NILSON JACOB GONZALEZ CAMPO- 22500452
YAN CARLOS CUARAN IMBACUAN - 22502591
LORENA PORTILLA DIAZ – 22500248

Desarrollo de una Solución ETL para la Gestión de Precios de Medicamentos en Colombia en 2025

Descripción del problema:

La gestión y el análisis de los precios de medicamentos en Colombia es un desafío crítico debido a la variabilidad y dispersión de los datos disponibles. Los precios de los medicamentos varían significativamente dependiendo del canal de distribución (institucional vs. comercial). Este hecho complica aún más el análisis, ya que los datos no solo están dispersos en diversas fuentes, sino que también presentan inconsistencias debido a la fuente de los datos y las diferencias entre los canales de distribución.

Los precios en el sector institucional suelen estar regulados o sujetos a acuerdos especiales, mientras que los precios en el canal comercial son generalmente más altos debido a márgenes comerciales y demanda en el mercado privado. Esta disparidad crea una barrera para obtener una visión clara de los precios de los medicamentos y dificulta la toma de decisiones por parte de los consumidores para comparar precios y obtener el mejor valor.

A pesar de la importancia de este análisis, no existe una solución centralizada que permita agrupar, transformar y almacenar eficientemente los datos de precios de medicamentos en Colombia. Las partes interesadas, como las autoridades de salud, fabricantes de medicamentos, y consumidores, se ven limitadas al no poder comparar precios y analizar patrones entre los canales, lo que afecta tanto la transparencia como la competitividad del mercado.

Justificación:

Para abordar la variabilidad en los precios y comprender cómo fluctúan en los diferentes canales de distribución, es necesario contar con una solución ETL que centralice, limpie, y estandarice los datos de los precios de los medicamentos. Esta solución no solo permitiría realizar comparaciones claras entre los precios en el canal institucional y el comercial, sino que también facilitaría la identificación de patrones y tendencias de precios a lo largo del tiempo y entre distribuidores.

La centralización de estos datos permitiría a las entidades sanitarias tomar decisiones informadas sobre la regulación de precios y el acceso a medicamentos, mientras que los consumidores podrían hacer elecciones más informadas, buscando los mejores precios disponibles para los medicamentos. Además, el análisis y la estandarización de los datos de precios mejorarían la competitividad y garantizarían un acceso más equitativo a los tratamientos para la población.

Objetivo General:

Desarrollar una solución ETL para centralizar, transformar y estandarizar los precios de los medicamentos en Colombia para mejorar la toma de decisiones informadas en el sector de la salud y por parte de los consumidores.

Objetivos Específicos:

- Identificar y seleccionar las fuentes de datos más relevantes para el análisis de precios de medicamentos en Colombia.
- Desarrollar el proceso ETL para integrar y limpiar los datos, eliminando inconsistencias y valores atípicos.
- Realizar un análisis exploratorio de los precios de los medicamentos, identificando patrones y diferencias entre los canales institucional y comercial.
- Generar un conjunto de datos estandarizado y centralizado para su posterior análisis y comparación de precios.
- Proporcionar recomendaciones basadas en los datos para mejorar la transparencia de precios y la competitividad en el mercado farmacéutico colombiano.

Análisis de los Datasets (N° Dataset):

1. **Clicsalud - Termómetro de Precios de Medicamentos (2024):**
    - **Descripción**: Este dataset contiene información relevante sobre los precios de medicamentos en 2024.
    - **Cumple con**: Las variables de precios están correctamente documentadas, lo que lo convierte en una fuente útil para el análisis de precios.
2. **Clicsalud - Termómetro de Precios de Medicamentos (2025 Actualización):**
    - **Descripción**: Similar al anterior, pero con informacion actualizada correspondientes al año 2024.
    - **Cumple con**: Contiene tanto las variables de precios como las de tipos de medicamentos, por lo que es útil para comparar precios entre diferentes tipos de medicamentos.
3. **Droguerías con expendio de medicamentos de control especial:**
    - **Descripción**: Contiene información sobre las droguerías con expendio de medicamentos de control especial.
    - **No cumple con**: Aunque el dataset proporciona información sobre la razón social de las droguerías y su ubicación, no contiene ninguna variable relacionada con los precios de los medicamentos, que es la variable clave para este análisis. Por lo tanto, se descarta.
4. **Medicamentos POS:**
    - **Descripción**: Proporciona información sobre los medicamentos en el POS (Plan Obligatorio de Salud).

- **No cumple con**: Aunque cuenta con varias variables importantes, como los códigos de medicamentos vigentes, no incluye la variable de precios, lo cual es esencial para este estudio. Por lo tanto, se descarta.

5. **Medicamentos vitales no disponibles:**

   - **Descripción**: Contiene datos sobre medicamentos que no están disponibles en el mercado.
   - **No cumple con**: Este dataset no tiene información sobre los precios de los medicamentos y se centra más en los lotes de medicamentos no disponibles. Por lo tanto, no es útil para este análisis.

6. **Listado de medicamentos en venta libre:**

   - **Descripción**: Incluye información sobre medicamentos de venta libre.
   - **No cumple con**: Solo proporciona detalles descriptivos de los medicamentos, como los compuestos, pero no contiene información sobre precios, lo cual lo hace irrelevante para el análisis.

7. **Códigos únicos de medicamentos para los registros sanitarios vigentes:**

   - **Descripción**: Presenta información sobre medicamentos registrados y vigentes en Colombia.
   - **No cumple con**: Aunque es un dataset detallado con características como fechas de vencimiento y estado de los medicamentos (activos/inactivos), no contiene información sobre los precios, por lo que no es útil para este estudio.

8. **Códigos únicos de medicamentos para los registros sanitarios vencidos:**

   - **Descripción**: Incluye datos sobre medicamentos que han vencido su registro sanitario.
   - **No cumple con**: Similar al dataset anterior, no contiene información sobre precios, por lo que se descarta.

9. **Códigos únicos de medicamentos para los registros sanitarios en trámite de renovación:**

   - **Descripción**: Información sobre medicamentos que están en proceso de renovación de su registro sanitario.
   - **No cumple con**: No contiene variables de precios, lo cual lo hace inadecuado para el análisis de precios.

