

ETL ENTREGA PROYECTO 2

Maestría en Ciencia de Datos e Inteligencia Artificial

Nelcy Lucia Zapata Gil–22502267 Valentina Isaza Ospina - 22502266

Resumen— El Fondo Mixto de Etnocultura y Desarrollo Social - FONPACÍFICO enfrenta dificultades en su gestión financiera debido a la ausencia de un flujo de caja proyectado, lo que impide tomar decisiones informadas sobre la disponibilidad de recursos para su sostenibilidad y la cofinanciación de proyectos. La dispersión de información financiera en múltiples fuentes y la dependencia de procesos manuales agravan esta situación, aumentando el riesgo de errores y limitando la generación de reportes fiables en tiempo real.

Este proyecto propone diseñar, preparar y construir flujos de datos eficientes mediante procesos de Extracción, Transformación y Carga (ETL) para integrar diversas fuentes de datos. Que permita a FONPACIFICO desarrollar un modelo predictivo para el análisis avanzado de la información y optimice su gestión financiera, mejorando la toma de decisiones estratégicas y operativas. La implementación de procesos ETL en la entidad, facilitará la consolidación de datos financieros, mejorará la precisión de los reportes y aumentará la capacidad de análisis en tiempo real, utilizando herramientas como Python, Pandas y MySQL, y establecerá las bases para el uso futuro de modelos predictivos.

1. DESCRIPCIÓN DEL PROBLEMA

FONPACÍFICO es una entidad pública sin ánimo de lucro que gerencia proyectos de interés público a nivel nacional. En la actualidad, enfrenta un desafío crítico en su gestión financiera debido a la falta de un flujo de caja proyectado, lo que genera incertidumbre sobre la sostenibilidad y disponibilidad de recursos para la ejecución de proyectos. Esta situación impacta la toma de decisiones estratégicas y operativas, y la capacidad de la entidad para anticiparse a periodos de déficit o superávit de liquidez y gestionar de manera eficiente los recursos.

La entidad opera en un entorno altamente regulado, donde los ciclos de contratación estatal, las restricciones normativas y los cambios en la administración pública influyen directamente en el flujo financiero. La dinámica presupuestaria gubernamental, marcada por la apertura y cierre de vigencias fiscales, así como las limitaciones impuestas en períodos electorales (Ley de Garantías), o las alteraciones (provocados por factores externos como inestabilidad social, condiciones climáticas adversas o emergencias nacionales) de cronograma de flujo de los proyectos financiados por el Sistema General de Regalías (SGR), compromete la estabilidad financiera de la entidad e

impactan la capacidad de planificación y operación administrativa, así como la ejecución de los proyectos a su cargo.

Desde una perspectiva técnica, la ausencia de un sistema automatizado de Extracción, Transformación y Carga de datos (ETL) dificulta la consolidación de información financiera dispersa en diferentes fuentes. La dependencia de procesos manuales aumenta la probabilidad de errores y dificulta la generación de reportes confiables en tiempo real, lo que limita la capacidad de análisis y toma de decisiones fundamentadas.

El desarrollo de un modelo de flujo de caja proyectado, mediante procesos de Extracción, Transformación y Carga (ETL) para integrar diversas fuentes de datos, permitiría a FONPACIFICO desarrollar un modelo predictivo para el análisis avanzado de la información y optimice su gestión financiera, mejorando la toma de decisiones estratégicas y operativas. La implementación de procesos ETL en la entidad, facilitará la consolidación de datos financieros, mejorará la precisión de los reportes y aumentará la capacidad de análisis en tiempo real, utilizando herramientas como Python, Pandas y MySQL, y establecerá las bases para el uso futuro de modelos predictivos.

1.1. Contribución al Desarrollo de Habilidades en ETL y Ciencia de Datos

La implementación de procesos ETL en la entidad, facilitará la consolidación de datos financieros, mejorará la precisión de los reportes y aumentará la capacidad de análisis en tiempo real, utilizando herramientas como Python, Pandas y MySQL, permitirá aplicar conocimientos clave en ETL y Ciencia de Datos para la estructuración y análisis de datos financieros. Además, servirá como base para futuras implementaciones de modelos de Machine Learning para predicción financiera, generando un impacto significativo en la eficiencia operativa de la organización.

Este proyecto no solo aborda un problema de gestión financiera real, sino que también representa una aplicación práctica de la Ciencia de Datos en la optimización de procesos organizacionales, fortaleciendo la capacidad analítica y de predicción financiera de FONPACÍFICO.

2. JUSTIFICACIÓN DEL PROYECTO

La implementación de procesos de Extracción, Transformación y Carga (ETL) permitirá a FONPACÍFICO optimizar la gestión financiera y mejorar la toma de decisiones. Este modelo proporcionará una visión consolidada y automatizada de los ingresos y egresos, alineándose con la planificación comercial y las restricciones estatales.

Desde la perspectiva de la Ciencia de Datos, este proyecto permite aplicar técnicas de ETL para estructurar la información financiera dispersa en diversas fuentes, reduciendo errores manuales y mejorando la eficiencia operativa. Además, facilitará el análisis de datos para anticipar posibles riesgos financieros.

Los beneficios clave incluyen:

- a. Mejor previsión financiera, permitiendo la identificación temprana de periodos críticos.
- b. Mayor transparencia y cumplimiento normativo.
- c. Automatización del análisis financiero, integrando datos heterogéneos en un sistema centralizado.
- d. Optimización de la planeación estratégica, comercial, y operativa considerando los ciclos de contratación estatal.

3. IDENTIFICACIÓN DE FUENTES

Para el proceso de Extracción, Transformación y Carga (ETL) de datos financieros, se identificarán y utilizarán diversas fuentes de datos, clasificadas según su tipo, formato y volumen:

3.1. Fuentes Internas

- a. <u>Presupuesto 20204 y 2025:</u> Archivo Excel (.xls) con proyecciones de ingresos y egresos.
- b. <u>Sistema Financiero y Contable Soul SIFP:</u>
 Documento Word (.docx) Políticas y criterios contables.
- c. <u>Historial de Contratos</u>: Base de datos interna con registros de contratación pública y privada.
- d. <u>Estados Financieros</u>: Documentos financieros en formato Excel y PDF (.xlsx, .pdf).
- e. <u>Matriz de información Consolidada:</u> Base de datos interna con registros de contratación pública y privada formato Excel y PDF (.xlsx, .pdf).

3.2. Fuentes Externas

- a. <u>Sistema de Contratación Estatal (SECOPI y II):</u>
 Datos abiertos (.csv, API) sobre contratos y licitaciones.
- b. <u>GESPROY del DNP</u>: Plataforma tecnológica que permite a las entidades ejecutoras gestionar los proyectos financiados con recursos del Sistema General de Regalías (SGR) (.csv, .xlsx).
- c. <u>SPGR del DNP</u>: el Sistema de Presupuesto y Giro de Regalías, es un sistema de información que

- gestiona el presupuesto y el giro de regalías del SGR (Sistema General de Regalías) administrado por el Ministerio de Hacienda y Crédito Público. (.csv, .xlsx).
- d. <u>Normatividad Estatal:</u> Documentación en PDF (.pdf) sobre Ley de Garantías y restricciones presupuestales.
- e. <u>Calendario Electoral y de Vigencias</u>: Datos estructurados en formato CSV (.csv) con fechas clave del ciclo gubernamental.

a. Tipo de Dato

Fuente: Excel (Drive)

Tipo de Dato: Semi-Estructurado

Información: Columnas con valores numéricos,

fechas y textos sobre contratos y pagos.

Fuente: MySQL

Tipo de Dato: Estructurado

Información: Tablas con registros normalizados de contratos, pagos y estados financieros.

b. Formato del Dato

Fuente: Excel (Drive)
Formato: .xlsx

Detalles: Archivo de hoja de cálculo con varias

hojas.

Fuente: MySQL

Formato: Tabla de MySQL

Detalles: Base de datos relacional con múltiples

tablas organizadas.

c. Volumen de los Datos

Fuente: Excel (Drive)

Volumen Aproximado: Medio (MB - GB) Frecuencia de Actualización: Actualización

manual mensual o trimestral.

Fuente: MySQL

Volumen Aproximado: Medio (MB - GB) Frecuencia de Actualización: Actualización frecuente, dependiendo de la operación diaria.

Excel: 1,12 MB

MySQL:

Base de datos	Tamaño (MB)	Tamaño (GB)
viaticos	0.27	0.00
soul_red_2024	26.58	0.03
soul_fon_2023	66.24	0.06
soul_fon_2022	86.39	0.08
soul_fon_2025	79.86	0.08
soul fon 2024	98.11	0.10

4. PROBLEMA

La falta de un sistema ETL para integrar y analizar la información financiera impide a FONPACÍFICO tener una visión clara y oportuna de su situación financiera, lo que afecta su capacidad para planificar y ejecutar sus proyectos de manera eficiente.

