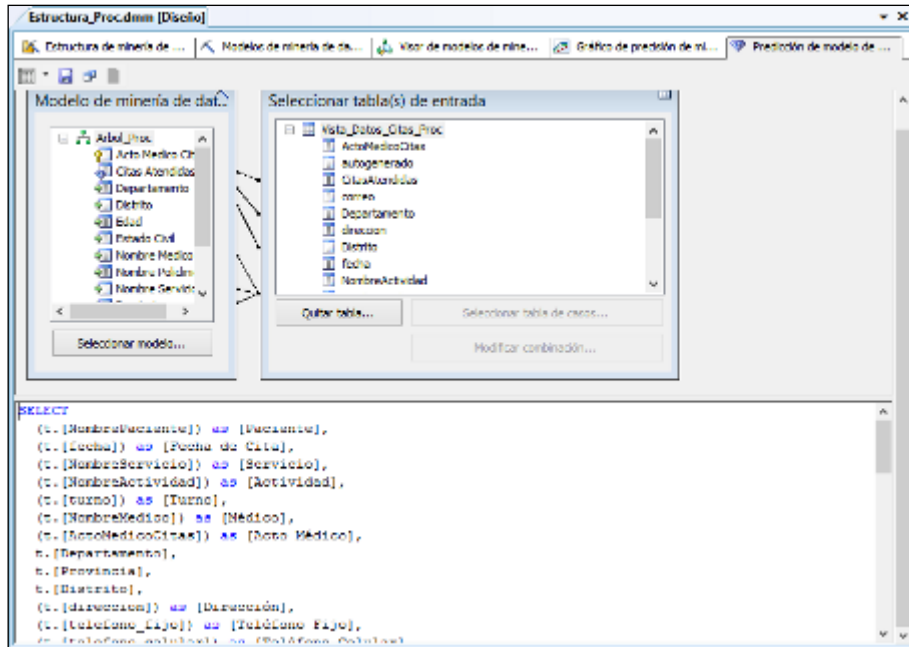


Figura 47: Elaboración de la consulta DMX – Actividad Procedimientos

Creada la consulta, se visualiza el resultado.

auto generado	ActoMedicoCitas	fecha	CitasAtendidas	codigoServicio	codigoActividad	cmp	cod_turno	Precio	% Deserción
430912SPBMD1...	9089979	21/02/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9095516	11/07/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9099907	28/02/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9269072	27/06/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9296225	13/06/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9242967	06/06/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9036587	24/01/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9018665	16/01/2014 12...	0	AD2	96	35047	13	42	99.99999477...
430912SPBMD1...	9263338	26/06/2014 12...	0	AD2	96	35047	13	42	99.99999477...
940228ICRAN...	9228886	23/05/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9089995	21/02/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9031235	22/01/2014 12...	0	AD2	96	35047	18	42	99.999997325...
940228ICRAN...	9188297	25/04/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9242976	06/06/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9267905	02/07/2014 12...	0	AD2	96	35047	18	42	99.999997325...
940228ICRAN...	9154875	04/04/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9295536	11/07/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9119695	14/03/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9211665	09/05/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9132353	21/03/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9301266	16/07/2014 12...	0	AD2	96	35047	18	42	99.999997325...
940228ICRAN...	9028477	17/01/2014 12...	0	AD2	96	35047	13	42	99.999997325...
940228ICRAN...	9142988	28/03/2014 12...	0	AD2	96	35047	13	42	99.999997325...

Figura 48: Relación de pacientes con probabilidad de deserción – Actividad Procedimientos

5. Una vez definida la consulta DMX a utilizar, en el entorno del ETL, se agrega 4 controles de *Tarea de Consulta de minería de datos*.

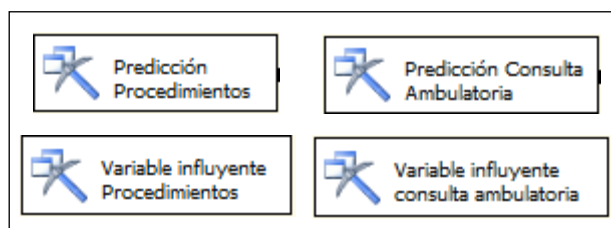


Figura 49: Flujo de Control - Tareas de Consulta de minería de datos

- **Predicción Procedimientos**
Mediante este flujo de control procedemos a poblar una tabla temporal llamada *Temporal_Proc*; con los datos de los posibles pacientes desertores.
Haciendo doble clic en el control, nos aparece la siguiente ventana:

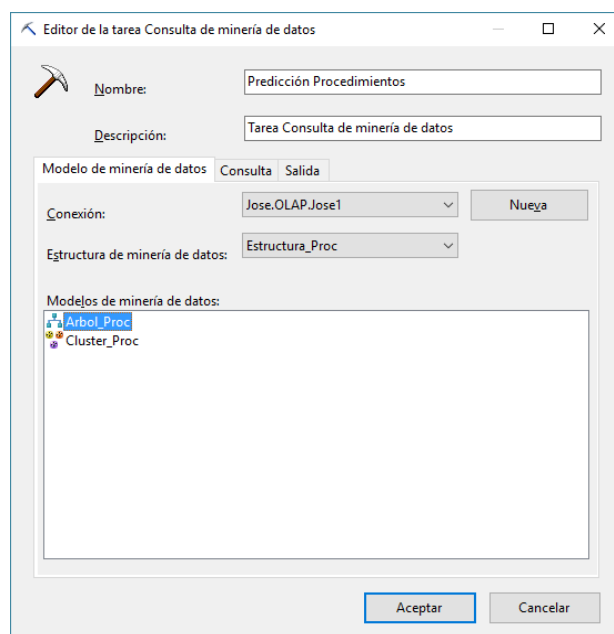


Figura 50: Selección del modelo de minería de datos – Actividad Procedimientos

- En la pestaña **Modelo de minería de datos**, elegimos nuestra conexión de Analysis Services, la estructura de minería de datos, *Estructura_Proc*, y nuestro modelo de minería de datos: *Arbol_Proc*.

- En la pestaña Consulta, pegamos la consulta DMX que elaboramos en el paso anterior, para este caso será la siguiente:

```

SELECT
    t.[autogenerado],
    t.[ActoMedicoCitas],
    t.[fecha],
    t.[CitasAtendidas],
    t.[codigoServicio],
    t.[codigoActividad],
    t.[cmp],
    t.[codigoTurno],
    t.[Precio],
    (PredictProbability([Arbol_Proc].[Citas Atendidas],0)*100) as [%
Deserción]
From
    [Arbol_Proc]
PREDICTION JOIN
    OPENQUERY([Origen Minería],
        'SELECT
            [autogenerado],
            [ActoMedicoCitas],
            [fecha],
            [CitasAtendidas],
            [codigoServicio],
            [codigoActividad],
            [cmp],
            [codigoTurno],
            [Precio],
            [Departamento],
            [Provincia],
            [Distrito],
            [NombreServicio],
            [NombreMedico],
            [turno]
        FROM
            [dbo].[Vista_Datos_Citas_Proc]
        ') AS t
    ON
        [Arbol_Proc].[Departamento] = t.[Departamento] AND
        [Arbol_Proc].[Provincia] = t.[Provincia] AND
        [Arbol_Proc].[Distrito] = t.[Distrito] AND
        [Arbol_Proc].[Citas Atendidas] = t.[CitasAtendidas] AND
        [Arbol_Proc].[Nombre Servicio] = t.[NombreServicio] AND
        [Arbol_Proc].[Nombre Medico] = t.[NombreMedico] AND
        [Arbol_Proc].[Turno] = t.[turno]
Where
    (PredictProbability([Arbol_Proc].[Citas
Atendidas],0)*100)>70
ORDER BY PredictProbability([Arbol_Proc].[Citas Atendidas],0)
desc

```

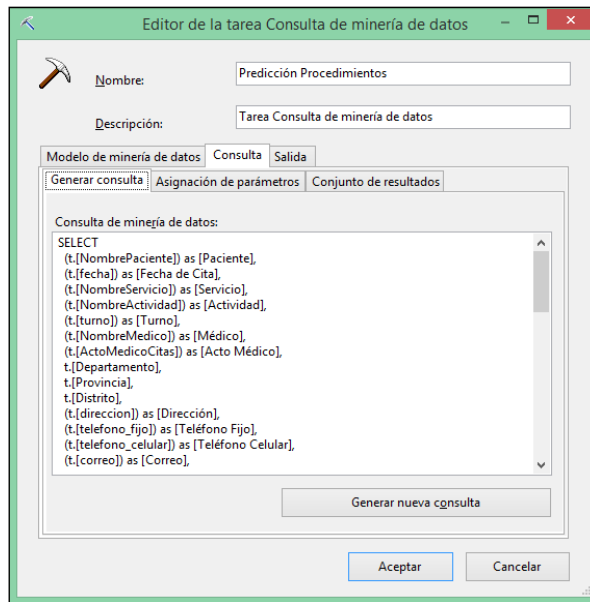


Figura 51: Consulta DMX – Actividad Procedimientos

- En la pestaña Salida, establecemos nuestra conexión de nuestro datamart “DatamartDesercion”, ingresamos el nombre de la tabla de salida (si no existiese la tabla, en este paso la crea), y por último, marcamos la casilla que dice “Quitar y volver a crear la tabla de salida”, pues es una tabla temporal; es por eso que la llamamos Temporal_Proc.