10. **Droguerías con expendio de medicamentos de control especial (Autorizadas):**

    - **Descripción**: Similar al dataset 3, pero con mayor detalle sobre la autorización de las droguerías.
    - **No cumple con**: Este dataset solo proporciona información sobre las droguerías y su ubicación. No contiene datos sobre los precios de los medicamentos ni sobre los tipos de medicamentos, por lo que se descarta.

11. **Clicsalud - Termómetro de Precios de Medicamentos (2024):**
    - **Descripción**: Proporciona una comparación de precios de medicamentos en el mercado en 2024.
    - **No cumple con**: Contiene información detallada sobre los precios de los medicamentos, sin embargo, no contiene variables como el canal de distribución.

12. **Regulación de precios de medicamentos:**
    - **Descripción**: Este dataset contiene información sobre la facturación de productos, pero no está bien regulado por INVIMA.
    - **No cumple con**: No proporciona información directa sobre los tipos de medicamentos ni sus precios, ya que los productos están definidos solo por un ID. Además, la fiabilidad de los datos es dudosa, por lo que no es adecuado para este análisis.

13. **Registros sanitarios de medicamentos homeopáticos:**
    - **Descripción**: Información sobre el registro sanitario de medicamentos homeopáticos.
    - **No cumple con**: No incluye variables de precios ni valores unitarios, por lo que no es relevante para el análisis de precios de medicamentos.

14. **Consulta pública de Precios de Medicamentos:**
    - **Descripción**: Proporciona una consulta pública sobre los precios de los medicamentos.
    - **No Cumple con**: Contiene tanto el tipo de medicamento como su respectivo precio, sin embargo, no contiene variables como el canal de distribución.

15. **Medicamentos (Comparativa de precios):**
    - **Descripción**: Permite comparar los precios equivalentes por tableta o cápsula de diferentes alternativas de un mismo medicamento.
    - **Cumple con**: Las variables de precios están correctamente documentadas, lo que lo convierte en una fuente útil para el análisis de precios.

16. **Medicamentos (Detalles y Estado):**
    - **Descripción**: Este dataset contiene información detallada sobre los medicamentos, como fechas de vencimiento y si están activos o inactivos.
    - **No cumple con**: Aunque tiene información detallada sobre el estado de los medicamentos, no contiene datos sobre los precios, lo que lo hace innecesario para este análisis.

Los datasets más relevantes para este análisis son:

- Clicsalud - Termómetro de Precios de Medicamentos (2024): Proporcionan información sobre precios y tipos de medicamentos.
- Clicsalud - Termómetro de Precios de Medicamentos (2025 Actualización): Proporcionan información sobre precios y tipos de medicamentos.
- Medicamentos (Comparativa de precios): Ofrece datos útiles para comparar precios entre diferentes opciones de medicamentos.

Los siguientes datasets se descartan debido a la falta de información sobre precios de medicamentos o su irrelevancia para el análisis:

- Droguerías con expendio de medicamentos de control especial
- Medicamentos POS
- Medicamentos vitales no disponibles
- Listado de medicamentos en venta libre
- Consulta pública de Precios de Medicamentos
- Códigos únicos de medicamentos (Vigentes, Vencidos, Renovación)
- Regulación de precios de medicamentos
- Registros sanitarios de medicamentos homeopáticos
- Medicamentos (Detalles y Estado)

1. Diagrama de flujo ETL

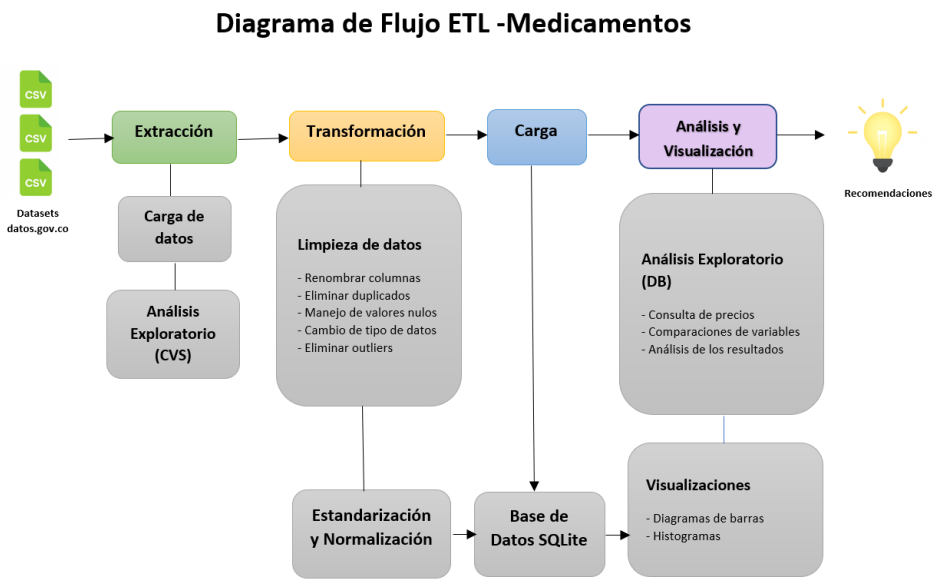

Figura 0. Diagrama de flujo ETL

2. Análisis exploratorio

Una vez seleccionados los datasets se escoge los datasets mas aptos para el análisis que son los siguientes:

Índice	Título	Fuente	Formato	Volumen	Descripción
1	Precio Medicamentos (2024)	Datos Colombia	csv	12.5k filas, 9 columnas	Contiene los precios de los medicamentos para el año 2024, incluyendo variables clave como principio activo, precio por tableta, y tipo de medicamento. Es crucial para entender la distribución de precios en 2024.
2	Clicsalud - Termómetro de Precios de Medicamentos (2025)	Clicsalud	csv	12.5k,12 columnas	Este dataset es similar al anterior, pero actualizado con datos correspondientes a 2025. Incluye los precios de los medicamentos con una actualización de la información.
15	Medicamentos (Comparativa de Precios)	Datos Colombia	csv	12,5 k filas, 9 columnas	Contiene información adicional sobre los medicamentos, como Expediente INVIMA, canal de distribución, y más detalles descriptivos de cada medicamento. Proporciona un análisis más profundo y completo.