5. EXPLORACIÓN INICIAL DE DATOS

5.1. DATOS INTERNOS

5.1.1. Archivo "Matriz de Gestión de Proyectos 04-03-2025"

a. Estructura del Archivo

- Contiene 243 columnas y varias filas con datos nulos.
- Muchas columnas están etiquetadas como "Unnamed", lo que indica un problema de encabezados o estructura del archivo. Las primeras filas contienen categorías como:
 - **Estado Principal del Proyecto**: "Sin Contratar", "Terminado", en ejecución, suspendido, etc.
 - **Entidad Auditora** y datos administrativos de seguimiento.
 - **Información Financiera**, incluyendo compromisos, órdenes de pago, deducciones y comisiones.

b. Problemas identificados

- Datos dispersos: Se observan muchas columnas con nombres genéricos como "Unnamed", lo que puede deberse a una mala estructuración del archivo.
- Valores faltantes: Existen múltiples valores vacíos, especialmente en campos de auditoría y seguimiento financiero.
- Formato irregular: Algunas columnas parecen ser una combinación de etiquetas y valores, lo que requiere limpieza antes de analizarlas.

c. Acciones de Transformación de Datos

Para mejorar la usabilidad de este archivo en el proceso ETL:

- **Renombrar columnas** para mejorar la comprensión de los datos.
- Eliminar columnas irrelevantes o sin datos significativos.
- Estandarizar estados de proyectos para facilitar su clasificación.
- Transformar la estructura del archivo en un formato más relacional para facilitar análisis en una base de datos.

5.1.2. Archivo "SOUL. PROYECTOS 04-03-2025.xlsx" 1. Estructura del Archivo

- Contiene 14 columnas y 398 filas, con información sobre proyectos financiados dentro del sistema SOUL.
- Algunas variables contienen datos en formato numérico (valores financieros) y otras en texto (códigos de proyectos y programas).
- Existen valores faltantes en varias columnas, especialmente en nombres de proyectos y sectores.

a. Principales Variables en el Archivo

Identificación de Proyectos:

- Código del Proyecto (codigo).
- Nombre del Proyecto (nombre), aunque muchos valores están vacíos.
- Sector y Nombre del Sector (sector, nombre sec).
- Código BPIN (cbnp), vinculado a la planeación de inversión pública.

Información Financiera:

- Valor del Proyecto (valor), indicando los montos asignados a cada iniciativa.
- Contratos Asociados (contratos), aunque casi todos están en 0.
- Disponibilidades, Compromisos, Obligaciones, IVA, Deducciones y Pagos, con información detallada sobre la ejecución presupuestal.

b. Problemas Identificados

Valores Faltantes:

- Muchas filas tienen nombres de proyectos vacíos (nombre).
- Algunas celdas de sector y nombre del sector están sin información.

Valores Numéricos con Cero:

 La mayoría de las columnas financieras como Contratos, Disponibilidades, Compromisos, Obligaciones y Pagos están en cero, lo que indica que los datos pueden no estar actualizados o representar proyectos inactivos.

Formato de Datos:

- La columna **Código BPIN** (**cbnp**) tiene registros mezclados entre números y texto (BPIN2023000060014 vs 2021000030042), lo que indica una posible inconsistencia en los formatos.
- Algunas columnas que deberían ser categóricas (nombre sec) tienen valores faltantes.

c. Acciones de Transformación de Datos

Para mejorar la usabilidad de este archivo en el proceso ETL

- Llenar valores faltantes en nombre y nombre_sec con información de referencia si es posible.
- Revisar los valores en cero, para determinar si corresponden a datos no actualizados o proyectos cancelados.
- Estandarizar el código BPIN, asegurando que todos tengan el mismo formato.
- Transformar columnas categóricas, asignando códigos únicos donde falten valores.
- Generar una variable de estado del proyecto, según si tiene asignación presupuestal activa o está en cero.

5.2. DATOS EXTERNOS

5.2.1. <u>Archivo "SECOP 1. CONSULTA PRINCIPALES"</u>

a. Estructura del Archivo

- Contiene 75 columnas y 114 filas, con una gran cantidad de datos administrativos y financieros sobre contratación estatal.
- Se enfoca en procesos de compra pública registrados en el SECOP I (Sistema Electrónico de Contratación Pública).
- La información está organizada por entidades contratantes, modalidades de contratación, montos y fechas clave.

b. Principales Variables en el Archivo

• Identificación de la Contratación:

UID: Código único de cada proceso de compra.

Año de Firma del Contrato y Año de Cargue en el SECOP.

Nivel de la Entidad: Territorial o Nacional. Nombre de la Entidad y su NIT.

• Modalidad y Tipo de Contratación:

Código y Descripción de Modalidad de Contratación: Ej., Contratación Directa, Régimen Especial.

Orden de la Entidad: Centralizado o descentralizado.

• Ubicación Geográfica:

Departamento y Municipio de la Entidad que ejecuta la contratación.

• Estado y Cumplimiento Normativo:

Fecha de Última Actualización.

Fecha de Liquidación del Contrato (NaN en muchos casos).

Cumplimiento del Decreto 248 (indica si cumple requisitos de contratación pública).

Otros Indicadores:

Si la empresa es una MiPyme. Si cumple con sentencias de contratación pública.

c. Problemas identificados

Valores Faltantes:

La columna "Fecha de Liquidación" tiene muchos valores nulos, lo que puede indicar que los contratos aún están activos o que faltan registros. Variables como "Cumple Decreto 248" y "Es MiPyme" tienen valores genéricos como "No definido", lo que sugiere que estos datos no están siendo bien reportados.

• Formato y Normalización:

Existen columnas con nombres en mayúsculas y caracteres especiales, lo que puede dificultar su procesamiento.

Algunas variables están en formato texto cuando deberían ser numéricas o categóricas.

• Fechas No Estructuradas Correctamente:

"Fecha de Última Actualización" está bien estructurada, pero "Fecha de Liquidación" tiene muchos valores en blanco o sin definir.

• Consistencia en Modalidades de Contratación:

Algunos registros de "Modalidad de Contratación" pueden ser normalizados, ya que contienen diferentes formas de escritura para el mismo tipo de contratación.

d. Acciones de Transformación de Datos

Para mejorar la usabilidad de este archivo en el proceso ETL

- Estandarizar nombres de columnas eliminando caracteres especiales y mayúsculas inconsistentes.
- Manejar valores faltantes, imputando fechas faltantes o marcándolas como contratos en proceso.
- Convertir datos categóricos en variables codificadas, especialmente en modalidades de contratación.
- Normalizar la información financiera, verificando si existen inconsistencias en los datos de contratación.
- Crear variables derivadas, como "Tiempo de Ejecución del Contrato", basado en la diferencia entre firma y liquidación.

5.2.2. Archivo "SECOP 2. CONSULTA PRINCIPALES"

a. Estructura del Archivo

- Contiene 59 columnas y 103 filas con información detallada sobre procesos de contratación pública en SECOP II.
- Se enfoca en datos de licitaciones, adjudicaciones y contratación estatal, con un alto nivel de detalle.
- Algunas columnas contienen valores repetidos, lo que puede indicar datos duplicados o contratos con múltiples registros.

b. Principales Variables en el Archivo

• Identificación de la Entidad y el Contrato: Entidad Contratante (Entidad, Nit Entidad).

Departamento y Ciudad de la Entidad.

ID del Proceso (ID del Proceso): Código único del proceso de compra.

Referencia del Proceso: Nombre del contrato.

• Detalles de Contratación:

Estado del Proceso (Estado de Apertura del Proceso), puede ser "Abierto", "Adjudicado", etc. Tipo de Contrato y Subtipo de Contrato. Código de Categoría de Compra.

Nombre del Proveedor Adjudicado y NIT del Proveedor.

Información de Publicación y Adjudicación:
 URL del Proceso (URLProceso), lo que permite acceder a detalles en SECOP II.

Código de la Entidad y estado de adjudicación (Estado Resumen).

c. Problemas identificados

• Valores Repetidos:

Existen registros duplicados de adjudicación para el mismo contrato (ID del Proceso y Referencia del Proceso aparecen varias veces para el mismo proveedor).

Puede ser necesario **de duplicar los datos**, dejando solo la adjudicación final.

• Valores Faltantes:

Algunas columnas como "Subtipo de Contrato" y "Categorías Adicionales" tienen valores en "No definido", lo que sugiere datos incompletos en el sistema de contratación.

Formato de Datos:

"Código Entidad" está separado por comas, lo que puede requerir limpieza y conversión a formato numérico.

"URLProceso" es un enlace funcional, pero podría ser estructurado para consulta más eficiente.

• Datos No Normalizados:

Nombres de proveedores pueden estar en diferentes formatos (mayúsculas/minúsculas). Estado del proceso tiene diferentes categorizaciones que podrían ser unificadas.

d. Acciones de Transformación de Datos

Para mejorar la usabilidad de este archivo en el proceso **ETL**

- Eliminar duplicados, dejando solo la adjudicación final de cada contrato.
- Completar valores faltantes, especialmente en tipos de contrato y categorías de compra.
- Convertir "Código Entidad" a formato numérico eliminando comas.
- Estandarizar los nombres de proveedores y entidades, asegurando consistencia en el análisis.
- Generar un indicador de estado de contratación, para clasificar contratos en "Adjudicado", "En proceso" o "Cancelado".

5.2.3. Archivo "GESPROY. CG-CG-CG-proy_04_marzo_2025_09-38-10.xlsx.xlsx"

a. Estructura del Archivo

- Contiene 22 columnas y 168 filas.
- La mayoría de las columnas tienen nombres genéricos como "Unnamed", lo que sugiere un problema con los encabezados o la estructura de los datos.
- Incluye información sobre proyectos financiados a través del Sistema General de Regalías (SGR).

b. Principales Variables en el Archivo

- **Ubicación Geográfica**: Departamento, región y entidad territorial que gestiona cada proyecto.
- Datos del Provecto:

BPIN: Código del proyecto en el Banco de Programas y Proyectos de Inversión Nacional.

Nombre del Proyecto.

