

**DASHBOARD PARA LA IDENTIFICACIÓN DE LOS POSIBLES INTENTOS  
DE FRAUDE DE LAS PRESTADORAS DE SERVICIO ASOCIADAS A LAS  
ENTIDADES ASEGURADORAS**

**Autor**

Carlos Alberto Alfaro Prieto

**Director**

Elio H. Cables Pérez, Ph.D

Universidad Antonio Nariño  
Facultad de Ingeniería de Sistemas  
Especialización en Gobierno de Datos  
Bogotá D.C  
2023

## Índice de contenidos

<b>Introducción</b> .....	1
<b>2.Descripción y formulación del problema</b> .....	2
2.1 Objetivo General .....	3
2.2 Objetivos Específicos.....	3
<b>3.Marco referencial</b> .....	4
3.1. Marco teórico .....	4
3.1.1. Calidad de datos .....	4
3.1.2. Limpieza de datos .....	5
3.1.3. Python .....	5
3.1.4. Microsoft Visual Studio Integración Servicios (SSIS) .....	6
3.1.5. Power BI .....	6
3.1.6. Fraude en la prestación de servicios de salud .....	7
3.2. Estado del Arte.....	7
3.3. Impacto .....	11
3.4. Componente de Innovación.....	11
<b>4.Metodología</b> .....	12
4.1. Obtención de los datos .....	13
4.2. Preparación de datos .....	13
4.3. Organización de los datos .....	14
4.4. Visualización de datos .....	14
<b>5.Desarrollo de la propuesta</b> .....	15
5.1. Obtención de los datos .....	15
5.2. Preparación de datos .....	21
5.3. Organización de los datos .....	28
5.4. Visualización de datos .....	35
5.4.1. Herramienta de Visualización .....	35
5.4.2. Definición de los Datos .....	35
5.4.3. Diseño Tableros .....	36

6. Conclusiones .....	40
7. Referencias .....	41
8. Anexos .....	42
8.1. Anexo Tablas fuentes Altos Consumos .....	42
8.2 Tablas fuentes Omisión .....	44
8.3 Repositorio de Fuentes Omisión Altos Estancias .....	47
8.4 Anexo Diccionario de Datos Fuentes .....	50
8.5 Anexo Tableros .....	53
8.6 Anexo Tablero Estancias .....	56
8.7 Anexo Diagrama De Componentes .....	59
8.8 Anexo 28 Proceso omisión .....	60
8.9 Anexo 29 Proceso Estancias Prolongadas .....	60
8.10 Anexo 30 Datos Críticos Estancias Prolongadas .....	61
8.11 Anexo 31 Datos Críticos Estancias Prolongadas .....	61
8.12 Anexo 32 Datos Críticos Estancias Prolongadas .....	62

## Índice de Figuras

Figura 1. Metodología del Proyecto .....	13
Figura 2. Proceso negocio Altos Consumos .....	15
Figura 3. Proceso negocio Estancias .....	15
Figura 4. Proceso negocio Omisión Preexistencias.....	16
Figura 5. Modelo De datos estancias .....	17
Figura 6. Modelo De datos Omisión.....	17
Figura 7. Modelo De datos altos consumos.....	18
Figura 8. repositorio de fuentes.....	18
Figura 9. repositorio de fuentes intercambio terceros.....	19
Figura 10. Cargue inicial.....	21
Figura 11. diagrama de componentes línea Bach.....	21
Figura 12. importación de informacion para limpieza.....	22
Figura 13. Exportación de terceros a datos.....	22
Figura 14. Logs.....	23
Figura 15. Error Log.....	23
Figura 16. Validación existencia de fuentes.....	24
Figura 17. Limpieza de datos.....	24
Figura 18. limpieza de datos.....	25
Figura 19. componentes.....	25
Figura 20. repositorio intercambio de datos.....	26

Figura 21. Limpieza.....	27
Figura 22. ejecución Jobs envía que tarea se está ejecutando.....	27
Figura 23. ejecución Jobs.....	27
Figura 24. Inicio ejecución ETL Importación landing zone.....	28
Figura 25. Inicio ejecución ETL Cargue de archivos a tablas físicas.....	28
Figura 26. Inicio ejecución ETL validación creación tablas.....	29
Figura 27. Inicio ejecución ETL cargue a tablas físicas.....	30
Figura 28. ejecución proceso ETL cargue sit minería.....	30
Figura 29. ejecución proceso ETL cargue de fuentes.....	31
Figura 30. ejecución proceso ETL cargue a tabla física.....	31
Figura 31. ejecución SP SQL cargue tablas transaccionales .....	32
Figura 32. ejecución SP SQL.....	32
Figura 33. ejecución proceso ETL (Sit Minería) Estancias.....	33
Figura 34. ejecución proceso ETL (Sit Minería) .....	33
Figura 35. ejecución disparador ETL (Sit Minería) .....	34
Figura 36. ejecución SP (Sit Minería) .....	34
Figura 37. Bases de datos.....	34
Figura 38. fuentes.....	35
Figura 39. Bases de datos Omisión.....	36
Figura 40. Casos pendientes por región y valor ahorrado.....	37
Figura 41. Tipos de casos Gestionados.....	37

Figura 42. Total, Ahorros.....	38
Figura 43. Casos por región.....	39
Figura 44. Indicadores de ahorro.....	39
Figura 45. Paginas.....	39

### Índice de tablas

Tabla 1. repositorio de fuentes Altos.....	16
Tabla 2. repositorio de fuentes Estancias.....	16
Tabla 3. Diccionario de datos enfermedades.....	19
Tabla 4. Diccionario de datos Gasto medico.....	20
Tabla 5. Diccionario de datos.....	20

## Resumen

El fraude en la actualidad afecta negativamente a las compañías aseguradoras de salud, por tal motivo se debe realizar una temprana detección por parte de los auditores, y con base a los hallazgos tomar decisiones que mejoren la detección de casos fraudulentos. Pero esta actividad cada día genera mayores retos en el área auditora ya que cada vez son más los casos de fraude que se deben revisar los cuales pueden ser desde Facturación y codificación fraudulenta, Falsificación de documentos, la presentación de reclamaciones fraudulentas y uso indebido de la identidad del paciente. Este proyecto permite proporcionar una herramienta automatizada para detectar y prevenir fraudes en los servicios de salud. El desarrollo del Dashboard se basa en la recopilación y análisis de datos relevantes, así como en la implementación de algoritmos y técnicas de detección de anomalías. A través de la integración de múltiples fuentes de datos, como facturas médicas, historias clínicas y registros de reclamaciones, se busca identificar patrones sospechosos y anomalías en los patrones de uso de los servicios médicos.

La metodología empleada en este proyecto se enfoca en el procesamiento de grandes volúmenes de datos y en la aplicación de algoritmos para identificar los posibles intentos de fraude. Además, se utilizan técnicas de visualización de datos para presentar los resultados de manera clara y accesible en el Dashboard.

Los resultados obtenidos demuestran la eficacia del Dashboard en la detección temprana de fraudes y en la identificación de prestadoras de servicios con comportamientos sospechosos. El sistema permite a la aseguradora tomar medidas preventivas y correctivas de manera oportuna, lo que contribuye a minimizar las pérdidas económicas y proteger los intereses de la compañía y sus asegurados.

Las implicaciones prácticas de este proyecto incluyen un mejor control de calidad en la prestación de servicios de salud, una reducción de los costos asociados a fraudes y una mejora en la confianza de los clientes hacia la aseguradora. Además, se recomienda continuar investigando y actualizando el Dashboard para adaptarse a las cambiantes tácticas de fraude utilizadas por las prestadoras de servicios.

Palabras clave: Dashboard, detección de fraude, prestadoras de servicios, aseguradora, análisis de datos.

## Abstract

Fraud currently negatively affects health insurance companies, for this reason an early detection must be made by auditors, and based on the findings, decisions must be made to improve the detection of fraudulent cases. But this activity every day generates greater challenges in the audit area as more and more cases of fraud to be reviewed which can be from fraudulent billing and coding, falsification of documents, submission of fraudulent claims and misuse of the patient's identity. This project provides an automated tool to detect and prevent fraud in healthcare services. The development of the Dashboard is based on the collection and analysis of relevant data, as well as the implementation of algorithms and anomaly detection techniques. Through the integration of multiple data sources, such as medical bills, medical records, and claims records, it seeks to identify suspicious patterns and anomalies in the usage patterns of medical services.

The methodology employed in this project focuses on processing large volumes of data and applying algorithms to identify potential fraud attempts. In addition, data visualization techniques are used to present the results in a clear and accessible way on the Dashboard.

The results obtained demonstrate the effectiveness of the Dashboard in the early detection of fraud and in the identification of service providers with suspicious behavior. The system allows the insurer to take preventive and corrective measures in a timely manner, which contributes to minimizing economic losses and protecting the interests of the company and its policyholders.

The practical implications of this project include improved quality control in the provision of health services, a reduction in the costs associated with fraud, and improved client confidence in the insurer. In addition, it is recommended that the Dashboard continue to be researched and updated to adapt to the changing fraud tactics used by service providers.

Keywords: Dashboard, fraud detection, service providers, insurer, data analysis.



## 1. Introducción

La automatización de procesos basada en datos es una práctica cada vez más relevante en la industria de seguros en la rama de la salud. Esta estrategia permite a las empresas de seguros automatizar tareas y procesos tediosos y repetitivos, aumentando la eficiencia y reduciendo los errores humanos.

Lo que conlleva a realizar con el presente proyecto un aporte importante en el área de Fraude en la salud aplicando herramientas de automatización, en el proceso de identificación de fraude en sus diferentes modalidades como lo son presentación de facturas falsas o infladas por servicios que no se proporcionaron o que fueron innecesarios. o la manipulación de códigos de diagnóstico y procedimientos médicos para obtener un reembolso mayor de lo correspondiente Altos consumos, Estancias Prolongadas y omisión de Preexistencias, simplificando el flujo de trabajo y optimizando los procesos de gestión de reclamaciones, auditorías y gestión de riesgos, entre otros.

Además, al integrar el análisis de datos en los procesos automatizados, se pueden obtener perspectivas valiosas para la toma de decisiones empresariales. Los datos pueden utilizarse para identificar tendencias, patrones y anomalías en la información, lo que ayuda a mejorar la precisión y eficacia de la gestión de riesgos, así como a la elaboración de estrategias de negocio.

Implementando herramienta de visualización de datos que puede ayudar a la empresa de seguros a tener una visión más clara y completa de sus procesos, lo que les permitirá identificar oportunidades de mejora y tomar decisiones informadas. incluyendo el rendimiento financiero, el comportamiento de los aliados y el estado de los procesos, permitiendo a la empresa tomar decisiones más objetivas y estratégicas.

## 2. Descripción y formulación del problema

La aseguradora es una compañía dedicada a la prestación de servicios de aseguramiento más importantes del país con más de 50 años de experiencia ofreciendo a sus afiliados uno de los más amplios portafolios del mercado, como seguros generales, seguros de vida, salud y ARL, le apuesta y cree en el fomento de la cultura de la prevención y por eso dispone en uno de sus más completos portafolios de servicio en aseguramiento a la salud, esto en pro de proteger al usuario contra los costos asociados en la atención y los tratamientos médicos de alto costo que se puedan presentar. Los productos pueden variar según la necesidad y preferencias de los afiliados incluyendo cobertura para gastos médicos de mayor cuantía, medicina prepagada, pólizas de hospitalización, atención ambulatoria, medicamentos recetados, servicios preventivos y servicios de emergencia.

Dado este gran portafolio se requiere de aliados estratégicos como IPS y Hospitales que permitan brindar el mejor servicio garantizando la cobertura y prestación de servicio de forma oportuna, teniendo en cuenta lo anterior se identifica la siguiente problemática.

Dando cumplimiento a las pólizas de salud se identificó en la prestación de los servicios, los últimos años se ha venido presentando un incremento en el número de casos de fraude por sobrecostos por un valor de 1.320.000.000, en los procesos de Altos Consumos, Omisión de Preexistencia y Estancias Prolongadas, los cuales cuentan con procesos de auditoria manual pero no son suficientes para cubrir la cantidad de casos que se vienen presentando generando que muchos de ellos no puedan ser detectados y por consiguiente sean pagados generando perdidas a la compañía.

Presentando así la siguiente interrogante.

¿Como identificar de forma temprana los posibles fraudes que se puede generar en la prestación de servicios de salud a partir de los elementos referidos a preexistencias, altos consumos y estancias prolongadas de los usuarios afiliados a los diferentes prestadores de salud asociados a la aseguradora?

## **2.1 Objetivo General**

Desarrollar un Dashboard que permita visualizar y consultar de una forma más clara y dinámica los posibles intentos de fraude a partir de la información suministrada por los diferentes prestadores de salud asociado a la aseguradora.

## **2.2 Objetivos Específicos**

- Caracterizar el sistema de información referente a la facturación de los servicios prestados por los prestadores de servicio de salud.
- Identificar los rasgos esenciales de la información referente a omisión, estancias prolongadas y altos consumos que permitan la clasificación de posibles intentos de fraude.
- Crear un modelo para la clasificación de los posibles intentos de fraude a partir de la información entregada por los prestadores de servicio.
- implementar un Dashboard para la visualización de información de los posibles intentos de fraude.

### 3. Marco referencial

En el sector de la salud, se presenta una alta incidencia de fraude, ya sea por parte de las IPS que facturan servicios no prestados o por parte de los pacientes que omiten información sobre enfermedades ya preexistentes, que al momento de adquirir un contrato con la aseguradora están obligados a declarar este problema implica grandes pérdidas económicas.

#### 3.1. Marco teórico

En el siguiente apartado se describirá los componentes teóricos que de acuerdo con la actividad son relevantes para el desarrollo del proyecto de igual forma se hace referencias a herramientas tecnológicas que son necesarias para el mismo.

##### 3.1.1 Calidad de datos

Como se describe en muchos textos la calidad de datos se refiere a la medida en que los datos son precisos, completos, consistentes y relevantes para su uso previsto. La calidad de los datos es un aspecto fundamental en cualquier organización, ya que los datos son una parte vital de la toma de decisiones y de la gestión efectiva de los procesos de negocio. (Silberschatz, 2012).

Algunos de los problemas más comunes relacionados con la calidad de los datos incluyen.

**Inconsistencia:** los datos pueden ser inconsistentes si se ingresan de forma diferente en diferentes sistemas o si los procesos no están estandarizados.

**Falta de precisión:** los datos pueden ser imprecisos si se ingresan con errores o si se utiliza una fuente no confiable.

**Falta de integridad:** los datos pueden ser incompletos si faltan campos o información crítica.

**Falta de relevancia:** los datos pueden ser irrelevantes si no están relacionados con el propósito para el que se están utilizando.

La mala calidad de los datos puede tener un impacto negativo en la toma de decisiones, ya que puede llevar a decisiones erróneas. También puede tener un impacto en la eficiencia y eficacia de los procesos de negocio, así como en la satisfacción del cliente y en la capacidad de una organización para cumplir con los requisitos regulatorios. (Wang, R. Y., & Strong, D. M. 2012).

Para mejorar la calidad de los datos, es importante establecer procesos y controles de calidad de datos rigurosos, que incluyan la definición clara de los estándares de calidad de datos, la implementación de controles para la entrada de datos, la validación de datos y la gestión de errores y excepciones. También

se deben definir responsabilidades claras y roles para la gestión de la calidad de los datos, y se deben establecer procesos para la medición continua de la calidad de los datos y la mejora continua. (DAMA International,2017).

### 3.1.2 Limpieza de datos

La limpieza de datos es el proceso de identificar y corregir o eliminar errores, inconsistencias, redundancias y valores atípicos en los datos. Es una parte importante del proceso de análisis de datos y es esencial para garantizar la precisión y la integridad de los resultados del análisis.

La limpieza de datos generalmente implica las siguientes tareas:

**Identificación de datos erróneos o faltantes:** se revisa los datos para identificar aquellos que contienen valores incorrectos o faltantes.

**Corrección de errores:** se corrigen los valores incorrectos mediante la eliminación de datos duplicados, la corrección de errores de entrada de datos y la actualización de la información.

**Eliminación de datos duplicados o redundantes:** se eliminan los datos duplicados o redundantes que pueden afectar negativamente la precisión del análisis.

**Normalización de datos:** se convierten los datos a un formato común para asegurarse de que sean comparables y coherentes.

**Manejo de valores atípicos:** se identifican y se evalúan los valores atípicos para determinar si son errores o valores legítimos. En caso de ser valores legítimos, se pueden mantener en los datos, y en caso contrario, se eliminan o se corrigen.

La limpieza de datos es una tarea importante en el análisis de datos, ya que los errores en los datos pueden llevar a resultados inexactos y a decisiones equivocadas. (DAMA International,2017).

### 3.1.3 Python

Python es un lenguaje de programación ampliamente utilizado en la limpieza de datos debido a su facilidad de uso y a la gran cantidad de bibliotecas disponibles para manejar datos. realizar operaciones matemáticas y estadísticas, manipular expresiones regulares y visualizar datos de manera efectiva. Algunas de las bibliotecas más populares en Python para la limpieza de datos son

**Pandas:** es una biblioteca de análisis de datos que proporciona estructuras de datos flexibles para trabajar con datos tabulares, como hojas de cálculo. Pandas es muy útil para limpiar y transformar datos, eliminando filas y columnas con

valores nulos, y manipulando datos numéricos y de cadena (Python Software Foundation. (s.f.)).

**NumPy:** es una biblioteca para computación científica en Python que proporciona herramientas para trabajar con matrices y arreglos de datos. NumPy es útil para realizar operaciones matemáticas y estadísticas en datos y para manipular y filtrar valores atípicos Python Software Foundation. (s.f.).

### 3.1.4 Microsoft Visual Studio Integración Servicios (SSIS)

herramienta ETL (Extracción, Transformación y Carga) que permite a los desarrolladores crear y ejecutar paquetes de integración de datos para la extracción, transformación y carga de datos desde y hacia diferentes fuentes de datos. SSIS cuenta con un diseñador de paquetes visual, un motor de ejecución y un catálogo de SSIS para almacenar y administrar paquetes de integración de datos. SSIS es compatible con una amplia gama de formatos de datos y proporciona una amplia gama de transformaciones de datos. En resumen, SSIS es una herramienta ETL muy útil para la integración de datos en diferentes fuentes de datos mediante un flujo de trabajo visual y componentes de extracción, transformación y carga de datos (Microsoft. (s.f.)2023).

### 3.1.5 Power BI

es una herramienta de análisis de datos de Microsoft que permite a los usuarios conectar, visualizar y compartir datos de diferentes fuentes en una sola plataforma. Algunos de los beneficios de usar Power BI incluyen:

**Análisis de datos en tiempo real:** Power BI permite a los usuarios analizar datos en tiempo real y tomar decisiones basadas en información actualizada y precisa.

**Conexión con múltiples fuentes de datos:** Power BI permite a los usuarios conectarse con diferentes fuentes de datos, como bases de datos, archivos de Excel y servicios en línea, lo que facilita la integración de datos de diferentes sistemas y fuentes. (Microsoft. (s.f.)2023).

**Visualización de datos:** Power BI proporciona una amplia gama de opciones de visualización de datos, como gráficos, tablas y mapas, lo que permite a los usuarios ver los datos de manera más clara y fácil de entender. (Microsoft. (s.f.)2023).

**Informes personalizados:** Power BI permite a los usuarios crear informes personalizados y compartirlos con otros usuarios, lo que facilita la colaboración y el trabajo en equipo. **Acceso desde cualquier dispositivo:** Power BI es accesible desde cualquier dispositivo con una conexión a Internet, lo que permite a los

usuarios ver y analizar datos desde cualquier lugar y en cualquier momento. (Microsoft. (s.f.)2023).

### **3.1.6 Fraude en la prestación de servicios de salud**

El fraude en la prestación de servicios de salud es un problema creciente que afecta tanto a las aseguradoras como a los pacientes. Se define el fraude como cualquier acción intencional que busca obtener beneficios económicos de manera ilegal o engañosa. En el contexto de las aseguradoras, el fraude puede manifestarse a través de facturación excesiva, servicios innecesarios, duplicación de reclamaciones, entre otros. En el sector de la salud, los procesos de reclamación y gestión de datos desempeñan un papel fundamental en la detección de posibles intentos de fraude. Estos procesos implican la recopilación, almacenamiento y procesamiento de datos relacionados con las reclamaciones de servicios médicos. Es crucial comprender cómo se llevan a cabo estos procesos y cómo se registran los datos pertinentes para identificar patrones y comportamientos sospechosos.

Existen diferentes métodos y técnicas para la detección de fraudes en el ámbito de la salud. Estos métodos pueden incluir el análisis de datos, la minería de datos, el aprendizaje automático y la detección de anomalías. El análisis de datos permite identificar patrones anormales o inconsistentes en los registros de reclamaciones y servicios médicos. La minería de datos ayuda a descubrir relaciones ocultas y patrones predictivos. El aprendizaje automático permite desarrollar modelos que aprenden de los datos históricos y pueden predecir comportamientos fraudulentos. La detección de anomalías se enfoca en identificar desviaciones significativas con respecto a los patrones normales García, M. A. (2020).

### **3.2. Estado del Arte**

Este apartado del documento menciona las investigaciones previas encontradas y que tienen relación en cuanto la implementación en procesos de automatización enfocados al procesamiento de datos, la detección de fraude y la implementación de herramientas que permiten proporcionar una visualización detallada de los datos clave de la empresa, incluyendo el rendimiento financiero, el comportamiento de los clientes y el estado de los procesos.

El presente proyecto “Análisis de datos para la toma de decisiones con el uso de tableros de control aplicado a los registros de ventas digitales de un E-Commerce” se centra en la implementación de una arquitectura de inteligencia

empresarial (BI) utilizando herramientas de Microsoft. Su objetivo es analizar datos de una operación de ventas digitales asistidas de un retail, con el propósito de identificar comportamientos de compra, detectar clientes fraudulentos y mejorar el proceso de control de fraude previo a una preventa. Los datos necesarios para este análisis provienen de una base de datos transaccional y otros orígenes, como archivos planos. Ango, J. (2020). Análisis de datos para la toma de decisiones con el uso de tableros de control aplicado a los registros de ventas digitales de un E-Commerce (Tesis de pregrado). Universidad de las Américas, Quito.

En el proyecto anteriormente descrito se identificó que no se tuvo en cuenta un factor importante para la identificación de casos fraudulentos que, si se aplica en el actual proyecto y es la comparación con perfiles de prestadores aliados con muchos casos de fraude, generando así una base de datos perfilada como casos de alto riesgo que permita cruzar y alertar si existe en los casos a analizar prestadores con ese perfil.