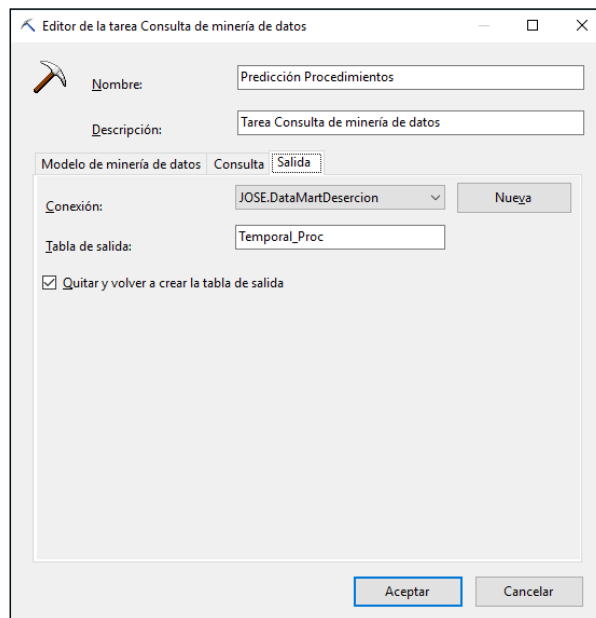


Figura 52: Salida del resultado de la consulta DMX – Actividad Procedimientos

- **Predicción Consulta Ambulatoria**
Mediante este flujo de control procedemos a poblar una tabla temporal llamada *Temporal_Con*; con los datos de los posibles pacientes desertores.
Haciendo doble clic en el control, nos aparece la siguiente ventana:

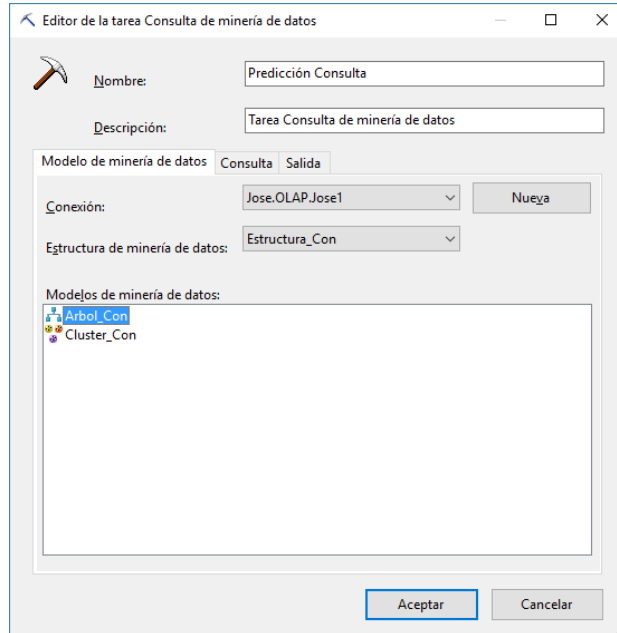


Figura 53: Selección del modelo de minería de datos – Actividad Consulta Ambulatoria

- En la pestaña Modelo de minería de datos, elegimos nuestra conexión de Analysis Services, la estructura de minería de datos, Estructura_Con, y nuestro modelo de minería de datos: Arbol_Con.
- En la pestaña Consulta, pegamos la consulta DMX que elaboramos en el paso anterior, para este caso será la siguiente:

```

SELECT
  t.[autogenerado],
  t.[ActoMedicoCitas],
  t.[fecha],
  t.[CitasAtendidas],
  t.[codigoServicio],
  t.[codigoActividad],
  t.[cmp],
  t.[codigoTurno],
  t.[Precio],
  (PredictProbability([Arbol_Con].[Citas Atendidas],0)*100) as [%
Deserción]
From
  [Arbol_Con]
PREDICTION JOIN
  OPENQUERY([Origen Minería],
    'SELECT
      [autogenerado],
      [fecha],
  
```

```

[codigoActividad],
[codigoServicio],
[codigoTurno],
[cmp],
[ActoMedicoCitas],
[Precio],
[CitasAtendidas],
[NombrePoliclinico],
[TipoAsegurado],
[Sexo],
[Edad],
[EstadoCivil],
[Departamento],
[Provincia],
[Distrito],
[NombreServicio],
[NombreMedico],
[turno]
FROM
    [dbo].[Vista_Datos_Citas_Con]
) AS t
ON
    [Arbol_Con].[Nombre Policlínico] = t.[NombrePoliclinico] AND
    [Arbol_Con].[Tipo Asegurado] = t.[TipoAsegurado] AND
    [Arbol_Con].[Sexo] = t.[Sexo] AND
    [Arbol_Con].[Edad] = t.[Edad] AND
    [Arbol_Con].[Estado Civil] = t.[EstadoCivil] AND
    [Arbol_Con].[Departamento] = t.[Departamento] AND
    [Arbol_Con].[Provincia] = t.[Provincia] AND
    [Arbol_Con].[Distrito] = t.[Distrito] AND
    [Arbol_Con].[Citas Atendidas] = t.[CitasAtendidas] AND
    [Arbol_Con].[Nombre Servicio] = t.[NombreServicio] AND
    [Arbol_Con].[Nombre Medico] = t.[NombreMedico] AND
    [Arbol_Con].[Turno] = t.[turno]
where
    (PredictProbability([Arbol_Con].[Citas
Atendidas],0)*100)>50
ORDER BY PredictProbability([Arbol_Con].[Citas Atendidas],0)
desc

```

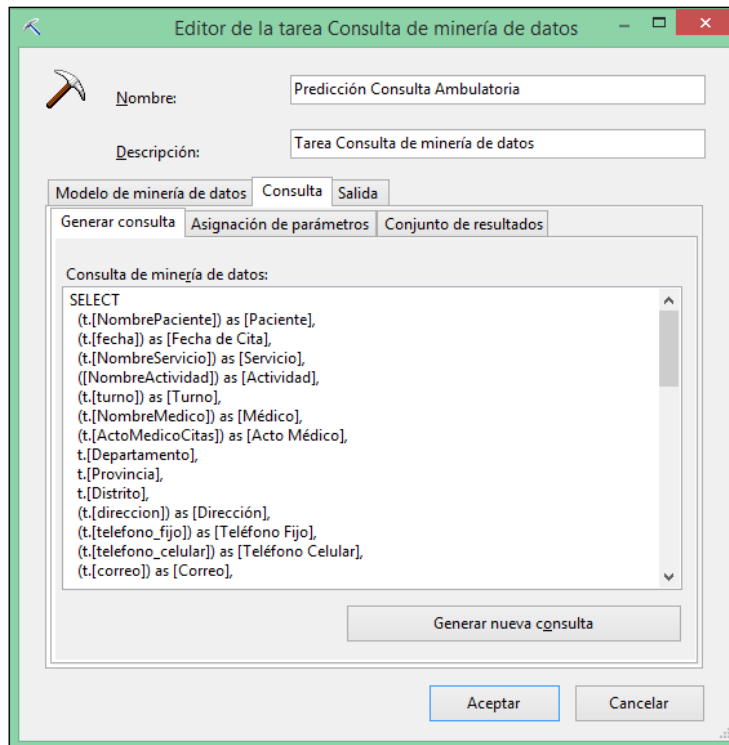


Figura 54: Consulta DMX – Actividad Consulta Ambulatoria

- En la pestaña Salida, establecemos nuestra conexión de nuestro datamart “DatamartDesercion”, ingresamos el nombre de la tabla de salida (si no existiese la tabla, en este paso la crea), y por último, marcamos la casilla que dice “Quitar y volver a crear la tabla de salida”, pues es una tabla temporal; es por eso que la llamamos Temporal_Con.

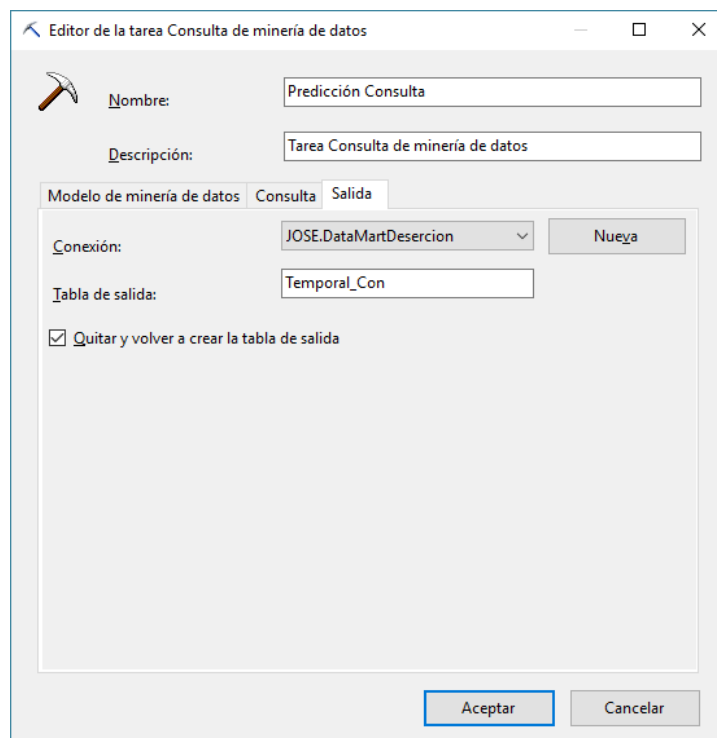


Figura 55: Salida del resultado de la consulta DMX – Actividad Consulta Ambulatoria

- Variable influyente Procedimientos
Mediante este flujo de control procedemos a poblar una tabla temporal llamada *Influyente_Proc*; con los datos de la variable más influyente.