Fecha de Acuerdo: Fecha en que se aprobó el proyecto.

Estado del Proyecto: Ej., "Cerrado", "Contratado en Ejecución", "Terminado".

• Financiamiento:

Valor SGR: Recursos asignados por el Sistema General de Regalías.

Valor Nación y Otros: Recursos adicionales. Valor Total del Proyecto.

• Ejecución:

Avance Físico (%): Grado de avance de la obra o ejecución del proyecto.

Avance Financiero (%): Relación entre el dinero ejecutado y el presupuesto total.

Valor Pagos: Dinero ya desembolsado en el proyecto.

c. Problemas identificados

 Encabezados mal estructurados: Muchas columnas tienen espacios o caracteres especiales (\t).

• Valores faltantes en algunas variables:

Algunas celdas de "Subsector" están vacías. En algunas filas, la columna de "Valor Otros" está en blanco.

• Valores inconsistentes en Avance Físico y Financiero:

Algunos proyectos muestran **0% de avance físico**, pero con desembolsos financieros significativos. Se requiere validar si estos valores reflejan correctamente la ejecución de cada proyecto.

• Conversión de Formatos:

Fechas almacenadas como texto deben convertirse a formato YYYY-MM-DD.

Valores numéricos como "Valor Total del Proyecto" tienen decimales que deben revisarse.

d. Acciones de Transformación de Datos

Para mejorar la usabilidad de este archivo en el proceso ETL

- Estandarizar nombres de columnas eliminando espacios y caracteres no deseados.
- Rellenar valores faltantes con estrategias de interpolación o categorización.
- Validar la consistencia de avances físicos y financieros, detectando posibles errores.
- Convertir datos financieros a formato numérico adecuado para cálculos precisos.
- Optimizar la base de datos, eliminando columnas irrelevantes o redundantes.

5.2.4. Archivo "SPGR. Exportar CRONOGRAMA DE FLUJOS 04-03-2025.xlsx"

a. Estructura del Archivo

- Contiene 38 columnas y 79 filas, con datos sobre la programación financiera de proyectos de inversión gestionados a través del Sistema de Presupuesto y Giro de Regalías (SPGR).
- La información está organizada por entidades ejecutoras, proyectos, códigos presupuestales y programación de recursos a lo largo del tiempo.

b. Principales Variables en el Archivo

• Identificación del Proyecto:

Vigencia Presupuestal (2025-2026 en este caso). **Código y Nombre de la Entidad** (Nombre de la Entidad, Identificación de la Entidad).

NIT del Ejecutor del Proyecto (NIT del Ejecutor del proyecto de Inversión).

Código y Nombre del Proyecto (Identificación del proyecto de Inversión, Nombre del proyecto de Inversión).

• Información Presupuestaria:

Código del recurso presupuestal (Código del recurso presupuestal).

Código y Descripción del Fondo (Código del Fondo, Descripción del Fondo).

Tipo de recurso (Descripción del recurso presupuestal).

• Cronograma de Flujos Financieros:

Fechas en formato "AAAA-MM-DD" que representan meses específicos desde el 2002 hasta 2025.

Montos programados en cada fecha, indicando desembolsos o giros esperados para los proyectos.

c. Problemas Identificados

Valores Faltantes:

En algunas columnas, especialmente en los códigos presupuestales, hay valores vacíos o incompletos.

• Formato de Fechas y Desembolsos:

Hay múltiples columnas con **fechas en formato de texto** (ejemplo: 2002-04-01 00:00:00), lo que puede dificultar el análisis temporal.

La mayoría de los valores en estas fechas están en **cero** (0), lo que podría indicar falta de programación efectiva o datos incompletos.

Nombres de Variables Largos e Inconsistentes:
 Algunos nombres de columnas como
 "Identificación del proyecto de Inversión"
 pueden acortarse para mayor claridad.

Los nombres de entidades y fondos deben estandarizarse para evitar inconsistencias en el análisis.

d. Acciones de Transformación de Datos

Para mejorar la usabilidad de este archivo en el proceso **ETL**

- Eliminar columnas irrelevantes y dejar solo aquellas con valores financieros relevantes.
- Convertir columnas de fechas a formato adecuado (YYYY-MM-DD) y organizarlas en un formato más estructurado.
- Filtrar datos con montos programados diferentes de cero, para enfocarse en proyectos con ejecución presupuestal real.
- Normalizar nombres de entidades y códigos presupuestales, asegurando consistencia con otros archivos financieros.
- Transformar el dataset a un modelo relacional, con una tabla para entidades, provectos y desembolsos por fecha.

5.2.5. Archivo "SPGR. ExportarPAGOS 04-03-2025.xlsx"

a. Estructura del Archivo

- Contiene 46 columnas y varios registros de pagos, con información sobre documentos de pago, fechas, montos y beneficiarios.
- Se enfoca en pagos realizados bajo el Sistema de Presupuesto y Giro de Regalías (SPGR).
- Los pagos incluyen detalles sobre deducciones, compromisos, cuentas por pagar y reintegros.
- Principales Variables en el Archivo

Identificación del Pago:

Número de Documento (Numero Documento).

Fecha de Registro y Fecha de Pago, que indican el ciclo de ejecución del desembolso.

Estado del Pago (Estado), donde la mayoría están marcados como "**Pagada**".

Valores Monetarios:

Valor Bruto (Valor Bruto).

Valor Deducciones (Valor Deducciones).

Valor Neto (Valor Neto), que indica el monto final después de descuentos.

Información del Beneficiario:

Tipo de Beneficiario (Tipo Beneficiario), usualmente "**Beneficiario final**".

Tipo de Identificación y Número de Identificación del beneficiario.

Estructura Presupuestaria:

Compromisos y **Obligaciones** (Compromisos, Obligaciones).

Cuentas por Pagar y Órdenes de Pago (Cuentas por Pagar, Órdenes de Pago).

Reintegros (Reintegros), que aparecen en **cero** (**0**) en la mayoría de registros.

b. Problemas identificados

• Valores Faltantes:

Algunas columnas como "Concepto Pago" y "Documento Masivo" están en blanco en varios registros.

Existen valores nulos en los números de cuenta bancaria.

• Formato de Fechas v Valores Monetarios:

Las fechas están en formato de texto (YYYY-MM-DD HH:MM:SS), lo que puede requerir conversión a datetime.

Algunos valores numéricos tienen espacios adicionales (Valor Bruto, Valor Neto), lo que puede generar problemas al hacer cálculos.

• Datos Repetidos:

Algunos registros tienen **idénticos valores en varias columnas**, lo que puede indicar duplicación de pagos.

• Información Financiera Incompleta:

Cuentas por Pagar y Obligaciones tienen valores que podrían no coincidir con los registros de compromisos.

Reintegros aparecen en **cero en todos los registros**, lo que puede indicar que la información de devoluciones no está siendo registrada.

c. Acciones de Transformación de Datos

Para mejorar la usabilidad de este archivo en el proceso ETL

- Limpiar y convertir las fechas a un formato de fecha legible.
- Eliminar espacios y caracteres adicionales en valores monetarios.
- Identificar registros duplicados y consolidarlos si es necesario.
- Completar valores faltantes en Concepto Pago y otros campos críticos.
- Estandarizar las cuentas por pagar, compromisos y obligaciones, verificando consistencia con los valores de pago.
- Crear una tabla de beneficiarios única con identificadores limpios.

6. EXTRACCION DE DATOS

En primera instancia enlistamos los dataframe en formatos. excel y .csv que han identificado previamente como insumo para el desarrollo del flujo de caja, para ello, se optó por dividirlo en 5 categorías según la fuente desde la cual se extrajo, quedando de la siguiente forma:

FUENTE	DATA BASE
GESPROY	GESPROY. CG-CG-CG-proy_04_marzo_2025_09-38-10.xlsx.xlsx
PROYECTOS	MATRIZ GESTIÓN DE PROYECTOS 04-03- 2025.csv
SECOP	SECOP 1. CONSULTA PRINCIPALES.csv SECOP 2. CONSULTA PRINCIPALES.csv
SOUL	EXCEL_FACTURACION_2024.csv RECAUDO2.csv
SPGR	SPGR. ExportarCRONOGRAMA DE FLUJOS 04- 03-2025 SPGR. ExportarPAGOS 04-03-2025.csv

Seguidamente cada una pasa por un proceso de ingesta, selección y limpieza, como se presenta a continuación:

SECOP

Para los archivos pertenecientes a esta categoría se lleva a cabo el mismo proceso, donde una vez sea importada se procede a imprimir el listado de las columnas de cada una, con el fin de seleccionar únicamente las de interés, siendo divididas de la siguiente manera

SECOP 1

- 'Anno Cargue SECOP',
- 'Nombre Entidad',
- 'NIT de la Entidad',
- 'Tipo De Contrato', 'Municipios Ejecucion',
- 'Numero de Contrato',
- 'Cuantia Proceso'.
- 'Cuantia Contrato',
- 'Valor Total de Adiciones',
- 'Valor Contrato con Adiciones'.
- 'Municipio Entidad',
- 'Departamento Entidad'

SECOP 2

- 'Entidad',
- 'Nit Entidad',
- 'Fecha de Publicacion del Proceso'

Teniendo presente que solo se necesita información del 2024 para cada una utilizamos las columnas; Anno Cargue SECOP y Fecha de Publicacion del Proceso para filtrar la información solo correspondiente a esa vigencia. Por último, se procede a guardar los datos de nuestro interés en las siguientes rutas:

"SECOP/df_filtrado_secop1_2024.csv" "SECOP/df_filtrado_secop2_2024.csv"

• <u>MATRIZ PROYECTOS</u>

Para este archivo en excel se realizo un proceso mas arduo, debido a que se presentaba muchos datos en formato "NaN" e incluso columnas vacías, una vez en listada las columnas, se seleccionan las siguiente para un posterior procesamiento:

- 'Estado Principal',
- 'Estado Derivados',
- 'Valor Aporte de la Entidad (\$)',
- 'N Proyecto / Contrato / Convenio / ',
- 'Entidad Contratante',
- 'Departamento',
- 'Año',
- 'Valor Aporte de la Entidad (\$).1',
- 'Porcentaje Aporte de la Entidad (%)',
- 'Valor Aporte de Fonpacífico (\$)',
- 'Porcentaje Aporte de Fonpacífico (%)',
- 'Entidad del Derivado'

Como se logra observar incluso los nombres de las columnas cuenta con un espacio antes o después del nombre lo que puede presentar errores y confusiones al momento de procesar, por lo cual se opta por renombrar dichas columnas más adelante, siguiendo con este proceso al igual que los .csv anteriores, este también se filtra por vigencia 2024 a través de la columna: 'Año' y se guarda de la siguiente forma: "MATRIZ_PROYECTOS/df_filtrado_matrizgestion proyectos_2024.csv"

Llevando a cabo el mismo proceso se realiza lo mismo para las plataformas de "*GESPROY*" y "*SPGR*", sin embargo, los datos proporcionados representan vigencia 2025, por lo cual, para este primer análisis, se opta por no tomar en cuenta dicha información, hasta que no se analice dicho año.

Finalmente, la ultima fuente que se procesa es la de SOUL, donde, se seleccionan las siguientes columnas para procesar:

FACTURACION

- 'codigo',
- 'fecha',
- 'subtotal',
- 'total'.
- 'saldo',
- 'nom_tercero',
- 'abonos',

RECAUDOS

- 'fecha',
- 'factura'.
- 'nombre',
- 'consignacion',
- 'deducciones'.
- 'nombre'

Y se almacenan en las siguientes rutas: "SOUL/df_filtrado_facturacion.csv" "SOUL/df_filtrado_recaudo.csv"

7. TRANSFORMACION

El principal objetivo de este primer análisis es concentrarse en el **flujo de caja de 2024**, lo cual requiere datos directamente relacionados con las transacciones financieras, pagos, aportes y facturación ocurridos en ese año. Los archivos provenientes de **SECOP**, aunque contienen información sobre los procesos de contratación pública, no tienen un impacto directo en el flujo de caja para el análisis de 2024, ya que los datos sobre contratos y procesos anteriores no reflejan transacciones o pagos realizados en dicho año.

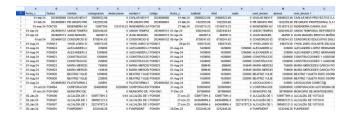
Por lo tanto, se decide **descartar el uso de SECOP en este primer acercamiento**, para evitar introducir datos

irrelevantes que puedan complicar el análisis y dificultar la identificación de los flujos de caja específicos de 2024. Esta decisión asegura que el análisis se enfoque de manera más directa en los aspectos financieros que son fundamentales para el análisis de flujo de caja.

Teniendo en cuenta lo anterior la transformación se llevará a cabo con las siguientes fuentes;

- "SOUL/df filtrado facturacion.csv"
- "SOUL/df_filtrado_recaudo.csv"
- "MATRIZ_PROYECTOS/df_filtrado_matrizgesti onproyectos_2024.csv"

Como primera etapa de la transformación se procede a realizar merge entre los dos DataFrames de facturación y recaudo; Esto a través de un inner mediante las llaves: "factura" y "codigo", obteniendo asi la siguiente tabla denominada: df_combinado_facturacion_recaudos.csv



Una vez realizado dicho cruce se hace el mismo procedimiento, pero ahora para incluir la matriz de proyectos, adquiriendo como nombre: df_combinado_facturacion_recaudos_matriz.csv

	A						101		1	4		L.			0	
3	Spirita_in	factors	norebre		debuccores	remore.f	cedige		NAME OF TAXABLE	total	14100	nors_tercero	abones		Estado Prevopo	Estado Derivad
	01-Aug-34	2024000082	PLATAFORMA I	245991		0 PLATAFORNIA	2024000362	01-Aug-24	245991	245991		D ASSOCIACION	245991	ASSOCIACION S	Ejecución	Ejecución
	03-Jan-24	540	ASSCIACION O	524300.93		G ASSOCIACION O	540	5 05-Jan-24	524300,93	\$24300.90		0 ASSOCIACION	524300,93	ASSOCIACION O	Decución	Ejeoución
4	55 par 24	2024000018	PLATAFORMA I	1792037		0 PLATAFORMA	2024000311	03-64-24	1792637	1790830		0 ASSOCIACION	0 1792031	ASSOCIACIÓN O	Ejecución	Ejeoución
	50-jun-24	2024000016	PLATAFORMA I	245991		0 PLATAFORMA	2024000316	03-94-24	245991	24500		0 ASSOCIACION	345991	ASOCIACIÓN O	Epecución	Ejeopolón
	00-jul-24	2024000317	PLATAFORMA I	3200000		O PLATAFORNIA	202400031	03-65-24	3200000	3200000		0 ASSOCIACION	2200000	ASSOCIACION (Djecución	Ejedyción
	06-jun-24	FON936	GP KANGURO	4000000		0 OF KANGURO	PONRIS	06-jun-04	4750000	4780000	7600	60 GP KANGURO	4000000	GP KANGURO	: Terminado	Terminado
	05-jun-24	PONS36	OF KANDURO	4000000		0 OF KANDURD	PONSS6	08-jun-24	4790000	4760000	7600	00 OF KANDURO	4000000	OF KANOURO	l Terminado Palta	Terminado pen
	05-jun-24	FONE36	GP KANGURO			g GP KANGURO	FON935	06-jun-24	4780000	4780000	7600	00 GP KANGURO	1 4000000	GP KANGURO	: Terminado	Terminado
12.	05-un-24	FOR636	OF KANGURO	4000000		0 OF KANGURO	FONS38	06-jun-24	4790000	4750000	7600	60 OF KANGURO	1 4000000	SP KANSURO	1 Terminado	Terrinade
*	06-km-24	FONE36	OF HANGURO	4000000		0 OF KANDURD	PONSSS	06-jun-24	4790000	4790000	7600	60 OF KANSURO	4000000	OF KANDURO	Terminado	Termiando
વ	06-jun-24	PON636	OF KANDURO	780000		0 OF KANGURD	FON836	06-jun-24	4790000	4790000	7600	00 OF KANSURO	4000000	SP KANSURO	(Terrinado	Terrinado
13.	05-jun-24	FCN936	GP KANGURO	760000		0 OF KANGURO	: FON036	06-jun-24	4780000	4750000	7600	00 GP KANGURO	1 4000000	OF KANGURO	Terminado Falta	Terminado pen
14.	06 un 24	FON836	OF HANGURO	700000		O OP KANGURO	FONE36	06-un-24	4700000	4760000	7600	60 OF KANGURO	4000000	OP KANGURO	Tanninado	Terminado
10	06-jun-24	FONS36	OF KANGURO	760000		0 OF KANDURD	PONESE	06-jun-24	4790000	4780000	Tecc	00 OF KANDURG	1 4000000	OF KANDURO	Terminado	Terminado
19	06-jun-24	FON836	GP KANGURO	780000		0 OF KANGURD	FON898	06-jun-24	4790000	4750000	7600	00 GP KANGURO	1 4000000	GP KANGURO	: Terminado	Termiando
127	07-may-24	2024000180	CONSCRICTO A	87236363		A CONSORCIO A	2024000190	29-Apr-24	57230363,9	\$7238363.0		0 CONSONCIO	97230363,6	CONSONCIO	Ejecución	Ejecución
-	08-Aug-24	2024000086	CONSCRORO	17485027		0 CONSORCIO A	2024000346	100 Aug 24	37570954,75	37570854.75	20085227	75 CONSORCIO	17485027	CONSORCIO	r Ejecución	Ejeosolón
10	10-34-24	2024000000	FUNDACION E	19847252.85		0 FUNDACION E	2024000300	25-jun-24	19847252.65	19547252.85		0 FUNDACION 6	19047252.05	FUNDACION E	* Terminado pend	Terminado pen
20	11-Apr-24	2024000175	PLATAFORMA I	245991		0 PLATAFORMA	2024000175	11-Apr-24	245991	24599		0 ASSOCIACION	245991	ASSOCIACION O	Ejecución	Ejeopolón
21	1247-24	PON831	CONSULTORA	20000000		0 CONSULTORA	PONS31	28-may-24	47467600.05	4749T000.00	200000	ee consultons	27467660.03	CONSULTORA	Terminado pend	Terminade pan
20	12-34-24	FON090	FUNDACIÓN M.	10022529,41		O FUNDACIÓN M	FONDE	1294-24	22400000	22400000	3876470.	SE FUNDACIÓN I	18823529.41	FUNDACIÓN V	Terminado	Terminado
21	12(4)-24	FON890	FUNDACIÓN M	3576470.59		0. FUNDACIÓN M	F01690	1246-26	22400000	22400000	3676470.	50 FUNDACIÓN I	18823529,41	FUNDACIÓN N	Terminado	Terrinado
24	12-Sep-24	2024000476	GF KANGURO	4400000		0 OF KANGURD	2024000479	12-Sep-24	4400000	4400000		D OF KANGURO	1 4400000	OF KANDURO	Terminado	Terminado
20	12-Sep-24	2024000476	OP KANGURO	4400000		0 OF KANGURO	2024000479	12-Sep-24	4400000	4400000		D OP HANDURG	4400000	OF KANGURO	(Terminado Fata	Terminade perv
24	12-Sep-24	2024000476	OF KANGURO	4400000		g of KANGURO	2024000475	12-Sep-24	4400000	4400000		0 OF KANGURO	1 4400000	OF KANDURO	Terminado	Terminado