El fraude se ha convertido en una grave amenaza para los sistemas de atención de salud de todo el mundo. Las instituciones de seguridad social tratan constantemente de optimizar los procesos con la ayuda de análisis avanzados y de la inteligencia artificial (IA) para detectar y supervisar el fraude de forma más eficaz y eficiente, si bien la adopción de esas tecnologías emergentes plantea desafíos importantes. Los presupuestos públicos de todo el mundo están sometidos a presión. El rápido aumento de los costos de la atención de salud, debido en parte al progreso tecnológico y al envejecimiento de la población, requiere que los países hagan un uso adecuado de los recursos escasos en materia de salud para atender a quienes más lo necesitan. El sector de la salud es una parte esencial de la seguridad social y representa una gran proporción del PIB. Se mueven grandes cantidades de dinero mediante numerosas transacciones individuales, por lo que este sector es un objetivo atractivo para los defraudadores. ISSA.(2022). Detectar el fraude en la atención de salud mediante las tecnologías emergentes. Recuperado de 4 de julio 2022 <https://ww1.issa.int/es/analysis/detecting-fraud-health-care-through-emerging-technologies>

La pandemia de COVID-19 afectó considerablemente a la prestación de atención de salud. Los cambios urgentes en las prácticas relativas a los códigos en la facturación, a la telesalud y a las prescripciones permitieron a los sistemas de salud adaptar satisfactoriamente la ejecución de los procesos de salud. Sin embargo, esta rápida adaptación generó posibles vulnerabilidades al fraude y al derroche. Aunque la mayor parte de los proveedores de servicios son honrados y bienintencionados, el comportamiento fraudulento (que se define más

adelante) tiene efectos negativos directos en la atención de salud. Produce un derroche de recursos limitados y puede poner en peligro a los pacientes al proporcionarles una atención innecesaria o al obstaculizarles el acceso a los servicios médicos que necesitan. Las cantidades defraudadas no están disponibles para financiar la prevención, reembolsar las innovaciones o invertir en programas que garanticen la igualdad de acceso a una atención de calidad. Las pérdidas causadas por fraudes o errores representan un promedio del 6 por ciento del gasto en salud (OCDE, 2017). La atención de salud y el seguro médico también son cada vez más vulnerables al fraude, que, por definición, es oculto y difícil de evaluar. Las cantidades objeto del fraude detectado en el ámbito de la atención de salud aumentan cada año (EHFCN, 2017). Los métodos tradicionales de detección del fraude en la atención de salud, que a menudo se limitan a detectarlo a posteriori en lugar de prevenirlo, hasta ahora no parecen haber sido eficientes ni eficaces. Los datos de salud son difíciles de remitir y los investigadores no pueden controlar manualmente las transacciones en tiempo real. Una manera más eficaz de prevenir el fraude y el abuso es detectarlo antes de que se abonen las cantidades reclamadas. Por lo tanto, el paradigma de la gestión del gasto indebido de salud se está transformando, al pasar del seguimiento a la prevención. Las tecnologías emergentes respaldan este cambio. Las tecnologías de detección y prevención del fraude han logrado grandes avances al reducir el tiempo de detección y permitir la creación de análisis más rápidos, avanzados y precisos. Se ha tratado de automatizar la detección del fraude mediante métodos informáticos que aplican la extracción de datos a las solicitudes de reembolso al seguro de salud, y los enfoques basados en las nuevas tecnologías facilitan la comprobación de las solicitudes de atención de salud. ISSA.(2022). Detectar el fraude en la atención de salud mediante las tecnologías emergentes. Recuperado de 4 de julio 2022 <https://ww1.issa.int/es/analysis/detecting-fraud-health-care-through-emerging-technologies>

En los dos apartados anteriores se da un resumen a alto nivel de la literatura existente sobre los problemas que aquejan e implican el fraude en las compañías aseguradoras y prestadoras de salud y nos permite evidenciar que a pesar de la importancia y el impacto económico que implica para las compañías no existen herramientas desarrolladas que mejoren la forma como actualmente se realizan los procesos de auditoría a las facturas entregadas por los prestadores, por eso la importancia del proyecto “DASHBOARD PARA LA IDENTIFICACIÓN DE LOS POSIBLES INTENTOS DE FRAUDE DE LAS PRESTADORAS DE SERVICIO ASOCIADAS A LA ASEGURADORA” ya que permite automatizar un proceso manual e incluir tableros para la revisión de casos gestionados.

Comprender adecuadamente la tipología del fenómeno del fraude en la atención de salud es fundamental para elaborar estrategias adecuadas para la buena gobernanza de los datos en el sistema de salud. Es importante normalizar las definiciones para mejorar la comunicación y el intercambio de datos, lo que permite establecer indicadores internacionales de referencia y elaborar medidas específicas. El fraude y el abuso en la atención de salud afecta a todos los sectores de la industria sanitaria, a saber, los fabricantes de medicamentos y de dispositivos, los hospitales, las farmacias, los médicos, los proveedores, los distribuidores, los laboratorios, los pacientes y los pagadores. Es posible que el grupo más afectado sean los pagadores, tanto de la sanidad pública como de la privada. El fraude en la atención de salud se produce de distintas formas, como el soborno, las solicitudes falsas y la derivación ilegal de pacientes. ISSA. (2022). Detectar el fraude en la atención de salud mediante las tecnologías emergentes. Recuperado de 4 de julio 2022 <https://ww1.issa.int/es/analysis/detecting-fraud-health-care-through-emerging-technologies>

En el texto anterior nos da una muestra de los primeros intentos por evaluar y entender cómo se puede abordar la problemática del fraude y que es necesario para combatirla estrategias que se abordan en nuestro proyecto y se aplican como normalización de los datos creación de medidas específicas, identificación de la información relevante y todo esto se realizó basado framework DAMA que nos brinda herramientas para lograr el objetivo.

### **3.3. Impacto**

Tras la puesta en marcha del proyecto Dashboard para la identificación de los posibles intentos de fraude de las prestadoras de servicio asociadas a la aseguradora, se estima un impacto positivo en el procesamiento de información aumentando el volumen de información procesada superior al 85% con un nivel de calidad y efectividad del 90% mejorando la auditoría y evitando revisión de casos de bajo impacto.

Se estima con la implementación de tableros se mejoren un 90% los tiempos de respuesta y toma de decisiones con relación a reglas o implementación de nuevos controles que mejoren la detección de casos de fraude.

### **3.4. Componente de Innovación**

El presente proyecto ofrecerá un nuevo método de revisión e identificación del fraude permitiendo a los auditores trabajar de forma más certera de casos con alto índice de fraude y no realizando reprocesos con casos que no generen valor, adicional se entregará un tablero En el cual se reflejará toda la información de la gestión realizada mostrando el ahorro obtenido cantidad de casos gestionados por prestadora de servicio.

## 4. Metodología

Con el propósito de alcanzar los objetivos definidos para el proyecto se tuvo como referente el marco teórico establecido en la metodología CRISP-DM (Cross- industry Standard Process for Data Mining) (IBM. 2021), además, del DAMA (DAMA International,2017).

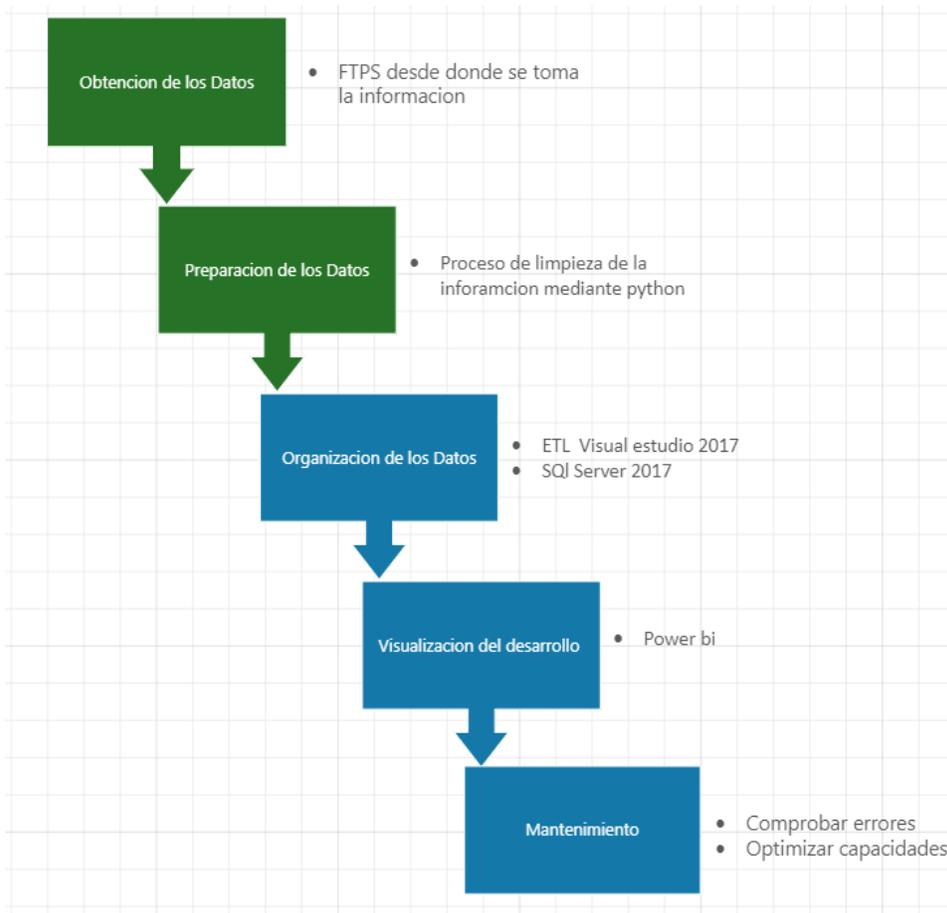
En el proyecto de automatización y limpieza de datos resulta fundamental para garantizar la calidad de los datos y el éxito del proyecto. CRISP-DM siendo un método estándar para la minería de datos que se centra en el proceso completo de extracción de conocimiento a partir de datos, desde la comprensión del negocio hasta la implementación de los modelos predictivos. Por otro lado, DAMA es una metodología específica para la gestión de datos que se enfoca en la definición de políticas, estándares y procedimientos para asegurar la calidad y el uso adecuado de los datos.

La combinación de ambas metodologías permite abordar de manera integral el proceso de automatización y limpieza de datos, desde la identificación de las necesidades del negocio, la recolección y limpieza de los datos, hasta la implementación de soluciones automatizadas para la gestión de estos.

La utilización de CRISP-DM y DAMA también garantiza una gestión adecuada del ciclo de vida de los datos, lo que implica que se mantengan actualizados y disponibles para su uso en las diferentes etapas del proyecto. Además, permite establecer un control de calidad de los datos, lo que implica la identificación y corrección de errores o inconsistencias en los mismos.

Esto permitió establecer las fases siguientes para el desarrollo del trabajo. (ver figura #1).

**Figura 1 Metodología del Proyecto**



Fuente: elaboración propia tomado de Visio (2023)

A continuación, se describe de forma detallada las fases correspondientes a la metodología aplicada.

#### 4.1 Obtención de los datos

Esta fase se enfoca en la recopilación de datos de diferentes fuentes, como bases de datos, archivos de texto, hojas de cálculo, sensores o páginas web. Los datos pueden ser estructurados o no estructurados y pueden requerir diferentes técnicas para su recopilación, como web scraping o API.

#### 4.2 Preparación de los datos

En esta fase, se realiza una limpieza y transformación de los datos recopilados para que sean utilizables en el análisis. Esto puede implicar la eliminación de valores atípicos, la eliminación de datos duplicados, la normalización de datos y

la conversión de formatos de datos. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

### **4.3 Organización de los datos**

En esta fase, los datos limpios y transformados se organizan en un formato que permita su análisis y exploración. Esto puede incluir la creación de tablas de datos, la unión de tablas de datos, la creación de índices o la agregación de datos. Muller, M. (2019). *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*. Technics Publications.

### **4.4 Visualización de los datos**

En esta fase, se utilizan herramientas de visualización de datos para representar los datos de manera gráfica y significativa. Esto puede implicar la creación de gráficos, diagramas de dispersión, mapas o tablas de resumen para facilitar la comprensión y la interpretación de los datos.

Cada fase de la metodología de procesamiento de datos es importante para garantizar que los datos se recopilen, limpien y presenten de manera efectiva para la toma de decisiones empresariales informadas y precisas. Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (2nd ed.). Analytics Press.

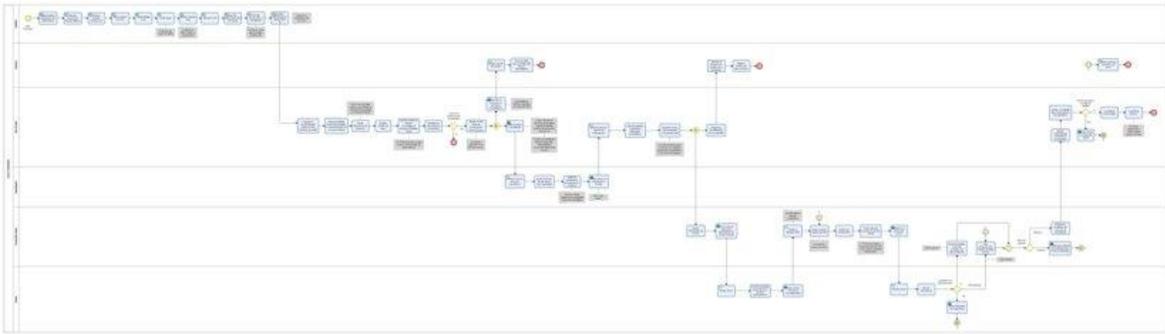
## 5. Desarrollo de la propuesta

A continuación, se describe cada uno de los procesos y sus respectivas actividades involucrando en cada etapa de la metodología planteada.

### 5.1 Obtención de los datos

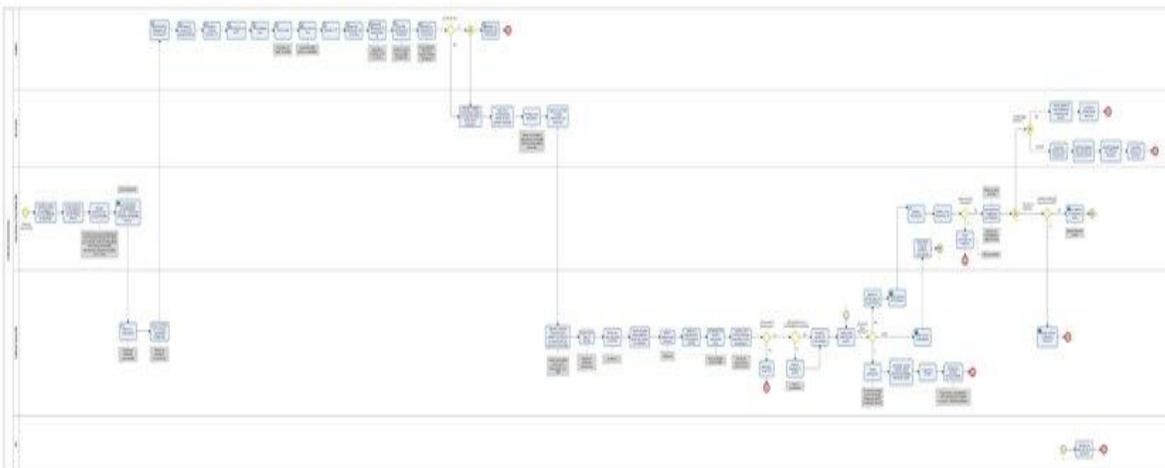
La aseguradora cuenta con proveedores externos que almacenan y consolidan información de las diferentes plataformas dispuestas a los aliados en salud, correspondiente a la radicación de facturación de servicios prestados por prestación de servicios de salud en las plataformas que recopila y almacena toda la información correspondiente a cobros generados por prestación de servicios en salud (EPS) y se dispone a la aseguradora mediante FTTP, como se muestra en figura #2, figura #3, y figura #4 en el cual se refleja los diferentes proceso de negocio. Se incluyen en los anexos para mayor claridad anexos 28 y 29

Figura 2 Proceso negocio Altos Consumos



Fuente: Elaboración propia Proceso de Negocio

Figura 3 Proceso negocio Estancias Prolongadas



Fuente: Elaboración propia Proceso de Negocio

Figura 4 Proceso negocio Omisión Preexistencias



Fuente: Elaboración propia Proceso de Negocio

Una vez identificado el proceso y las fuentes de esta información, se asigna a los encargados de cada área la tarea de extraer los datos y almacenarlos en el repositorio designado” Intercambio terceros”, son 3 procesos los cuales se encuentra distribuidos así.

En cantidad de fuentes de información para el proceso de **omisión de preexistencias** es necesario cargar (10) fuentes en archivos .xlsx nombrados así.

- STR\_SAL\_AHORROS\_INICIATIVA
- STR\_SAL\_ENFERMEDADES
- STR\_SAL\_GASTO\_MEDICO
- STR\_SAL\_HISTORICO\_OMISION
- STR\_SAL\_HISTORICOS\_INVESTIGADOS
- STR\_SAL\_HYC\_USUARIOS\_VIGENTES\_2021
- STR\_SAL\_MPP\_USUARIOS\_VIGENTES\_2021
- STR\_SAL\_NEGACIONES
- STR\_SAL\_PARAMETROS
- STR\_SAL\_PROCEDIMIENTOS

para el proceso de **Altos Consumos** es necesario cargar (3) fuentes en archivos .xlsx nombrados así.

- GASTO\_MEDICO\_HISTORICO\_ALTOS
- STR\_SAL\_AHORROS\_INICIATIVA\_ALTOS\_CONSUMOS
- STR\_SAL\_PARAMETOS\_ALTOS\_CONSUMOS

para el proceso de **Estancias Prolongadas** es necesario cargar (5) fuentes en archivos .xlsx nombrados así.

- STR\_SAL\_AHORROS\_ESTANCIAS
- STR\_SAL\_PARAMETOS\_ESTANCIA
- STR\_SAL\_PARAMETOS\_ESTANCIA\_DIARIO
- STRL\_SAL\_CENSO\_HOSPITALARIO\_DIARIO
- STRL\_SAL\_CENSO\_HOSPITALARIO\_MENSUAL

es importante que la informacion alojada en la ruta cuente con la estructura y nombre definido como se evidencia en tablas #1 y tabla #2. la relación completa de las tablas se encuentra en el anexo 1,2,3,4,5,6.

tabla 1 repositorio de fuentes Altos

FECHA_PRC	DIAGNO	Regional	PROVEEDOR	EPS	FECHA_ASIC	FECHA_ENT	DIAS	ESTADO_IN	ESTADO_P	VALOR_AHO	COSTO	MES_CUAN	ID
28/09/2021		BOGOTÁ	CONFYGES	FAMISANAR	28/09/2021	08/10/2021	8	Alertado	Dx indirectos	\$ 0			8E+07
28/09/2021		CARTAGENA	CONFYGES	NUEVA EPS	28/09/2021	08/10/2021	8	Alertado	Positivo	\$ 6.813.102		01/10/2021	1E+09
28/09/2021		BOGOTÁ	CONFYGES	SANITAS	28/09/2021	08/10/2021	8	Sin Historia	Sin Historia	\$ 0			1,1E+09
28/09/2021		CALI	CONFYGES	SSS COLMED	28/09/2021	08/10/2021	8	Sin Historia	Sin Historia	\$ 0			3,2E+07
28/09/2021		BUCARAMAN	CONFYGES	SANITAS	28/09/2021	08/10/2021	8	Sin Historia	Sin Historia	\$ 0			1,1E+09
28/09/2021		BUCARAMAN	INVIPRO	SANITAS	28/09/2021	08/10/2021	8	Alertado	Positivo	\$ 8.552.398		01/12/2021	1,1E+09
28/09/2021		IBAGUE	INVIPRO	SANITAS	28/09/2021	08/10/2021	8	Alertado	Positivo	\$ 6.542.130		01/10/2021	6,6E+07
28/09/2021		BARRANQUI	INVIPRO	SURAMERICA	28/09/2021	08/10/2021	8	Sin Historia	Sin Historia	\$ 0			3,3E+07
28/09/2021		BOGOTÁ	INVIPRO	SURAMERICA	28/09/2021	08/10/2021	8	Sin Historia	Sin Historia	\$ 0			1,0E+07

Fuente: Elaboración repositorio de fuentes a cargar

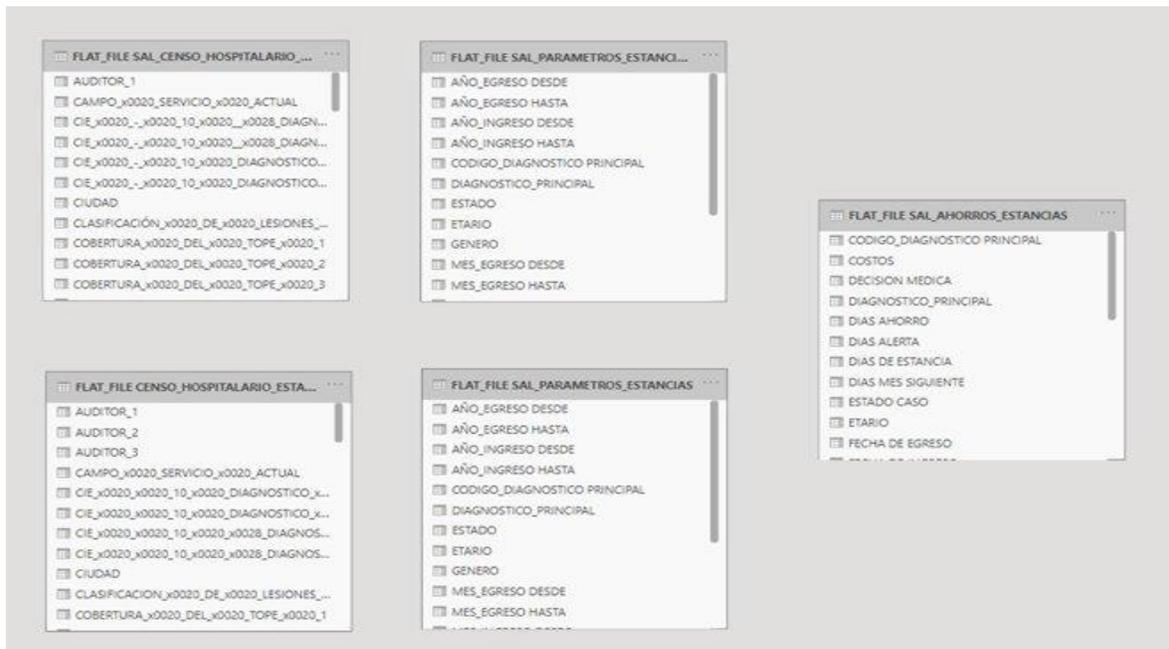
tabla 2 repositorio de fuentes Estancias

NUMERO	FECHA	AREA	QUE	ID	NUMEF	SUCURSAI	CLAVE	NOMB	DIAGNOS	CÓDIGO	CI	ESPECI	DECISION	FECHA	MODUL	BH	FECHA	OBSERVACIO
1	44615	FRAUDE	CC	1E+09				RINITIS AL	J303J343S6	ORTOF	FIJAR PRE	44663		43866				
2	44585	AUTORIZACK	CC	1E+08				HIPERTENI	110X	PSIQUI	FIJAR PRE	44677						
3	44610	FRAUDE	CC	7E+07				HIPERLIP	E782I872I8	NEURC	FIJAR PRE	44685		43851			CONTRATO C	
4	44642	AUTORIZACK	CC	7E+07				HIPERPLA	N40XK760J	PSIQUI	CANCELA	44685		43854				
5	44669	AUTORIZACK	TI	1E+09				PROTEINL	N069L932F	NEURC	CANCELA	44726		43912				

Fuente: Elaboración repositorio asignado de fuentes a cargar

A continuación, se relaciona el modelo de datos de las tres iniciativas que componen el proyecto Estancias Prolongadas Omisión de Preexistencia y altos Consumos.