Tabla 1. Fuentes de los datasets

3. Extracción de los datos

Análisis de la Estructura de los Datasets

Se realiza una inspección inicial de los datos para entender su estructura y calidad. Se utilizaron Pandas para cargar y visualizar los datasets, los cuales tienen diferentes columnas con variables clave.

Al revisar los tres datasets, se observó que comparten muchas columnas similares, como principio activo, precio por tableta, y fabricante. Sin embargo, el Dataset 2 tiene la columna adicional medicamento y canal, mientras que el Dataset 15 contiene tres columnas extras con más detalles sobre el medicamento, como Expediente INVIMA, canal de distribución, y numerofactor.

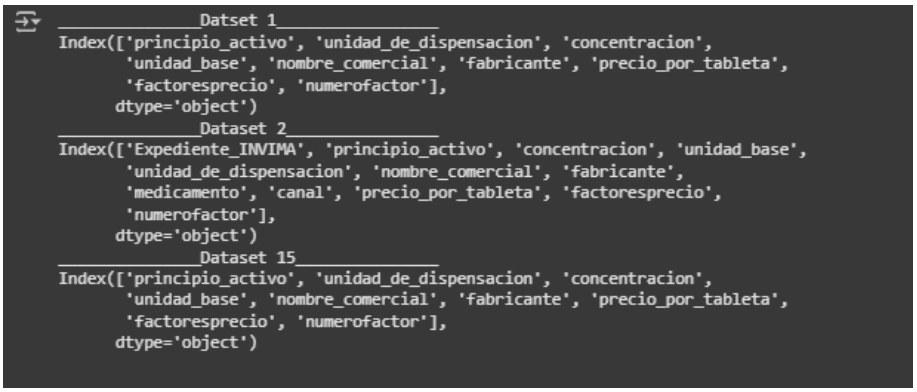

Figura 1. Estructura de los datasets

4. Transformación de los datos

Renombrado de columnas

Se realizó el renombrado de columnas, lo que se busca es que todas las columnas tengan nombres consistentes en todos los datasets. Esto es especialmente importante cuando se tienen datasets con nombres de columnas similares, pero ligeramente diferentes. En este caso, se renombraron las columnas de los tres datasets para que todos tengan una estructura de columna unificada, facilitando su análisis y combinación.

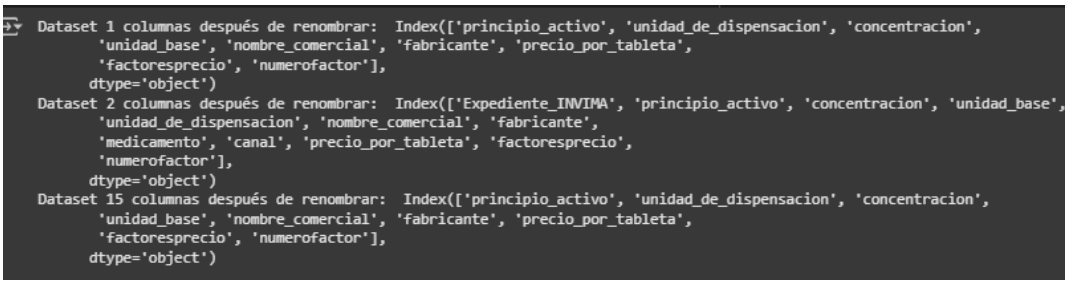

Figura 2. Renombrado de variables

Verificación de valores numéricos

Se verifico los tipos de datos para asegurarnos de que los valores numéricos, como los precios por tableta, estén en formato numérico, para que sean tratados correctamente en las siguientes fases de análisis. En este caso, se observó que los precios estaban inicialmente como texto (tipo object), lo que podría causar problemas al realizar cálculos, análisis estadísticos o incluso al generar gráficos.

Utilizando Pandas, se transformaron las columnas de precio por tableta en numerales de tipo float64. Esto se realizó con la función pd.to_numeric (), que convierte los valores de texto a tipo float64, un tipo adecuado para análisis numérico.

Se utilizó el parámetro errors='coerce' para que cualquier valor no numérico se convierta en NaN (Not a Number), lo cual puede ser útil para identificar y manejar valores no válidos en el futuro.

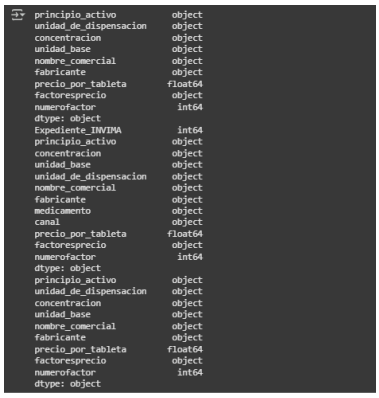

Figura 3. Tipo de valores en los datasets

Limpieza de valores duplicados en los datasets

Se utilizo el método duplicated () de Pandas para identificar registros exactos duplicados en cada uno de los tres datasets. Esta función devuelve una serie de booleanos que indica si una fila es un duplicado de una fila anterior. En el Dataset 1, se encontraron 126 duplicados, En el Dataset 2, no se encontraron duplicados (resultado de 0), En el Dataset 15, también se encontraron 126 duplicados.

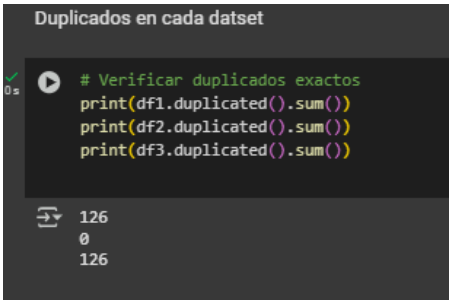

Figura 4. Datos duplicados

Después de identificar los duplicados, se utilizó el método drop_duplicates () para eliminar las filas duplicadas en los datasets df1 y df3. Este método elimina todas las filas duplicadas basándose en todas las columnas de cada dataset, dejando solo las filas únicas. Este paso es importante para garantizar que no haya información redundante que pueda distorsionar los resultados del análisis.

Después de eliminar los duplicados en df1 y df3, la verificación mostró que no hay duplicados restantes en ambos datasets (ambos resultaron en 0 duplicados).