- En la pestaña Modelo de minería de datos, elegimos nuestra conexión de Analysis Services, la estructura de minería de datos, Estructura_Proc, y nuestro modelo de minería de datos: Arbol_Proc.
- En la pestaña Consulta, pegamos la consulta DMX que elaboramos en el paso anterior, para este caso será la siguiente:
SELECT TOP 3 NODE_CAPTION
FROM [Arbol_Proc].content
WHERE [PARENT_UNIQUE_NAME]>'0' and
[PARENT_UNIQUE_NAME]<='00000000001'
ORDER BY MSOLAP_NODE_SCORE DESC

Donde:

- NODE_CAPTION = etiqueta del nodo
- PARENT_UNIQUE_NAME = Nombre único del nodo primario del nodo.
- MSOLAP_NODE_SCORE = importancia del nodo en el modelo.
- En la pestaña Salida, establecemos nuestra conexión de nuestro datamart “DatamartDesercion”, ingresamos el nombre de la tabla de salida (si no existiese la tabla, en este paso la crea), y por último, marcamos la casilla que dice “Quitar y volver a crear la tabla de salida”, pues es una tabla temporal, pues cada vez que se ejecute el modelo de minería de datos determinará la variable más influyente, según sea el caso. A la tabla la llamamos *Influyente_Proc*.

- Variable influyente Consulta Ambulatoria
Mediante este flujo de control procedemos a poblar una tabla temporal llamada *Influyente_Con*; con los datos de la variable más influyente.

- En la pestaña Modelo de minería de datos, elegimos nuestra conexión de Analysis Services, la estructura de minería de datos, Estructura_Con, y nuestro modelo de minería de datos: Arbol_Con.
- En la pestaña Consulta, pegamos la consulta DMX que elaboramos en el paso anterior, para este caso será la siguiente:
SELECT TOP 3 NODE_CAPTION
FROM [Arbol_Con].content
WHERE [PARENT_UNIQUE_NAME]>'0' and
[PARENT_UNIQUE_NAME]<='00000000001'
ORDER BY MSOLAP_NODE_SCORE DESC

Donde:

- NODE_CAPTION = etiqueta del nodo
- PARENT_UNIQUE_NAME = Nombre único del nodo primario del nodo.
- MSOLAP_NODE_SCORE = importancia del nodo en el modelo.

- En la pestaña Salida, establecemos nuestra conexión de nuestro datamart “DatamartDesercion”, ingresamos el nombre de la tabla de salida (si no existiese la tabla, en este paso la crea), y por último, marcamos la casilla que dice “Quitar y volver a crear la tabla de salida”, pues es una tabla temporal, pues cada vez que se ejecute el modelo de minería de datos determinará la variable más influyente, según sea el caso. A la tabla la llamamos Influyente_Con.

Al ejecutarse las tareas se pueblan las tablas antes mencionadas.

6. Al tener las tablas pobladas, se procede a poblar las tablas con el resultado final, tanto para la actividad de Consulta Ambulatoria como para Procedimientos.

El proceso se realiza en el ETL, de la siguiente manera:

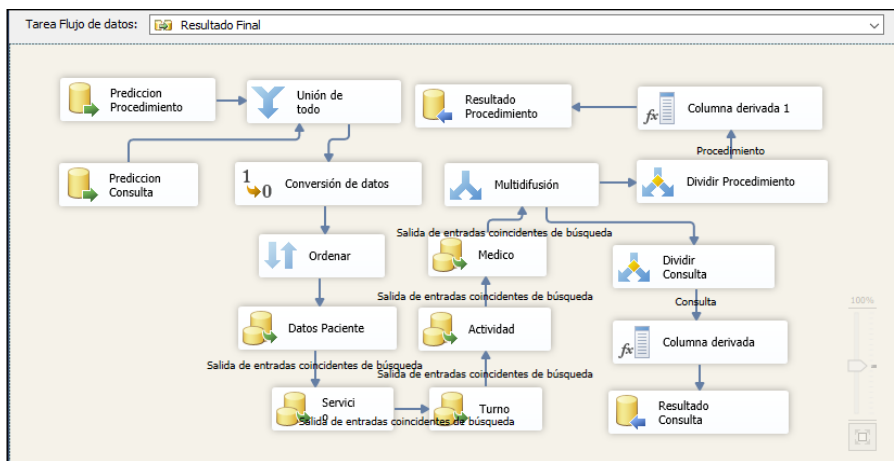


Figura 56: Flujo de datos “Resultado Final”

Donde:

- Predicción Consulta
Consulta los registros de la tabla Temporal_Con.
- Predicción Procedimientos
Consulta los registros de la tabla Temporal_Proc.
- Unión de todo
Une los registros de las tablas Temporal_Con y Temporal_Proc.
- Luego sigue el proceso ETL, combinando datos con las dimensiones.
- Resultado Procedimiento
Hace la población de la tabla *Resultado Procedimiento*, donde se almacenará el resultado final de la predicción de la actividad procedimiento.
- Resultado Consulta

Hace la población de la tabla *Resultado Consulta*, donde se almacenará el resultado final de la predicción de la actividad consulta ambulatoria.

Con todo el proceso seguido hasta el momento, el Flujo de Control, resultó como lo muestra la siguiente imagen:

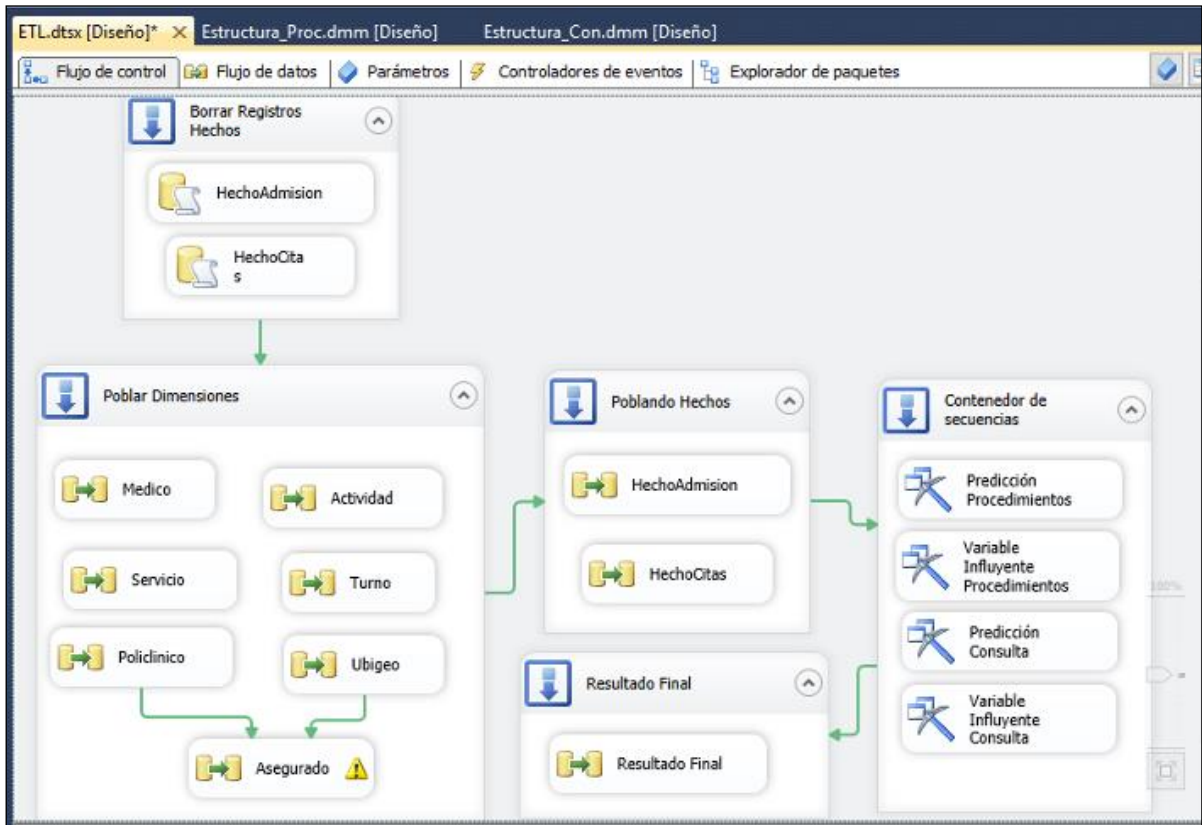


Figura 57: Flujo de Control Final

4.4.4. Evaluar el modelo

Los modelos de árboles de decisión, de ambas estructuras (Estructura_Con y Estructura_Proc) se han acercado mucho al modelo ideal, superando al modelo de clustering con el cual se comparó su probabilidad de predicción.