Figura 5 Modelo De datos estancias



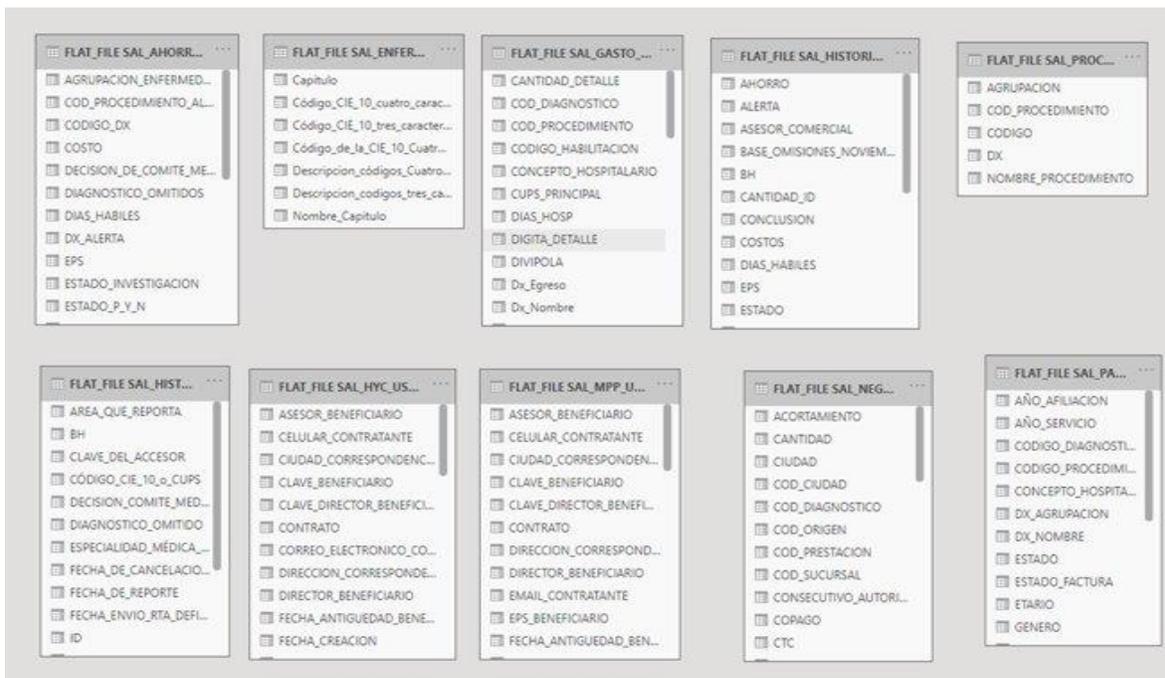
Fuente: Elaboración propia Modelo Estancias

Figura 6 Modelo De datos Omisión



Fuente: Elaboración propia Modelo Omisión

Figura 7 Modelo De datos altos consumos



Fuente: Elaboración propia Modelo De datos altos consumos

Relación fuentes requeridas para el cargue y posterior proceso de limpieza de uno de los procesos, ya que uno de los controles establecidos desde línea Bach es la validación antes de ejecutar el proceso, el programa valida que la totalidad de las fuentes se encuentre en el repositorio con el nombre establecido y fechas actualizadas. Como se evidencia en figura #8 y figura #9. Anexo

Figura 8 repositorio de fuentes

Nombre	Fecha de modificación	Tipo	Tamaño
Dashboards	10/03/2023 9:14 p. m.	Carpeta de archivos	
HISTORICO	8/03/2023 5:00 p. m.	Carpeta de archivos	
STR_SAL_AHORROS_INICIATIVA	4/05/2023 8:50 p. m.	Hoja de cálculo d...	200 KB
STR_SAL_ENFERMEDADES	22/02/2023 8:12 a. m.	Hoja de cálculo d...	633 KB
STR_SAL_GASTO_MEDICO	4/05/2023 8:50 p. m.	Hoja de cálculo d...	18.375 KB
STR_SAL_HISTORICO_OMISION	4/05/2023 8:50 p. m.	Hoja de cálculo d...	65 KB
STR_SAL_HISTORICOS_INVESTIGADOS	4/05/2023 8:50 p. m.	Hoja de cálculo d...	148 KB
STR_SAL_HYC_USUARIOS_VIGENTES_2021	4/05/2023 3:58 p. m.	Hoja de cálculo d...	4.375 KB
STR_SAL_MPP_USUARIOS_VIGENTES_2021	4/05/2023 8:50 p. m.	Hoja de cálculo d...	21.195 KB
STR_SAL_NEGACIONES	8/05/2023 9:10 a. m.	Hoja de cálculo d...	9.029 KB
STR_SAL_PARAMETROS	4/05/2023 8:50 p. m.	Hoja de cálculo d...	310 KB
STR_SAL_PROCEDIMIENTOS	4/05/2023 8:50 p. m.	Hoja de cálculo d...	975 KB

Fuente: Elaboración ubicación repositorio de fuentes a cargar

Figura 9 repositorio de fuentes intercambio terceros

ntercambioTerceros > SALUD > FRAUDE\_SALUD > Altos\_Consumos > input > Usuarios > BK10

Buscar en BK10

Nombre	Fecha de modificación	Tipo	Tamaño
Gasto_Medico_Historico_Altos	16/03/2023 10:45 a. m.	Hoja de cálculo d...	32.504 KB
STR_SAL_AHORROS_INICIATIVA_ALTOS_C...	16/03/2023 11:07 a. m.	Hoja de cálculo d...	14 KB
STR_SAL_PARAMETOS_ALTOS_CONSUMOS	16/03/2023 9:53 a. m.	Hoja de cálculo d...	346 KB

patria Seguros

Fuente: Elaboración ubicación repositorio de fuentes a cargar

Para poder definir la estructura y tipo de información que se carga en los repositorios fue necesario la creación y socialización de un diccionario de datos que permitió facilitar la parametrización de fuentes e identificación de campos esenciales para la ejecución del proceso y prevenir que el programa presente errores al intentar cargar fuentes con estructura diferente, adicional se logró tener un mayor entendimiento de la información contenida en cada una de las fuentes en tabla #3, Tabla #4 y tabla #5 Anexos 10,11,12,13,14,15,16 se muestra la estructura creada para el diccionario de datos de varias de las fuentes que hacen parte del proceso.

tabla 3 Diccionario de datos enfermedades

N°	Nombre del Campo	Ejemplo	Descripción del Campo
1	FECHA_PROCESO	01/02/2023 23:51	fecha ejecucion
2	NUMERO_INTERNO	19978122	numero_interno
3	UNIDAD_NEGOCIO	AXA Colpatría-Medicina prepagada MPP	unidad_negocio
4	FOLIO	80868198	folio
5	SUC_PAGO		suc_pago
6	SUC_CONTRATO_COLPATRIA	640	suc_contrato_colpatria
7	DIVIPOLA	5615	divipola
8	CODIGO_HABILITACION	56151204401	codigo_habilitacion
9	PLAN	11	plan
10	TIPO_CONTRATO		tipo_contrato
11	NRO_CONTRATO	1,12856E+11	nro_contrato
12	TIPO_IDEN_BENEF	CC	tipo_iden_benef
13	NRO_IDENTIFICACION_BENEF	43757350	nro_identificacion_benef
14	VALOR_BRUTO_FACTURA	0	valor_bruto_factura
15	VALOR_NETO	0	valor_netto
16	VALOR_VALES	0	valor_vales
17	VALOR_DESCUENTO	0	valor_descuento
18	COD_DIAGNOSTICO	R030	cod_diagnostico
19	COD_PROCEDIMIENTO	39145	cod_procedimiento
20	NOMBRE_PROCEDIMIENTO	Consulta de urgencias	nombre_procedimiento
21	VALOR_UNITARIO	55800	valor_unitario

Fuente: Elaboración propia Diccionario de datos Enfermedades

tabla 4 Diccionario de datos Gasto medico

N°	Nombre del Campo	Ejemplo	Descripción del Campo
1	NUMERO_INTERNO	20848751	Número interno del registro
2	UNIDAD_NEGOCIO	AXA Colpatría-Medicina prepagada MPP	Descripción de la unidad de negocio de salud a la que pertenece el contrato
3	FOLIO	81212723	Número de folio del gasto medico
4	SUC_PAGO		Nombre de la sucursal de donde proviene el pago del contrato
5	SUC_CONTRATO_COLPATRIA	610	Nombre de la sucursal desde donde se vendió el contrato
6	DIVIPOLA	11001	Nomenclatura estandarizada para la identificación de Entidades Territoriales (departamentos, distritos y municipios), Áreas No Municipalizadas y Centros Poblados de Colombia
7	CODIGO_HABILITACION	110010561801	Código de la habilitación del contrato
8	PLAN_S	15	Nombre del plan de salud al que pertenece el cliente
9	TIPO_CONTRATO		Nombre del tipo de contrato de salud asociado al cliente
10	NRO_CONTRATO	152100000000	Número del contrato de salud asociado al cliente

Fuente: Elaboración propia Diccionario de datos Gasto medico

Tabla 5 Diccionario de datos

N°	Nombre del Campo	Ejemplo	Descripción del Campo	Criticidad	Tipo del Cam	Longitu	Puede ir vacío	Cuándo puede ir va
1	CANTIDAD_ID	1	Conteo de casos	ALTO	nvarchar	510	NO	N/A
2	INPUT	OMS	Clasificación del caso	ALTO	nvarchar	510	NO	N/A
3	ID_USUARIO	80074843	Número del documento de identidad del usuario	ALTO	nvarchar	510	NO	N/A
4	NOMBRE_USUARIO	JORGE MARIO MERCADO DE LA ROSA	Nombre completo del usuario	ALTO	nvarchar	510	NO	N/A
5	ASESOR_COMERCIAL	ADOLFO BOLIVAR MENESES	Nombre del asesor comercial	ALTO	nvarchar	510	NO	N/A

Fuente: Elaboración propia Diccionario de datos Histórico-Investigados

Para iniciar el proceso de cargue de información, se asignó una carpeta nombrada dependiendo de cada proceso para el cargue de información en repositorios diferentes como se evidencia en figura #10.

Figura 10 Cargue inicial

IntercambioTerceros > SALUD > FRAUDE_SALUD			
Nombre	Fecha de modificación	Tipo	
Altos_Consumos		Carpeta de archivos	
Estancias_Prolongadas		Carpeta de archivos	
Omision_Preexistencias		Carpeta de archivos	

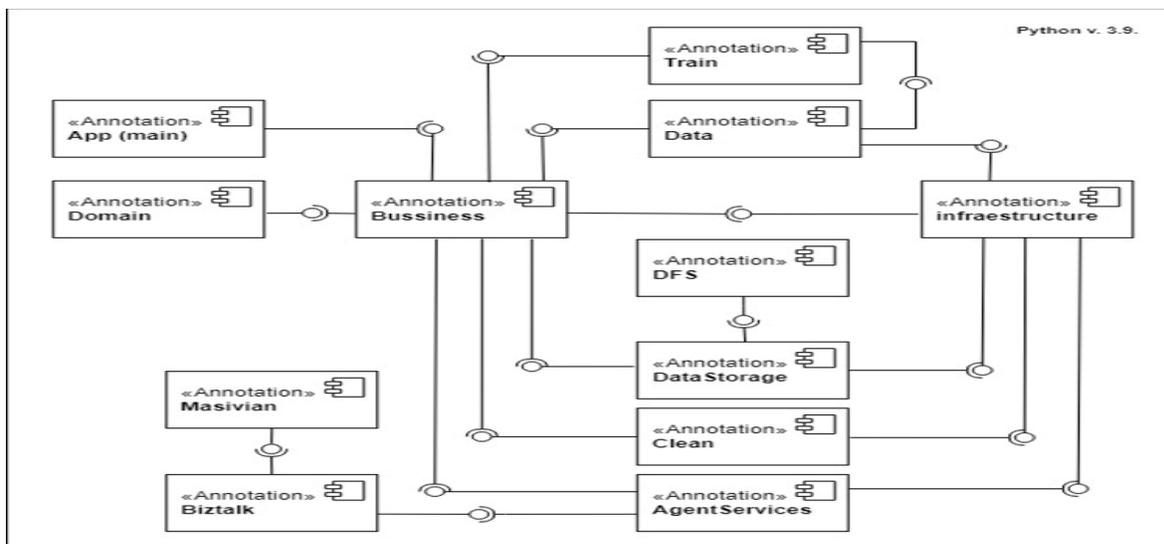
Fuente: Elaboración repositorio de fuentes a cargar

## 5.2 Preparación de datos

La preparación de los datos inicia con la ejecución de código en lenguaje de programación Python desarrolladas por capas y denominada como línea Bach en el cual inicia con la consulta de la información en el repositorio asignado, según ruta, validando que las fuentes se encuentren almacenadas, completas y estén actualizadas al momento de realizar la ejecución del proceso, como se evidencia en figura # 11, figura # 12 y figura # 13. Adicional en el proceso de limpieza que se realiza en línea Bach se identifican, los datos críticos en anexo se encuentran información crítica a modificar Anexos 28,29,30,31,32.

Los datos varían desde campos tipo fecha que presentan complejidad por el tiempo de formato el cual en el procesamiento de información debe ser unificado ya que se requieren para cálculos posteriores de cantidad tiempo afiliación

Figura 11 diagrama de componentes línea Bach



Fuente: Elaboración propia diagrama de componentes línea Bach

En las imágenes siguientes se evidencia las líneas de código en la cual se hace la importación de los datos desde la carpeta intercambio de terceros por cada proceso y extrae la información basado en parámetros específicos cargados ejemplo: nombre del archivo, ruta ubicación, tipo de archivo.

Iniciando así el proceso de limpieza en el cual se pueden identificar diferentes tipos de datos y variables relevantes. A continuación, se presenta un ejemplo de la naturaleza de los datos, el tipo de variables y el dominio correspondiente:

Gasto Medico:

**Variable:** VALOR\_BRUTO\_FACTURA.

Tipo de variable: Numérica.

Dominio: Por ejemplo, valores positivos que representen el monto facturado en la moneda utilizada (por ejemplo, dólares, euros, etc.).

**Variable:** COD\_PROCEDIMIENTO.

Tipo de variable: Categórica (nominal).

Dominio: Por ejemplo, códigos estándar utilizados en el campo de la salud, como el Sistema de Codificación de Procedimientos (CPT) o el Sistema de Clasificación de Diagnósticos (ICD), que representan procedimientos y servicios médicos específicos.

**Variable:** Edad del paciente.

Tipo de variable: Numérica (discreta).

Dominio: Por ejemplo, valores enteros positivos que representen la edad del paciente, como 18, 25, 40.

**Variable:** Genero Paciente.

Tipo de variable: Categórica (nominal).

Dominio: Por ejemplo, dos categorías posibles: masculino o femenino (M, F)

**Variable:** COD\_DIAGNOSTICO.

Tipo de variable: Categórica (nominal).

Dominio: Por ejemplo, diferentes condiciones médicas o diagnósticos que se obtienen de sistemas de clasificación médica, como el Sistema Internacional de Clasificación de Enfermedades (ICD), que asigna códigos a diferentes enfermedades y trastornos.

**Variable:** NIT\_PRESTADOR

Tipo de variable: Categórica (nominal).

Dominio: Por ejemplo, identificadores únicos asignados a cada proveedor de servicios de salud.

**Variable:** NOMBRE\_PRESTADOR.

Tipo de variable: Categórica (nominal).

Dominio: Por ejemplo, diferentes categorías que representen los tipos de proveedores de servicios de salud.

**Variable:** PLAN.

Tipo de variable: Categórica (ordinal).

Dominio: Por ejemplo, diferentes categorías que representen el nivel de cobertura o el tipo de plan de seguro, como básico, intermedio, premium, etc.

Figura 12 importación de información para limpieza

```

src > {} config.json > ...
22
23 "DebugConfiguration": {
24   "_comment": "Debug : 0 ejecuta debug, de lo contrario no",
25   "DEBUG" : 0,
26   "LOG_FILE_PATH" : "
27     \\IntercambioTerceros\\SALUD\\FRAUDE_SALUD\\Omisión_Preexistencias\\Logs\\
28   "PROJECT_NAME": "FSALUD"
29 }
30 "DefaultEmailSender": {
31   "END_POINT": "https:
32     ones/Notificaciones.svc?wsc
33   "CANAL" : "r -",
34   "ID_TRANSACCION": "0",
35   "USER": "FSALUD",
36   "SUBJECT": "NOTIFICACION FSALUD - TERMINACION INCORRECTA DE PROCESO",
37   "NOTIFICATION_SENDER": "
38   "MSJ" : "ESTA ES EL MENSAJE DE ENVIO DE CORREO POR DEFECTO",
39   "RECIPIENTS": "carlo:
40   "PROCESO": " " ,
41   "NOMBRE_PROYECTO" : "FSALUD"
42 },
43 "paths_input_output": {
44   "INPUT": "
45     \\IntercambioTerceros\\SALUD\\FRAUDE_SALUD\\Omisión_Preexistencias\\Input\\Usuarios
46   "OUTPUT": "
47     \\IntercambioTerceros\\SALUD\\FRAUDE_SALUD\\Omisión_Preexistencias\\Output\\RESULT

```

Fuente: Editor Python importación de información para limpieza

Figura 13 Exportación de terceros a datos

```

1  from src.domain.response.fileParametersResponse import FileParametersResponse
2  from pandas.core.frame import DataFrame
3  from src.dataStorage.file import File
4
5  class FileManagement:
6
7      @classmethod
8      def importFile(cls, process:str, parameters:FileParametersResponse) -> DataFrame:
9          print('Inicio Importación del archivo: ', parameters.inputFileName or 'None')
10         file:DataFrame = File.getFile(parameters=parameters, process=process)
11         return file
12
13     @classmethod
14     def exportFile(cls, file:DataFrame, process:str, parameters:FileParametersResponse) -> DataFrame:
15         print('Inicio Exportación del archivo: ' + parameters.outputFileName or 'None')
16         File.writeFile(file = file, process = process, parameters = parameters)
17

```

Fuente: Editor Python importación de información para limpieza

si las fuentes no se encuentran actualizadas el sistema no realiza la ejecución del código enviando un log de alerta en la cual indica la no actualización de fuentes almacenadas al momento de ejecutar el proceso como se evidencia en figura #15, si por el contrario las fuentes se encuentran actualizadas el proceso inicia con la toma de la información de todas las fuentes alojadas en la carpeta y uno a uno con cada archivo realiza un proceso de limpieza de información como se evidencia en la figura #14 y figura #15 y figura #16.

Figura 14 Logs

Vista

ic2tvaftp1 > IntercambioTerceros > SALUD > FRAUDE\_SALUD > Altos\_Consumos > Logs

guros

Nombre	Fecha de modificación	Tipo	Tamaño
FSALUDALTOS20221213-112332	13/12/2022 11:23 a. m.	Documento de te...	5 KB
FSALUDALTOS20221213-114308	13/12/2022 11:43 a. m.	Documento de te...	5 KB
FSALUDALTOS20221214-170547	14/12/2022 5:05 p. m.	Documento de te...	5 KB
FSALUDALTOS20221214-170827	14/12/2022 5:08 p. m.	Documento de te...	5 KB
FSALUDALTOS20221221-115303	21/12/2022 11:53 a. m.	Documento de te...	4 KB
FSALUDALTOS20221221-115656	21/12/2022 11:56 a. m.	Documento de te...	3 KB
FSALUDALTOS20221222-084222	22/12/2022 8:42 a. m.	Documento de te...	3 KB
FSALUDALTOS20221222-084950	22/12/2022 8:49 a. m.	Documento de te...	3 KB
FSALUDALTOS20221222-085040	22/12/2022 8:50 a. m.	Documento de te...	3 KB
FSALUDALTOS20221227-121619	27/12/2022 12:16 p. m.	Documento de te...	1 KB
FSALUDALTOS20221227-125035	27/12/2022 12:50 p. m.	Documento de te...	1 KB
FSALUDALTOS20221227-133416	27/12/2022 1:34 p. m.	Documento de te...	1 KB
FSALUDALTOS20230112-114457	12/01/2023 11:44 a. m.	Documento de te...	3 KB
FSALUDALTOS20230112-140838	12/01/2023 2:08 p. m.	Documento de te...	4 KB
FSALUDALTOS20230112-150717	12/01/2023 3:07 p. m.	Documento de te...	2 KB
FSALUDALTOS20230113-123625	13/01/2023 12:36 p. m.	Documento de te...	1 KB
FSALUDALTOS20230126-222522	26/01/2023 10:25 p. m.	Documento de te...	3 KB
FSALUDALTOS20230214-131604	14/02/2023 1:16 p. m.	Documento de te...	3 KB

Fuente: ruta carpeta Log de errores

Figura 15 Error Log de errores

FSALUDALTOS20221213-112332: Bloc de notas

rchivo Edición Formato Ver Ayuda

Objeto BD	Fallo	Descripción
	NO	
	NO	
PARAMETOS_ALTOS_CONSUMOS.xlsx	NO	
PARAMETOS_ALTOS_CONSUMOS.xlsx	NO	
PARAMETOS_ALTOS_CONSUMOS.csv	NO	
	NO	
AHORROS_INICIATIVA_ALTOS_CONSUMOS.xlsx	NO	
AHORROS_INICIATIVA_ALTOS_CONSUMOS.xlsx	NO	
AHORROS_INICIATIVA_ALTOS_CONSUMOS.csv	NO	
	NO	
dico_Historico_Altos.xlsx	NO	
dico_Historico_Altos.xlsx	NO	
dico_Historico_Altos.csv	NO	
	NO	
	SI	(pymssql.pymssql.OperationalError) (50000, b'No job named "ETL_FSALUD_ALTOS_CONSUMOS"DB-Lib error message [SQL: [dbo].[SP_EXECUTE_JOB_SSIS] %(JobName)s, %(TimeLimit)s, %(Debug)s, %(Status)s] [parameters: {'JobName': 'ETL_FSALUD_ALTOS_CONSUMOS', 'TimeLimit': '30', 'Debug': '0', 'Status': '1'}] (Background on this error at: https://sqlalche.me/e/14/e3q8)
	SI	FALTA EL PARAMETRO SUBJECT_FUENTE_AUSENTE: 1 validation error for ParameterResponse: parameter: none is not an allowed value (type=type_error.none.not_allowed)
	SI	'ErrorCategory' object has no attribute 'end_point'
	SI	expected a simple type, a tuple or a list
	NO	

Fuente: editor código Python error Log de errores

Figura 16 Validación existencia de fuentes

```

12 class verify_date():
13
14
15 def def_verify_input(path:str):
16     today_date = date.today()
17     name_file_0 = path+"Gasto_Medico_Historico_Altos.xlsx"
18     name_file_1 = path+"STR_SAL_AHORROS_INICIATIVA_ALTOS_CONSUMOS.xlsx"
19     name_file_2 = path+"STR_SAL_PARAMETOS_ALTOS_CONSUMOS.xlsx"
20
21
22     if str(time.strftime("%Y-%m-%d",time.localtime(getmtime(name_file_0)))) == str(today_date):
23         | print("sí se encuentra Gasto_Medico_Historico_Altos.xlsx")
24     else:
25         | print("no se encuentra actualizado Gasto_Medico_Historico_Altos.xlsx")
26
27
28     if str(time.strftime("%Y-%m-%d",time.localtime(getmtime(name_file_1)))) == str(today_date):
29         | print("sí se encuentra STR_SAL_AHORROS_INICIATIVA_ALTOS_CONSUMOS.xlsx")
30     else:
31         | print("no se encuentra actualizado STR_SAL_AHORROS_INICIATIVA_ALTOS_CONSUMOS.xlsx")
32
33
34     if str(time.strftime("%Y-%m-%d",time.localtime(getmtime(name_file_2)))) == str(today_date):
35         | print("sí se encuentra STR_SAL_PARAMETOS_ALTOS_CONSUMOS.xlsx")
36     else:

```

Fuente: editor código Python error Log de errores

A continuación, se describen los pasos generales para el alistamiento y limpieza de información del archivo xlsx. Lo primero es verificar la calidad de los datos, Esto implica la revisión de la información para asegurarse de que no haya celdas vacías, filas o columnas que no contengan datos relevantes. Posterior a esto es Identificar y eliminar datos duplicados ya que es común encontrar datos duplicados en los archivos XLS, como se evidencia en imagen figura #17, figura #18 y figura #19.