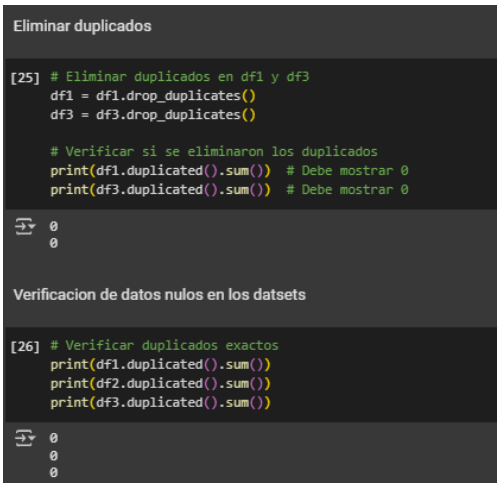

Figura 5. Eliminación de datos duplicados

Verificación de registros (Post limpieza de duplicados)

En este paso, se verifica la cantidad de registros o filas de cada dataset utilizando df. shape, una propiedad de Pandas que devuelve el número de filas y columnas en un Dataframe. Se Utilizo df. shape [0] para obtener la cantidad de filas (registros) en cada uno de los datasets.

El número de registros de cada dataset es similar, lo que indica que no hay grandes discrepancias en la cantidad de datos, pero es importante notar que el Dataset 2 tiene ligeramente más registros, lo cual podría ser por la inclusión de más información (como el canal y el Expediente INVIMA).

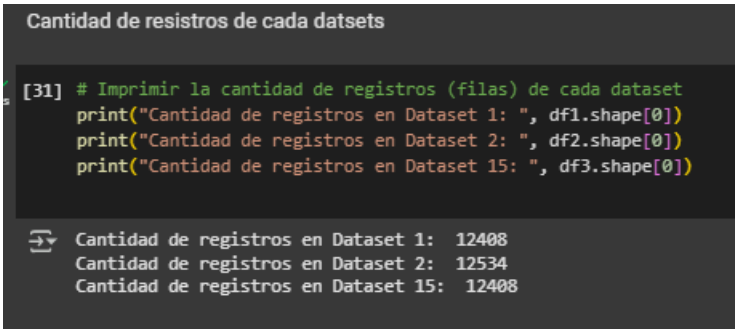

Figura 6. Conteo de registros

En este paso, se verifica si existen duplicados en las columnas clave que se utilizarán para combinar los datasets, basadas en las variables más importantes como principio_activo, concentracion, precio_por_tableta, entre otras. Esto asegura que al combinar los datasets no se introduzcan registros duplicados que afecten la calidad del análisis.

Se usó el método duplicated () de Pandas en las columnas claves seleccionadas antes de combinar los datasets. Los tres datasets (df1, df2, df3) no tienen registros duplicados basados en esas combinaciones clave. El resultado fue 0 duplicados para todos ellos.

Figura 7. Verificación de duplicados

Unificación de los datasets

En este paso, se realizó la unión de los datasets df1 y df15, ya que ambos contienen valores muy similares. El método utilizado es pd.merge() de Pandas, que permite combinar los datasets utilizando claves compartidas. Se especifica el parámetro on para indicar las columnas en las que basar la unión.

Se utilizó un outer join, lo que garantiza que todos los registros de ambos datasets se incluyan en el resultado, incluso si no tienen coincidencias exactas entre las claves de unión.

El dataset combinado tiene 12,408 registros, lo que indica que no se han perdido registros importantes después de la unión de df1 y df15.

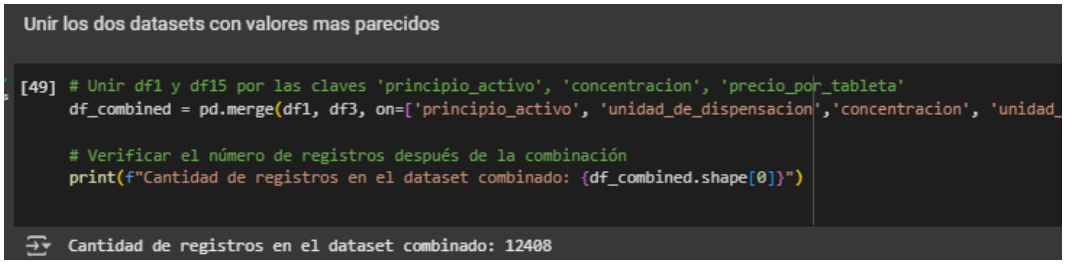

Figura 8. Unificación de los datasets

Posteriormente, se unió el dataset combinado (que incluye df1 y df15) con df2. Este paso fue necesario para integrar valores adicionales de df2, que incluyen variables nuevas como medicamento, canal, y Expediente_INVIMA, que estaban ausentes en los otros datasets.

El merge () de Pandas se utilizó nuevamente para combinar los datasets, utilizando las mismas claves o variables, en este caso que estaban disponibles y que tenían relación con el dataset df2 para asegurar que los registros se alineen correctamente.

Después de combinar los tres datasets, el dataset final contiene 12,534 registros. Este número es ligeramente mayor que el combinado de df1 y df15, lo que sugiere que df2 aportó registros adicionales sin duplicados en las claves de unión.

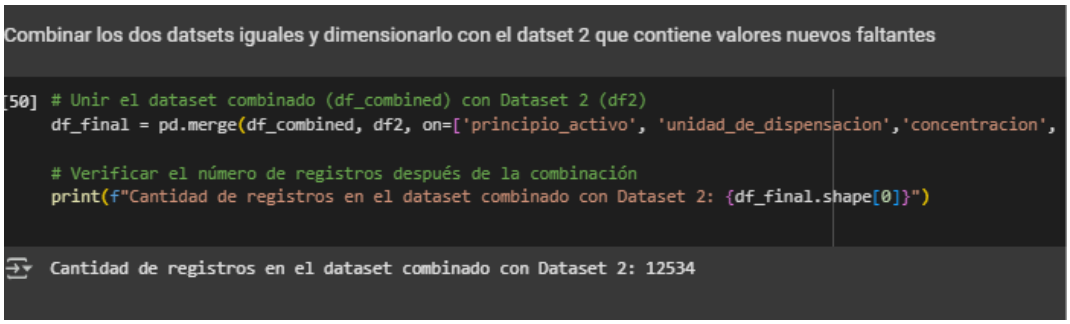

Figura 9. Conteo de registros en el dataset combinado

Verificación de valores nulos (Post Unificación de datasets)

Se verifica si hay valores nulos en el dataset final combinado. Esta es una parte importante, ya que puede darse la posibilidad de haber ocurrido errores en la combinación de los datasets y crearse valores nulos que pueden afectar el análisis y las visualizaciones posteriores.

