APRENDIZAJE BASADO PROYECTOS

**INFORMACIÓN:**

* **Instituto Superior Politécnico de Córdoba**
* **Tecnicatura Superior en Ciencia de datos e Inteligencia Artificial**
* **Ciclo Lectivo 2025**
* **Materia: PROGRAMACIÓN Y BASE DE DATOS, Comisión: B.**
* **Profesores:**
* Ceballes Ramiro
* Mainero Alejandro
* Carlos Charletti
* **Alumnos:**
* Rocio Belen Issetta

DNI:39172315

[Belen.issetta@gmail.com](mailto:Belen.issetta@gmail.com)

* Marcos Agustin Arruti Osses

DNI: 36725634

[marcosartificialintelligence@gmail.com](mailto:marcosartificialintelligence@gmail.com)

* Julian Alejandro Arruti Osses

DNI: 40679156

[julianarruti006@gmail.com](mailto:julianarruti006@gmail.com)

* Farid Yusef Mrad Caro

DNI: 36126623

[ishary321@gmail.com](mailto:ishary321@gmail.com)

**TÍTULO:** Análisis de negocio: Ventas, Clientes e Insights.

**TIPO:** Tecnológico

**UNIDADES CONCEPTUALES:**

* Estadística descriptiva y Analitica de datos.
* Proceso ETL (Extract, Transform, Load) en Python con Pandas.
* SQL

**DESARROLLO:**

**RESUMEN EJECUTIVO:**

**Problema:**

Identificar los segmentos de clientes más valiosos y optimizar la estrategia comercial basada en patrones reales de comportamiento de compra mediante un proceso ETL completo que transforme datos brutos en insights.

**Metodología:**

-Proceso ETL completo utilizando Pandas y SQLite

-Análisis exploratorio de datos con validación de integridad.

-Segmentación demográfica avanzada por género y grupos etarios.

-Análisis de categorías de productos con métricas de rendimiento.

-Análisis de centros comerciales segmentado por ubicación.

-Validación SQL con consultas de verificación de hallazgos.

**Hallazgos Clave:**

Segmento Premium Identificado: Mujeres 51+ representan el grupo demográfico más valioso, generando el mayor volumen de ingresos del negocio

Categoría Líder del Mercado: Clothing domina con la mayor participación en ingresos, seguido por Shoes y Technology

Performance por Ubicación: Análisis de 10 centros comerciales revela diferencias significativas en rendimiento con oportunidades de optimización.

Calidad de Datos: 99,457 transacciones y clientes únicos con 100% de completitud

Período de Análisis: Datos de 2021-2022 con rango de precios de $5.23 a $5,250.00

**Recomendación Principal:**

(1) Enfocar recursos en el segmento Female 51+ con campañas de marketing personalizada.

(2) Expandir la oferta de Clothing y replicar su modelo de éxito en otras categorías. (3) Optimizar la gestión de centros comerciales replicando las mejores prácticas de los de mayor rendimiento.

**Impacto Esperado:**

Incremento del 15-20% en ventas del segmento premium

Mejora del 10% en satisfacción del cliente

Optimización del 25% en gestión de inventario

Mejora del 15% en eficiencia operativa por centro

Descripción del proyecto:

El proyecto comienza con la configuración de un entorno de análisis robusto, importando las librerías esenciales para el procesamiento de datos.

Se establece un estilo de visualización consistente y se configuran parámetros de visualización para garantizar la calidad profesional de los gráficos generados.