Figura 17 Limpieza de datos

```

14 @classmethod
15 def cleanData(cls, df : DataFrame, process :str, parameters : FileParametersResponse) -> DataFrame:
16     print('Inicio Limpieza del archivo:' + parameters.inputFileName)
17     try:
18
19         objectPy = parameters.inputFileName + parameters.inputFileType
20         validerow : int = 0
21         for col_name in df.columns:
22             if col_name in parameters.textColumns:
23                 CleanString.cleanString(df, col_name)
24                 df[col_name] = [x[:200] for x in df[col_name]]
25             elif col_name in parameters.dateColumns:
26                 CleanDate.cleanDate(df, col_name)
27             elif col_name in parameters.numericColumns:
28                 CleanNumeric.cleanNumeric(df, col_name)
29             else:
30                 validerow = 1
31                 LoggingHandler.emit(*LogCodification.cleanUnspecifiedDataType(process=process, objectPy=objectPy)
32
33         if validerow == 0: #Si no ocurrieron errores, se crea el registro de que no falló
34             LoggingHandler.emit(*LogCodification.clean_success(process=process, objectPy=objectPy))
35         else:
36             raise #No sé que hacer aún
37         return df
38
39 except Exception as ex:

```

Fuente: editor código Python limpieza de datos

Figura 18 limpieza de datos

```

1 from numpy import inf
2 import pandas as pd
3
4 class CleanNumeric():
5
6     @classmethod
7     def cleanNumeric(cls, df, col_name) -> pd.DataFrame:
8
9         df[col_name] = df[col_name].astype(str)
10
11         df.loc[:, col_name]= df.loc[:, col_name].str.lstrip()
12         df.loc[:, col_name]= df.loc[:, col_name].str.rstrip()
13         df.loc[:, col_name]= df.loc[:, col_name].str.replace(";", "")
14         df.loc[:, col_name]= df.loc[:, col_name].str.split('.').str[0]
15
16         df[col_name].replace({'nan': '0'}, regex=True, inplace=True)
17         df[col_name].replace({'NAN': '0'}, regex=True, inplace=True)
18         df[col_name].replace({' ': '0'}, regex=True, inplace=True)
19         df[col_name].replace({' ': ''}, regex=True, inplace=True)
20         df[col_name].replace({'-': ''}, regex=True, inplace=True)
21         df[col_name].replace({'SINDATOS': '0'}, regex=True, inplace=True)
22
23         df[col_name] = df[col_name].astype('int64')
24
25         return df

```

Fuente: editor código Python limpieza de datos

Por lo tanto, se identifican y eliminan los datos duplicados para garantizar que los datos sean precisos y confiables. Se corrigen los errores de formato ya que es muy común, que los datos pueden tener problemas de formato, como valores de fecha incorrectos e invertidos o valores de moneda que se muestran como texto.

Figura 19 componentes

```

class CleanString():
    @classmethod
    def cleanString(cls, df, col_name) -> pd.DataFrame:
        df[col_name] = df[col_name].astype(str)
        df.loc[pd.isna(df[col_name]), col_name]=''
        df.loc[:, col_name]= df.loc[:, col_name].str.upper()
        df.loc[:, col_name]= df.loc[:, col_name].str.lstrip()
        df.loc[:, col_name]= df.loc[:, col_name].str.rstrip()
        replaceDict = {";":"", "Ã": "O", "Ä": "O", "Ï½": "O", "Á": "A", "É": "E", "Í": "I", "Ó": "O", "Ú": "U", "À": "A"}
        for key,value in replaceDict.items():
            df.loc[:, col_name]= df.loc[:, col_name].str.replace(key, value, regex=True)
        df[col_name]= df[col_name].apply(lambda x: x.replace('NAN','') if 'NAN' in x and len(x)==3 else x)
        df[col_name]= df[col_name].apply(lambda x: x.replace('nan','') if 'nan' in x and len(x)==3 else x)
        df[col_name]= df[col_name].apply(lambda x: x.replace('0x2A','') if 'nan' in x and len(x)==4 else x)
        df[col_name]= df[col_name].apply(lambda x: x.replace('0x17','') if 'nan' in x and len(x)==4 else x)

```

Fuente: editor código Python limpieza de datos

Una vez se realiza el proceso de limpieza y preparaciones los datos con la ejecución de la línea Bach se dispone a dejar las fuentes en un nuevo repositorio nombrado intercambio de datos esto con el fin de garantizar que la información migrada a esta carpeta ya se encuentra limpia y normalizada cambiando la extensión de los archivos a CSV esto para que al momento de disparar la ETL tome los archivos y no presente problemas con el cargue a tablas físicas como se evidencia en figura #20.

Figura 20 repositorio intercambio de datos

Nombre	Fecha de modificación	Tipo	Tam
STR_SAL_AHORROS_INICIATIVA_2023-05-04	4/05/2023 4:14 p. m.	Archivo de valores...	
STR_SAL_HISTORICO_OMISION_2023-05-04	4/05/2023 4:14 p. m.	Archivo de valores...	
STR_SAL_HISTORICOS_INVESTIGADOS_2023-05-04	4/05/2023 4:14 p. m.	Archivo de valores...	
STR_SAL_PROCEDIMIENTOS_2023-05-04	4/05/2023 4:14 p. m.	Archivo de valores...	
STR_SAL_ENFERMEDADES_2023-05-04	4/05/2023 4:14 p. m.	Archivo de valores...	
STR_SAL_NEGACIONES_2023-05-04	4/05/2023 4:14 p. m.	Archivo de valores...	2
STR_SAL_GASTO_MEDICO_2023-05-04	4/05/2023 4:13 p. m.	Archivo de valores...	3
STR_SAL_HYC_USUARIOS_VIGENTES_2023-05-04	4/05/2023 4:10 p. m.	Archivo de valores...	
STR_SAL_PARAMETROS_2023-05-04	4/05/2023 4:09 p. m.	Archivo de valores...	
STR_SAL_AHORROS_INICIATIVA_2023-04-11	11/04/2023 8:00 a. m.	Archivo de valores...	
STR_SAL_HISTORICO_OMISION_2023-04-11	11/04/2023 8:00 a. m.	Archivo de valores...	
STR_SAL_HISTORICOS_INVESTIGADOS_2023-04-11	11/04/2023 8:00 a. m.	Archivo de valores...	
STR_SAL_PROCEDIMIENTOS_2023-04-11	11/04/2023 8:00 a. m.	Archivo de valores...	
STR_SAL_ENFERMEDADES_2023-04-11	11/04/2023 8:00 a. m.	Archivo de valores...	
STR_SAL_NEGACIONES_2023-04-11	11/04/2023 7:59 a. m.	Archivo de valores...	1
STR_SAL_GASTO_MEDICO_2023-04-11	11/04/2023 7:57 a. m.	Archivo de valores...	2
STR_SAL_HYC_USUARIOS_VIGENTES_2023-04-11	11/04/2023 7:53 a. m.	Archivo de valores...	

Fuente: repositorio intercambio de datos

Posteriormente, se envía la orden desde Python para la ejecución de la ETL la cual extraerá la información de la carpeta intercambio de datos y la carga en tablas físicas, en la base Landing Zone o “Zona de aterrizaje”, esta base de datos se creó con el fin de realizar alistamientos adicionales a la información ya que, desde el área de negocio, las tablas iniciales pueden ser reutilizadas para diferentes procesos una vez cargados en la base de datos Landing Zone.

se almacenan con el nombre definido en el diccionario de datos sin espacio ni caracteres especiales como se evidencia en figura #21, figura # 22 y figura # 23

Figura 21 código limpieza

```

46 SsisParameters = SsisRepository.getSsisParameters(process='ETL')
47 if isinstance(SsisParameters, SsisParametersResponse):
48     resultEtl = SsisRepository.executeEtl(SsisParameters)
49     print(resultEtl)
50 if resultEtl.result == '1':
51     LoggingHandler.emit(*LogCodification.ssis_ejecucion_success())
52     allfiles= os.listdir(parameters.outputPath)
53     for f in allfiles:
54         shutil.move(parameters.outputPath+ f, parameters.archiveDes+ f)
55

```

Fuente: editor Python orden ejecución ETL

Figura 22 ejecución Jobs envía que tarea se está ejecutando

```

46 @classmethod
47 def executeEtl(cls, ssisParam:SsisParametersResponse) -> SsisResponse:
48     try:
49         print('Inicio Ejecución ETL: ', ssisParam.jobname)
50         validerows:int = 0
51         response = Connection.session.execute('[[dbo].[SP_EXECUTE_JOB_SSIS] :JobName, :TimeLimit, :Debug, :Statu
52         for row in response:
53             validerows = 1
54             etlResponse: SsisResponse = SsisResponse(result = row['RESULT'])
55
56         if validerows == 0 :
57             etlResponse: SsisResponse = SsisResponse(result = None)
58
59     except exc.SQLAlchemyError as errorBD:
60         etlResponse: SsisResponse = None
61         LoggingHandler.emit(*LogCodification.bd_error(errorBD))
62
63     except Exception as ex:
64         etlResponse: SsisResponse = None
65         Connection.session.rollback()
66         LoggingHandler.emit(*LogCodification.parameters_other_exception(process='ETL', parameter='ETL', error =
67

```

Fuente: editor Python ejecución Jobs

Figura 23 Envía notificación de ejecución proceso Jobs

```

46 SsisParameters = SSISRepository.getSSISParameters(process='ETL')
47 if isinstance(SsisParameters, SsisParametersResponse):
48     resultEtl = SSISRepository.executeEtl(SsisParameters)
49     print(resultEtl)
50 if resultEtl.result == '1':
51     LoggingHandler.emit(*LogCodification.ssis_ejecution_sucess())
52     allfiles= os.listdir(parameters.outputPath)
53     for f in allfiles:
54         shutil.move(parameters.outputPath+ f, parameters.archiveDes+ f)
55

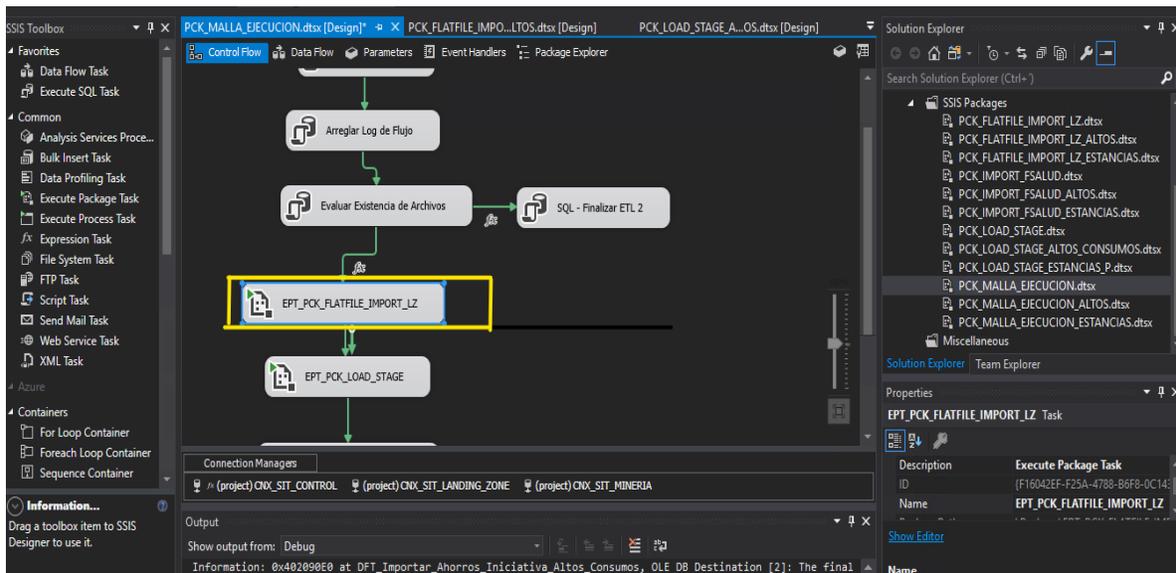
```

Fuente: editor Python ejecución Jobs

### 5.3 Organización de los datos

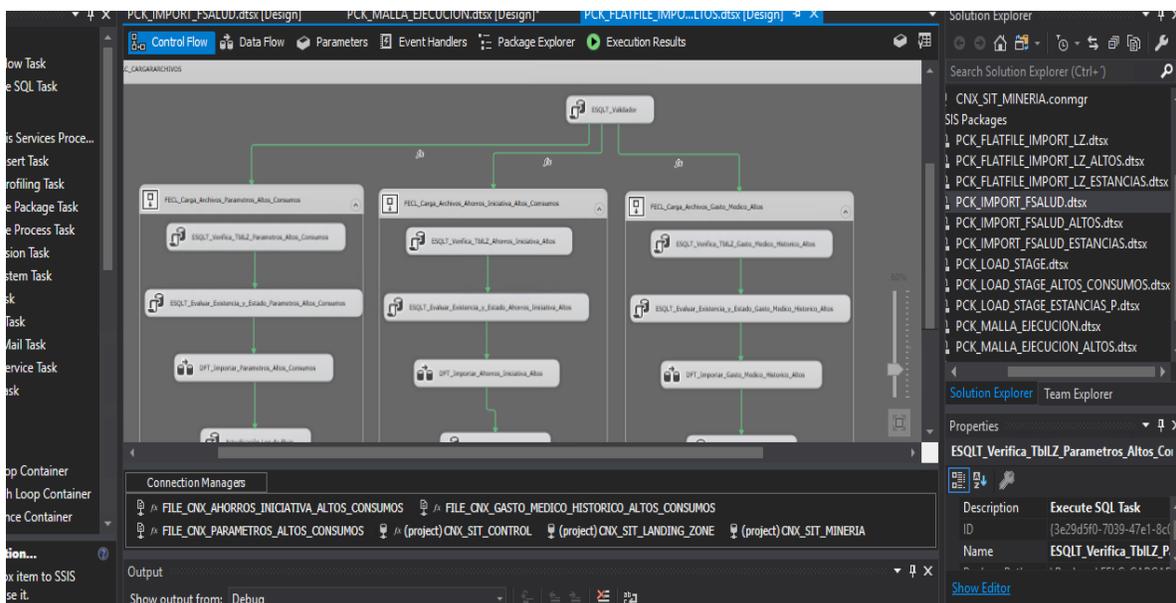
Una vez ya ejecutada la malla de la ETL donde se realizó la creación de las tablas en el esquema y base de datos específica se procede a validar si es el primer cargue se inserta la información, pero si por el contrario ya se habían realizado cargues anteriores se trunca la tabla para el nuevo cargue como se evidencia en las figuras #24, figura#25, figura#26, figura#27.

Figura 24 Inicio ejecución ETL Importación landing zone



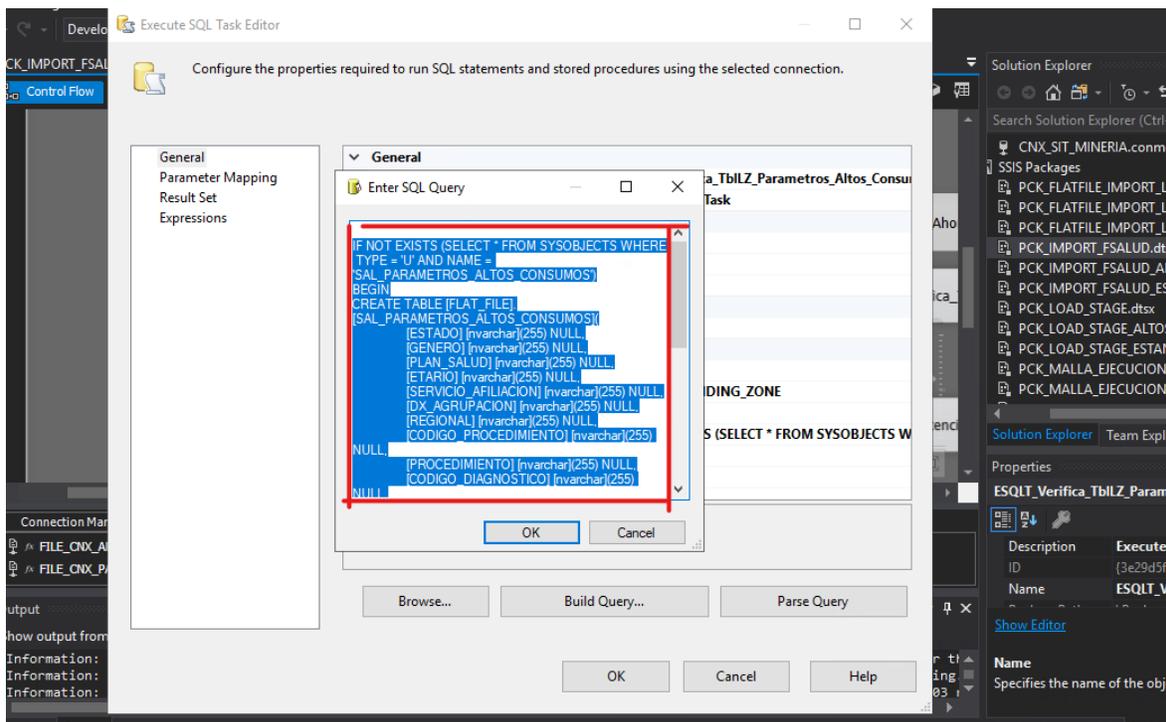
Fuente: Elaboración propia Línea de comando ejecución ETL línea Bach

Figura 25 Inicio ejecución ETL Cargue de archivos a tablas físicas



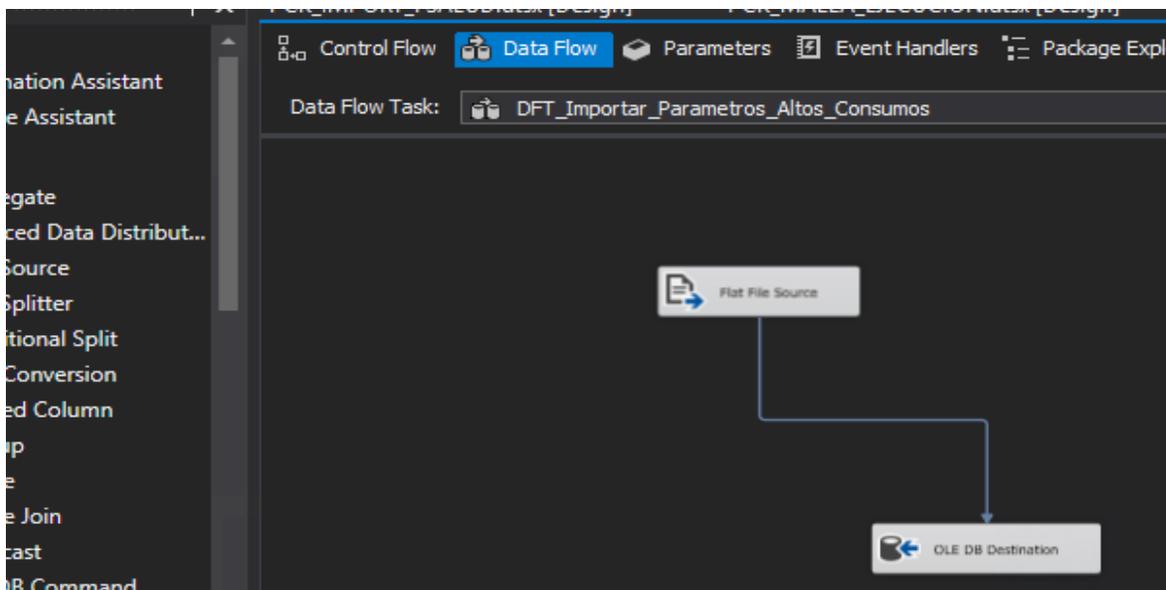
Fuente: Elaboración propia ejecución ETL línea Bach

Figura 26 Inicio ejecución ETL validación creación tablas



Fuente: Elaboración propia ejecución ETL comando creación tablas línea Bach

Figura 27 Inicio ejecución ETL cargue a tablas físicas



Fuente: Elaboración propia ejecución ETL Inserción de datos en tabla Landing Zone

Se procede a realizar un nuevo cargue, este proceso se ejecuta desde líneas de comandos en la ETL donde se migra la información de Landing Zone a una nueva base de datos denominada [Sit Minería] que llevará un esquema según el proyecto y en esta zona si se realizar procesos de minería que puede implicar la creación de tablas transaccionales que permitan lograr el objetivo del área, ejemplo, cruces entre tablas, actualización de datos, eliminación de campos y creación de tablas resultado como se evidencia en la figura #28, figura #29, figura #30.

Figura 28 ejecución proceso ETL cargue sit mineria

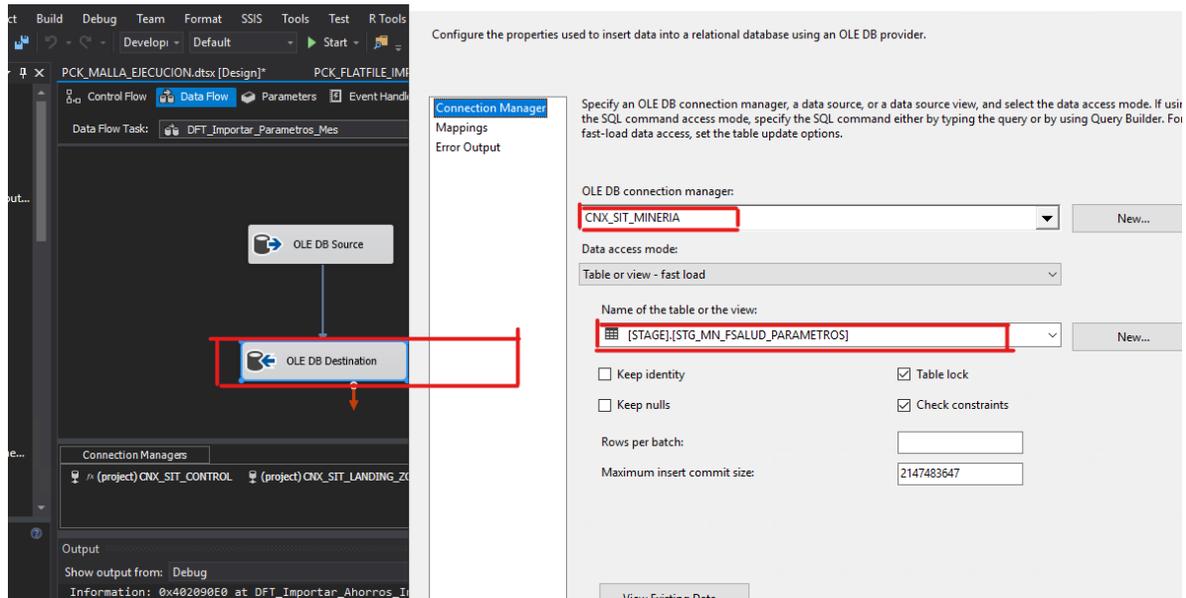


Fuente: Elaboración propia Ejecución ETL Sit Minería

Figura 29 ejecución proceso ETL cargue de fuentes

Fuente: Elaboración propia Ejecución ETL Sit Minería

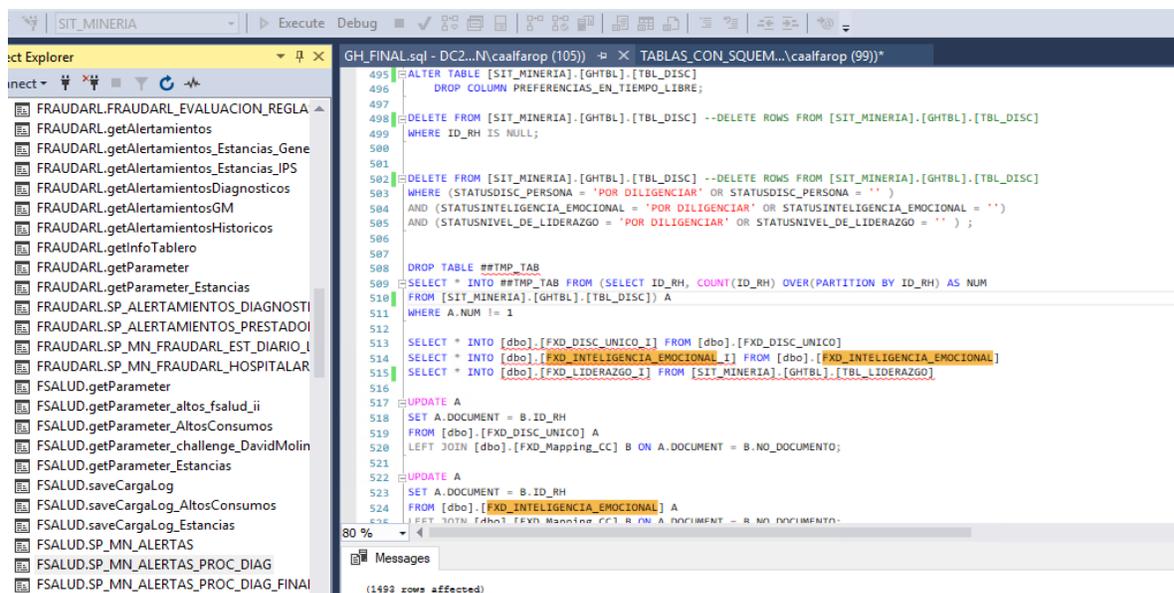
Figura 30 ejecución proceso ETL cargue a tabla física



Fuente: Elaboración propia Ejecución ETL Sit Minería

En el proceso de organización de los datos que se realiza en la base de datos Sit Minería, se ejecutan cruces de tablas, tales como, [FSALUD].[TBL\_MN\_FSALUD\_VIGENTES], [STAGE].[STG\_MN\_FSALUD\_GASTO\_MEDICO] y se aplican reglas de negocio con el objetivo de lograr identificar los casos atípicos que podrían ser casos de fraude como se evidencia en figura #31, figura #32.