La verificación muestra que no hay valores nulos en ninguna de las columnas, lo que significa que los datos están completos y listos para el análisis.

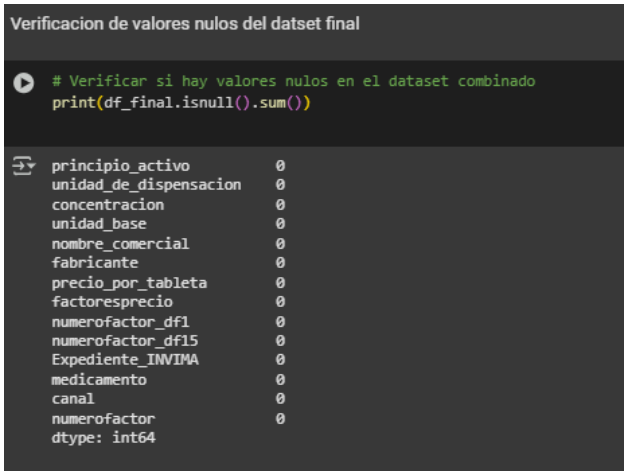

Figura 10. Verificación de valores nulos en el dataset combinado

Verificación de valores numéricos del dataset unificado

En esta etapa, se verifica las columnas numéricas del dataset final utilizando el método describe () de Pandas. Este método proporciona una visión general de las columnas numéricas, como el precio por tableta y los factores de precio. El resultado incluye información sobre el conteo, media, desviación estándar, mínimo, máximo y los percentiles de cada columna.

No hay valores nulos en ninguna de las columnas numéricas, lo que sugiere que los datos están completos y listos para el análisis posterior.

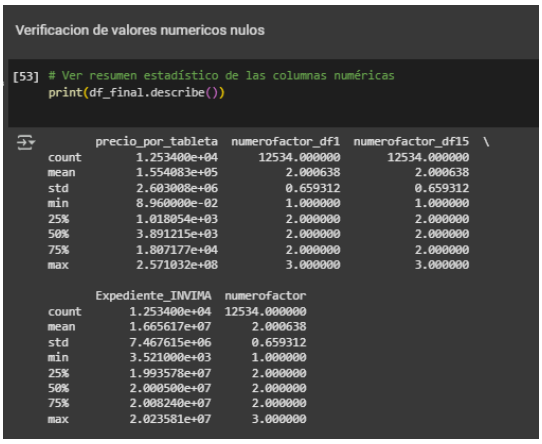

Figura 11. Verificación de valores numéricos

Pruebas para detectar Outliers en el dataset unificado

Para cerciorarnos de que nuestro datasets podía someterse al análisis final, había que verificar si los valores de la variable más importante, que en este caso es el precio de los medicamentos. Se opto por usar el método de rango intercuartílico (IQR) utilizado para detectar outliers en esta variable. Este método calcula la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1), y cualquier valor fuera del rango definido por Q1 - 1.5 * IQR y Q3 + 1.5 * IQR se considera un outlier.

La detección de outliers reveló que varios precios por tableta se encontraban fuera de los valores normales, con algunos precios extremadamente altos, lo que podría ser producto de registros erróneos o medicamentos de alto costo. Estos outliers fueron identificados para ser gestionados en la etapa de limpieza.

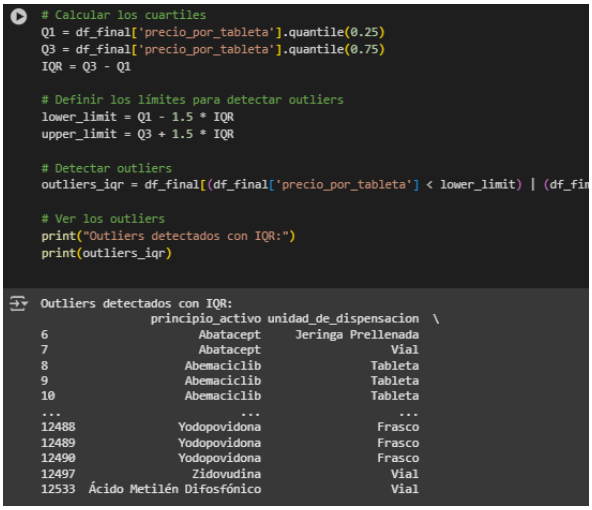

Figura 12. Detección de outliers – Método Intercuartílico

Para una mejor validación en cuanto a la detección de outliers se utilizó otro método el Z-score para detectar outliers en los precios de los medicamentos. Este método evalúa qué tan lejos están los valores de un conjunto de datos respecto a la media, expresado en términos de desviaciones estándar. Se detecto outliers como es el caso de estos medicamentos Remodulin (fabricante Ferrer),

con varios registros con precios atípicos o como también en estos medicamentos Stelara (fabricante Janssen) y Entyvio (fabricante Baxalta) también tienen precios que se consideran outliers en este análisis.