				16	- 4	-				44	AM
Males Assessed also	M Supranta / Ca	Entided Cooked	Decompose	150	Males America de l	Summerica Anna	Official Assessed the	(Remettin Anna	Entided del De	alumate.	1777
\$ 151,920,000,	Convenio Intera	MUNICIPIO DE	QUINDIO	2024.0	\$ 151,920,000,0	95.24%	\$7.595.000,00	4,78%	GP KANGURO	SAS	
\$ 249.461.479,	CONVENIO INT	MUNICIPIO DE	QUINDIO	2024.0	\$ 249,481,479,0	89,39%	\$29,600,000,00	10,61%	GP KANGURO	SAS	
\$ 180,000,000)	CONV. INT. Nº 0	MUNICIPIO DE	QUINDIO	2024.0	\$ 180,000,000,0	94,74%	\$10,000,000,00	5,28%	GP KANGURO	SAS	
\$ 240,000,000,	Convenio Intera	MUNICIPIO DE	QUINDIO	2024.0	\$ 240,000,000,0	93,75%	\$16,000,000,00	6,25%	GP KANGURO	SAS	
\$ 240,000,000,	Convenio Intera	MUNICIPIO DE	QUINDIO	2024.0	\$ 240,000,000,0	90,91%	\$24,000,000,00	9,09%	GP KANGURO	SAS	
\$ 151.920.000.	Convenio Intera	MUNICIPIO DE	QUINDIO	2024.0	\$ 151,920,000,0	95,24%	\$7,596,000,00	4,76%	GP KANGURO	SAS	
\$ 249.401.479,	CONVENIO INT	MUNICIPIO DE	QUINDIO	2024.0	\$ 249,461,479,0	89,39%	\$29,600,000,00	10,61%	GP KANGURO	SAS	
\$ 3,066,060,90	EJECUTOR DE	MUNICIPIO DE	HULA	2024.0	\$ 3,066,068,906	100,00%	\$0.00	0.00%	CONSORCIO	AGUAS DE COLO	MBIA 2
\$ 3,066,068,90	EJECUTOR DE	MUNICIPIO DE	HULA	2024.0	\$ 3,066,068,906	100,00%	\$0.00	0.00%	CONSORCIO	AGUAS DE COLO	MBIA 2
\$ 1,385,109,36	EJECUTOR DE	MUNICIPIO LIT	сносо	2024.0	\$ 1,385,109,361	100.00%	\$0.00	0.00%	FUNDACION I	ENLACE COLOME	BIA
\$ 118,660,000,0	CONVENIO INT	MUNICIPIO DE	QUINDIO	2024.0	\$ 118,660,000,0	93,00%	\$8,036,000,00	0,34%	ASOCIACION	CONECTO	
\$ 2,178,654,10	EJECUTOR DE	MUNICIPIO DE	PUTUMAYO	2024.0	\$ 2,178,654,105	100,00%	\$0.00	0.00%	CONSULTOR	A Y CONSTRUCT	ORA CS S.A.S.
\$ 800,000,000	Convenio de As	GOBERNACIO	SAN ANDRES	2024.0	\$ 800,000,000,0	95.24%	\$40,000,000,00	4.78%	FUNDACIÓN I	MADANA BLUE	
				2024.0							
	\$ 118.000.000.1 \$ 118.000.000.1 \$ 118.000.000.1 \$ 118.000.000.1 \$ 118.000.000.1 \$ 118.000.000.1 \$ 118.000.000.1 \$ 128.000.000.1	VIEW INJURIES OF THE STATE OF T	View Agents of 1 M Proyects Cell Emission Contents 1 1148-0000 CONVENTION OF MANAGERO CE 2 1148-0000 CONVENTION OF MANAGERO 2 1148-0000 CONVENTION OF MANAGERO 3 1148-0000	6 THE WAY AND A	VIEW Auton to 11 Physics Code Distract Corress Description (Autonomous Section Code) (Autonomous	\(Value April 10 Physics Coll Emiss Control Description \(\text{Value April 10 Physics Coll Emiss Control Description \(\text{Value April 10 Physics Coll Emiss Control Description \(\text{Value April 10 Physics Coll Emiss Control Description \(\text{Value April 10 Physics Control Description \(\text{Value April 10 Ph	VIEW Auton in J. Physiatric Coll Edited Corress Description Adv. View June in Processing 2014 1989 (2014)	Value Agent Na I Prepatat Cali Effect Common Destruction 2044 2 14 16 2004 2 1 1 1 1 1 1 1 1 1	Value Applies March Projection Commission Com	Value Agent to 1 Projects Commission Section Province Pr	Value Agencies 1 Projectio Commission Profession Value Profession P

Luego de ello, y de eliminar las columnas que cuentan con la misma información, haciendo que la matriz sea redundante, se identifica que es necesarios los proyectos que se encuentren en estado de ejecución, por lo cual, haciendo uso de esta variable "Estado Principal" se trae únicamente los proyectos que correspondan;

		- 8	- 6		- 1			- 14			1 1 1 1 1 1	1.	68	N.	0	
1	feefra_x	facture	nombre	consignation	dedunctores	cedige	feetra_y	\$5(a)	14'00	attones	Estado Principal	Valor Aporte de	N Proyesto / C	or Emildad Contrata	Percentage Aper	Vision Aportie I
2	01-Aug-24	202400000	PLATAFORMA	245991		2024000082	01-Aug-34	245991		245991	Ejecución	\$ 112,660,000,	CONVENDOR	IT MUNICIPIO DE	92,00%	\$8,000,000,0
5	03-Jan-24	54	S ASOCIACION O	524300.93		545	03-Jan-24	524300.90		524300.93	Decución	\$ 110,550,000	CONVENCE IS	T MUNICIPIO DE	93,66%	\$8,006,000.0
4	03-64-24	202400001	PLATAFORMA	1792037	939	2024000315	03-jul-24	1790837	0	1792837	Ejesusión:	\$ 118,660,000	E CONTINUO I	T MUNICIPIO DE	93,66%	\$8,006,000.2
	03-jul-24	202400001	S FLATAFORMA S	245991		2024000016	03-jul-24	245991	. 0	245991	Ejecución	\$ 119,660,000	CONWENCE IN	IT MUNICIPIO DE	93,00%	\$8,036,000.0
4	03-jul-24	202400021	PLATAFORMA D	2200000		2024000317	93-66-24	3200000	. 0	2290000	Sjeoución	\$ 112,660,000.	COMMINO I	T MUNICIPIO DE	93,65%	52,006,000:0
1	07-may-04	202400018	A GONGORGIO S	57236363	0.6	2024000182	29-Apr-24	\$7236363.6	0	57236383.9	Ejecución	\$ 3.000 000.00	R EJECUTOR D	E MUNICIPIO DE	100,00%	93.00
٠	05-Aug-24	202400000	e CONSIDECIO A			2024000000	02-Aug-24	37570954.75	20088227.75		Ejecución	\$ 3.000.008.90	A BUBOUTON D	E MUNICIPIO DE	100,00%	93.00
2	11-Apr-24	202400017	E PLATAFORMA I	245991		2024000175	11-Apr-24	245991	0	245991	Ejesusión	\$ 118 860 000.	(COMPAND)	F MUNICIPIO DE	63,00%	\$8,000,000,0
12	15-Aug-24	202400000	E CONSCINCIO A	17435627		20240000006	02-Aug-24	27570054.75	20088227.75	17405627	Ejeoución	\$ 3,095,065,00	к вивоитом о	E MUNICIPIO DE	100,00%	92.00
11	18-000-24	202400063	B ASOCIACION O	171250		2024000636	09-nov-24	171250	. 0	171250	Epeciation	\$ 118,660,000.	CONVENOR	IT MUNICIPIO DE	83,00%	\$6,006,000
12	19-00-24	202400003	ASOCIACION O	901570,5		2024000638	09-nov-24	951570.5	0	901570,5	Ejecución :	\$ 118,060,000	I OWNERO IS	IT MUNICIPIO DE	93,00%	\$4,036,000.0
12	19-jun-04	202400000	CONSCROOR A	29192079.27		2024000290	19-jun-04	29192079.27		29102079;27	Ejecución	\$ 2,000,000.00	K FUEDUTOR D	E MUNICIPIO DE	100,00%	93.00
14	1944-24		D PLATAFORMA D		4	2024000360	1994-04	1100000	0		Ejeousión	\$ 118.660,000	CONVENIO I	IT MUNICIPIO DE	93,00%	98 036 000
15	20-nov-24	202400067	ASOCIACION C	1333750		2024000677	22-rev-24	1333750	0	1222750	Ejecución	\$ 118 660 000	CONFINO I	IT MUNICIPIO DE	93,00%	98,036,000
Ħ.	20-Dec-24	ATE155	CONSCRICTO	13178644,01	455354.84	ATE166	17-Dec-24	13535925.9		13636628,0	Ejecución	5 2, 193, 225, 54	EJECUTOR D	E MUNICIPIO DE	100,00%	93.00
17	22-Apr-24	202400017	PLATAFORMA	1933129		2024000176	22-Apr-24	1933125	0	1955125	Depution	\$ 118,660,000.	CONVENIO I	T MUNICIPIO DE	93,66%	58.036.000.0
18	22-pai-24	202400006	# PLATAFORMA S	500000		2024000084	22-jul-24	500000	. 0	600000	Ejecución	\$ 118,660,000.	CONVENO II	IT MUNICIPIO DE	93,00%	58,036,000.0
13	24-jun-24	202400029	1 ANGULARO CO	29143635,08		2024000291	25-jun-24	28142625.06		20140605,06	Ejecución	\$ 2,410,344,25	P EJECUTOR O	E MUNICIPIO DE	100,00%	90,00
20	24-009-24		CONSCRIGO		1004526,1	ATES!	09-aut-24		. 0	29004043,95		\$ 2,193,265,54	o soruciação	E MUNICIPIO DE	100,00%	90.00
81	25-Apr-04	202400018	PLATAFORMA D	1607200		2024000187	29-Apr-24	1607200		1007200	Ejecución	\$ 118,660,000	E CONTRACO I	IT MUNICIPIO DE	90,00%	54 036 000 0
#	25-Aug-24	202400043	E PLATAFORMA	150000		2024000455	28-Aug-24	150000	. 0	150000	Ejeousién.	\$ 118,860,000,	CONVENOR	IT MUNICIPIOSE	93,00%	\$8,036,000.0
22	28 Aug-24	202400043	PLATAFORMA S	2113359.7		2024000432	29 Aug-24	2113399.7	. 0	2113369.7	Esecución	\$ 118,660,000.	CONTINUO	IT MUNICIPIO DE	93,00%	98.036.000.0
24	25-Aug-24	202400043	1 PLATAFORMA I	940800		2024000431	25-Aup-24	\$60600	0.	860600	Ejesución	\$ 118,660,000.	CONVENIOR	IT MUNICIPIO DE	93,00%	88.036.000.0
24	27-Aug-04	FORGS	ANGULARO CO	24424007	849532.56	FON955	25-Aug-24	25273599.58	. 0	25272569.55	Decusión	5 2,410,344,25	O ROTUGBLE H	E MUNICIPIO DE	100.00%	92.00