Figura 31 ejecución SP SQL cargue tablas transaccionales



Fuente: Elaboración propia Ejecución Query Sit Minería

Figura 32 ejecución SP SQL

```

ALTER procedure [FSALUD].[SP_MN_ALERTAS_PROC_DIAG]

--CREATE PROCEDURE [FSALUD].[SP_MN_ALERTAS_PROC_DIAG]

AS
DECLARE @var_plan_salud varchar(50)
, @var_cod_procedimiento varchar(50)
, @var_nombre_procedimiento varchar(200)
, @var_diagnostico varchar(200)
, @var_dias varchar(2)
, @var_meses varchar(2)
, @var_anios varchar(2)
, @var_edad_ini varchar(3)
, @var_edad_fin varchar(3)
, @var_genero varchar(10)
, @var_regional varchar(30)
, @var_cur_reglas NVARCHAR(max)
, @var_select NVARCHAR(max)
, @var_select_neg NVARCHAR(max)
, @var_date_proceso datetime
, @var_filtro_proc_diag NVARCHAR(max)

SELECT RTRIM(LTRIM(data)) as PLAN_SALUD, CODIGO_PROCEDIMIENTO, PROCEDIMIENTO, COD_DIAGNOSTICO, DIAS2, MESES2, ANIOS2, EDAD_INICIAL, EDAD_FINAL, GENERO, REGIONAL
SET @var_date_proceso = GETDATE()

SET @var_cur_reglas = 'DECLARE CUR_Reglas CURSOR FOR SELECT RTRIM(LTRIM(data)) as PLAN_SALUD, CODIGO_PROCEDIMIENTO, PROCEDIMIENTO, COD_DIAGNOSTICO, DIAS2, MI
EXECUTE sp_executesql @var_cur_reglas

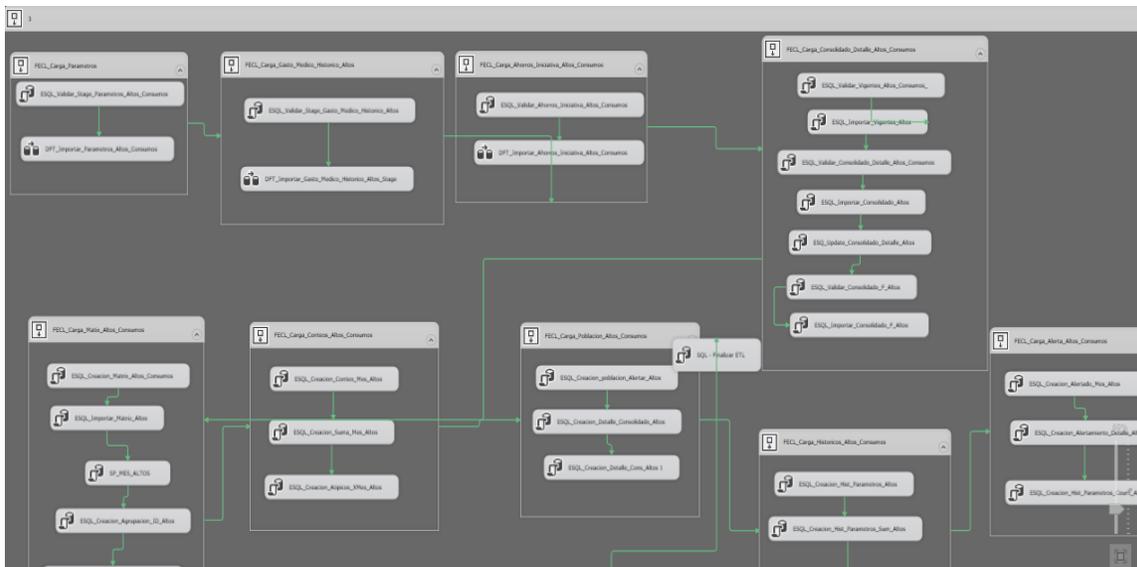
IF NOT EXISTS (SELECT * FROM SYSOBJECTS WHERE TYPE = 'U' AND NAME = 'TBL_MN_FSALUD_ALERTA_GASTO_MEDICO')
BEGIN

```

Fuente: Elaboración propia Ejecución SP Sit Minería

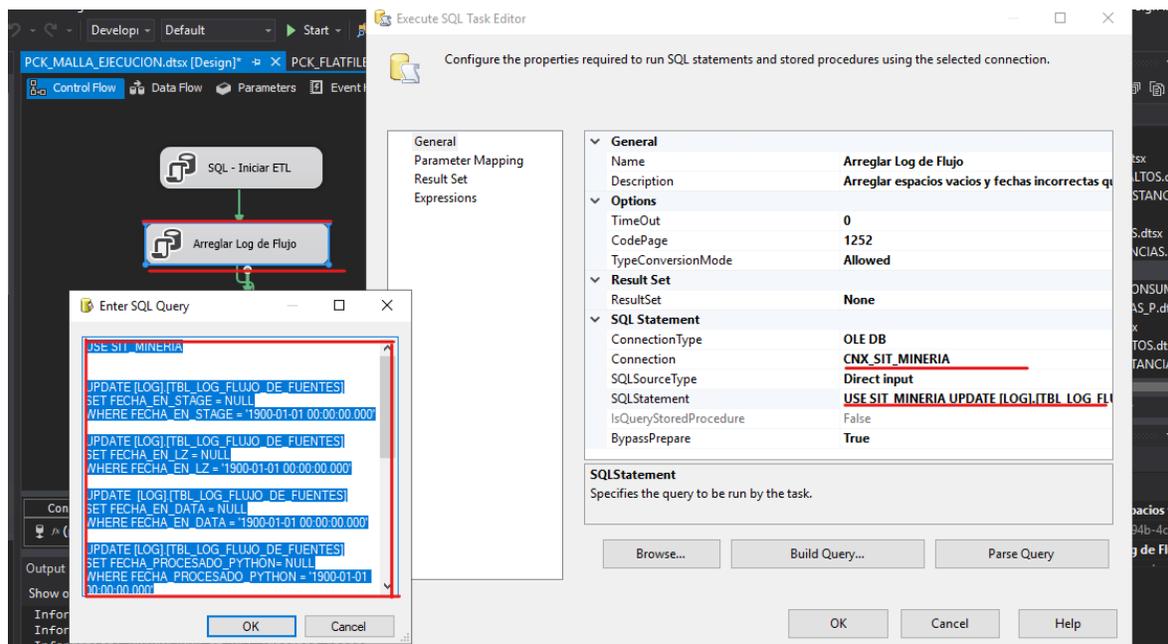
En el desarrollo del proceso de organización se crean tablas transaccionales que son incluidas en la ETL con dos propósitos, uno es cargar datos en estas tablas con el resultado de los cruces y el segundo es enviar un log en el cual se identifique en que paso de la ejecución se encuentra el proceso cómo se evidencia en la figura #33, figura #34, figura #35.

Figura 33 ejecución proceso ETL (Sit Minería) Estancias



Fuente: Elaboración propia Ejecución ETL (Sit Minería)

Figura 34 ejecución proceso ETL (Sit Minería) Estancias



Fuente: Elaboración propia Ejecución ETL Log (Sit Minería)

Figura 35 ejecución disparador ETL (Sit Minería)

```

AS
DECLARE @var_plan_salud varchar(50)
,@var_cod_procedimiento varchar(50)
,@var_nombre_procedimiento varchar(200)
,@var_diagnostico varchar(200)
,@var_dias varchar(2)
,@var_meses varchar(2)
,@var_anios varchar(2)
,@var_edad_ini varchar(3)
,@var_edad_fin varchar(3)
,@var_genero varchar(10)
,@var_regional varchar(30)
,@var_cur_reglas nvarchar(max)
,@var_select nvarchar(max)
,@var_select_neg nvarchar(max)
,@var_date_proceso datetime
,@var_filtro_proc_diag nvarchar(max)

SELECT RTRIM(LTRIM(data)) as PLAN_SALUD, CODIGO_PROCEDIMIENTO, PROCEDIMIENTO, COD_DIAGNOSTICO, DIAS2, MESES2, ANIOS2, EDAD_INICIAL, EDAD_FINAL, GENERO, REGIONAL_DE_AFILIACION from STAGE.STG_MN_FSALUD_PARAMETROS cross apply dbo.sp1itString2(
SET @var_date_proceso = GETDATE())

SET @var_cur_reglas = 'DECLARE CUR_Reglas CURSOR FOR SELECT RTRIM(LTRIM(data)) as PLAN_SALUD, CODIGO_PROCEDIMIENTO, PROCEDIMIENTO, COD_DIAGNOSTICO, DIAS2, MESES2, ANIOS2, EDAD_INICIAL, EDAD_FINAL, GENERO, REGIONAL_DE_AFILIACION from STAG
EXECUTE sp_ejecutesql @var_cur_reglas

IF NOT EXISTS (SELECT * FROM SYSOBJECTS WHERE TYPE = 'U' AND NAME = 'TBL_MN_FSALUD_ALERTA_GASTO_MEDICO')
BEGIN
CREATE TABLE [FSALUD].[TBL_MN_FSALUD_ALERTA_GASTO_MEDICO]
(
[PLAN_x002F_TIPO_SINIESTRO][nvarchar](255) NULL,

```

Fuente: Elaboración propia Ejecución ETL (Sit Minería)

Una vez el proceso de negocio se cumple y genera las tablas resultado con el esquema definido para el área se procede a exportar los resultados a un repositorio asignado u output, donde el usuario final tendrá acceso a las tablas resultado, las cuales se encuentran en archivos xlsx, esto con el fin de enviar información de alertamiento a los auditores que desarrollaran tareas de revisión y glosas según los hallazgos encontrados en la información esta tarea se realiza como se evidencia en la figura #36, figura #37, figura #38.

Figura 36 ejecución SP (Sit Minería)

```

AS
DECLARE @var_plan_salud varchar(50)
,@var_cod_procedimiento varchar(50)
,@var_nombre_procedimiento varchar(200)
,@var_diagnostico varchar(200)
,@var_dias varchar(2)
,@var_meses varchar(2)
,@var_anios varchar(2)
,@var_edad_ini varchar(3)
,@var_edad_fin varchar(3)
,@var_genero varchar(10)
,@var_regional varchar(30)
,@var_cur_reglas NVARCHAR(max)
,@var_select NVARCHAR(max)
,@var_select_neg NVARCHAR(max)
,@var_date_proceso datetime
,@var_filtro_proc_diag NVARCHAR(max)

SELECT RTRIM(LTRIM(data)) as PLAN_SALUD, CODIGO_PROCEDIMIENTO, PROCEDIMIENTO, COD_DIAGNOSTICO, DIAS2, MESES2, ANIOS2, EDAD_INICIAL, EDAD_FINAL, GENERO, REGIONAL_DE_AFILIACION from STAGE.STG_MN_FSALUD_PARAMETROS cross apply dbo.splitString(
SET @var_date_proceso = GETDATE()

SET @var_cur_reglas = 'DECLARE CUR_Reglas CURSOR FOR SELECT RTRIM(LTRIM(data)) as PLAN_SALUD, CODIGO_PROCEDIMIENTO, PROCEDIMIENTO, COD_DIAGNOSTICO, DIAS2, MESES2, ANIOS2, EDAD_INICIAL, EDAD_FINAL, GENERO, REGIONAL_DE_AFILIACION from STAG
EXECUTE sp_mecutesql @var_cur_reglas

IF NOT EXISTS (SELECT * FROM SYSOBJECTS WHERE TYPE = 'U' AND NAME = 'TBL_MN_FSALUD_ALERTA_GASTO_MEDICO')
BEGIN
CREATE TABLE [FSALUD].[TBL_MN_FSALUD_ALERTA_GASTO_MEDICO]
(
[PLAN_x002F_TIPO_SINIESTRO][nvarchar](255) NULL,

```

Fuente: Elaboración propia Ejecución SP (Sit Minería)

Figura 37 Bases de datos



Fuente: Motor de base de datos SQL server

Figura 38 fuentes

ro	Nombre	Fecha de modificación	Tipo
d	RESULTADO	4/05/2023 4:05 p. m.	Archivo compri...
	TOTAL_DETALLE_ALERTA_ALTOS_CONSUMOS	2/05/2023 2:03 p. m.	Hoja de cálculo d...
	ALERTADOS_ALTOS_CONSUMOS_MES	2/05/2023 2:03 p. m.	Hoja de cálculo d...
	HISTORICO_ALERTADOS_ALTOS_CONSUMOS	2/05/2023 2:03 p. m.	Hoja de cálculo d...
XA Colpatría Seguros	HISTORICO_ANALIZADOS_ALTOS_CONSUMOS	2/05/2023 2:02 p. m.	Hoja de cálculo d...
	HISTORICO_PARAMETROS_ALTOS_CONSUMOS_SUM	2/05/2023 2:00 p. m.	Hoja de cálculo d...

Fuente: ventana de explorador de Windows

## 5.4 Visualización de los datos

En esta etapa se establece los siguientes procesos de desarrollo para crear bajo Power bi la visualización de los datos procesados.

- Establecer la herramienta de visualización de información.
- Definición de los datos a utilizar y como se requerían visualizar.
- Diseño de los tableros.

### 5.4.1 Herramienta de Visualización

Se determinó el uso de la herramienta Power bi para la visualización de datos por las razones a continuación descritas.

- Power BI permite conectar y consolidar datos de diversas fuentes, como bases de datos, hojas de cálculo, servicios en la nube, entre otros. Esta capacidad de integración facilita la creación de un tablero de visualización completo.
- Ofrece una amplia gama de visualizaciones interactivas y dinámicas, como gráficos, tablas, mapas y diagramas. Estas visualizaciones permiten a los usuarios explorar y analizar los datos de manera efectiva.
- Se integra con otras herramientas de Microsoft, como Excel, Azure, SQL Server, SharePoint, lo que permite aprovechar las funcionalidades

### 5.4.2 Definición de los Datos

Se definió el uso de los siguientes datos como resultado de los procesos lógicos y que por su construcción cuentan con información importante para análisis de la operatividad diaria.

- Alertamiento\_Dia\_Gasto\_Medico
- Alertamiento\_Dia\_Negaciones
- Alertamiento\_Historico\_Gasto\_Medico
- Alertamiento\_Historico\_Negaciones
- TOTAL\_DETALLE\_ALERTA\_ALTOS\_CONSUMOS
- ALERTADOS\_ALTOS\_CONSUMOS\_MES
- HISTORICO\_ALERTADOS\_ALTOS\_CONSUMOS
- HISTORICO\_ANALIZADOS\_ALTOS\_CONSUMOS
- HISTORICO\_PARAMETROS\_ALTOS\_CONSUMOS\_SUM
- HISTORICO\_PARAMETROS\_ALTOS\_CONSUMOS\_COUNT
- HISTORICO\_PARAMETROS\_ALTOS\_CONSUMOS
- SAL\_GASTO\_MEDICO\_HISTORICO\_ALERTADOS\_ESTANCIAS\_ALTOS
- SAL\_GASTO\_MEDICO\_HISTORICO\_ANALIZADOS\_ESTANCIAS\_ALTOS
- SAL\_GASTO\_MEDICO\_HISTORICO\_PARAMETROS\_ESTANCIAS\_ALTOS

### 5.4.3 Diseño Tableros

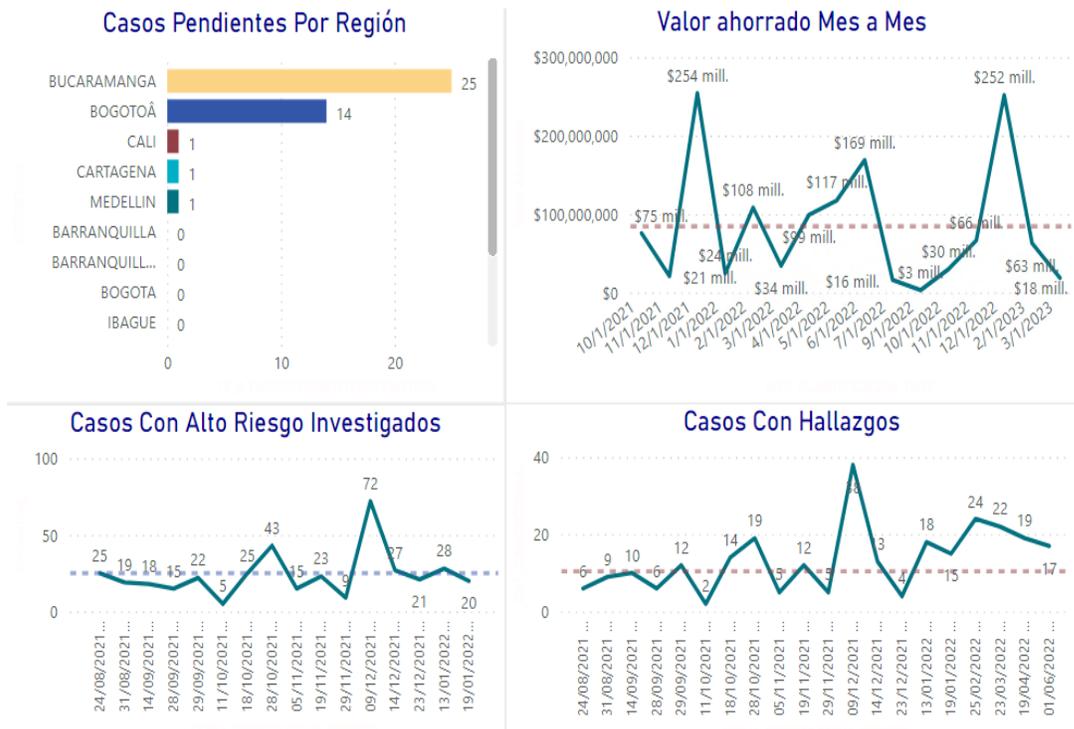
La creación de tableros se realiza con la herramienta power bi ya que nos ofrece una amplia gama de opciones de visualización de datos, desde gráficos simples hasta tablas dinámicas, mapas y gráficos interactivos, ofrece capacidades de conexión y consolidación de datos desde una variedad de fuentes, como bases de datos, hojas de cálculo, servicios en la nube, entre otros. permite integrar datos de diferentes fuentes en un solo a continuación se evidencia la creación de tableros con base a las fuentes resultado del proceso de análisis de información lugar figura #39

Figura 39 Bases de datos Omisión

Fuente: Elaboración propia Bases de Datos Omisión

Imagen tablero Omisión de Preexistencia utilizando las fuentes Históricas resultado de ejecución cómo se evidencia en la figura #40, figura #41 y figura #42. nos muestran al detalle los resultados obtenidos del análisis y aplicación de algoritmos que permitió identificar los casos de fraude y entregarlos en archivos xlsx

Figura 40 Casos pendientes por región y valor ahorrado



Fuente: Herramienta power bi Gráficos propios

Cada tablero nos muestra un resumen completo de la información manejada y el resultado obtenido del análisis para obtener los casos fraudulentos

Figura 41 Tipos de casos Gestionados

Casos Analizados - Alto Riesgo			Casos Alto Riesgo - Sin Historia			Casos Positivos Investigados - Desc. Médica		
FECHA_PROCESO	ANALIZADOS	ALTO RIESGO MODELO	FECHA_PROCESO	ALTO RIESGO INVESTIGADOS	SIN HISTORIA	FECHA_PROCESO	POSITIVOS INVESTIGADOS	DESC. MEDICA POSITIVOS
12/10/2022	45		24/08/2021	25		24/08/2021	4	0
23/03/2022	40		31/08/2021	19		31/08/2021	5	0
19/04/2022	42		14/09/2021	18		14/09/2021	5	0
09/12/2021	72		28/09/2021	14		28/09/2021	3	0
<b>Total</b>	<b>833</b>	<b>1</b>	<b>Total</b>	<b>833</b>	<b>23</b>	<b>Total</b>	<b>140</b>	<b>29</b>

Casos Alto Riesgo Modelo E Investigados			Casos Alto Riesgo - Negativos			Casos Indirectos Investigados Y Desc. Médica		
FECHA_PROCESO	ALTO RIESGO MODELO	ALTO RIESGO INVE	FECHA_PROCESO	ALTO RIESGO INVESTIGADOS	CASOS NEGATIVOS	FECHA_PROCESO	INDIRECTOS INVESTIGADOS	DESC. MEDICA INDIR
12/10/2022	397		13/04/2023	14	6	23/03/2022	13	
23/03/2022	235		23/03/2023	20	9	28/11/2022	11	
19/04/2022	204		06/03/2023	8	2	<b>Total</b>	<b>207</b>	
<b>Total</b>	<b>10485</b>		<b>Total</b>	<b>833</b>				

Casos Analizados - No Investigados			Casos Estado - Alto Riesgo Investigados			Casos Alto Riesgo Investigados - Ahorro Indirecto		
FECHA_PROCESO	ANALIZADOS	NO INVESTIGADOS	FECHA_PROCESO	ESTADO	ALTO RIESGO INVESTIGADOS	FECHA_PROCESO	ALTO RIESGO INVESTIGADOS	AHORROS INDIREC
12/10/2022	45	352	28/11/2022	Pendiente Finalizar	97	24/08/2021	25	
23/03/2022	40	195	09/12/2021	Pendiente Finalizar	72	31/08/2021	19	
19/04/2022	42	162	01/07/2022	Pendiente Finalizar	52	<b>Total</b>	<b>833</b>	
23/12/2021	21	105	12/10/2022	Pendiente Finalizar	45			
<b>Total</b>	<b>833</b>	<b>9652</b>	<b>Total</b>	<b>Pendiente Finalizar</b>	<b>833</b>			

Casos Investigados Alto Riesgo, Indirectos y Positivos				Casos Indirectos Investigados y Ahorros Indirectos			
FECHA_PROCESO	ALTO RIESGO INVESTIGADOS	INDIRECTOS INVESTIGADOS	CASOS POSITIVOS INVESTIGADOS	FECHA_PROCESO	INDIRECTOS INVESTIGADOS	AHORROS INDIRECTOS	
24/08/2021	25	2	4	09/12/2021	27	12	
31/08/2021	19	4	5	01/07/2022	13	7	
14/09/2021	18	5	5	13/01/2022	11	6	
28/09/2021	14	3	3	19/04/2022	13	6	
<b>Total</b>	<b>72</b>	<b>14</b>	<b>17</b>	<b>Total</b>	<b>70</b>	<b>31</b>	

Fuente: Herramienta power bi Gráficos propios

Figura 42 Total Ahorros



Fuente: Herramienta power bi Gráficos propios

Creación de páginas: Una vez que se han creado las visualizaciones, se organizan en páginas que permiten que los usuarios naveguen fácilmente a través del tablero. Esto puede incluir la creación de páginas de resumen, páginas de detalles y páginas de comparación cómo se evidencia en la figura #43.