Detección de Outliers con el Método del Z-score

```
from scipy import stats

# Calcular el Z-score de la columna 'precio_por_tableta'
z_scores = stats.zscore(df_final['precio_por_tableta'])

# Detectar outliers si el Z-score es mayor a 3 o menor a -3
outliers_zscore = df_final[(z_scores > 3) | (z_scores < -3)]

# Ver los outliers
print("Outliers detectados con Z-score:")
print(outliers_zscore)
```

```
11918          Remodulin     Ferrer
11919          Remodulin     Ferrer
11920          Remodulin     Ferrer
11921          Remodulin     Ferrer
12019            Stelara    Janssen
12020            Stelara    Janssen
12189            Entyvio    Baxalta
```

Figura 13. Detección de outliers – Método Z-score

Visualización de Outliers mediante histogramas y Boxplot

Una identificación más clara de los outliers más entendibles es por medio de gráficas, se hace uso de la biblioteca Matplotlib, para la visualización de lo que causan los outliers, el histograma muestra la distribución de los precios por tableta. A partir de la visualización, es evidente que los precios están altamente sesgados hacia valores muy bajos, con una gran concentración de datos cercanos a 0. Esto se debe principalmente a la presencia de outliers en los datos, ya que los precios extremadamente altos de algunos medicamentos están distorsionando la distribución.

```
import matplotlib.pyplot as plt
# Crear histograma de 'precio_por_tableta'
plt.figure(figsize=(10, 6))
plt.hist(df_final['precio_por_tableta'], bins=30, color='skyblue', edgecolor='black')
plt.title('Distribución de Precio por Tableta')
plt.xlabel('Precio por Tableta')
plt.ylabel('Frecuencia')
plt.show()
```

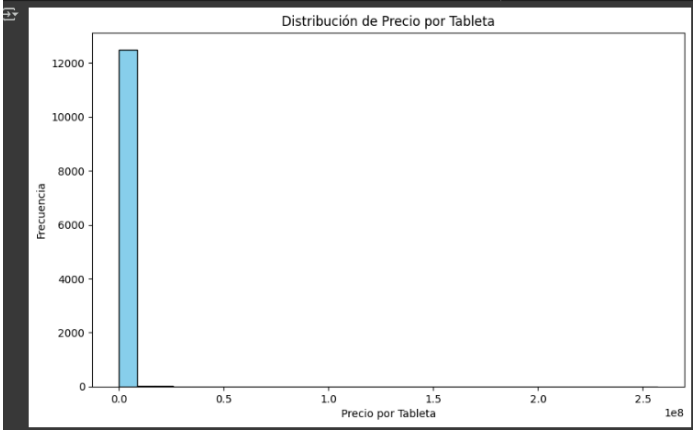

Figura 13. Visualización de outliers – Diagrama de barras

El siguiente grafico es un Boxplot muestra que la mayoría de los precios están dentro de un rango estrecho, pero hay algunos valores extremos fuera de los bigotes, que son los valores atípicos (outliers). Estos puntos indican que algunos medicamentos tienen un precio mucho mayor en comparación con el resto de los datos.

Los puntos fuera de los bigotes, son los outliers, que, en este caso, están representados por puntos alejados de la caja. Estos valores extremos probablemente distorsionan la distribución.

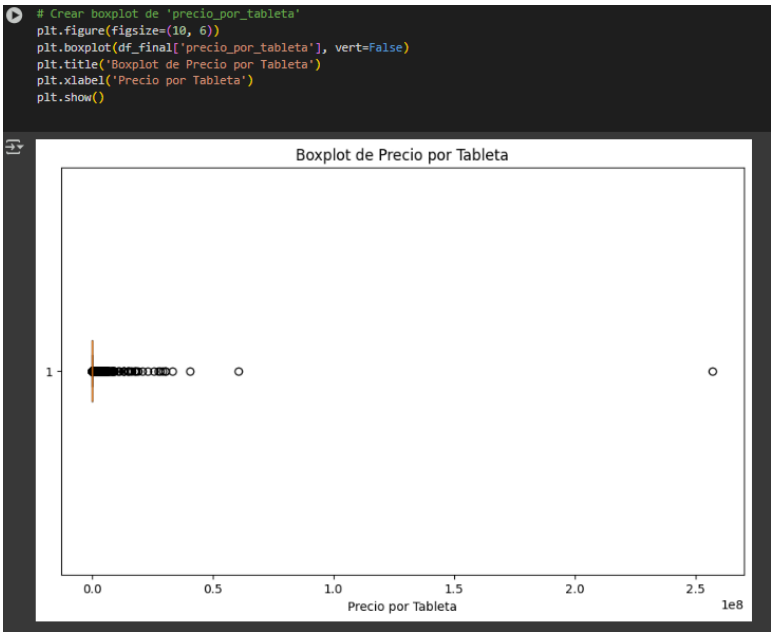

Figura 13. Visualización de outliers – Boxplots

Limpieza de Outliers

Después de aplicar los métodos IQR y Z-score para detectar los outlier, se procedió a limpiarlos, se ha logrado reducir el número de registros en el dataset de 12,534 a 10,534, eliminando así los valores atípicos que afectaban la distribución.

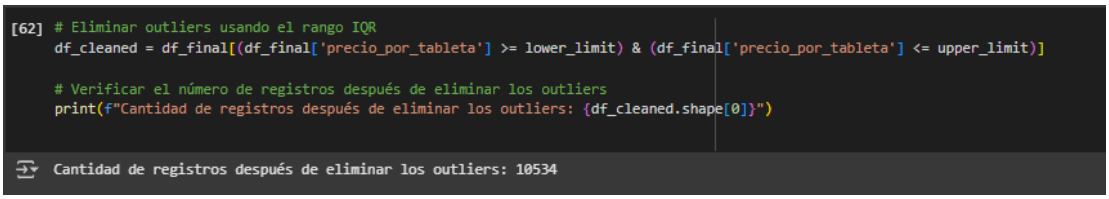

Figura 14. Eliminación de outliers

Verificación de datos (Post limpieza de Outliers)

Una vez realizado la limpieza de los outliers se verifica nuevamente, si se aplicaron los cambios para ello ejecutamos nuevamente las gráficas, anteriores y observamos los cambios. Se puede observar que los datos de los precios están más normalizados que antes.

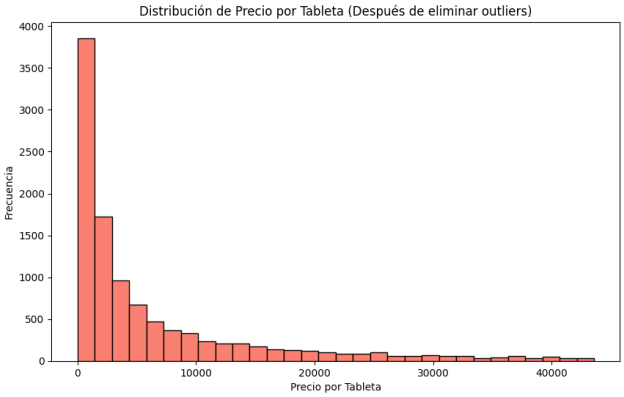

Figura 15. Verificación de outliers diagrama de barras

Hacemos lo mismo con la grafica de boxplots y observamos que, aunque hay una gran cantidad de precios concentrados en un rango determinado, también hay una pequeña cantidad de medicamentos con precios mucho más altos, lo que sugiere que a comparación de las anteriores graficas hay un cambio significativo, puesto que antes en su mayoría todos los productos tenían precios exagerados, lo que sesgaba el análisis de esta variable.