Seguidamente, se re ordenan las columnas, con el fin de que sea más fácil su visualización e interpretación de datos interrelacionados



Una vez lista la información se verifica el tipo de formato en que se encuentran los datos numéricos; Los valores numéricos están almacenados como cadenas de texto (object) en lugar de números (int o float). Esto ocurre porque los valores incluyen símbolos de moneda, comas como separadores de miles y puntos como separadores decimales. Para realizar cálculos correctamente, es necesario convertirlos a números eliminando los caracteres no numéricos.

Como se oberva en cuanto a las columnas de valores numéricos, cada una cuenta con un formato diferente, por lo cual, se identifican cuales comparten el mismo, y se agrupan, como se muestra a continuación:

"Valor Aporte de la Entidad (\$)" y "Valor Aporte de Fonpacífico (\$)"

- Ambos valores incluyen el signo \$, tienen separador de miles con punto (.) y separador decimal con coma (,).
- Ejemplo: \$ 118.660.000,00 y \$8.036.000,00.

"total", "abonos", "consignacion", "deducciones"

- Estos valores parecen ser numéricos, pero algunos registros tienen coma como separador decimal (524300,93), mientras que otros solo tienen números enteros (245991).
- En algunos casos, los valores no tienen separadores de miles, lo que indica que hay inconsistencias en el formato.
- Ejemplo: 245991, 524300,93, 1792837.

"saldo"

 En la muestra proporcionada, todos los valores de la columna son 0, lo que sugiere que podrían ser enteros o ser tratados como string. Para eliminar los formatos iniciales, se aplica el siguiente procedimiento;

Para estandarizar los valores de "Valor Aporte de la Entidad (\$)" y "Valor Aporte de Fonpacífico (\$)", eliminaremos el formato monetario, es decir:

- 1. Se suprime el símbolo \$.
- 2. Eliminaremos los puntos que separan los miles.
- 3. Reemplazaremos la coma decimal por un punto (para convertirlo en un formato numérico estándar).

```
    Limpieza de montos:
        Valor Aporte de la Entidad ($) Valor Aporte de Fonpacífico ($)
    118660000.0 8036000.0
    118660000.0 8036000.0
    118660000.0 8036000.0
    118660000.0 8036000.0
    118660000.0 8036000.0
```

Para el segundo grupo de columnas: "total", "abonos", "consignacion", "deducciones", vamos a:

- 1. Eliminar los separadores de miles (si existen).
- Reemplazar la coma decimal por un punto para mantener un formato numérico estándar.
- 3. Convertir los valores a tipo float.
- Limpieza de números: total abonos consignacion deducciones 245991.00 245991.00 245991.00 0.0 1 524300.93 524300.93 524300.93 0.0 1792837.00 1792837.00 1792837.00 0.0 2 245991.00 245991.00 0.0 3 245991.00 3200000.00 3200000.00 3200000.00 0.0

Para garantizar que la columna "saldo" tenga el mismo formato que los demás valores numéricos (dos decimales), aplicaremos una limpieza específica asegurando que todos los valores sean tratados como float con dos decimales.

```
• Limpieza de saldo:
0  0.0
1  0.0
2  0.0
3  0.0
4  0.0
Name: saldo, dtype: float64
```

El proceso de limpieza de datos aplicado a la matriz de facturación y recaudos consistió en la sustitución de valores incorrectos o mal formateados en las columnas monetarias y numéricas, asegurando la coherencia y precisión de la información. Para ello, se creó una copia del DataFrame original con el fin de preservar la estructura de los datos y evitar modificaciones no deseadas. Posteriormente, se reemplazaron las columnas monetarias con montos corregidos, las columnas numéricas con valores limpios y la columna de saldo con datos ajustados. Finalmente, el DataFrame procesado se exportó en formato CSV con codificación UTF-8 y sin incluir índices, almacenándose en un repositorio en Google Drive.

Valor Aporte de la E	Porcentaje Aporte di	Valor Aporto de Fon	Porcentaje Aporte d	nombre	codigo	fecha_y	total	saldo	abonos	factura
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000382	01-Aug-24	245991	0	245991	2024000382
118660000	93,66%	8036000	6,34%	ASOCIACION CONE	545	03-Jan-24	524300.93	0	524300.93	545
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000315	03-Jul-24	1792837	0	1792837	2024000315
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000316	03-Jul-24	245991	0	245991	2024000316
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000317	03-Jul-24	3200000	0	3200000	2024000317
3066068906	100,00%	0	0,00%	CONSORCIO AGUA	2024000192	29-Apr-24	57238383.9	0	57238383.9	2024000192
3066068906	100,00%	0	0,00%	CONSORCIO AGUA:	2024000396	02-Aug-24	37570854.75	20085227.75	17485627	2024000396
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000175	11-Apr-24	245991	0	245991	2024000175
3066068906	100,00%	0	0,00%	CONSORCIO AGUA:	2024000396	02-Aug-24	37570854.75	20085227.75	17485627	2024000396
118660000	93,66%	8036000	6,34%	ASSCIACION CONE	2024000636	08-Nov-24	171250	0	171250	2024000636
118660000	93,66%	8036000	6,34%	ASOCIACION CONE	2024000639	08-Nov-24	961570.5	0	961570.5	2024000639
3066068906	100,00%	0	0,00%	CONSORCIO AGUA:	2024000280	18-Jun-24	29192079.27	0	29192079.27	2024000280
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000360	19-Jul-24	1100000	0	1100000	2024000360
118660000	93.66%	8036000	6,34%	ASSCIACION CONE	2024000677	22-Nov-24	1333750	0	1333750	2024000677
2193285542	100,00%	0	0,00%	CONSORCIO CUBIE	ATE158	17-Dec-24	13636928.9	0	13636928.9	ATE158
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000176	22-Apr-24	1933125	0	1933125	2024000176
118660000	93,66%	8036000	6,34%	PLATAFORMA DE C	2024000364	22-Jul-24	500000	0	500000	2024000364
2410344254	100,00%	0	0,00%	ANGULARQ CONST	2024000291	20-Jun-24	28143635.06	0	28143635.06	2024000291
2193285542	100,00%	0	0,00%	CONSORCIO CUBIE	ATE61	09-011-24	29884948.96	0	29884948.96	ATE61

Luego de ello renombramos las columnas con el fin de evitar el error mencionado previamente y ayudar a una mejor interpretación de los datos presentes