Figura 43 Casos por región

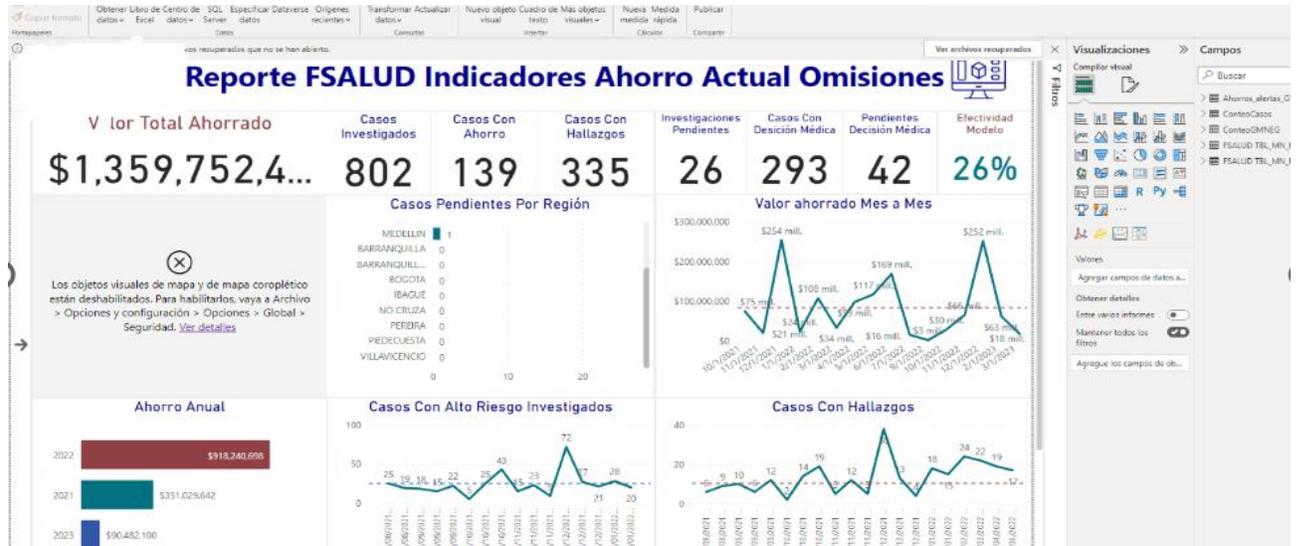


Fuente: Herramienta power bi Gráficos propios

Creación de filtros y selecciones: Para ayudar a los usuarios a interactuar con el tablero, se pueden agregar filtros que les permitan explorar los datos en más detalle. Esto puede incluir filtros de fecha, filtros de ubicación y filtros de

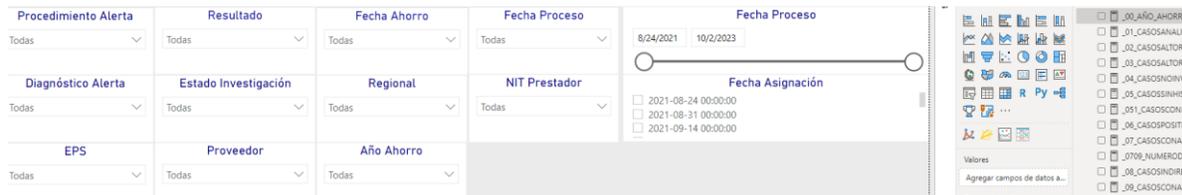
categoría cómo se evidencia en la figura #44 y figura #45, anexo 17,18,19,20,21,22,23,24,25,26,27

Figura 44 Indicadores de ahorro



Fuente: Elaboración propia Filtros

Figura 45 Paginas



Fuente: Herramienta power bi Gráficos propios

Creación de paneles: Una vez que se han creado las páginas y las visualizaciones, se organizan en paneles que se pueden compartir con otros usuarios. Los paneles pueden incluir una o varias páginas y pueden ser compartidos en línea o a través de la aplicación móvil de Power BI.

## 6. Conclusiones

En base a los objetivos, se puede concluir que la implementación de la automatización de procesos basados en datos ha sido efectiva en la detección temprana de casos fraudulentos. Al crear un proceso de alertamiento automático, se ha logrado reducir el número de casos enviados a investigar, lo que ha generado un ahorro de 1.200.000 millones de pesos en nuevos casos detectados efectivos.

casos investigados en un 90% por cada uno de los procesos, además, la creación de un Dashboard para el seguimiento de los principales KPIs acordados con la línea de negocio ha permitido tener una visión clara de la evolución del negocio en tiempo real. Esto ha permitido a la empresa tomar decisiones y ajustar su estrategia para mejorar su desempeño.

En general, la automatización de procesos basados en datos ha permitido mejorar la eficiencia en la detección de casos fraudulentos, reducir los costos y mejorar el desempeño de la empresa en general. Esto demuestra la importancia de la automatización de procesos basados en datos y su impacto positivo en las empresas que buscan mejorar su rendimiento y competitividad.

## 7. Referencias

Silberschatz, A., Galvin, P. B., & Gagne, G. (2012). *Operating System Concepts* (9th ed.). Wiley.

Aula21. (08 de 2018). Aula21. Obtenido de Aula21: <https://www.cursosaula21.com/quees-python/>

Wang, R. Y., & Strong, D. M. (2012). *Database Systems: Concepts, Design, and Applications*. McGraw-Hill Education.

DAMA International. (2017). Home. Recuperado el 24 de mayo de 2023, de <https://www.dama.org/>

Python Software Foundation. (s.f.). *Python Programming Language*. Recuperado el 24 de mayo de 2023, de <https://www.python.org/>

Microsoft. (s.f.). *Integration Services (SSIS)*. Recuperado el 24 de mayo de 2023, de <https://docs.microsoft.com/en-us/sql/integration-services/>

Microsoft. (s.f.). *Power BI*. Recuperado el 24 de mayo de 2023, de <https://powerbi.microsoft.com/>

García, M. A. (2020). *Fraude en la prestación de servicios de salud: Estrategias de detección y prevención*. Ciudad de México, México: Editorial Salud.

Ango, J. (2020). *Análisis de datos para la toma de decisiones con el uso de tableros de control aplicado a los registros de ventas digitales de un E-Commerce* (Tesis de pregrado). Universidad de las Américas, Quito.

ISSA.(2022).Detectar el fraude en la atención de salud mediante las tecnologías emergentes. Recuperado de 4 de julio 2022 <https://ww1.issa.int/es/analysis/detecting-fraud-health-care-through-emerging-technologies>

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Muller, M. (2019). *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*. Technics Publications.

Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (2nd ed.). Analytics Press.

## Anexos

### 1. Anexos

A continuación, se relacionan tablas, diccionario de datos y flujogramas del proyecto.

#### 8.1 Anexo Tablas fuentes Altos Consumos

##### Anexo 1. Gasto Medico

NUMERO	UNIDAD_NEGOCIO	FOLIO	SUC_PAGO	SUC	DIVIPOLA	CODIGO_H	PLA	TIPO	NRO_C	TIP	NRO_IDENTI	VALOR	VALOR	VALOR	VALOR	COD	COD_PROG	NOMBRE_PROG
22898355	AXA Colpatría-Medi	81290868		610	11001	1100105618	15		152093	CC	80470090	118337	118337	0	0	R688	871121	RADIOGRAFÍA
22901954	AXA Colpatría-Medi	81293133		610	11001	1100105618	15		152081	CC	19376004	0	0	0	0	Z017	954100	AUDIOMETRIA
22901954	AXA Colpatría-Medi	81293133		610	11001	1100105618	15		152081	CC	19376004	0	0	0	0	Z017	954302	INMITANCIA A
22901957	AXA Colpatría-Medi	81291829		630	76001	7600101111	11		112796	CC	16729355	0	0	0	0	N40	902210	1. HEMOGRAM
22901957	AXA Colpatría-Medi	81291829		630	76001	7600101111	11		112796	CC	16729355	0	0	0	0	N40	904108	12. PROLACTI
22901963	AXA Colpatría-Medi	81293518		610	11001	1100105618	11		111011	RC	1011259999	0	0	0	0	K529	110900075	LACTATO RING
22901963	AXA Colpatría-Medi	81293518		610	11001	1100105618	11		111011	RC	1011259999	0	0	0	0	K529	902210	HEMOGRAMA
22901963	AXA Colpatría-Medi	81293518		610	11001	1100105618	11		111011	RC	1011259999	0	0	0	0	K529	903883	GLUCOSA SEM
22901969	AXA Colpatría-Medi	81293517		610	11001	1100105618	11		112869	RC	1014904481	0	0	0	0	R501	123101026	GUANTE ESTER
22901969	AXA Colpatría-Medi	81293517		610	11001	1100105618	11		112869	RC	1014904481	0	0	0	0	R501	906301	ADENOVIRUS
22901969	AXA Colpatría-Medi	81293517		610	11001	1100105618	11		112869	RC	1014904481	0	0	0	0	R501	907106	UROANALISIS
22901972	AXA Colpatría-Medi	81291880		630	76001	7600101111	11		112779	CC	1116234127	0	0	0	0	Z017	898201	1. ESTUDIO DE
22901973	AXA Colpatría-Medi	81293046		610	11001	1100105618	15		152063	CC	1018510843	0	0	0	0	R688	903016	FERRITINA
22901957	AXA Colpatría-Medi	81291829		630	76001	7600101111	11		112796	CC	16729355	0	0	0	0	N40	904921	6. TIROXINA LI
22901963	AXA Colpatría-Medi	81293518		610	11001	1100105618	11		111011	RC	1011259999	0	0	0	0	K529	110900031	DEXTOSA+SO
22901963	AXA Colpatría-Medi	81293518		610	11001	1100105618	11		111011	RC	1011259999	0	0	0	0	K529	120101005	CATETER INTR

##### Anexo 2. Ahorros Altos

CHA_PROCE	ID	REGIONAL	ROVEEDO	EPS	FECHA_AS	FECHA_ENTR	DIAS_HABILE	ESTADO_INV	ESTADO_P_Y	VALOR_AHOR	COSTO	MES_CUA
#####	#####	CARTAGENA	ALIANZA	NA DE SERV	#####	26/12/2022	19	Alertado	Positivo	#####		#####
#####	#####	CALI	INVIPRO	NUEVA EPS	#####	20/12/2022	15	Alertado	Positivo	#####		#####
#####	#####	MEDELLIN	CONFYGES	RIODOS CON	#####	15/01/2023	33	Alertado	Positivo			
#####	#####	CARTAGENA	CONFYGES	RIODOS CON	#####	15/01/2023	33	Alertado	Positivo			
#####	#####	BOGOTÁ	ALIANZA	D TOTAL S.A	#####	12/12/2022	9	Alertado	Negativo			
#####	#####	BOGOTÁ	ALIANZA	NA DE SERV	#####	24/12/2022	18	Alertado	Negativo			
#####	#####	CALI	CONFYGES	RIODOS CON	#####	15/12/2022	12	Alertado	Negativo			
#####	#####	BOGOTÁ	ALIANZA	D TOTAL S.A	#####	15/01/2023	33	Alertado	Positivo			
#####	#####	CARTAGENA	CONFYGES	RIODOS CON	#####	23/12/2022	18	Alertado	Dx indirectos			
#####	#####	BOGOTÁ	ALIANZA	NA DE SERV	#####	24/12/2022	18	Alertado	Positivo			
#####	#####	BOGOTÁ	CONFYGES	NSACION FA	#####	23/12/2022	18	Alertado	Positivo			
#####	#####	BOGOTÁ	CONFYGES	P.S. SANITA	#####	23/12/2022	18	Alertado	Dx indirectos			
#####	#####	MEDELLIN	ALIANZA	NA DE SERV	#####	12/12/2022	9	Sin Historia	Sin Historia			
#####	#####	BOGOTÁ	CONFYGES	P.S. SANITA	#####	12/12/2022	9	Alertado	Negativo			
#####	#####	CALI	ALIANZA	D TOTAL S.A	#####	12/12/2022	9	Sin Historia	Sin Historia			
#####	#####	BOGOTÁ	CONFYGES	RIODOS CON	#####	12/12/2022	9	Sin Historia	Sin Historia			

### Anexo 3. parámetros

CODIGO_DIA	Dx_Nombre	AÑO_SERVICIO	AÑO_AI	MES_SERVICIO	MES_AFILIACION	NIT_IPS	IPS	ESTADO_FACTURA	TIPO_CONCEPTO	TIPO_ATENCION
TODOS	TODOS	TODOS	2022	TODOS	TODOS	TODOS	TODOS	Radicada para Pago	Todos	Todos
TODOS	TODOS	TODOS	2023	TODOS	TODOS	TODOS	TODOS	Radicada para Pago	Todos	Todos

### 8.2 Tablas fuentes Omisión

#### Anexo 4. Ahorros iniciativa

ESTADO_INVESTADO	P_Y	VALOR_A COSTO MES_CUANTIA	ID	NOMBRE	CODIGO	DX_ALERTA	DECISION_DICION	COD_PRC	NOMBRE_PROCESO
Alertado	Dx indirecto	\$ 0		8E+07	R072	DOLOR PREC	NO SE CONS	881202	ECOCARDIOGRA
Alertado	Positivo	#####	#####	1E+09	R074	DOLOR EN EL FIJAR PREEXI		879301	TOMOGRAFIA C
Sin Historia	Sin Historia	\$ 0		1E+09	R011	SOPLO CARDIACO, NO ESF		881202	ECOCARDIOGRA
Sin Historia	Sin Historia	\$ 0		3E+07	Z014	EXAMEN GINECOLOGICO (		881401	ECOGRAFIA PEL
Sin Historia	Sin Historia	\$ 0		1E+09	K219	ENFERMEDAD DEL REFLUJ		881302	ECOGRAFIA DE J
Alertado	Positivo	#####	#####	1E+09	N939	HEMORRAGI	FIJAR PREEXI	881401	ECOGRAFIA PEL
Alertado	Positivo	#####	#####	7E+07	R073	OTROS DOLC	FIJAR PREEXI	895100	ELECTROCARDIC
Sin Historia	Sin Historia	\$ 0		3E+07	I158	OTROS TIPOS DE HIPERTE		879111	TOMOGRAFIA C
Sin Historia	Sin Historia	\$ 0		2E+07	Z125	EXAMEN DE PESQUISA ESF		881332	ECOGRAFIA DE Y
Alertado	Dx indirecto	\$ 0		1E+09	N913	OLIGOMENO	NO SE CONS	881401	ECOGRAFIA PEL
Sin Historia	Sin Historia	\$ 0		1E+09	M214	PIE PLANO PES PLANUS (A		873305	RADIOGRAFIA P
Alertado	Dx indirecto	\$ 0		7E+07	R688	OTROS SINT	NO SE CONS	881332	ECOGRAFIA DE Y
Sin Historia	Sin Historia	\$ 0		1E+07	R69	CAUSAS DE MORBILIDAD C		881302	ECOGRAFIA DE J
Alertado	Dx indirecto	\$ 0		5E+07	J040	LARINGITIS A	NO SE CONS	306001	NASOLARINGOS
Error al envi	Error al envi	\$ 0		1E+09	S835	ESGUINCES Y TORCEDURA		883522	RESONANCIA M
Sin Historia	Sin Historia	\$ 0		5E+06	J370	LARINGITIS CRONICA		306001	NASOLARINGOS
Error al envi	Error al envi	\$ 0		4E+07	H186	QUERATOCONO		951501	TOPOGRAFIA CC
Sin Historia	Sin Historia	\$ 0		1E+07	M224	CONDROMALACIA DE LA F		883522	RESONANCIA M
Alertado	Dx indirecto	\$ 0		4E+07	M942	CONDROMA	NO SE CONS	879520	TOMOGRAFIA C
Sin Historia	Sin Historia	\$ 0		1E+09	R074	DOLOR EN EL PECHO, NO E		881202	ECOCARDIOGRA

## Anexo 5. Enfermedades

Código	Capitul	Nombre_Cap	Código	Descripción	Código	Descripción_códigos_Cuatro_Caract
A000	1	Ciertas enfer	A00	Colera	A000	Colera debido a Vibrio cholerae 01,
A001	1	Ciertas enfer	A00	Colera	A001	Colera debido a Vibrio cholerae 01,
A009	1	Ciertas enfer	A00	Colera	A009	Colera, no especificado
A010	1	Ciertas enfer	A01	Fiebres Tif	A010	Fiebre tifoidea
A011	1	Ciertas enfer	A01	Fiebres Tif	A011	Fiebre paratifoidea A
A012	1	Ciertas enfer	A01	Fiebres Tif	A012	Fiebre paratifoidea B
A013	1	Ciertas enfer	A01	Fiebres Tif	A013	Fiebre paratifoidea C
A014	1	Ciertas enfer	A01	Fiebres Tif	A014	Fiebre paratifoidea, no especificada
A020	1	Ciertas enfer	A02	Otras Infec	A020	Enteritis debida a Salmonella
A021	1	Ciertas enfer	A02	Otras Infec	A021	Sepsis debida a Salmonella
A022	1	Ciertas enfer	A02	Otras Infec	A022	Infecciones localizadas debidas a Sa
A028	1	Ciertas enfer	A02	Otras Infec	A028	Otras infecciones especificadas con
A029	1	Ciertas enfer	A02	Otras Infec	A029	Infeccion debida a Salmonella, no e
A030	1	Ciertas enfer	A03	Shigelosis	A030	Shigelosis debida a Shigella dysent
A031	1	Ciertas enfer	A03	Shigelosis	A031	Shigelosis debida a Shigella flexner
A032	1	Ciertas enfer	A03	Shigelosis	A032	Shigelosis debida a Shigella boydii
A033	1	Ciertas enfer	A03	Shigelosis	A033	Shigelosis debida a Shigella sonnei
A038	1	Ciertas enfer	A03	Shigelosis	A038	Otras shigelosis
A039	1	Ciertas enfer	A03	Shigelosis	A039	Shigelosis de tipo no especificado
A040	1	Ciertas enfer	A04	Otras Infec	A040	Infeccion debida a Escherichia coli e
A041	1	Ciertas enfer	A04	Otras Infec	A041	Infeccion debida a Escherichia coli e
A042	1	Ciertas enfer	A04	Otras Infec	A042	Infeccion debida a Escherichia coli e

## Anexo 6. Históricos

NUMERO	FECHA	AREA_QUE	ID	NUMERO	SUCURS	CLAVE	NOMBRE	DIAGNOS	CÓDIGO	ESPECIAL	DECISION	FECHA	MODULO	BH	FECHA	OBSERV
1	44615	FRAUDE	CC	1,2E+09				RINITIS A	J303J34	ORTOPEC	FIJAR PRE	44663			43866	
2	44585	AUTORIZA	CC	9,9E+07				HIPERTEN	I10X	PSIQUIAT	FIJAR PRE	44677				
3	44610	FRAUDE	CC	7,1E+07				HIPERLIP	E782I87	NEUROL	FIJAR PRE	44685			43851	CONTR.
4	44642	AUTORIZA	CC	7,3E+07				HIPERPLA	N40XK7	PSIQUIAT	CANCELAI	44685			43854	
5	44669	AUTORIZA	TI	1,1E+09				PROTEINI	N069L93	NEUROL	CANCELAI	44726			43912	
6	44706	AUTORIZA	CC	4,3E+07				TIROTOXI	E058H62	NEUROL	FIJAR PRE	44734				
7	44679	FRAUDE	CC	4,4E+07				HIPOTIRC	E039E04	ORTOPEC	FIJAR PRE	44749			43973	
8	44701	AUTORIZA	CC	1E+09				SECUNDA	A514B24	CIRUGIA	FIJAR PRE	44853				ACTUAL
9	44706	FRAUDE	CC	1E+09				ENFERME	N602	NEUROL	FIJAR PRE	44853			43961	
10	44727	AUTORIZA	CC	3,2E+07				INCONTIN	N394	CARDIOL	FIJAR PRE	44873			43969	
11	44740	AUTORIZA	CC	3,2E+07				HIPERTEN	I10X0241E782M51	CANCELAI	44873					
12	44784	AUTORIZA	CC	2,6E+07				PSORIASI	L408		FIJAR PRE	44873				
13	44785	AUTORIZA	CC	8E+07				EPIDIDIM	N450N512		NO SE CO	44873				
14	44799	AUTORIZA	CC	7,2E+07				HIPERTEN	I10XI120		FIJAR PRE	44873				
15	44824	FRAUDE	CC	1,6E+07				Covid-19	U071R102	N40X	NO SE CO	44873				
16	44859	AUTORIZA	CC	5,3E+07				FRACTUR	S824 M725E669I7E		NO SE CO	44887				
17	44776	FRAUDE	TI	1E+09				INFECCIO	N390R55XI058		CANCELAI	44888				
18	44776	AUTORIZA	CC	3E+07				HEMANG	D180		FIJAR PRE	44888				
19	44826	AUTORIZA	CC	7,1E+07				SINDROM	M751M511G560		CANCELAI	44888				
20	44826	AUDITORIA	CC	4,2E+07				HIPERTEN	I10XD695M350N1		FIJAR PRE	44888				
21	44826	AUTORIZA	CC	4,4E+07				HIPERTEN	I10XD695M350N1		FIJAR PRE	44888				

## 8.3 Repositorio de Fuentes Omisión Altos Estancias

### Anexo 7 Omisión

» FRAUDE\_SALUD » Omision\_Preexistencias » Input » Usuarios

Nombre	Fecha de modificación	Tipo
📁 Dashboards	10/03/2023 9:14 p. m.	Carpeta de archivos
📁 HISTORICO	17/05/2023 9:17 a. m.	Carpeta de archivos
📄 STR_SAL_AHORROS_INICIATIVA	19/05/2023 2:41 p. m.	Hoja de cálculo d...
📄 STR_SAL_ENFERMEDADES	22/02/2023 8:12 a. m.	Hoja de cálculo d...
📄 STR_SAL_GASTO_MEDICO	17/05/2023 7:45 a. m.	Hoja de cálculo d...
📄 STR_SAL_HISTORICO_OMISION	4/05/2023 8:50 p. m.	Hoja de cálculo d...
📄 STR_SAL_HISTORICOS_INVESTIGADOS	4/05/2023 8:50 p. m.	Hoja de cálculo d...
📄 STR_SAL_HYC_USUARIOS_VIGENTES_2021	17/05/2023 9:10 a. m.	Hoja de cálculo d...
📄 STR_SAL_MPP_USUARIOS_VIGENTES_2021	17/05/2023 9:16 a. m.	Hoja de cálculo d...
📄 STR_SAL_NEGACIONES	17/05/2023 7:43 a. m.	Hoja de cálculo d...
📄 STR_SAL_PARAMETROS	4/05/2023 8:50 p. m.	Hoja de cálculo d...
📄 STR_SAL_PROCEDIMIENTOS	4/05/2023 8:50 p. m.	Hoja de cálculo d...