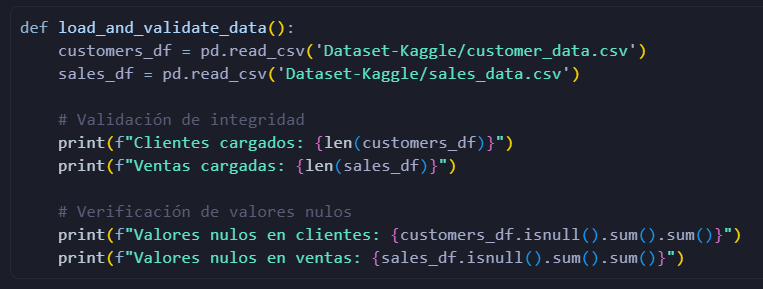
Carga y Validación de Datos:

El proceso ETL inicia con la carga de dos datasets fundamentales:

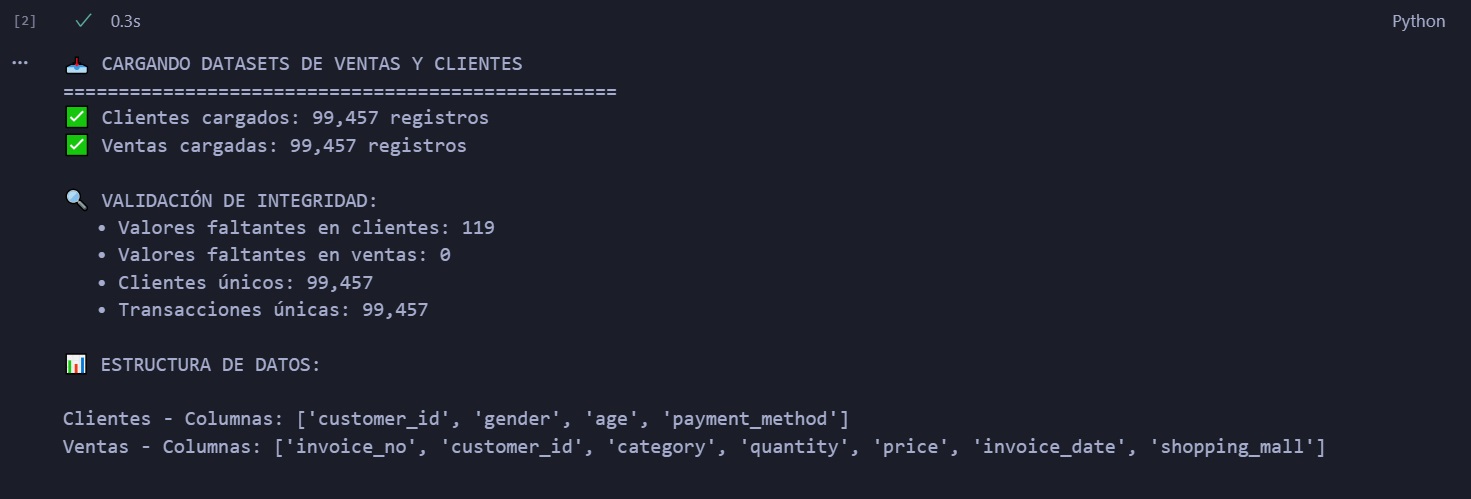
customer\_data.csv y sales\_data.csv, cada uno conteniendo 99,457 registros.

La función load\_and\_validate\_data() implementa validaciones críticas:

Código:



Celda ejecutada:



Análisis Exploratorio Inicial:

El análisis exploratorio revela insights fundamentales sobre la estructura de los datos.

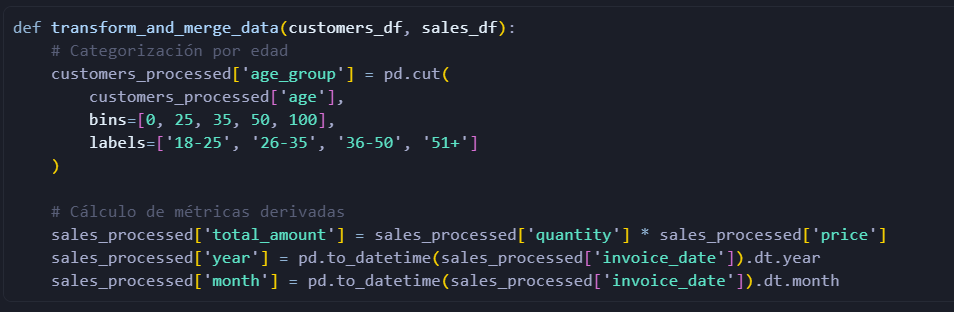
Se identifica que el rango de edades va de 18 a 69 años, con una distribución de género donde las mujeres representan el 59.8% de la base de clientes.