				4		- 10	- 4	1.0	2	K:	1	- 12	- 1	0	
O PROYECTO	ESTADO	ENTIDAD CONT	TVALOR APORTS	PORCENTAJE A	VALOR APOR	KTI PORCENTAJE /	ENTIDAD DEL CO	CODICO PACTI	FECHA FACT	UFVALOR FACT	UT VALOR EN SA	L VALOR DE AS	O CODIOO RECA	PECHA DE R	BC VALOR CONS
ON/ENO INT	Ejecución	MUNICIPIO DE	:118860000.0	\$3,06%	8038000.0	6,34%	PLATAFORMA.E.	2024000382	D1-Rug-24	245991.0	0.0	245991.0	2024000382	01-Aug-24	245991.0
COMMENO INT	Ejecución	MUNICIPIO DE	112660000.0	P3,66%	8036000.0	5,34%	ASSOCIACION C	545	03-Jan-24	524300.95	0.0	624300.93	545	03-Jan-24	524300.93
CONVENIO INT	Ejecution	MUNICIPIODE	110660000.0	P3.66%	0.0000000	6,34%	PLATHFORMAT	2004000315	03-pp	24 1792637.0	0.0	1792837.0	2024000315	93-66-	24 1792537.0
CONVENO INT	Ejecución	MUNICIPIO DE	118660000.0	P3.98%	0.0000008	6.34%	PLATAFORMAC	2024000316		24 245991.0	0.0	245991.0	2024000316	03-jul-	24 245991.0
ON/ENO IN	Ejecución	MUNICIPIODE	:1196600000:0	89,00%	8098000.0	0.34%	PLATAFORMA.E	2024000317	03-jui	24 3200000.0	0.0	2200000.0	2024000317	93-54	24 3200000.0
LIECUTOR DE	Ejecución	MUNICIPIODE	3066060906.0	100,00%	0.0	0,00%	CONSCROON	2024000192	29-Apr-24	57230303.9	0.0	57230303.9	2024000192	CT-may-	24 57239383.0
SUBDUTOR DE	Ejecución	MUNICIPIO DE	0.80908080808.0	100.00%	0.0	0.00%	CONSORCIO AI	2024000396	00-Aug-24	37970894.75	20086227.75	17405627.0	2024000398	06-Aug-24	17485627.0
COMENOIN	Ejecoción	MUNICIPIO DE	118660000.0	\$3,00%	8038000.0	6.34%	PLATAFORMA E	2024000176	11-Apr-24	245991.0	0.0	245991.0	2024000175	11-Apr-24	245991.0
LIECUTOR DE	Ejecución	MUNICIPIODE	2000000000000	100,00%	0.0	0,00%	CONSORCIO AI	2024000396	50-Aug-24	27570954.75	20095227.75	17405627.0	2024000398	15-Aug-24	17495627.0
CONVENIO INT	Securite	MUNICIPIO DE	119660000.0	P2.66%	6006000.0	6,34%	ASSOCIACION C	2024000000	06-nev	24 171250.0	0.0	171250.0	2024000838	10-00	24 171250.0
OWENO IN	Ejecución	MUNICIPIO DE	1118660000.0	\$3.00%	80080000.0	6,34%	ASOCIACION C	2024000639	08-nov-	24 991570.5	0.0	961570.6	2024000839	15-001-	24 961570.5
SUSCUTOR DE	Ejecución	MUNICIPIO DE	2066069905.0	100,00%	0.0	0,00%	CONSORCIO A/	2024000200	15-jun-	24 29192079:27	0.0	29192079.27	2024000280	19-541-	24 29192279.27
OWENO IN	Decusion	MUNICIPIO DE	118660000.0	P3.66%	0.0000003	6,34%	PLATAFORMA I	2024000360	1990	24 1100000.0	0.0	1100000.0	2024000393	19/44	24 0.0
ON/ENO INT	Ejecución	MUNICIPIO DE	118660000.0	P3.99%	8038000.0	6.34%	ASOCIACION C	2024000677	22-nov-	24 1333760.0	0.0	1333790.0	2024000677	25-nov-	24 1333760.0
SUBDUTOR DE	Ejecución	MUNICIPIO DE	12193285542.0	100,00%	0.0	0,00%	CONSORCIO C.	CTE 158	17-Dec-24	13030928.9	0.0	13030928.9	ATE159	20-Dec-24	13178544.01
COMPNO IN	Ejecución	MUNICIPIO DE	110660000.0	R2,00%	00080000	0,34%	PLATAFORMA E	2024000176	22-4pr-24	1933125.0	0.0	1933125.0	2024000175	22-Apr-24	1933125.0
CONVENIO INT	Ejecución	MUNICIPIO DE	118660000.0	93.00%	8036000.0	0.34%	PLATAFORMA C	2024000094	22-05-	24 900000.0	0.0	500000.0	2024000384	22-04	24 900000.0
SJECUTOR DE	Ejecución	MUNICIPIO DE	12410344254.0	100.00%	0.0	0.00%	ANGULARQ CC	2024000291	20 mm	24 28143635.00	0.0	28143635.00	2024000291	24-50-	24 28143635.00
SUBCUTOR DE	Ejecupión	MUNICIPIO DE	2192225542.0	100,00%	0.0	0,00%	CONSIGNOO	KTES1	29-oct-	24 29204940.95	0.0	29004943.90	ATER!	24-005	24 20000412.05
CONVENIO INT	Decusion	MUNICIPIO DE	110660000.0	\$3,00%	8036000.0	6,34%	PLATAFORMA C	2004000107	29-Apr-24	1907200.0	0.0	1607200.0	2024000187	25-Apr-24	1927200.0
CONVENIO INT	Ejecución	MUNICIPIO DE	118660000.0	63,00%	6008000.0	6,34%	PLATAFORMA C	2024000433	26-Aug-24	150000 D	0.0	160000.0	2024000433	26-Aug-24	150000.0
ON/ENO IN	Fjecución	MUNICIPIODE	119660000.0	92,00%	90080000	5,34%	PLATAFORMA I	2024000422	25-Aug-24	2112258.7	0.0	2113359.7	2024000432	26-Aug-24	2113359.7
OWENO IN	Decusio	MUNICIPIO DE	110660000.0	\$0.00%	0.0000000	6,74%	PLATAPORMAC	2024000431	25-Aug-24	980900.0	0.0	980900.0	2024000431	26-Aug-24	980800.0
JECUTOR DE		MUNICIPIO DE	2410344254.0	100.00%	0.0	0.00%	ANDULAND CO		26-Aug-24	29273599.56	0.0	25273599 56		27-Aug-24	24424067.0

El proceso de tratamiento de datos en la matriz de facturación y recaudos consistió en la carga de un archivo CSV previamente renombrado, seguido de la eliminación de columnas irrelevantes para el análisis. Posteriormente, se aplicó un formato monetario a las columnas financieras para garantizar la legibilidad de los valores numéricos. Se generó una copia del DataFrame original para evitar modificaciones sobre los datos de origen y se transformaron los valores en las columnas correspondientes mediante una función que les asignó el formato de moneda con separación de miles y dos decimales. Finalmente, los datos formateados se almacenaron en un nuevo archivo CSV y en un archivo Excel, donde además se aplicó alineación a la derecha y un formato específico de moneda en las celdas numéricas.