### Anexo 8 Estancias

» FRAUDE\_SALUD » Estancias\_Prolongadas » Input » Usuarios »

Nombre	Fecha de modificación	Tipo
📁 Dashboard	4/04/2023 2:15 p. m.	Carpet
📁 Otro	20/02/2023 10:35 a. m.	Carpet
📄 STR_SAL_AHORROS_ESTANCIAS	18/05/2023 7:49 a. m.	Hoja d
📄 STR_SAL_PARAMETOS_ESTANCIA	4/04/2023 12:16 p. m.	Hoja d
📄 STR_SAL_PARAMETOS_ESTANCIA_DIARIO	15/11/2022 12:18 p. m.	Hoja d
📄 STRL_SAL_CENSO_HOSPITALARIO_DIARIO	24/05/2023 8:33 a. m.	Hoja d
📄 STRL_SAL_CENSO_HOSPITALARIO_MENSUAL	25/04/2023 8:23 a. m.	Hoja d

## Anexo 9 Altos consumos

 BK10	16/03/2023 11:10 a. m.	Carpe
 Dashboard	10/02/2023 5:09 p. m.	Carpe
 Gasto_Medico_Historico_Altos	11/05/2023 8:27 a. m.	Hoja c
 STR_SAL_AHORROS_INICIATIVA_ALTOS_CONSUMOS	11/05/2023 8:30 a. m.	Hoja c
 STR_SAL_PARAMETOS_ALTOS_CONSUMOS	11/05/2023 8:29 a. m.	Hoja c

## 8.4 Anexo Diccionario de Datos Fuentes

### Anexo 10 Enfermedades

<b>Nombre</b>	SAL_ENFERMEADES							
<b>Periodicidad</b>	SEMANAL							
<b>Origen</b>	\\dc2vafpo1\IntercambioTerceros\SALUD\FRAUDE_SALUD\Omisión_Preexistencias\Input\Usuarios							
N°	Nombre del Campo	Ejemplo	Descripción del Campo	Criticidad	Tipo del Cam	Longit	Puede ir vacío	Cuándo puede ir vacío
1	Código_CIE_10_cuatro_caracteres	A000	Código CIE 10 de cuatro caracteres	MEDIO	nvarchar	510	NO	N/A
2	Capitulo	1	Número del capítulo de CIE 10	MEDIO	nvarchar	510	NO	N/A
3	Nombre_Capitulo	Ciertas enfermedades infecciosas y paras	Nombre del capítulo de CIE 10	MEDIO	nvarchar	510	NO	N/A
4	Código_CIE_10_tres_caracteres	A00	Código CIE 10 de tres caracteres	MEDIO	nvarchar	510	NO	N/A
5	Descripcion_codigos_tres_caracteres	Colera	Nombre de la enfermedad en general seg	MEDIO	nvarchar	510	NO	N/A
6	Código_de_la_CIE_10_Cuatro_Caracteres	A000	Código CIE 10 de cuatro caracteres	MEDIO	nvarchar	510	NO	N/A
7	Descripcion_codigos_Cuatro_Caracteres	Colera debido a Vibrio cholerae 01, biotip	Nombre de la enfermedad en específico se	MEDIO	nvarchar	510	NO	N/A

### Anexo 11 Gasto medico

<b>Nombre</b>	SAL_GASTOMEDICO							
<b>Periodicidad</b>	SEMANAL							
<b>Origen</b>	\\dc2vafpo1\IntercambioTerceros\SALUD\FRAUDE_SALUD\Omisión_Preexistencias\Input\Usuarios							
N°	Nombre del Campo	Ejemplo	Descripción del Campo	Criticidad	Tipo del Cam	Longit	Puede ir vacío	Cuándo puede ir vacío
1	NUMERO_INTERNO	22800218	Número interno del registro	ALTO	nvarchar	500	NO	N/A
2	UNIDAD_NEGOCIO	AXA Colpatría-Medicina prepagada MPP	Descripción de la unidad de negocio de	ALTO	nvarchar	500	NO	N/A
3	FOLIO	81286469	Número de folio del gasto medico	ALTO	nvarchar	500	NO	N/A
4	SUC_PAGO		Nombre de la sucursal de donde provie	ALTO	nvarchar	500	NO	N/A
5	SUC_CONTRATO_COLPATRIA	610	Nombre de la sucursal desde donde se	ALTO	nvarchar	500	NO	N/A
6	DIVIPOLA	11001	Nomenclatura estandarizada para la id	BAJO	nvarchar	500	NO	N/A
7	CODIGO_HABILITACION	110010561801	Código de la habilitación del contrato	BAJO	nvarchar	500	NO	N/A
8	PLAN	15	Nombre del plan de salud al que pert	BAJO	nvarchar	500	NO	N/A
9	TIPO_CONTRATO		Nombre del tipo de contrato de salud a	BAJO	nvarchar	500	SI	N/A
10	NRO_CONTRATO	152011560000	Número del contrato de salud asociado	MEDIO	nvarchar	500	NO	N/A
11	TIPO_IDEN_BENEF	CC	Tipo de documento de identidad del be	BAJO	nvarchar	500	NO	N/A
12	NRO_IDENTIFICACION_BENEF	20170960	Número de documento de identidad de	MEDIO	nvarchar	500	NO	N/A
13	VALOR_BRUTO_FACTURA	0	Valor bruto de la factura pagada por el	MEDIO	nvarchar	500	NO	N/A
14	VALOR_NETO	0	Valor neto (bruto + impuestos) de la fa	MEDIO	nvarchar	500	NO	N/A
15	VALOR_VALES	0	Valor del vale que el cliente debe pagar	ALTO	nvarchar	500	NO	N/A
16	VALOR_DESCUENTOS	0	Valor descuentos aplicados al pago del	ALTO	nvarchar	500	NO	N/A
17	COD_DIAGNOSTICO	H811	Número único usado para identificar el	ALTO	nvarchar	500	NO	N/A
18	COD_PROCEDIMIENTO	S41605	Número único usado para identificar el	ALTO	nvarchar	500	NO	N/A
19	NOMBRE_PROCEDIMIENTO	MANEJO INTRAHOSPITALARIO EN SALA DE OB	Nombre del procedimiento que el clie	ALTO	nvarchar	-1	SI	N/A
20	VALOR_UNITARIO	48632	Valor unitario del procedimiento que e	ALTO	nvarchar	500	SI	N/A
21	CANTIDAD_DETALLE	1	Número de procedimientos que el clie	MEDIO	nvarchar	500	SI	N/A
22	VALOR_DETALLE	48632	Valor de la totalidad de los procedimie	ALTO	nvarchar	500	SI	N/A
23	CUPS_PRINCIPAL		Número que describe la Clasificación U	ALTO	nvarchar	500	SI	N/A

24	CONCEPTO_HOSPITALARIO	Procedimientos NO QX	Grupo de procedimientos al cual hace	ALTO	nvarchar	500	SI	N/A
25	DIAS_HOSP		Número de días que el paciente requiere	ALTO	nvarchar	500	SI	N/A
26	TIPO_CUENTA	Hospitalarios	Tipo de cuenta a la que pertenece el m	BAJO	nvarchar	500	SI	N/A
27	FECHA_SERVICIO	20/02/2023	Fecha de cuando el cliente fue atendid	BAJO	nvarchar	500	SI	N/A
28	TIPO_IDENT_PRESTADOR	NIT	Tipo de documento del prestador que	ALTO	nvarchar	500	NO	N/A
29	NIT_PRESTADOR	860037950	Número de NIT del prestador que aten	ALTO	nvarchar	500	SI	N/A
30	NOMBRE_PRESTADOR	FUNDACION SANTA FE DE BOGOTA	Nombre del prestador que atendió al p	BAJO	nvarchar	500	SI	N/A
31	FECHA_RAD_FACT	27/02/2023	Fecha de radicación de la factura que e	MEDIO	nvarchar	500	SI	N/A
32	NUMERO_FACT	FS7430318	Número asignado a la factura de que e	MEDIO	nvarchar	500	SI	N/A
33	FECHA_FACTURA	21/02/2023	Fecha de la factura que el prestador ra	BAJO	nvarchar	500	SI	N/A
34	ESTADO_FACTURA	Radicada para Pago.	Estado de la factura que el prestador ra	BAJO	nvarchar	500	NO	N/A
35	NUMERO_SINIESTRO		Número asociado al siniestro reportado	ALTO	nvarchar	500	SI	N/A
36	NUMERO_AUTORIZACION		Número asociado a la autorización del	BAJO	nvarchar	200	SI	N/A
37	TIPO_SERVICIO		Tipo de servicio autorizado	MEDIO	nvarchar	500	SI	N/A
38	TIPO_ATENCION		Tipo de atención autorizada	BAJO	nvarchar	500	SI	N/A
39	DIGITA_DETALLE	BobnotRDE	Nombre del usuario que digita el detall	BAJO	nvarchar	500	NO	N/A

## Anexo 12 Histórico investigado

Nº	Nombre del Campo	Ejemplo	Descripción del Campo	Criticidad	Tipo del Campo	Longitud	Puede vacío	Cuándo puede ir a
1	CANTIDAD_ID	1	Conteo de casos	ALTO	nvarchar	510	NO	N/A
2	INPUT	OMS	Clasificación del caso	ALTO	nvarchar	510	NO	N/A
3	ID_USUARIO	80074843	Número del documento de identidad del usuario	ALTO	nvarchar	510	NO	N/A
4	NOMBRE_USUARIO	JORGE MARIO MERCADO DE LA ROSA	Nombre completo del usuario	ALTO	nvarchar	510	NO	N/A
5	ASESOR_COMERCIAL	ADOLFO BOLIVAR MENESES	Nombre del asesor comercial	ALTO	nvarchar	510	NO	N/A
6	REGIONAL	BOGOTÁ	Regional a la cual pertenece el contrato de salud	BAJO	nvarchar	510	NO	N/A
7	PROVEEDOR	NA	Nombre del proveedor	BAJO	nvarchar	510	NO	N/A
8	FECHA_ASIGNACION	NA	Fecha de asignación del caso	BAJO	nvarchar	510	NO	N/A
9	FECHA_ENTREGA	NA	Fecha de entrega del caso	BAJO	nvarchar	510	NO	N/A
10	DIAS_HABILES	NA	Días entre fecha entrega y fecha asignación	MEDIO	nvarchar	510	NO	N/A
11	COSTOS	0	Costos del caso en pesos colombianos	MEDIO	nvarchar	510	SI	En algunos estados puede no presentarse un valor
12	ESTADO	Alertado	Estado del caso	MEDIO	nvarchar	510	NO	N/A
13	POSITIVO	Dx indirecto	Resultado de la investigación	ALTO	nvarchar	510	NO	N/A
14	ALERTA	OMISIÓN DE LITIASIS BILIAR	Notas sobre el paciente y su caso	ALTO	nvarchar	510	NO	N/A
15	CONCLUSION	se envía a suscripción para fijar preexistencias	Conclusión del caso del paciente	ALTO	nvarchar	510	NO	N/A
16	AHORRO	3958600	Valor del ahorro en pesos colombianos	ALTO	nvarchar	510	NO	N/A
17	BH	FIJADA PE	Estado del paciente en Beyond Health	ALTO	nvarchar	510	NO	N/A
18	MES_CUANTIFICACION	44287	Mes de cuantificación del ahorro del caso	MEDIO	nvarchar	510	NO	N/A
19	BASE_OMISIONES_NOVIEMBRE		Estado del paciente sobre omisiones al mes de noviembre	ALTO	nvarchar	510	NO	N/A
20	EPS	SURAMERICANA	EPS del paciente	ALTO	nvarchar	510	NO	N/A
21	MES_REPORTO		Mes en el que se realizó el reporte del paciente	BAJO	nvarchar	510	NO	N/A
22	RANGO_PROVEEDOR	Mayor a 8 Días Habiles	Días hábiles restantes entre la fecha de asignación y fecha de entrega del caso	MEDIO	nvarchar	510	NO	N/A

Fuente: ventana de explorador de Windows

## Anexo 13 Histórico Omisión

Nombre	SAL_HISTORICO_OMISION
Periodicidad	SEMANAL
Origen	\\dc2\vaftp1\IntercambioTerceros\SALUD\FRAUDE_SALUD\Omission_Preexistencias\input\Usuarios

N°	Nombre del Campo	Ejemplo	Descripcion del Campo	Criticidad	Tipo del Campo	Longitud	Puede ir vacío	Cuándo puede ir vacío
1	NUMERO_CASO	1	Numeración de los casos	ALTO	nvarchar	510	NO	N/A
2	FECHA_DE_REPORTE	44615	Fecha en la que se realizó el reporte al usuario	ALTO	nvarchar	510	NO	N/A
3	AREA_QUE_REPORTA	FRAUDE	Área que reporta el caso	ALTO	nvarchar	510	NO	N/A
4	ID	CC	Tipo de documento de identificación del usuario	ALTO	nvarchar	510	NO	N/A
5	NUMERO_ID	1152690767	Número del documento de identificación del usuario	ALTO	nvarchar	510	NO	N/A
6	SUCURSAL_DEL_CONTRATO		Sucursal de donde proviene el contrato	BAJO	nvarchar	510	SI	N/A
7	CLAVE_DEL_ACCESOR		Código numérico del asesor del caso	BAJO	nvarchar	510	SI	N/A
8	NOMBRE_DEL_ASESOR		Nombre del asesor del caso	BAJO	nvarchar	510	SI	N/A
9	DIAGNOSTICO_OMITIDO	TESFRACTURA DEDO DE LA MANO COVID 19	Nombre del diagnóstico presuntamente omitido	BAJO	nvarchar	510	NO	N/A
10	CÓDIGO_CIE_10_o_CUPS	J303J3435626B342	Código CIE 10 o código CUP del diagnóstico presuntamente omitido	MEDIO	nvarchar	510	NO	N/A
11	ESPECIALIDAD_MÉDICA_DE_LA_OMISION	ORTOPEDIA	Especialidad médica pertinente al diagnóstico presuntamente omitido	MEDIO	nvarchar	510	SI	N/A
12	DECISION_COMITE_MEDICO	FIJAR PREEXISTENCIAS	Decisión por de parte del Comité Médico	MEDIO	nvarchar	510	NO	N/A
13	FECHA ENVIO RTA DEFINITIVA A CLIENTE		Fecha de envío de la respuesta	ALTO	nvarchar	510	NO	N/A

## Anexo 14 Usuarios Vigentes

Nombre	SAL_MPP_USUARIOS_VIGENTES
Periodicidad	SEMANAL
Origen	\\dc2\vaftp1\IntercambioTerceros\SALUD\FRAUDE_SALUD\Omission_Preexistencias\input\Usuarios

N°	Nombre del Campo	Ejemplo	Descripcion del Campo	Criticidad	Tipo del Campo	Longitud	Puede ir vacío	Cuándo puede ir vacío
1	TIPO	504	Tipo de contrato (colectivo o familiar)	ALTO	nvarchar	510	NO	N/A
2	GRUPO_ASOCIADO	CARTAGENA	Grupo al cual pertenece el cliente	ALTO	nvarchar	510	NO	N/A
3	REGIONAL	F	Regional a la cual pertenece el contrato de salud	ALTO	nvarchar	510	NO	N/A
4	PLAN_SALUD	Original plus	Plan de salud contratado por el cliente	ALTO	nvarchar	510	NO	N/A
5	CONTRATO	112901690000	Número del contrato asociado al cliente	ALTO	nvarchar	510	NO	N/A
6	FECHA_INICIO	25/01/2023 12:00:00 a. m.	Fecha inicio del contrato del cliente	BAJO	nvarchar	510	SI	N/A
7	FECHA_FIN	24/01/2024 12:00:00 a. m.	Fecha fin del contrato del cliente	BAJO	nvarchar	510	SI	N/A
8	FRECUENCIA_DE_PAGO	MENSUAL	Frecuencia del pago del contrato	BAJO	nvarchar	510	SI	N/A
9	FORMA_DE_PAGO	CHEQUE	Forma del pago del contrato	BAJO	nvarchar	510	NO	N/A
10	TARIFA		Número del tarifa asociado al cliente	MEDIO	nvarchar	510	NO	N/A
11	NOMBRE_COLECTIVO		Nombre de colectivo que incluye el contrato del cliente	MEDIO	nvarchar	510	SI	N/A
12	TIPO_IDENTIFICACION_COLECTIVO		Tipo identificación del colectivo que incluye el contrato del cliente	MEDIO	nvarchar	510	NO	N/A
13	NRO_IDENTIFICACION_COLECTIVO		Numero identificación del colectivo que incluye el contrato del cliente	ALTO	nvarchar	510	NO	N/A
14	TIPO_IDENTIFICACION_CONTRATANTE	CC	Tipo identificación del cliente contratante	ALTO	nvarchar	510	SI	N/A
15	NRO_IDENTIFICACION_CONTRATANTE	45556160	Número identificación del cliente contratante	ALTO	nvarchar	510	SI	N/A
16	PRIMER_APELLIDO_CONTRATANTE	LINDO	Primer apellido del cliente contratante	ALTO	nvarchar	510	NO	N/A
17	SEGUNDO_APELLIDO_CONTRATANTE	LUGO	Segundo apellido del cliente contratante	ALTO	nvarchar	510	SI	N/A
18	NOMBRES_CONTRATANTE	LADILYS MARIA	Nombres del cliente contratante	MEDIO	nvarchar	510	NO	N/A
19	EMAIL_CONTRATANTE	DALYLUZ52@HOTMAIL.COM	Correo electrónico del cliente contratante	MEDIO	nvarchar	510	NO	N/A
20	CELULAR_CONTRATANTE	3109006585	Celular del cliente contratante	MEDIO	nvarchar	510	NO	N/A
21	FECHA_NACIMIENTO_CONTRATANTE	26/10/1982	Fecha nacimiento del cliente contratante	MEDIO	nvarchar	510	NO	N/A
22	TIPO_IDENTIFICACION_BENEFICIARIO	CC	Tipo identificación del beneficiario del contrato	MEDIO	nvarchar	510	NO	N/A

Fuente: Elaboración Propia repositorio share point

## Anexo 15 Negaciones

<b>Nombre</b>	SAL_NEGACIONES					
<b>Periodicidad</b>	SEMANAL					
<b>Origen</b>	\\dc2tvaftp1\IntercambioTerceros\SALUD\FRAUDE_SALUD\Omission_Preexistencias\Input\Usuarios					
N°	Nombre del Campo	Ejemplo	Descripción del Campo	Criticidad	Tipo del Campo	Longitud
1	CTC		Campos vacío general o en 0	ALTO	nvarchar	
2	ACORTAMIENTO	58777	Campos vacío general o en 0	ALTO	nvarchar	510
3	CONSECUTIVO_AUTORIZACION	05/03/2023 IMPRESA APROBADA	Número de la autorización	ALTO	nvarchar	510
4	FECHA_EXPEDICION	HyC Elite Falabella	Fecha de expedición de la autorización	ALTO	nvarchar	10

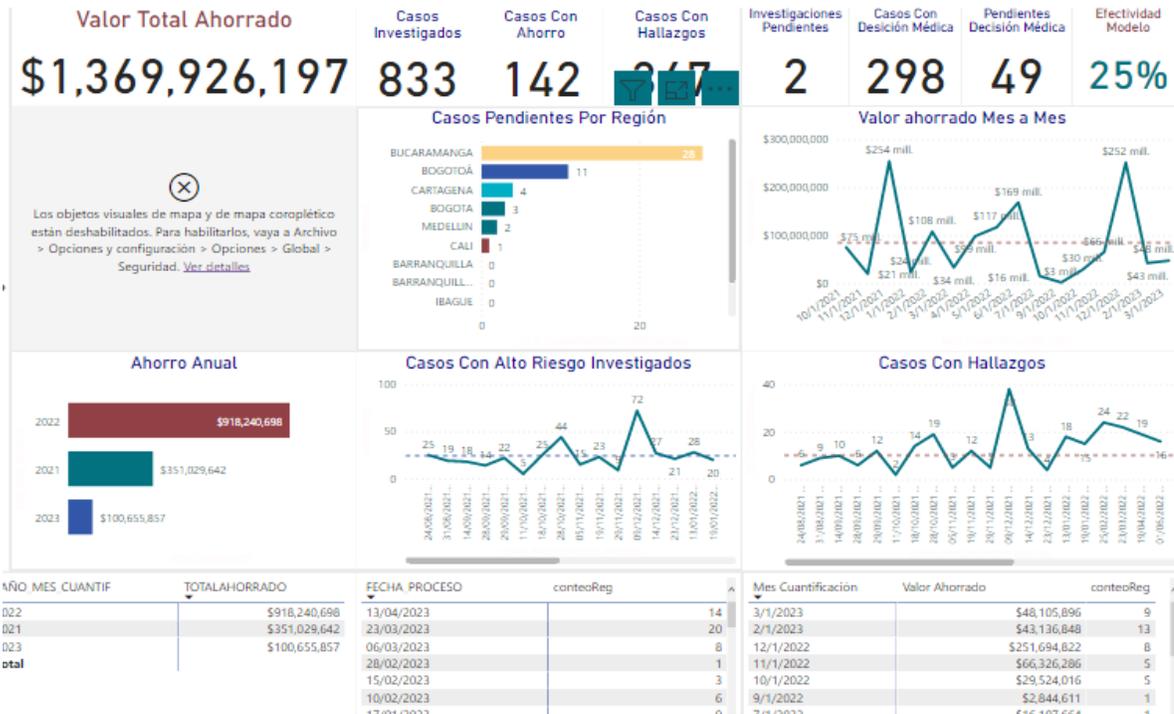
## Anexo 16 Negaciones

<b>Nombre</b>	SAL_PARAMETROS					
<b>Periodicidad</b>	SEMANAL					
<b>Origen</b>	\\dc2tvaftp1\IntercambioTerceros\SALUD\FRAUDE_SALUD\Omission_Preexistencias\Input\Usuarios					
N°	Nombre del Campo	Ejemplo	Descripción del Campo	Criticidad	Tipo del Campo	Longitud
1	ESTADO		Número del estado asociado al cliente	ALTO	nvarchar	
2	PLAN_SALUD	A	Nombre del plan de salud	ALTO	nvarchar	10
3	CODIGO_PROCEDIMIENTO	Todos	Código del procedimiento de salud	ALTO	nvarchar	10
4	PROCEDIMIENTO	452301	Nombre del procedimiento de salud	ALTO	nvarchar	10
5	DIAGNOSTICO_AGRUPACION	452301	Agrupación a la cual pertenece el diagnóstico del paciente	ALTO	nvarchar	10