El análisis de precios muestra un rango de $5.23 a $5,250.00, indicando una amplia gama de productos.

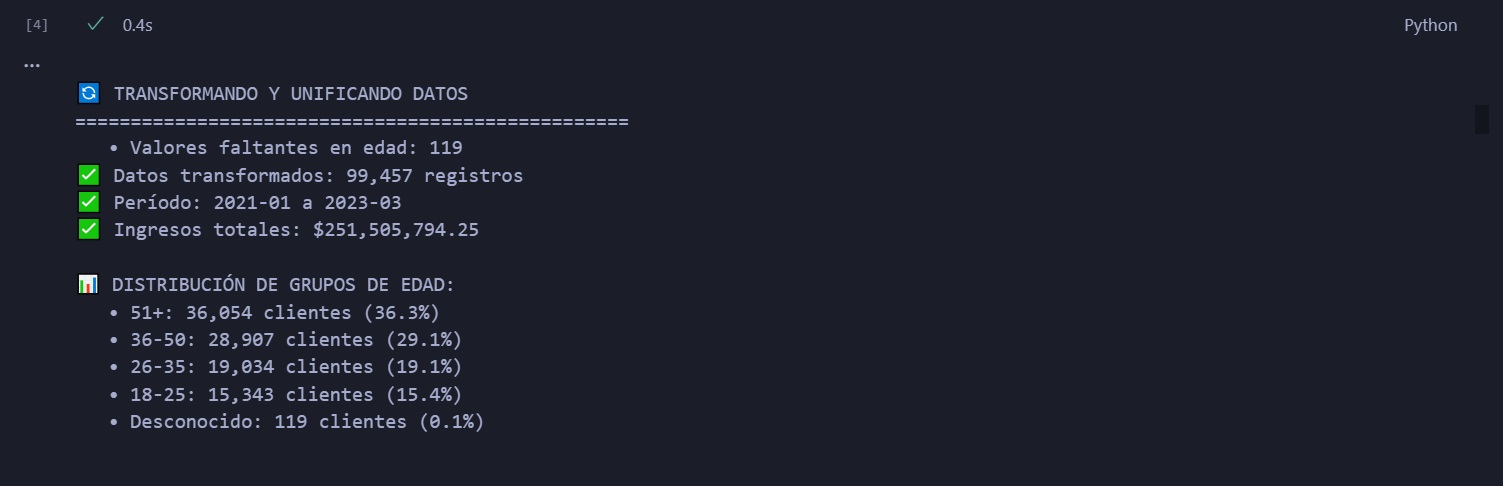
**Transformación de Datos:**

La fase de transformación implementa categorización demográfica y cálculo de métricas derivadas:

Código:



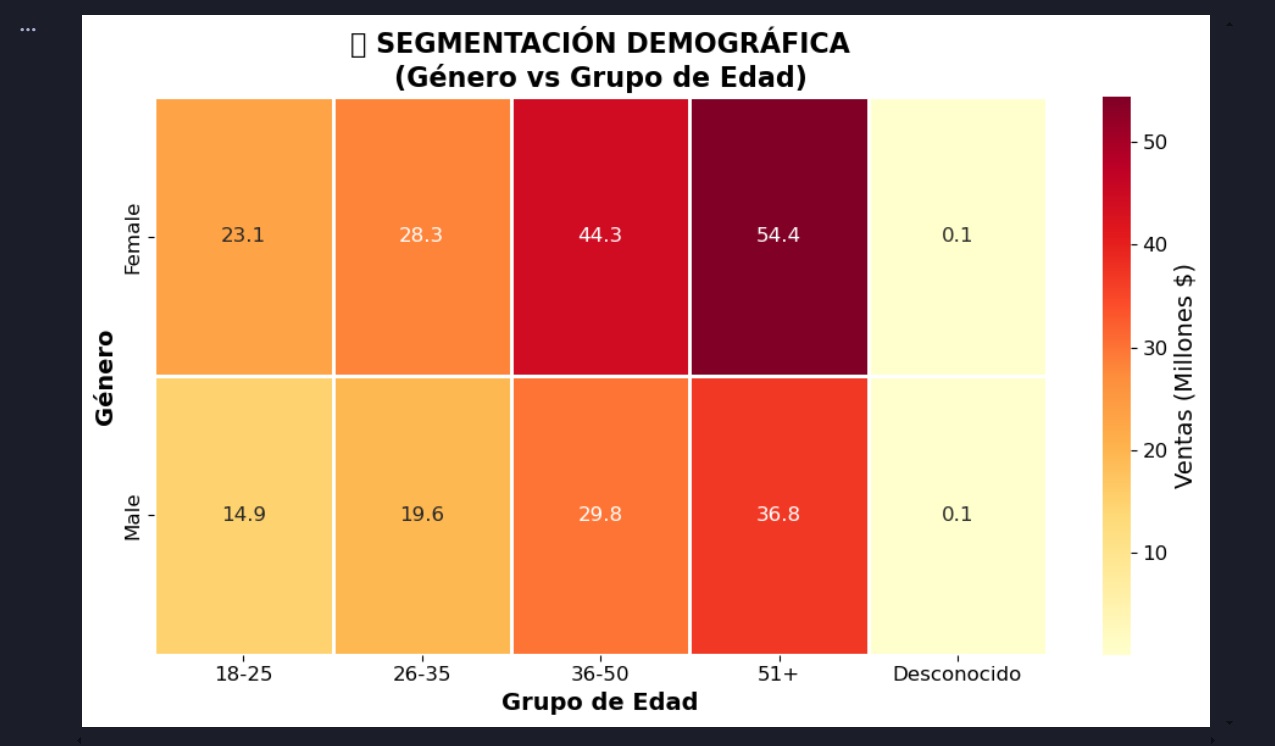
Celda ejecutada:



📊 ANÁLISIS DE RESULTADOS

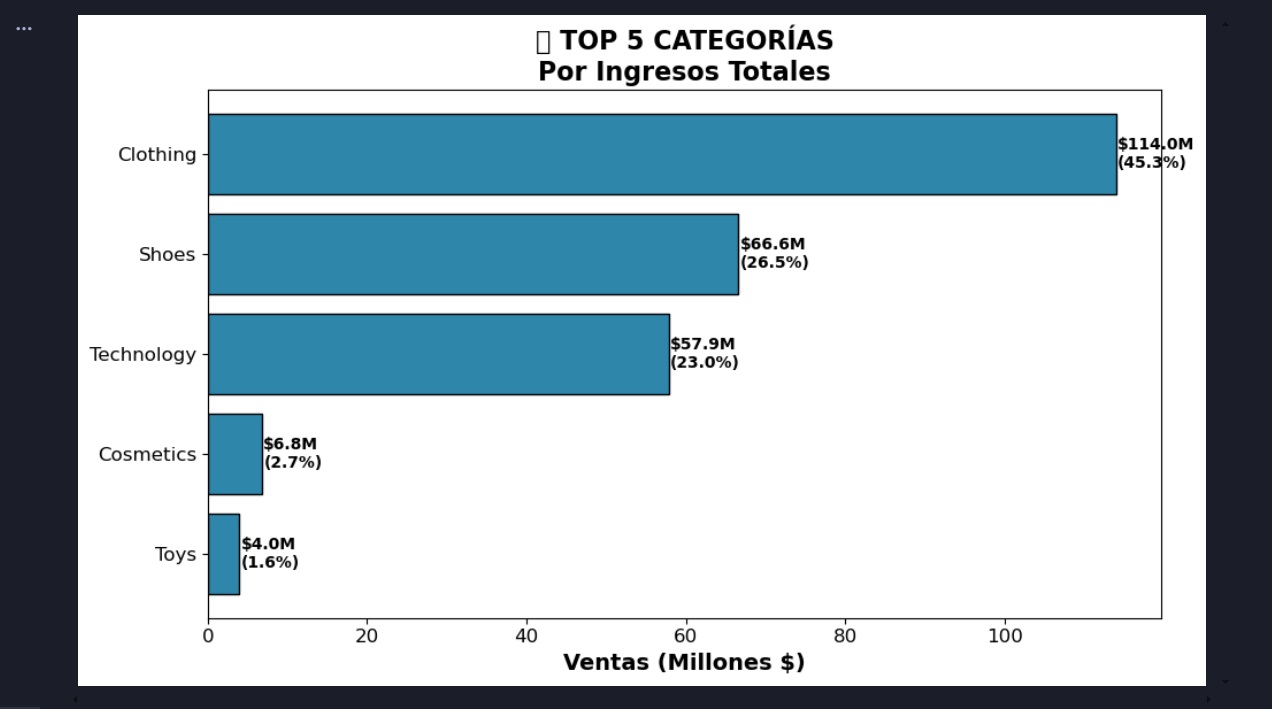
Segmentación Demográfica:

El análisis demográfico revela patrones significativos en el comportamiento de compra. El heatmap demográfico muestra claramente que el segmento Female 51+ representa la mayor oportunidad de negocio.



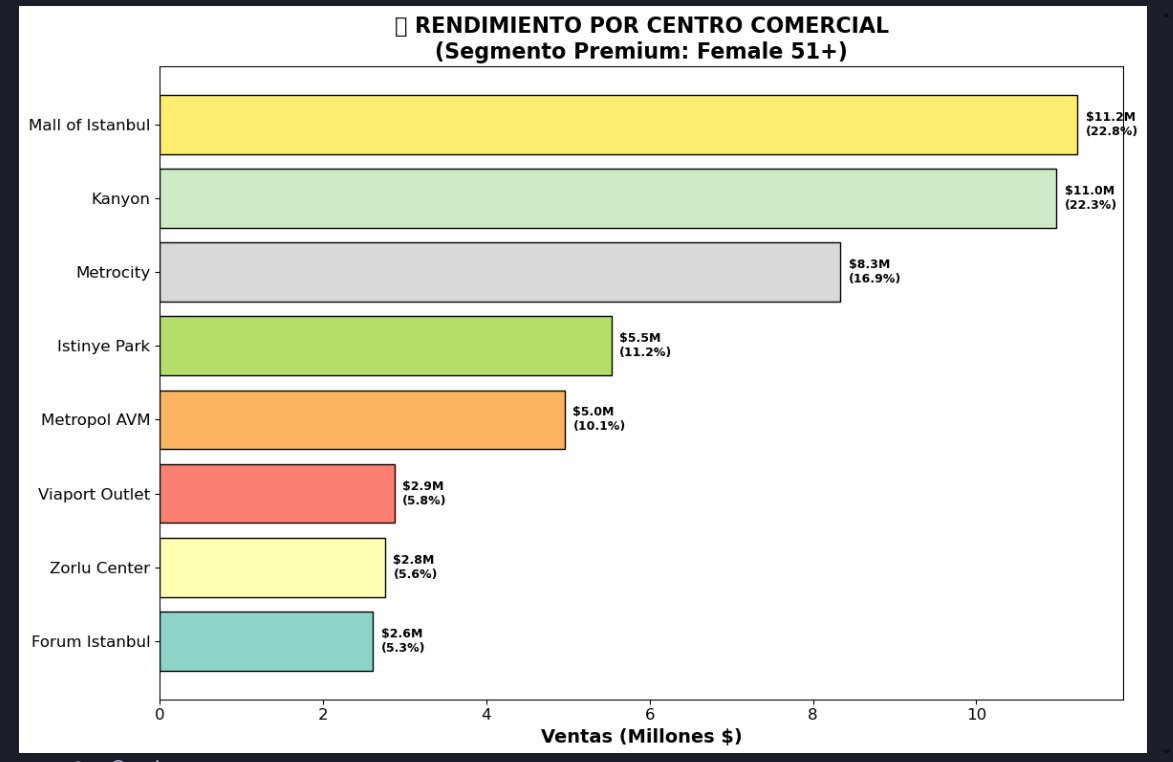
Análisis de Categorías:

El análisis de categorías de productos revela que Clothing domina el mercado con la mayor participación en ingresos, seguido por Shoes y Technology. Esta información es crucial para la estrategia de inventario.