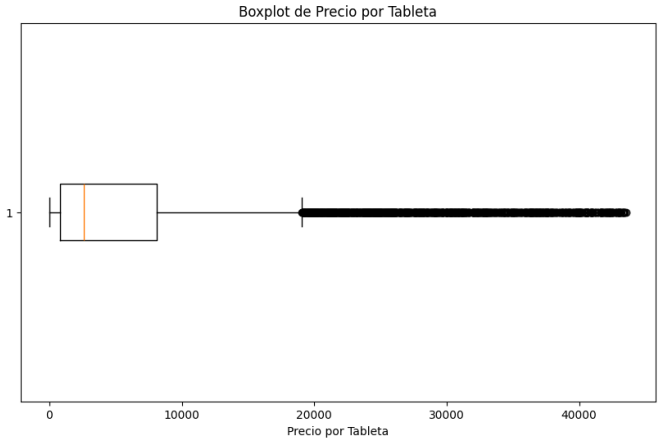

Figura 16. Verificación de outliers Boxplots

Esto nos permite identificar que nuestro dataset se ha, procesado correctamente y que se le puede aplicar el análisis siguiente. En este caso se verifica nuevamente que no se identifiquen valores nulos en las columnas ni duplicaciones en los datos antes de guardar el dataset en un archivo CSV.

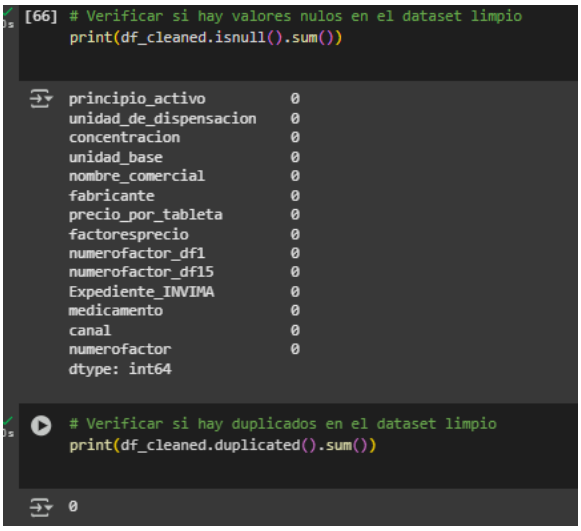

Figura 17. Verificación valores nulos

5. Carga de datos

Almacenamiento de los datos limpios en una base de datos

Una vez realizadas las validaciones se procede a guardar el dataset final y se lo almacena en una base de datos de SQLite y le asignamos el nombre de **medicamentos.db**.

```
import sqlite3

# Conectar a la base de datos SQLite (se creará si no existe)
conn = sqlite3.connect('medicamentos.db')
cursor = conn.cursor()

# Crear la tabla 'precios' si no existe (esta tabla almacenará los datos limpios)
cursor.execute('''
CREATE TABLE IF NOT EXISTS precios (
    id INTEGER PRIMARY KEY,
    principio_activo TEXT,
    unidad_de_dispensacion TEXT,
    concentracion TEXT,
    unidad_base TEXT,
    nombre_comercial TEXT,
    fabricante TEXT,
    precio_por_tableta REAL,
    factoresprecio TEXT,
    numerofactor_df1 INTEGER,
    numerofactor_df15 INTEGER,
    Expediente_INVIMA INTEGER,
    medicamento TEXT,
    canal TEXT,
    numerofactor INTEGER
)
''')
```

Figura 18. Creación de la base de datos SQLite

Después de la transformación, los datos fueron cargados en la base de datos de SQLite. La base de datos ahora contiene la información centralizada de los precios de medicamentos, con todas las variables relevantes como canal de distribución, principio activo, nombre comercial, precio por tableta, y otros detalles adicionales que permiten una comparación exhaustiva que se realizara posteriormente.

6. Análisis de los Resultados

Una vez, realizado el proceso de ETL a los datos de los medicamentos, se realizó el análisis de los datos, para dar las correspondientes recomendaciones para el cumplimiento de los objetivos, planteados. Algunos de los análisis que se realizaron fueron los siguientes:

Promedio de precios por medicamento

En este análisis, se identificaron los 10 medicamentos con los precios promedio más altos. Los resultados muestran medicamentos como Zebesten, Biovisc y Doxorubicina Clorhidrato entre los más caros.

Nombre comercial	Precio promedio (COP)
Zebesten	43,569
Biovisc	43,359
Doxorubicina Clorhidrato	43,209
Budek Plus	43,282
Lotinercan	43,083

Figura 18. Medicamentos con precios mas caros

Los medicamentos más caros parecen ser de marcas conocidas, como Zebesten (aproximadamente 43,569 COP por tableta), lo que indica que pueden ser medicamentos de alto costo o de especialidad. Los precios de estos medicamentos probablemente están influidos por factores como el principio activo, el fabricante o si están dirigidos a enfermedades crónicas o especializadas.

Distribución de los precios de medicamentos

Para identificar las tendencias de los precios se identificó, primero la distribución de los precios de los medicamentos.