Sin embargo en la Estructura_Con, la diferencia entre el modelo de árbol de decisión con el de clustering es muy pequeña, con un 83.26% frente a un 81.05%, respectivamente; por lo que sería recomendable verificar esta diferencia cuando se utilice la presente aplicación, asimismo tuvo un 39.24% de certeza de pronóstico en el grupo de prueba.

Por otro lado el modelo de árbol de decisión de la Estructura_Proc, obtuvo un 96.52% de probabilidad de predicción frente a un 79.36% del modelo clustering, asimismo se encuentra más cerca al modelo ideal y tuvo un 39.97% de certeza de pronóstico en el grupo de prueba.

4.5. EVALUACIÓN

Evaluar los resultados realizar el proceso de revisión y determinar futuras fases, son las tareas a detallar en esta fase.

4.5.1. Evaluar los resultados

Teniendo en cuenta los resultados obtenidos en el plan de prueba, se puede determinar que:

- Actividad Consulta Ambulatoria

En el periodo del grupo de prueba, se pudo haber disminuido el índice de deserción de un 39.89% a un 38.89%, lo que representaría un aprovechamiento de S/. 110,670.00 del presupuesto destinado a la consulta ambulatoria y se pudo haber brindado citas a 2,635 pacientes que se encuentren en lista, brindando así una atención oportuna.

- Actividad Procedimientos

En el periodo del grupo de prueba, se pudo haber disminuido el índice de deserción de un 40.03% a un 10.24%, lo que representaría un aprovechamiento de S/. 1'247,400.00 del presupuesto destinado a la actividad de procedimientos y se pudo haber brindado citas a 29,700 pacientes que se encuentren en lista, brindando así una atención oportuna.

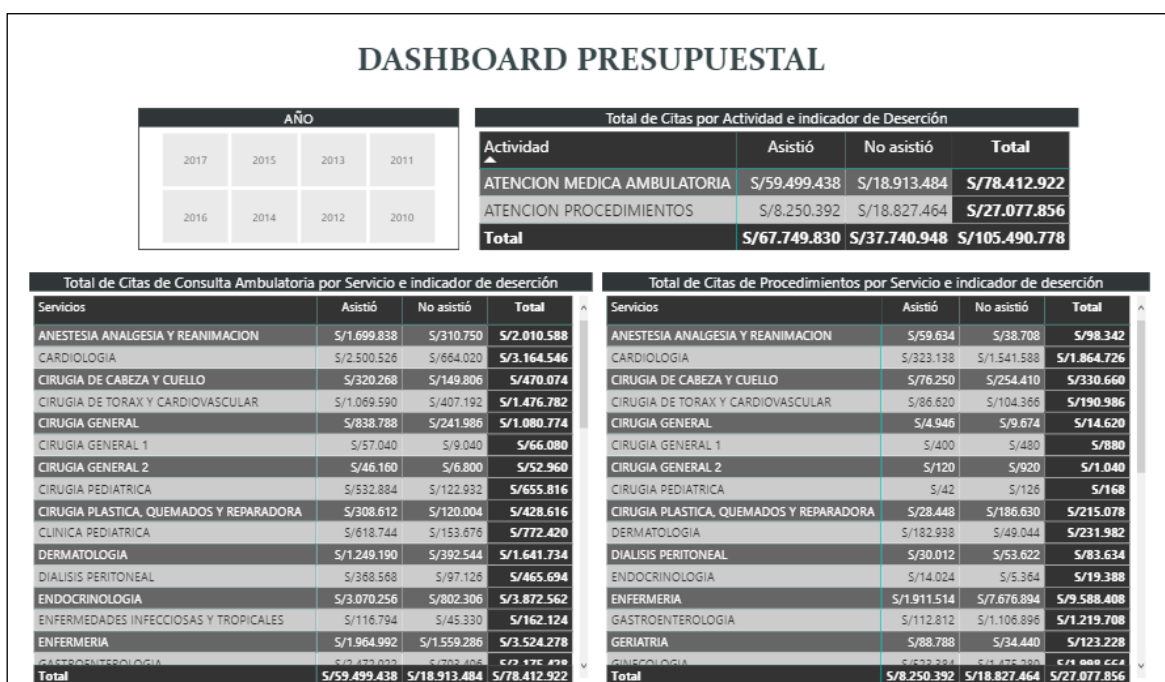


Figura 58: Dashboard Presupuestal

4.5.2. Proceso de Revisión

Se ha desarrollado el proceso de minería de datos, siguiendo la metodología CRISP-DM, respetando cada una de sus fases, las cuales, según mi apreciación, son muy detalladas, pues permite alinear los objetivos del negocio con los objetivos de minería de datos, incluyendo como tareas la evaluación de los resultados del modelo y de la consecución de los objetivos del negocio.

4.5.3. Determinación de futuras fases

Según lo desarrollado, todas las fases han contribuido a desarrollar satisfactoriamente la alternativa de solución que se había planteado; con la implementación y la puesta en marcha en la institución, se determinará si es necesario determinar nuevas fases.

4.6. IMPLEMENTACIÓN

Esta última fase está compuesta por el plan de implementación, plan de monitoreo y mantenimiento, el informe final y la revisión del proyecto.

4.6.1. Plan de implementación

El resultado de la presente aplicación de minería de datos, se ha desplegado a través de hojas de cálculo de la aplicación Microsoft Excel 2013:

- Hoja de Cálculo Consulta / Procedimientos

En esta hoja de cálculo se muestra el resultado de los pacientes con un 50% a más de probabilidad de desertar a su cita en la actividad de Consulta Ambulatoria / Procedimientos.

Asimismo se cuenta con un menú que sirve de filtro a través del campo "Servicio".

A esta hoja de cálculo tendrá acceso el personal de Módulo, el cual llamará a los pacientes para confirmar su asistencia; si el paciente confirma su asistencia el personal de Módulo editará el campo "Respuesta" con el valor 1, si por el contrario, el paciente no asistirá a su cita se modifica el mismo campo con el valor 0, procederá a eliminar la cita del paciente y llamará a un paciente que se encuentre en lista de espera.

El resultado de la aplicación demostró que, mediante la aplicación del presente modelado, se pudo haber reducido la tasa de deserción a la consulta médica, en el periodo del grupo de control, de un 39.89% a un 38.89%, en la actividad de Atención de Consulta Ambulatoria; y de un 40.03% a un 10.24% en la actividad de Atención de Procedimientos.

4.6.4. Revisión del proyecto

Durante el desarrollo del proyecto, se tuvo en consideración algunos puntos:

- En el proceso ETL, la limpieza fue muy específica, por ejemplo en el caso de limpiar y/o corregir colegiaturas erradas se tuvo que verificar uno a uno en las diversas páginas web de los diferentes colegios profesionales (médico, enfermeras, obstétricas, psicólogos, tecnólogos médicos, entre otros); incluso buscar coincidencias entre las colegiaturas registradas con las verdaderas.
- En los datos registrados no se puede identificar, por ejemplo, a qué tipo de procedimiento se le ha citado a los pacientes, por lo que el precio por esa actividad (procedimiento) sea cual sea el servicio se generaliza; siendo, en la realidad, precios diferentes según el tipo de procedimiento o servicio, por lo que puede que los reportes no se ajusten 100% a la realidad con respecto a este punto. Con respecto a este tipo de inconveniente se trató de respetar el precio que estaba registrado en cada cita.
- El desarrollo del proyecto se enfocó sólo en las actividades de Consulta Ambulatoria y Procedimientos, pues son las más demandadas.
- Para determinar la edad del paciente, no se consideró el dato consignado en el campo edad de la tabla adtcitas, pues no se encuentra actualizado; por el contrario se consideró la estructura del autogenerado para calcular la edad de cada paciente.
- En el resultado del modelado, se puede observar que no todos los pacientes tienen registrado algún número telefónico y/o correo, esto debido a que no se realizó el registro de dichos datos en la tabla adtelcom.
- Se consideró registrar los datos en el HechoAdmision, provenientes de las dimensiones creadas, en 2 instancias: una en la que, en una fecha determinado, se registren todos los datos históricos, y la otra, dos días posterior a la fecha en la que se registró el histórico para así poder registrar con 2 días de retraso los datos. Esto permitirá tener una “ventana de tiempo” para prevenir el registro de datos desactualizados o estar preparados ante contingencias como falta de registro de datos por fallas de energía eléctrica.
- En el HechoCitas, se registra los datos con 2 días de anticipación, pues se consideró un tiempo razonable para realizar las llamadas de confirmación de citas y la asignación de éstas a pacientes en lista de espera.

V. DISCUSIÓN

En este capítulo se hará un análisis de los resultados de la aplicación del sistema desarrollado, en el hospital en estudio. El análisis estará enfocado a evaluar los indicadores planteados en el capítulo III. Estos indicadores son:

- Índice de citas no atendidas.
- Presupuesto desaprovechado para la consulta externa.
- Índice de citas otorgadas por la deserción.