D	E		0							N N			
VALOR APORTS	PORCENTAJE /	VALOR APORT	PORCENTAJE A	ENTIDAD DEL I	CODIGO FACTI	FECHA FACTU	VALOR FACTU	VALOR EN SAL	VALOR DE ABO	CODIGO RECA	FECHA DE REC	VALOR CONSIG	VALOR DEDL
118,660,000.00	93,66%	\$8,036,000.00	6,34%	PLATAFORMA E	2024000382	01-Aug-24	3245,991.00	80.00	\$245,991.00	2024000382	01-Aug-24	9245,991.00	\$0.00
9118,660,000.00	93,00%	\$8,096,000.00	0.34%	ASOCIACION C	545	03-Jan-24	\$524,300.93	90.00	\$524,300.93				90.00
118,660,000.00	93,00%	\$8,036,000.00	6,34%	PLATAFORMA D	2024000315	03-jul-24	\$1,792,837.00	30.00	\$1,792,837.00	2024000315	03-jul-24	\$1,792,837.00	\$0.00
3118,660,000.00	93,66%	\$8,638,000.00	6,34%	PLATAFORMA (2024000316	03-jul-24	\$245,991.00	80.00	\$245,991,00	2024000316	03-jul-24	\$245,991.00	50.00
8118,660,000.00	93,66%	\$8,036,000.00	6.34%	PLATAFORMA (2024000317	03-jul-24	\$3,200,000.00	90.00	\$3,200,000.00	2024000317	03-jul-24	\$3,200,000.00	90.00
33,066,068,906	100,00%	\$0.00	0,00%	CONSORCIO A	2024000192	29-Apr-24	\$57,238,383.90	90.00	\$57,238,383.90	2024000192	07-may-24	\$57,238,383.00	90.90
33,066,068,906	100,00%	50.00	0,00%	CONSORCIO A	2024000396	02-Aug-24	\$37,570,854.75	820,085,227.76	\$17,485,627.00	2024000396	08-Aug-24	\$17,485,627.00	50.00
8118,660,000.00	93,00%	\$8,038,000.00	6,34%	PLATAFORMA (2024000175	11-Apr-24	\$245,991.00	90.00	\$245,991.00	2024000175	11-Apr-24	\$245,991.00	90.00
33,066,068,906	100,00%	90.00	0,00%	CONSORCIO A	2024000398	02-Aug-24	\$37,570,854.75	820,005,227.75	\$17,485,627.00	2024000396	15-Aug-24	\$17,495,627.00	90.00
\$118,660,000.00	93,00%	\$8,038,000.00	5,34%	ASOCIACION C	2024000636	08-nov-24	\$171,250.00	\$0.00	\$171,250,00	2024000636	19-oct-24	\$171,250.00	\$0.00
118,980,000.00	93,00%	\$8,035,000.00	6,34%	ASOCIACION C	2024000639	08-nov-24	9961,570.50	90.00	\$961,570.50	2024000639	18-oct-24	9961,570.50	50.00
33,066,066,906.	100,00%	90.00	0,00%	CONSORCIO A	2024000280	18-jun-24	929, 192,079,27	80.00	\$29,192,079.27	2024000280	19-jun-24	\$29,192,079,27	90.00
9118,660,000.00	93,66%	\$8,038,000.00	8,34%	PLATAFORMA [2024000380	19-jul-24	\$1,100,000.00	80.00	\$1,100,000:00	2024000380	19-jul-24	\$0.00	90.00
8118,660,000.00	93,00%	\$8,035,000.00	0,34%	ASOCIACION C	2024000677	22-nov-24	\$1,333,750.00	\$0.00	\$1,333,750.00	2024000677	20-nov-24	\$1,333,750.00	90.00
12,193,285,542	100,00%	50.00	0,00%	CONSORCIO C	ATE158	17-Dec-24	\$13,636,926.90	\$0.00	\$13,630,928.90	ATE158	20-Dec-24	\$13,178,544.01	\$450,384.89
118,660,000.00	93,00%	\$8,038,000.00	6.34%	PLATAFORMA [2024000176	22-Apr-24	\$1,933,125.00	80.00	\$1,933,125.00	2024000178	22-Apr-24	51,933,125.00	\$0.00
9118,980,000.00	93,00%	\$8,038,000.00	0.34%	PLATAFORMA (2024000364	22-jul-24	9500,000.00	80.00	9500,000,00	2024000364	22-jul-24	9500,000.00	90.00
12,410,344,254	100,00%	90.00	0,00%	ANGULARG CO	2024000291	20-jun-24	928,143,635,06	90.00	\$28,143,635.06	2024000291	24-jun-24	\$26,143,635.06	\$0.00
12,193,285,542	100,00%	50.00	0,00%	CONSORCIO C	ATES1	09-oct-24	\$29,884,948.06	90.00	\$29,884,948.98	ATE61	24-oct-24	\$26,880,412,86	81,004,536.10
9118,000,000.00	93,00%	\$8,036,000,00	0.34%	PLATAFORMA (2024000187	29-Apr-24	\$1,607,200.00	90.00	\$1,607,200.00	2024000187	25-Apr-24	\$1,607,200.00	90.00
118,660,000.00	93,00%	\$8,036,000.00	0.34%	PLATAFORMA (2024000433	26-Aug-24	\$150,000.00	80.00	\$150,000.00	2024000433	20-Aug-24	\$150,000.00	90.00
8118,660,000.00	93,00%	\$8,035,000.00	6,34%	PLATAFORMA (2024000432	26-Aug-24	92,113.359.70	\$0.00	\$2,113,359.70	2024000432	28-Aug-24	\$2,113,359.70	\$0.00
118,660,000.00	93,00%	\$8,038,000.00	6,34%	PLATAFORMA I	2024000431	26-Aug-24	9960,800.00	90.00	\$980,800.00	2024000431	26-Aug-24	9980,800.00	\$0.00
12,410,344,254	100,00%	90.00	0.00%	ANGULARO CO	FON056	28-Aug-24	\$25,273,500.56	80.00	\$25,273,599.56	FON056	27-Aug-24	\$24,424,057,00	\$849,532.56
32,410,344,254	100,00%	50.00	0.00%	ANGULARO CO	ATE3	12-Sep-24	324,657,383,25	80.00	\$24,657,383,25	ATE3	28-oot-24	\$24,657,383.00	50.25

Como se logra observar ya las columnas;

- 'VALOR APORTE DE LA ENTIDAD',
- 'VALOR APORTE FONPACIFICO'.
- 'VALOR FACTURACION',
- 'VALOR EN SALDO',
- 'VALOR DE ABONOS',
- 'VALOR CONSIGNADO'.
- 'VALOR DEDUCCIONES'

Presentan el mismo formato, lo que permite poder Generar nuevas columnas basadas en los datos existentes, con cálculos matemáticos;

En este fragmento realiza varios cálculos sobre los datos de facturación y recaudos, organizándolos en un resumen mensual. Primero, convierte la columna de fecha a un formato de año y mes (%Y-%m) para agrupar los datos por período. Luego, calcula tres métricas clave para cada mes: total facturado (suma de VALOR FACTURACION), total recaudado (suma de VALOR CONSIGNADO) y saldo pendiente (suma de VALOR EN SALDO). Después de obtener estos valores agregados, se les aplica un formato de moneda para mejorar la legibilidad. Finalmente, el resumen mensual se guarda en archivos CSV y Excel, aplicando formato numérico y alineación a la derecha en las celdas monetarias del archivo Excel para una mejor presentación de los datos.

MES	TOTAL_FACTURAD	TOTAL_RECAUDAD	SALDO_PENDIENT
2024-01	\$524,300.93	\$524,300.93	\$0.00
2024-04	\$61,024,699.90	\$61,024,699.00	\$0.00
2024-06	\$57,335,714.33	\$57,335,714.33	\$0.00
2024-07	\$7,084,819.00	\$5,984,819.00	\$0.00
2024-08	\$104,651,450.76	\$63,631,462.70	\$40,170,455.50
2024-09	\$24,657,383.25	\$24,657,383.00	\$0.00
2024-10	\$29,884,948.96	\$28,880,412.86	\$0.00
2024-11	\$2,958,551.50	\$2,958,551.50	\$0.00
2024-12	\$13,636,928.90	\$13,178,544.01	\$0.00

8. VALIDACION DE DATOS

Para validar la correcta generación del archivo de resumen mensual de facturación, se utilizó un procedimiento de carga y verificación de datos en un entorno de Python con la librería Pandas. Primero, se definió la ruta del archivo CSV almacenado en Google Drive y se montó el entorno para permitir su acceso. Posteriormente, el archivo fue leído utilizando pd.read_csv(), transformándolo en un DataFrame para su manipulación y análisis. Finalmente, se visualizaron las primeras filas del conjunto de datos mediante df_resumen.head(), lo que permitió verificar la estructura y coherencia de la información. Esta validación asegura que el archivo se generó correctamente, sin errores en la conversión ni en la integridad de los datos.

	MES	TOTAL_FACTURADO	TOTAL_RECAUDADO	SALDO_PENDIENTE
0	2024-01	\$524,300.93	\$524,300.93	\$0.00
1	2024-04	\$61,024,699.90	\$61,024,699.00	\$0.00
2	2024-06	\$57,335,714.33	\$57,335,714.33	\$0.00
3	2024-07	\$7,084,819.00	\$5,984,819.00	\$0.00
4	2024-08	\$104,651,450.76	\$63,631,462.70	\$40,170,455.50

Este fragmento de código realiza una validación de la calidad de los datos en el DataFrame df_resumen. Primero, utiliza df_resumen.isnull().sum() para contar y mostrar la cantidad de valores nulos en cada columna, lo que permite identificar posibles inconsistencias o datos faltantes. Luego, con df_resumen.dtypes, se verifica el tipo de dato de cada columna para asegurarse de que los valores tienen el formato esperado (por ejemplo, números en columnas monetarias y fechas en formato datetime).

```
Valores nulos por columna:
                           Tipos de datos de cada columna:
                   0
                             MES
                                                object
TOTAL FACTURADO
                   0
                             TOTAL FACTURADO
                                                object
TOTAL_RECAUDADO
                   0
                             TOTAL_RECAUDADO
                                                object
SALDO PENDIENTE
                   0
                             SALDO_PENDIENTE
                                                object
dtype: int64
                             dtype: object
```

En este fragmento se implementa validaciones adicionales para verificar la consistencia de los datos en el resumen mensual. Primero, df_resumen["MES"].unique() muestra los valores únicos en la columna "MES" para asegurarse de que los datos están correctamente agrupados por mes y no contienen valores atípicos. Luego, se recalculan las sumas "TOTAL_FACTURADO", totales de "TOTAL_RECAUDADO" y "SALDO_PENDIENTE" en el resumen mensual y se comparan con los valores originales dataset cargado $df_combinado_facturacion_recaudos_matriz_renombrado.$ csv. Esto permite detectar discrepancias y validar que no haya errores en el procesamiento de datos.

```
Valores únicos en la columna MES:
['2024-01' '2024-04' '2024-06' '2024-07' '2024-08' '2024-09' '2024-10'
  2024-11' '2024-12']
Suma de valores en el resumen mensual:
TOTAL_FACTURADO
                  $524,300.93$61,024,699.90$57,335,714.33$7,084,...
TOTAL RECAUDADO
                   $524,300.93$61,024,699.00$57,335,714.33$5,984,...
SALDO PENDIENTE
                   $0.00$0.00$0.00$0.00$40,170,455.50$0.00$0.00$0...
dtype: object
Suma de valores en el dataset original:
VALOR FACTURACION
                     3.017588e+08
VALOR CONSIGNADO
                     2.581759e+08
                     4.017046e+07
VALOR EN SALDO
dtype: float64
```

Este fragmento de código tiene como objetivo eliminar el formato de moneda de las columnas financieras y convertirlas a valores numéricos para facilitar cálculos posteriores. La función limpiar_moneda(valor) elimina los símbolos de dólar y las comas en caso de que los valores sean cadenas de texto, convirtiéndolos en números flotantes. Luego, se aplica esta función a las columnas "TOTAL_FACTURADO", "TOTAL_RECAUDADO" y "SALDO_PENDIENTE" para asegurarse de que los datos sean de tipo numérico. Finalmente, df_resumen.dtypes permite verificar que la conversión se haya realizado correctamente.

```
MES object
TOTAL_FACTURADO float64
TOTAL_RECAUDADO float64
SALDO_PENDIENTE float64
dtype: object
```