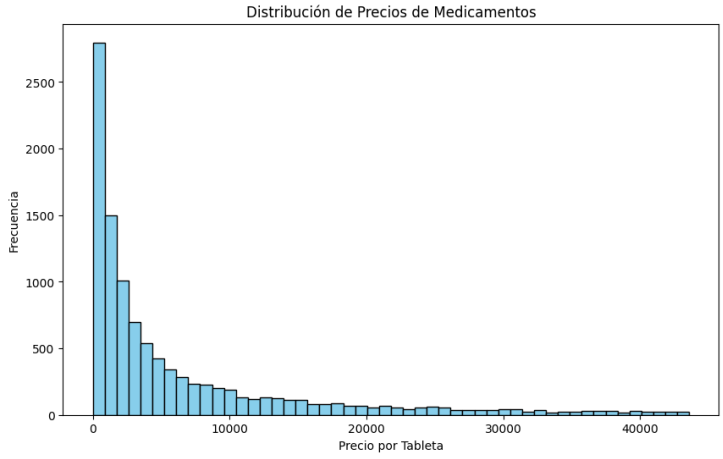

Grafico 1. Distribución de precios de los medicamentos

El histograma muestra que la mayoría de los precios de los medicamentos están concentrados en el rango más bajo, con un pequeño número de medicamentos con precios muy altos (columna a la derecha).

La gran mayoría de los precios de medicamentos se encuentra en el rango bajo (cerca de cero), lo que puede indicar que muchos medicamentos tienen precios accesibles. Hay una larga cola a la derecha, lo que sugiere que existen algunos medicamentos que tienen precios mucho más altos (posiblemente de especialidad).

Precio promedio por canal

Se procedió a realizar, el análisis para identificar las tendencias, en los precios como es el caso de los canales de distribución.

Los resultados muestran que el canal comercial tiene un precio promedio más alto (6,926 COP por tableta) en comparación con el canal institucional (5,792 COP por tableta). Esto confirma que los precios en el canal comercial son generalmente más altos que en el canal institucional.

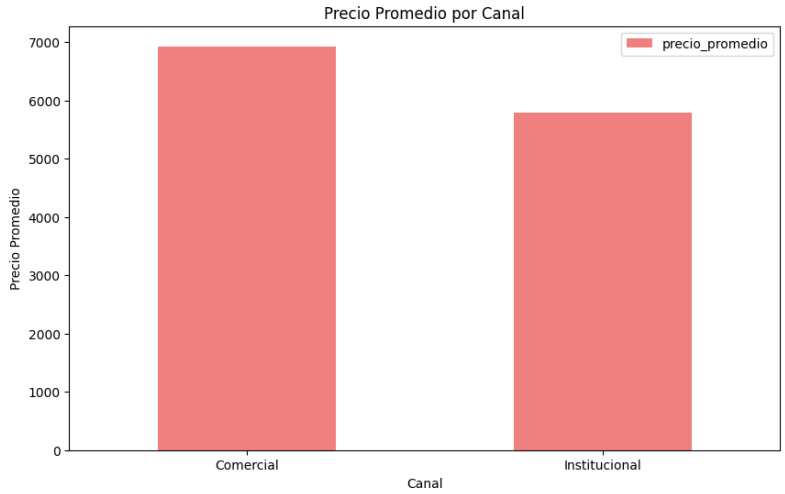

Gráfico 2. Precio promedio por canal de distribución

Los precios más altos en el canal comercial pueden reflejar márgenes de ganancia mayores o costos adicionales asociados con la distribución en farmacias privadas, marketing y otras actividades comerciales.

El canal institucional parece ofrecer precios más bajos, lo que podría estar relacionado con la compra a granel o políticas gubernamentales para reducir los costos de medicamentos en instituciones de salud pública.

7. Recomendaciones

Revisión de la Regulación de Precios de Medicamentos de Alto Costo

Los medicamentos más caros identificados en el análisis (como Zebesten y Biovisc) podrían tener precios significativamente más altos en comparación con otros medicamentos en el mercado, lo que sugiere que estos productos podrían estar influidos por factores de comercialización, patentes o monopolios.

Una de las recomendaciones podría, ser implementar políticas de control de precios para medicamentos de alto costo, particularmente aquellos que no están claramente justificados por la

complejidad de su principio activo o por la falta de alternativas genéricas. Establecer límites máximos de precio para ciertos medicamentos en función de su concentración y principio activo, de acuerdo con estudios de costos y análisis de mercado.

El impacto que se espera es, reducir los costos de medicamentos especializados para los pacientes, especialmente aquellos con enfermedades crónicas o complejas que dependen de medicamentos de alto costo.

Fomento de la Competencia en el Canal Comercial

El análisis de precios reveló que los canales comerciales (como las farmacias privadas) tienden a tener precios más altos que los canales institucionales (como los hospitales o farmacias públicas).

Una de las recomendaciones seria, crear incentivos para mejorar la competitividad en el canal comercial, como la promoción de la compra conjunta de medicamentos a precios preferenciales o el establecimiento de políticas de precios transparentes. Establecer un sistema de precios de referencia en las farmacias privadas y hacer que los precios sean accesibles, ajustándolos según el nivel de concentración de medicamentos y su disponibilidad genérica.

El impacto que se espera es, mejorar el acceso a medicamentos en el canal privado, reduciendo la brecha de precios entre los diferentes canales de distribución.

Establecer un Sistema de Información Transparente y Accesible sobre Precios

A pesar de las políticas de control de precios, los consumidores no siempre tienen acceso a información transparente sobre los precios de los medicamentos, lo que limita su capacidad de tomar decisiones informadas.

Una alternativa para solucionar estos inconvenientes es, desarrollar una plataforma pública de comparación de precios de medicamentos, en línea o móvil, que permita a los consumidores verificar los precios de diferentes productos y alternativas en tiempo real. El gobierno podría asociarse con organizaciones de salud para crear una base de datos accesible sobre precios de medicamentos, con información detallada sobre precios en diferentes canales y regiones.

El impacto que se espera es, empoderar a los consumidores con información transparente y actualizada, lo que les permitirá tomar decisiones más informadas sobre qué medicamento comprar y dónde.

8. Conclusión

El análisis realizado, basado en el proceso ETL de los precios de los medicamentos, ha identificado áreas clave donde se pueden implementar mejoras para promover una mayor transparencia de precios y competitividad en el mercado farmacéutico colombiano. Las recomendaciones proporcionadas están orientadas a reducir la brecha de precios entre los diferentes canales de distribución, promover mejoras en cuanto a la competitividad y crear políticas de monitoreo de precios. Estas iniciativas pueden mejorar significativamente el acceso de los consumidores a medicamentos a precios más justos y competitivos, contribuyendo a un sistema de salud más equitativo y accesible para todos.