La población de 2'168,828 citas (104,378 pacientes), considerada entre los años 2010 a 2016; ayudó a definir el perfil de los pacientes desiertos.

Para la evaluación del trabajo, se analiza los datos de estos indicadores, con el pronóstico de deserción, utilizando la minería de datos.

El resultado es analizado en las 2 actividades con mayor demanda, que son Consulta Ambulatoria y Procedimientos; considerando la muestra comprendida en el periodo de enero a diciembre de 2017 representada por 365,416 citas médicas y 47,228 pacientes.

- Actividad Consulta Ambulatoria: con un total de 264,213 citas (45,694 pacientes).
- Actividad de Procedimientos: con un total de 99,704 citas (21,337 pacientes). Ver Figura 35.

Índice de deserción.

El total de citas en la muestra, como ya vimos líneas superiores, es de 365,416; el índice de deserción en nuestra muestra es de 39.92% que hace un total de 145,878 citas perdidas a causa de la deserción.

A continuación se detalla el comportamiento de este indicador en las dos actividades seleccionadas: Consulta ambulatoria y procedimientos.

Antes de aplicar minería de datos en el periodo de la muestra, la deserción fue de 105,393 que representa un 25.21%.

Tabla 31: Índice de deserción antes de minería de datos

	Consulta Ambulatoria	Procedimientos
Total de Citas	264,213	99,704
Citas no atendidas (deserción)	105,393	39,909
Porcentaje de deserción	39.89%	40.03%

Con la aplicación de minería de datos se obtuvo los siguientes resultados:

Tabla 32: Índice de deserción después de aplicar minería de datos

	Consulta Ambulatoria	Procedimientos
Total de Citas	264,213	99,704
Citas no atendidas (deserción)	105,393	39,909
Deserción identificada	2,635	29,700
Total de deserción	102,758	10,209
Porcentaje de deserción	38.89%	10.24%

Como se puede apreciar en las Tablas 31 y 32, el índice de deserción disminuye de 105,393 a 102,758 citas no atendidas, representando una disminución de 1.00% en la actividad de Consulta Ambulatoria; mientras que en la actividad de Procedimientos este índice disminuye de 39,909 a 10,209 citas no atendidas, representando una disminución de un 29.79%.

El total de citas y la cantidad citas no atendidas se pueden apreciar en la Figura 35.

Presupuesto desaprovechado para la consulta externa.

El presupuesto desaprovechado es calculado según el total de citas no atendidas por deserción y teniendo como precio referencial por consulta S/. 42.00.

Tabla 33: Presupuesto desaprovechado para la Consulta Externa antes de la minería de datos

	Consulta Ambulatoria		Procedimientos	
	Citas	Presupuesto no aprovechado	Citas	Presupuesto no aprovechado
Total de Citas	264,213	-	99,704	-
Citas no atendidas (deserción)	105,393	S/.4,426,506.00	39,909	S/.1,676,178.00
Deserción identificada	2,635	S/.110,670.00	29,700	S/.1,247,400.00
Total de deserción	102,758	S/.4,315,836.00	10,209	S/.428,778.00
Porcentaje de deserción	38.89%		10.24%	

Según la tabla anterior se determinó que, sin la aplicación de minería de datos, el presupuesto desaprovechado en la muestra (enero a diciembre de 2017) fue de S/.4'426,506.00 en consulta ambulatoria y S/.1'676,178.00 en procedimientos. Con la aplicación de minería de datos, se demuestra que, durante el mismo periodo se hubiese, el presupuesto desaprovechado hubiese sido S/.4'315,836.00 y S/.428,778.00, respectivamente.

Índice de citas otorgadas por la deserción.

Antes de la aplicación, no se era capaz de predecir la probabilidad de deserción de los pacientes citados.

Con la aplicación de la minería de datos se pudo identificar que 2,635 citas no serían concretadas en consulta ambulatoria y 29,700 en procedimientos. Estas cifras se pueden observar en la Tabla 32.

VI. CONCLUSIONES

- 1.** Durante el periodo de prueba, se demostró que mediante el uso de la aplicación desarrollada, se pudo disminuir 1.00% el índice de deserción en Consulta Ambulatoria, y un 29.79% en Procedimientos.
- 2.** En lo que respecta al presupuesto, quedó demostrado que aplicando minería de datos durante el periodo de la muestra (enero a diciembre de 2017), se hubiese aprovechado S/.110,670.00 en Consulta Ambulatoria y S/.1'247,400.00 en Procedimientos. Ver Tabla 33.
- 3.** Asimismo, se determinó que, mediante el uso de la minería de datos, se hubiese aprovechado 2,635 citas en Consulta Ambulatoria y 29,700 en Procedimientos, destinados a pacientes en lista de espera. Ver Tabla 32.
- 4.** Con los resultados anteriores e identificando las posibles citas a desaprovecharse, se podrá disminuir el diferimiento de cita, que actualmente hay casos que son de hasta 2 meses; esto repercutirá favorablemente al paciente cuando solicite una cita. Asimismo esta identificación de deserción permite a los pacientes reprogramar o anular su cita, según sea el caso.

Recomendaciones

1. Se debe determinar acciones en los periodos de huelga, en donde las citas que se encuentren en ese periodo no sean consideradas como deserción, pues ya no compete al paciente.
2. Se recomienda realizar un buen registro de datos en el sistema transaccional, que alimentará la presente aplicación; esto para garantizar su calidad.
3. Tener presente los posibles cambios en el sistema transaccional, para dar un oportuno mantenimiento al presente sistema y garantizar su compatibilidad.
4. Por último, se recomienda implementar el sistema propuesto, por los beneficios demostrados en los puntos antes mencionados.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Belichón, Yolanda. (2010). “Minería de datos. Ingeniería de Telecomunicación”. Universidad Carlos III. Madrid. <http://www.it.uc3m.es/~jvillena/irc/practicas/10-11/15mem>
- Candela, Julio. (2015). “Proceso de descubrimiento de conocimiento para predecir el abandono de tratamiento en una entidad de salud pública”. Tesis de grado, Ingeniería Informática, Pontificia Universidad Católica del Perú. http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/6049/CANDELA_JULIO_PROCESO_DESCUBRIMIENTO_TRATAMIENTO.pdf
- Caniupan, Marileo y Gloria Toro. (2000). “Aplicación del Proceso de Data Mining en el Área Docente”. Tesis de Grado, Ingeniería Civil en Informática, Universidad del Bio Bio Chile, citado por Gallardo, José. (2009), en “Metodología para la definición de requisitos en proyectos de data mining (ER-DM)”. Departamento de lenguajes y sistemas informáticos e ingeniería de software. Universidad Politécnica de Madrid. http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf
- Chapman, Peter (NCR), Julian Clinton (SPSS), Randy Kerber (NCR), Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler), Colin Shearer (SPSS), and Rüdiger Wirth (DaimlerChrysler). 2000. “CRISP-DM 1.0 step-by-step data mining guide”, Technical report, citado por Gallardo, José. (2009), en “Metodología para la definición de requisitos en proyectos de data mining (ER-DM)”. Departamento de lenguajes y sistemas informáticos e ingeniería de software. Universidad Politécnica de Madrid. http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf
- Consejo Directivo, Seguro Social de Salud – EsSalud. (2012). “Plan Estratégico Institucional 2012 – 2016”. Lima.
- Corso, Cynthia. (2009). “Aplicación de algoritmos de clasificación supervisada usando Weka”. Universidad Tecnológica Nacional http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cynthia/CNIT_2009_Aplicacion_Algoritmos_Weka.pdf (último acceso junio 10, 2018).
- Curto, Josep y Jordi Conesa. (2010). “Introducción al Business Intelligence”. Barcelona: Editorial UOC.
- Date, Christopher. (2001). “Introducción a los Sistemas de Bases de Datos”. México: Pearson Educación.
- Ferrer, Francisco, Jesús Aguilar y Joaquín Peña. (2000). “Data Mining”. Departamento de Lenguajes y Sistemas. Informáticos. <https://www.lsi.us.es/docs/informes/lsi-2000-08.pdf> (último acceso junio 10, 2018).
- Gallardo, José. (2009). “Metodología para la definición de requisitos en proyectos de data mining (ER-DM)”. Departamento de lenguajes y sistemas informáticos e ingeniería de software. Universidad Politécnica de Madrid. http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf
- George A, Rubin G. (2003). Non-attendance in general practice: a systematic review and its implications for access to primary health care. Family Practice.
- Gerencia Central de Planificación y Desarrollo - EsSalud. (2012). “Glosario de términos normativos institucional”
- Gerencia Central de Salud – EsSalud. (2002). “Compendio de indicadores de salud”: 15, (último acceso junio 10 de 2018)