## 8.5 Anexo Tableros

### Anexo 17 Tablero valor total ahorrado

#### Omisión Preexistencia



Fuente: Elaboración Propia power bi

### Anexo 18 Filtros por diagnostico

Procedimiento Alerta: Todas

Resultado: Todas

Fecha Ahorro: Todas

Fecha Proceso: Todas

Diagnóstico Alerta: Todas

Estado Investigación: Todas

Regional: Todas

NIT Prestador: Todas

Fecha Asignación: 8/24/2021 - 4/13/2023

EPS: Todas

Proveedor: Todas

Año Ahorro: Todas

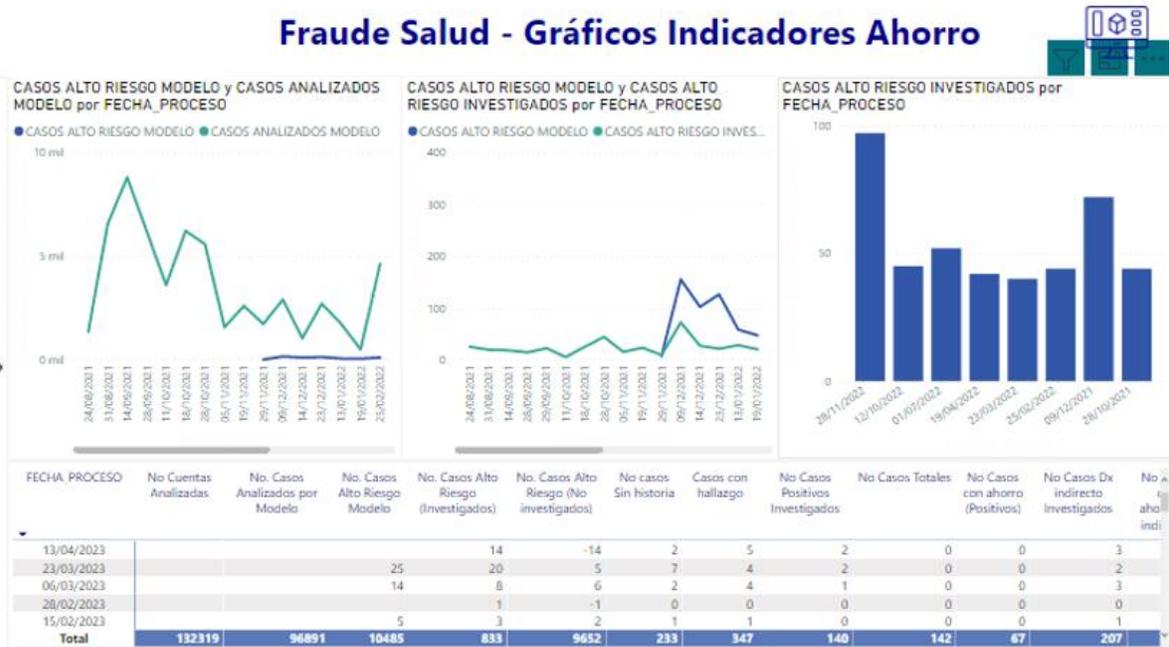
2021-08-24 00:00:00  
 2021-08-31 00:00:00  
 2021-09-14 00:00:00

## Anexo 19 Indicadores de ahorro

Fraude Salud Indicadores Ahorro												
FECHA PROCESO	No Cuentas Analizadas	No. Casos Analizados por Modelo	No. Casos Alto Riesgo Modelo	No. Casos Alto Riesgo (Investigados)	No. Casos Alto Riesgo (No investigados)	No casos Sin historia	Casos con hallazgo	No Casos Positivos Investigados	No Casos Totales	No Casos con ahorro (Positivos)	No Casos Dx indirecto Investigados	No Casos ahorro indirecto
24/08/2021	4473	1357		25	-25	7	6	4	1	1		2
31/08/2021	24558	6562		19	-19	3	9	5	7	5		4
14/09/2021	22957	8797		18	-18	4	10	5	4	1		5
28/09/2021	14799	6200		14	-14	8	6	3	3	3		3
29/09/2021				22	-22	7	12	3	4	2		9
11/10/2021	6356	3599		5	-5	1	2	2	1	1		0
18/10/2021	6794	6216		25	-25	9	14	4	6	3		10
28/10/2021	7425	5573		44	-44	10	19	7	10	5		12
05/11/2021	1874	1566		15	-15	4	5	1	0	0		4
19/11/2021	3686	2991		20	-20	7	12	3	2	1		6
29/11/2021	2468	1723	7	9	-2	1	5	1	3	1		4
09/12/2021	2868	2903	155	72	83	13	38	11	20	8		27
14/12/2021	894	1027	102	27	75	11	13	3	6	2		10
23/12/2021	2246	2703	126	21	105	12	4	0	2	0		4
13/01/2022	1348	1727	58	28	30	6	18	7	10	4		11
19/01/2022	519	493	47	20	27	1	15	9	7	5		6
25/02/2022	2748	4625	105	44	61	9	24	14	8	6		10
23/03/2022	2468	1723	235	40	195	11	22	9	9	5		13
19/04/2022	3638	2591	204	42	162	11	19	6	8	2		13
01/06/2022	2246	5573	123	29	94	8	16	7	5	3		9
01/07/2022				52	-52	13	20	7	9	2		13
02/09/2022			65	29	36	9	1	1	0	0		0
06/09/2022				1	-1	0	1	1	1	1		0
12/10/2022			397	45	352	17	16	7	8	3		9
28/11/2022				97	-97	31	23	12	6	1		11
04/12/2022				1	-1	0	1	1	1	1		0
<b>Total</b>	<b>132319</b>	<b>96891</b>	<b>10485</b>	<b>833</b>	<b>9652</b>	<b>233</b>	<b>347</b>	<b>140</b>	<b>142</b>	<b>67</b>		<b>207</b>

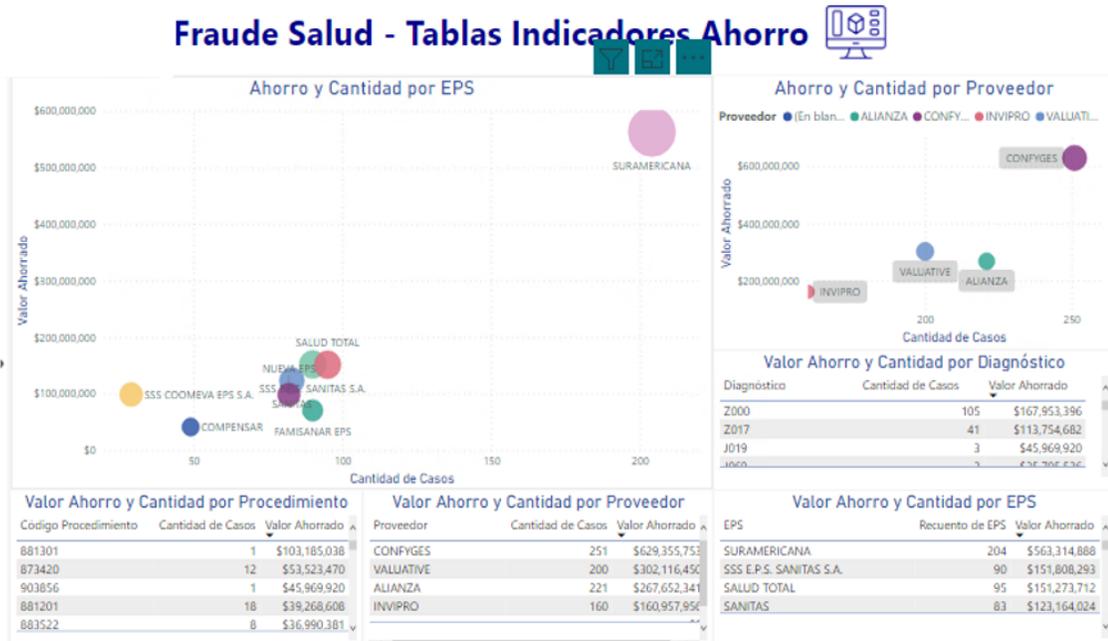
Fuente: Elaboración Propia power bi

## Anexo 20 Indicadores de ahorro gráficos



Fuente: Elaboración Propia power bi

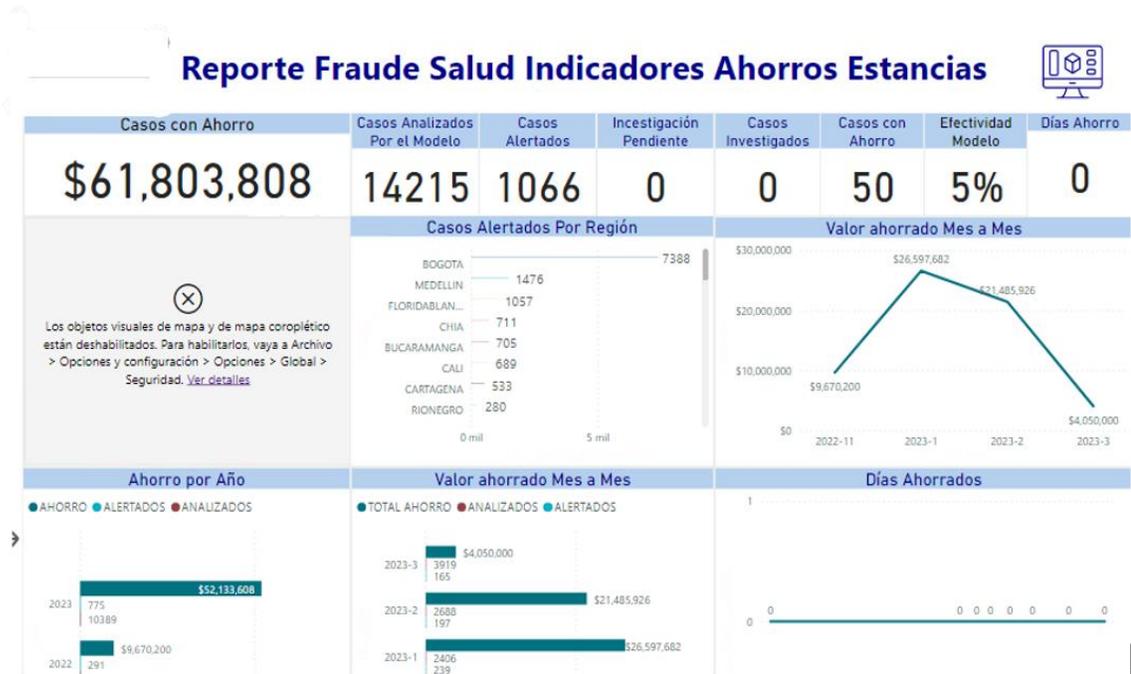
## Anexo 21 Tablas indicador ahorro



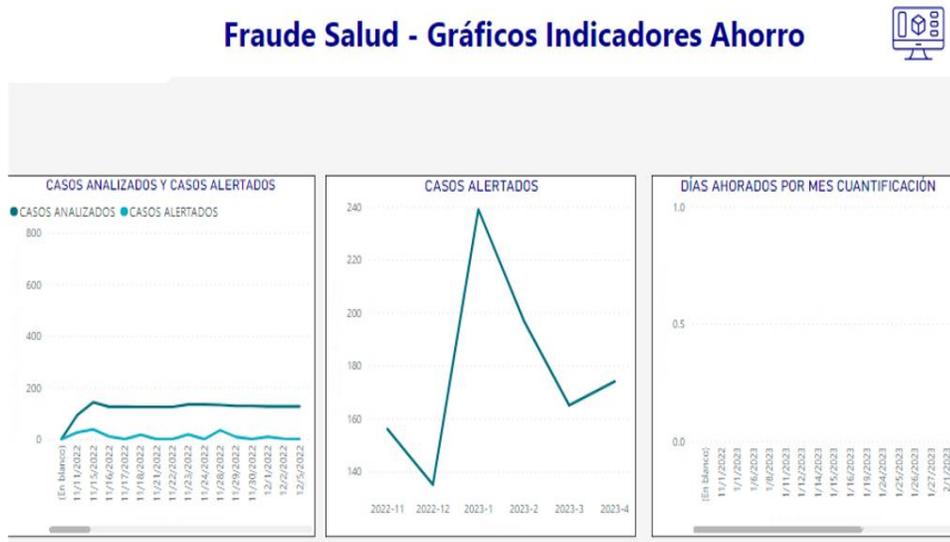
Fuente: Elaboración Propia power bi

## 8.6 Anexo Tablero Estancias

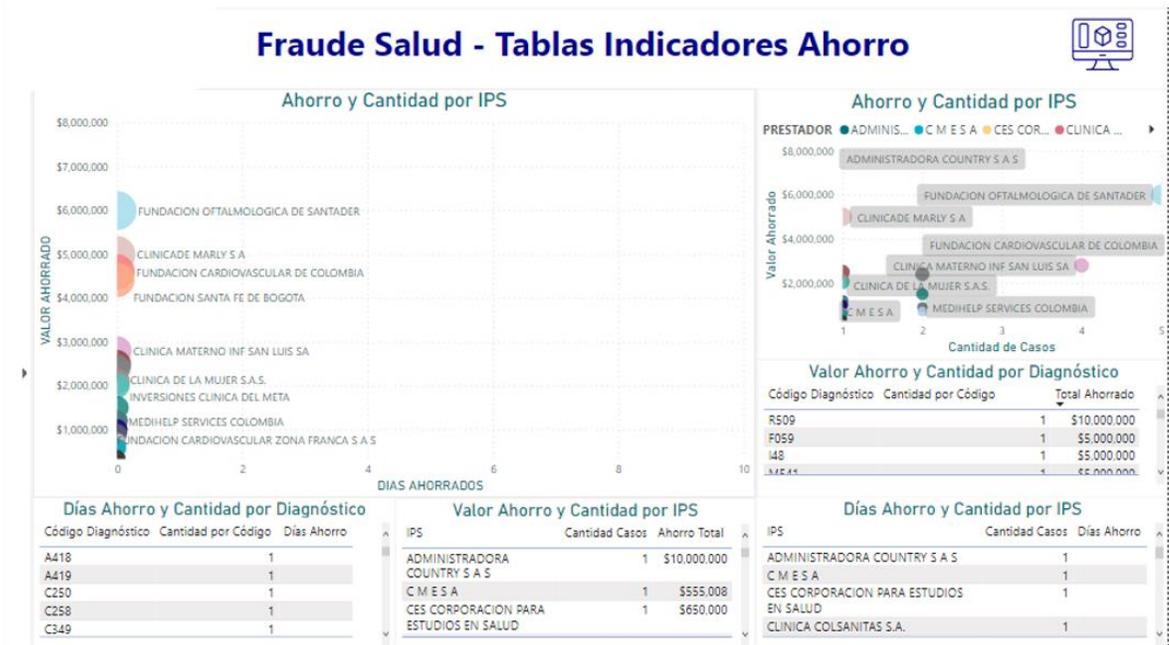
### Anexo 22 ahorro estancias



## Anexo 23 Casos Alertados

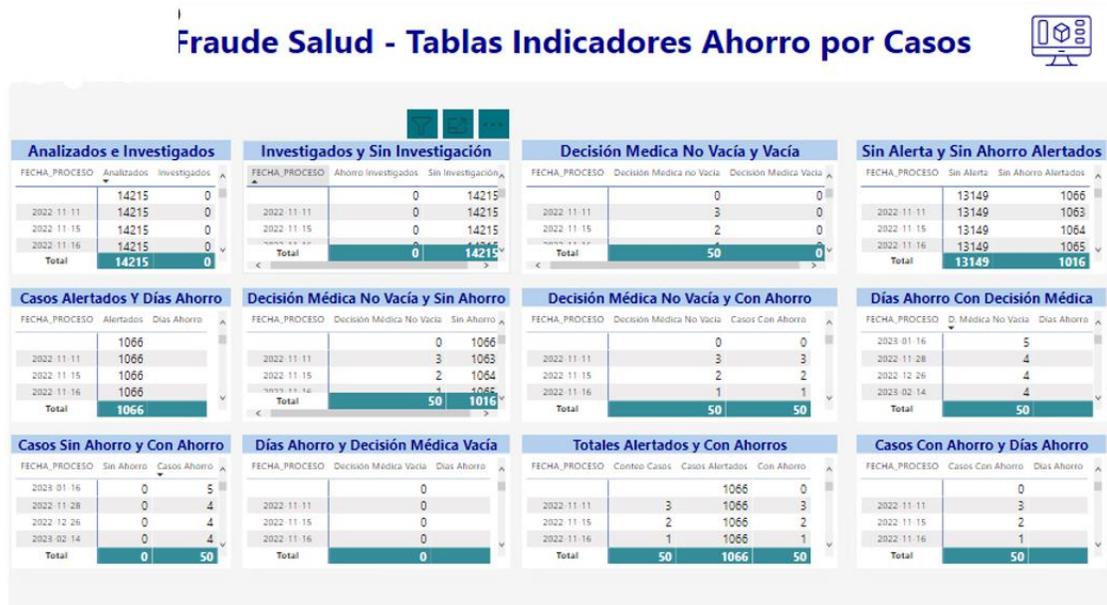


## Anexo 24 Indicadores de ahorro



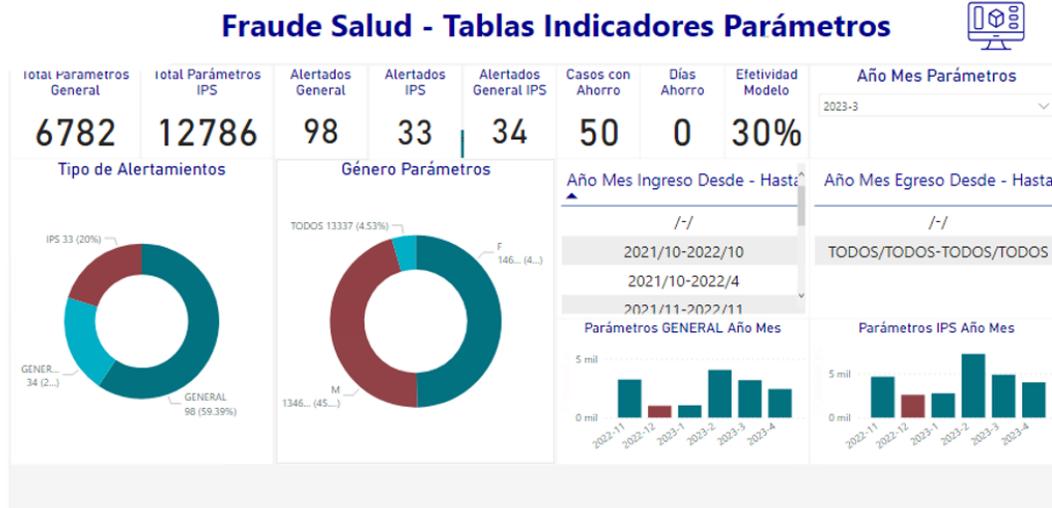
Fuente: Elaboración Propia power bi

## Anexo 25 Indicadores de ahorro por caso



Fuente: Elaboración Propia power bi

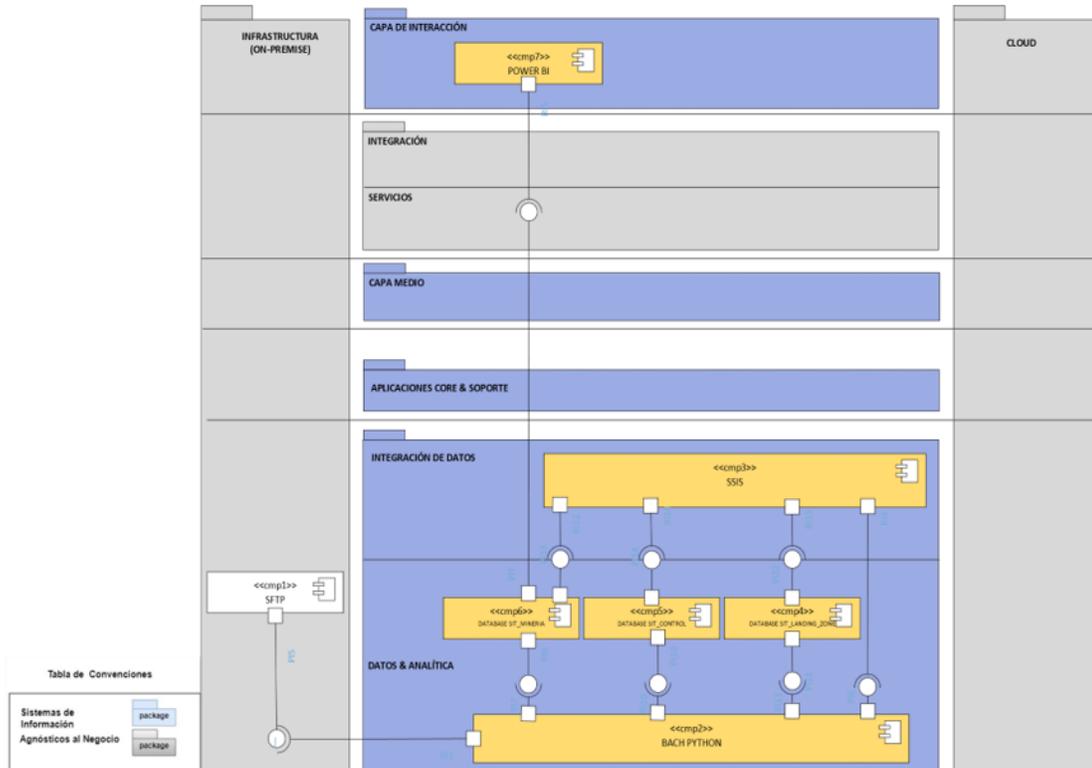
## Anexo 26



Fuente: Elaboración Propia power bi

## 8.7 Anexo Diagrama De Componentes

### Anexo 27 Diagrama de Componentes



Fuente: Elaboración Propia desarrollado en draiwo





## 8.10 Anexo 30 Datos Críticos Estancias Prolongadas

Fuente	Estancias prolongadas	
	CAMPO	NOM_CAMPO
Ahorros Estancias	3	GENERO
	6	CODIGO_DIAGNOSTICO PRINCIPAL
	7	FECHA DE INGRESO
	9	DIAS DE ESTANCIA
	10	PRESTADOR
Censo Hospitalario Diario	1	PLAN_x002F_TIPO_SINIESTRO
	10	CODIGO_HABILITACION
	11	FECHA_DE_INGRESO
Censo Hospitalario Mensual	1	PLAN_x002F_TIPO_SINIESTRO
	6	CODIGO_HABILITACION
	10	PRESTADOR
	12	FECHA_DE_INGRESO
	14	AUDITOR_1

Fuente: Elaboración Propia Datos Críticos Estancias Prolongadas

### 8.11 Anexo 31 Datos Críticos Estancias Prolongadas

Fuente	Omision	
	CAMPO	NOM_CAMPO
Ahorros Iniciativa	1	FECHA_PROCESO
	2	DIAGNOSTICOS_OMITIDOS
	6	FECHA_ASIGNACION_CASO
	7	FECHA_ENTREGA_CASO
	17	DX_ALERTA
enfermedades	3	Nombre_Capitulo
	4	Descripcion_codigos_tres_caracteres
	6	Descripcion_códigos_Cuatro_Caracteres
Usuarios vigentes	2	FECHA FIN
	3	FECHA ANTIGÜEDAD
	6	TELEFONO CORRESPONDENCIA
	7	DIRECCION CORRESPONDENCIA
	9	CIUDAD CORRESPONDENCIA
Negaciones	2	Fecha_Expedición
	3	Numero_Contrato
	5	Inicio_Vigencia
	10	Diagnostico
	14	Nombre
	16	Observaciones

Fuente: Elaboración Propia Datos Críticos Omisión de preexistencias

### 8.12 Anexo 32 Datos Críticos Estancias Prolongadas

Fuente	Altos Consumos	
	CAMPO	NOM_CAMPO
Gasto Medico	1	NUMERO_INTERNO
	6	FOLIO
	7	CODIGO_HABILITACION
	9	COD_DIAGNOSTICO
	10	NOMBRE_PROCEDIMIENTO
Ahorros Altos	1	FECHA_ASIGNACION_CASO
	2	ESTADO_P_Y_N
	4	DX_ALERTA

Fuente: Elaboración Propia Datos Críticos Altos Consumos