- González, Ana. (2007). “Inteligencia empresarial para la toma de decisiones en la PyME enfocada en la administración de la relación con el cliente (CRM) utilizando análisis de la canasta de compra (MBA)”. Tesis de Licenciatura, Universidad de las Américas Puebla. http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/gonzalez_m_ac/capitulo3.pdf
- Hernández, José, María Ramírez y César Ferri. (2004). “Introducción a la Minería de Datos”. Madrid: Pearson Educación S.A.
- Jabalera Mesa, Lourdes. (2015). Factores determinantes del absentismo en consultas externas de la Agencia Sanitaria Costa del Sol. Tesis doctoral. Universidad de Málaga. https://riuma.uma.es/xmlui/bitstream/handle/10630/10149/TD_Jabalera_Mesa.pdf?sequence=1
- Kroenke, David. (2003). “Procesamiento de Bases de Datos: Fundamentos, Diseño e Implementación”. México: Pearson Educación.
- La Tercera. (2017). “El 30% de consultas agendadas en centros privados se pierden por inasistencia de paciente”. <http://www2.latercera.com/noticia/30-consultas-agendadas-centros-privados-se-pierden-inasistencia-paciente/>
- Martínez de Pisón, Javier. (2003). “Optimización mediante técnicas de minería de datos del ciclo de recocido de una línea de galvanizado”. Tesis Doctoral, Universidad de La Rioja, Servicio de Publicaciones. <http://dialnet.unirioja.es/descarga/tesis/81.pdf> (último acceso junio 10 de 2018).
- Méndez del Rio, Luis. (2000.) “Más Allá del Business Intelligence”. Barcelona: Ediciones Gestión.
- Oficina de Gestión de la Calidad y Control Interno. (2012). “Motivos de Deserción a la Consulta Externa del Hospital Nacional Almanzor Asenjo. Seguro Social de Salud – Essalud”. Chiclayo.
- Pentaho Business Analytics. (2014). <http://www.pentaho.com/product/business-visualization-analytics>. (último acceso junio 10 de 2018).
- Pereira Victorio, César, Vázquez Molinero, Antonio, otros. (2016). Absentismo de pacientes a la consulta externa especializada en un hospital de tercer nivel en España. <https://www.sciencedirect.com/science/article/pii/S1889543316000372>
- Pérez, Milena y Rendón Margeri. (2016). Características asociadas con la inasistencia a la consulta de promoción y prevención en salud en una IPS de la ciudad de Medellín. http://bdigital.ces.edu.co:8080/repositorio/bitstream/10946/4394/1/Caracteristicas_asociadas_inasistencia.pdf
- Piatetski-Shapiro, Gregory, William Frawley y Christopher Matheus. (1992). “Knowledge discovery in databases: an overview”. California: AAAI-MIT Press, Menlo Park. Citado por Ferrer, Francisco, Jesús Aguilar y Joaquín Peña. (2000), en “Data Mining”. Departamento de Lenguajes y Sistemas. Informáticos
- Rodríguez, María, Valeriano Álvarez, José y Adolfo González. (2003) “Metodologías para la realización de proyectos de data mining”. http://www.aepro.com/files/congresos/2003pamplona/ciip03_0257_0265.2134.pdf. (último acceso junio 10 de 2018).
- Seguro Social de Salud. 2013. “Glosario de Términos de la Seguridad Social”
- Sevilla, Juncal, (2014). “Minería de datos en el análisis de itinerarios y costes en Salud Mental”. Tesis Doctoral, Universidad Autónoma de Madrid.

- Stair, Ralph y George Reynolds. (2000). “Principios de Sistemas de Información: Enfoque Administrativo”. México: Cengage Learning Editores.
- Viera, Luis, Santiago Ramírez, Luis Ortíz y Aduino da Costa Medeiros. (2009). “Introducción a la Minería de Datos”. Río de Janeiro: E-papers Serviços Editoriais Ltda.
- Villanueva, Álvaro. (2008). “Análisis, diseño e implementación de un datawarehouse de soporte de decisiones para un hospital del sistema de salud público”. Tesis de grado, Ingeniería Informática, Pontificia Universidad Católica del Perú.

VIII. ANEXOS

Anexo I: Porcentaje de encuestados por especialidad que desertaron a la atención en consultorios externos.

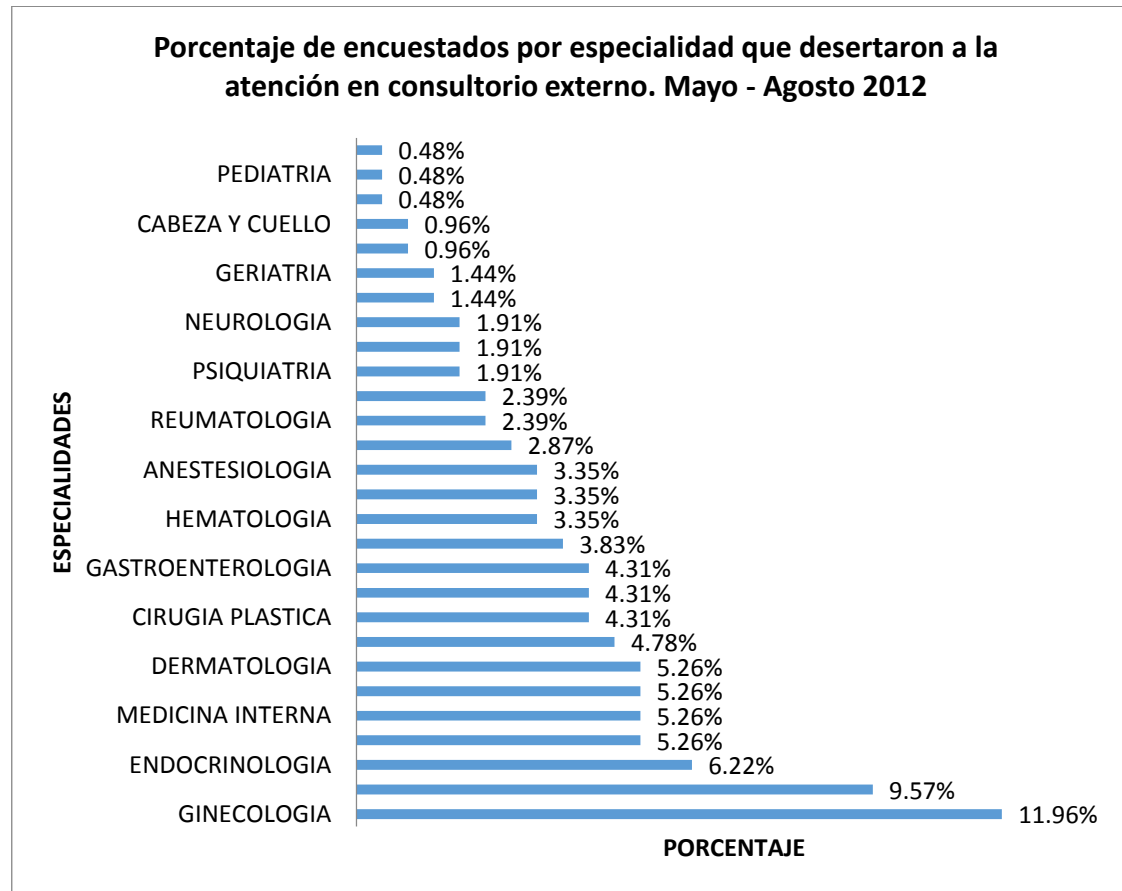


Figura 61: Porcentaje de desertores a la consulta externa
Fuente: Oficina de Gestión de la Calidad y Control Interno. 2012

Anexo II: Porcentaje de encuestados por sexo que desertaron a la atención en consultorios externos.

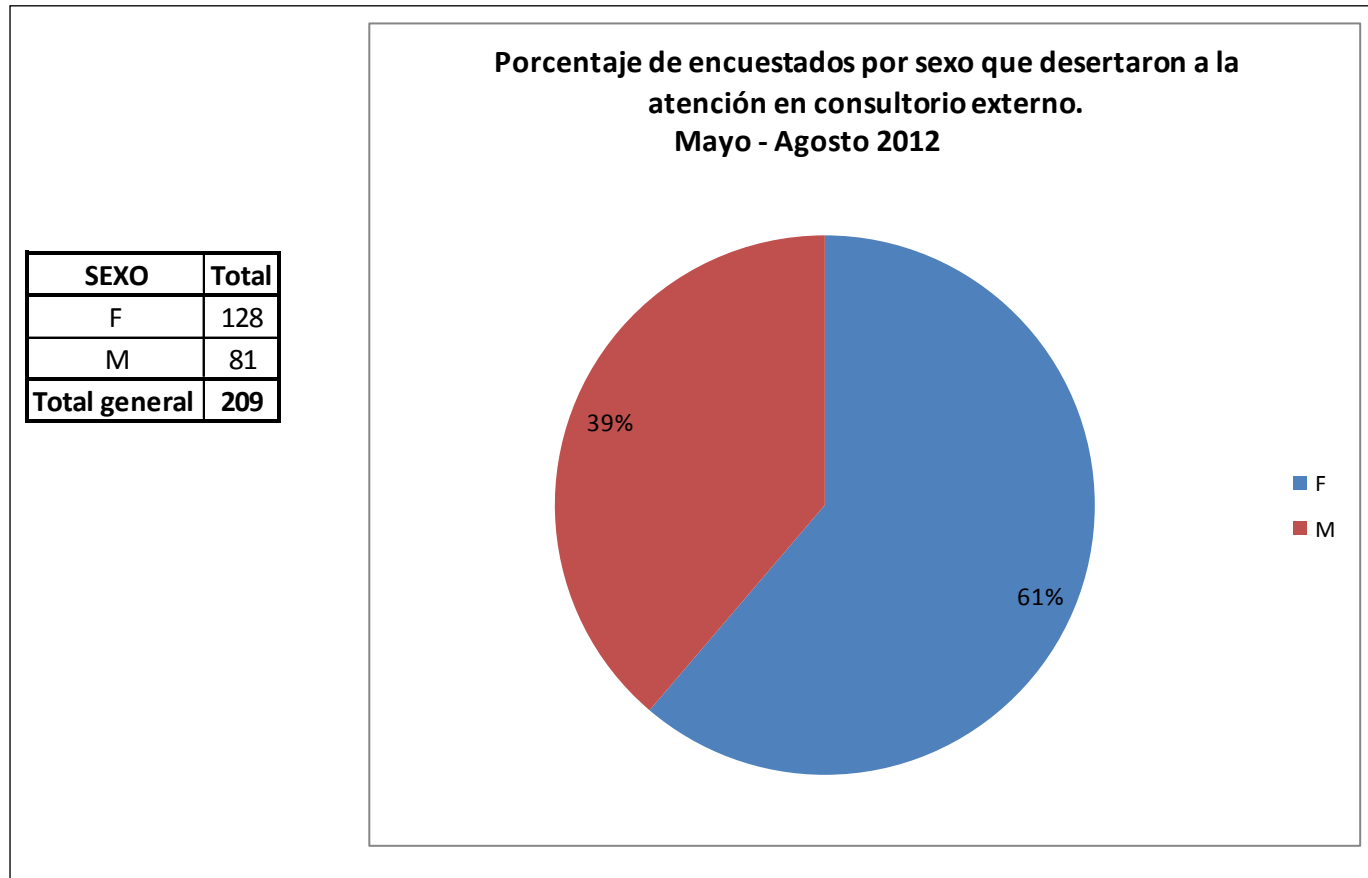


Figura 62: Porcentaje de desertores por sexo
Fuente: Oficina de Gestión de la Calidad y Control Interno. 2012.

Anexo III: Porcentaje de encuestados por rango de edades que desertaron a la atención en consultorios externos.

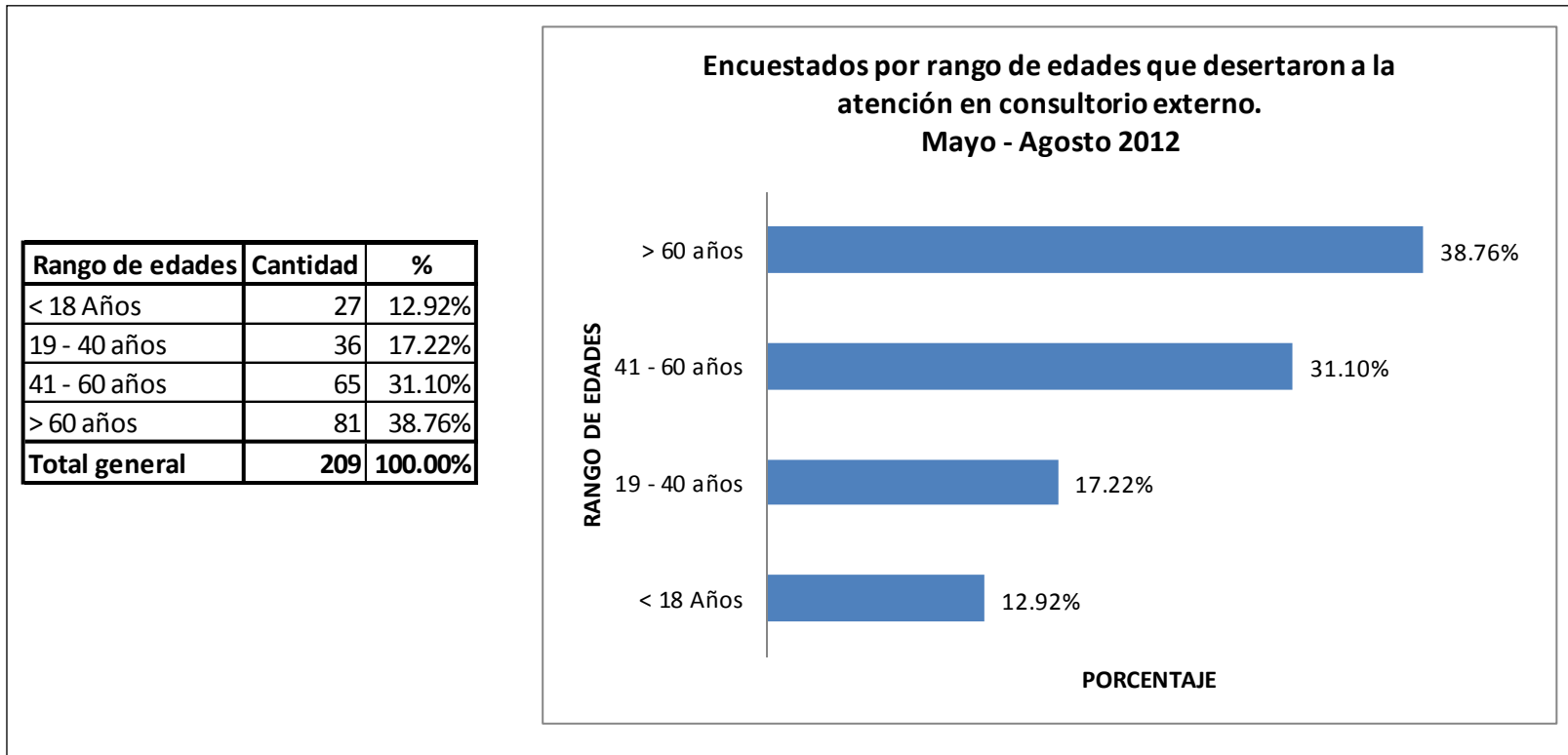


Figura 63: Porcentaje de desertores por rango de edades.
Fuente: Oficina de Gestión de la Calidad y Control Interno. 2012.

Anexo IV: Porcentaje de encuestados por tipo de seguro que desertaron a la atención en consultorios externos.

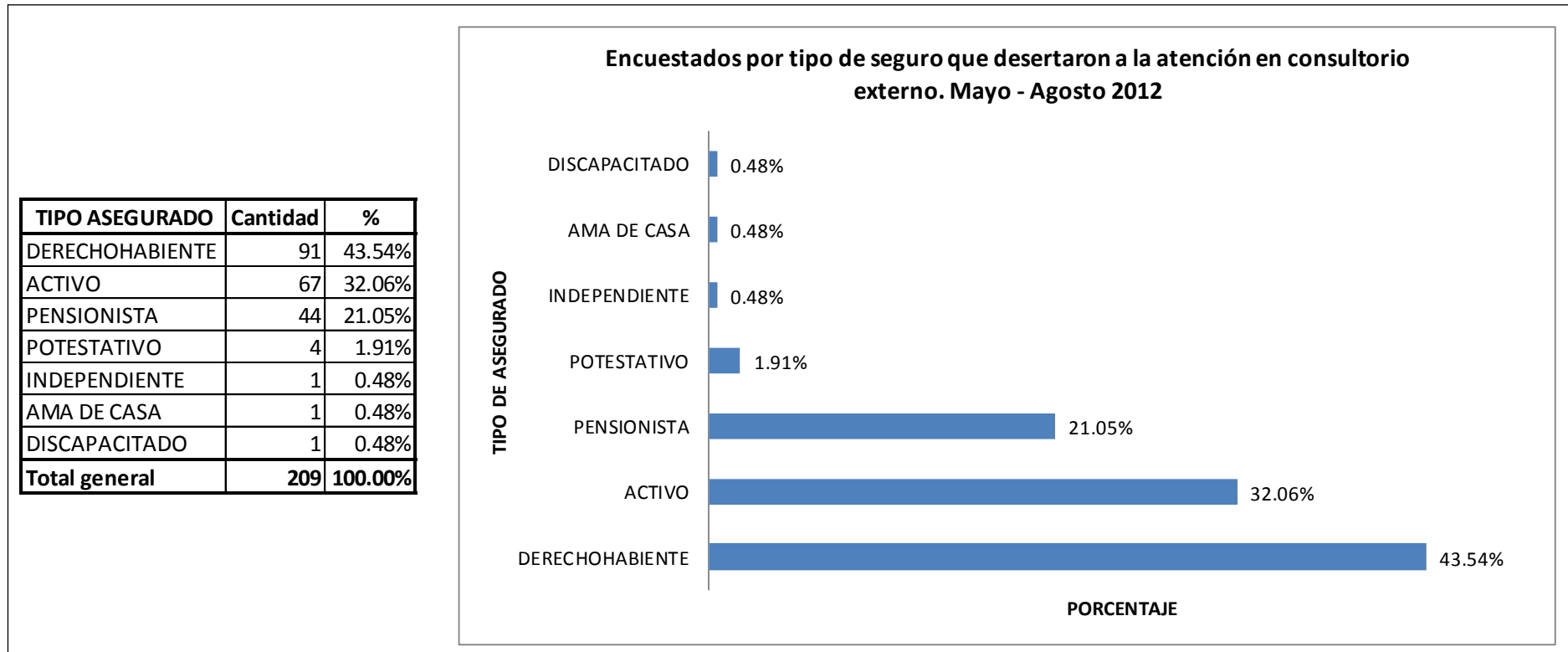


Figura 64: Porcentaje de desertores por tipo de asegurado
Fuente: Oficina de Gestión de la Calidad y Control Interno. 2012.

Anexo V: Porcentaje de encuestados por procedencia que desertaron a la atención en consultorios externos.

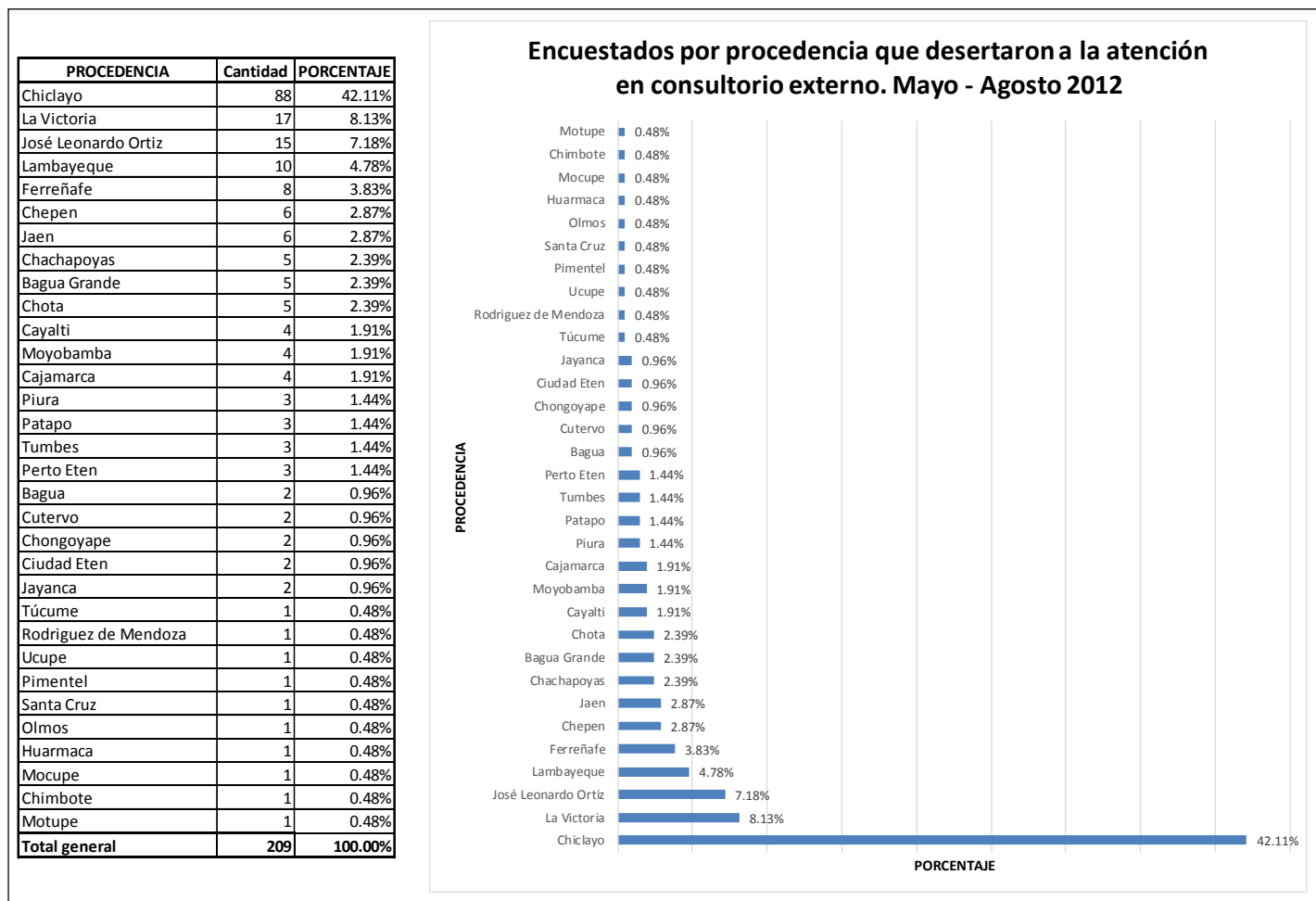


Figura 65: Porcentaje de desertores por procedencia
Fuente: Oficina de Gestión de la Calidad y Control Interno. 2012.

Anexo VI: Porcentaje de motivos manifestados por los encuestados que desertaron a la atención en consultorios externos.

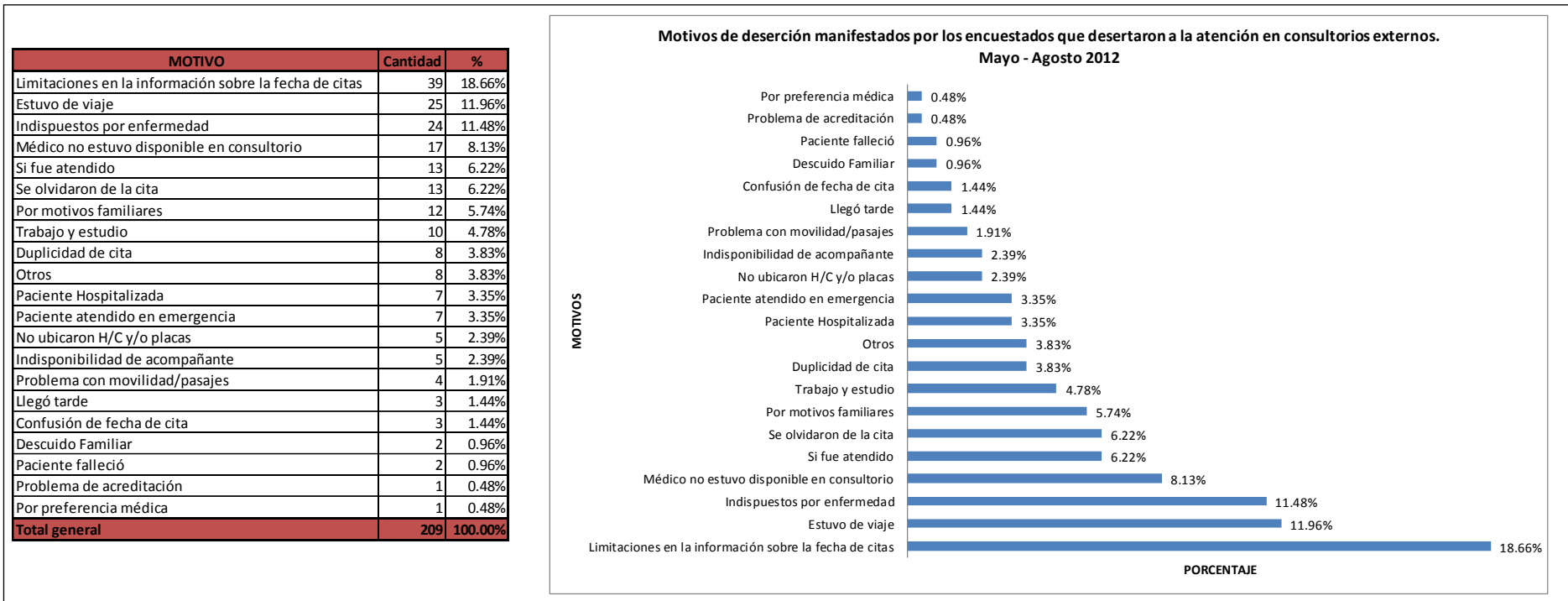


Figura 66: Porcentaje de motivos de deserción
Fuente: Oficina de Gestión de la Calidad y Control Interno. 2012.

Anexo VII: Cálculo de presupuesto desaprovechado.

Tabla 34: Total de presupuesto desaprovechado.

Área	Habitación	Turno	Precio	Promedio	Total Deserción	Total
Consulta Externa	-	Diario	S/. 49.56	S/. 49.56	209	S/. 10,358.04
Emergencia	-	Diurna	S/. 62.54	S/. 143.57	7	S/. 1,004.97
	-	Nocturna	S/. 73.16			
Observación	-	Diario	S/. 295.00	S/. 301.69	7	S/. 2,111.81
Hospitalización	Cama Individual		S/. 350.46			
	Cama Doble		S/. 307.98			
	Cama Múltiple		S/. 246.62			

Fuente: Oficina de Gestión de la Calidad y Control Interno. 2012.

Los datos mostrados en la Tabla 34, fue tomado como parte de la problemática, donde la muestra de la encuesta realizada, se tuvo una muestra de 209 pacientes; generando un desaprovechamiento de presupuesto, calculado en S/.10,358.04.

Asimismo se calculó que por acudir por los servicios de emergencia y hospitalización se generó un gasto adicional calculado en S/.3,116.77.