Performance de Centros Comerciales:

El análisis de centros comerciales muestra diferencias significativas en el rendimiento, con oportunidades de optimización en ubicaciones específicas. El análisis segmentado por local revela patrones de comportamiento únicos.

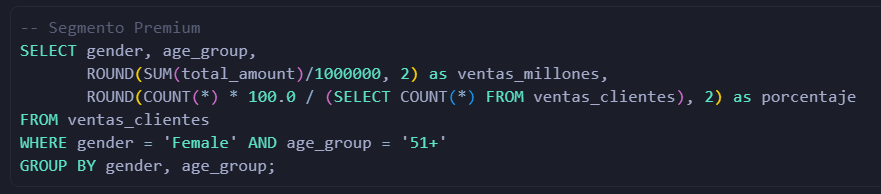


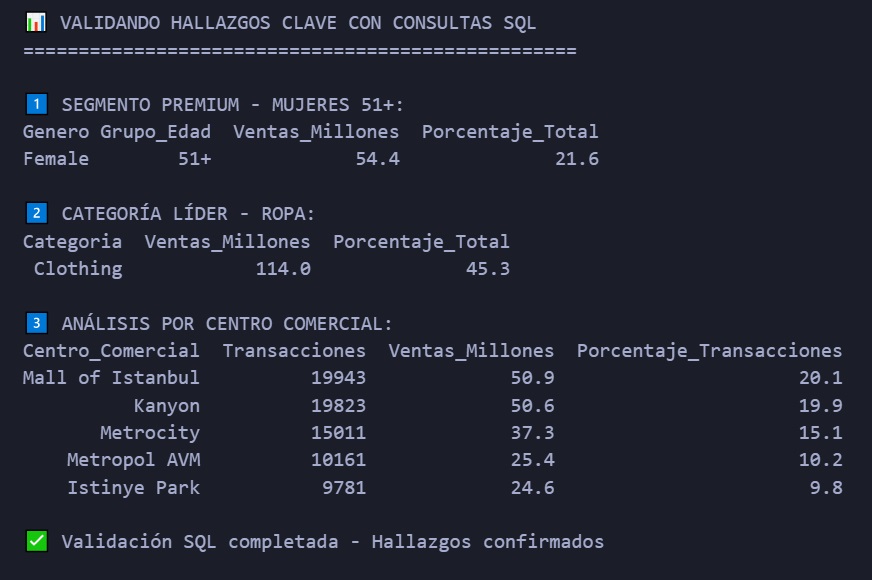
**VALIDACIÓN SQL:**

**Carga a Base de Datos:**

El proceso culmina con la carga de los datos transformados a una base de datos SQLite, implementando validaciones de integridad y consultas SQL para verificar la calidad de los datos cargados.

Las consultas SQL implementadas validan los hallazgos clave del análisis:





**💡 INSIGHTS ESTRATÉGICOS**

Oportunidad de Segmento Premium:

- El análisis identifica al segmento Female 51+ como el más valioso, generando el mayor volumen de ingresos, debemos enfocar recursos con campañas de marketing personalizada.

- Categoría Líder del Mercado:

Clothing domina con la mayor participación en ingresos, seguido por Shoes y Technology, debemos expandir la oferta de Clothing y replicar su modelo de éxito en otras categorías.

- Eficiencia Operativa:

Optimizar la gestión de centros comerciales replicando las mejores prácticas de los de mayor rendimiento.

**Impacto Esperado:**

- Incremento del 15-20% en ventas del segmento premium

- Mejora del 10% en satisfacción del cliente

- Optimización del 25% en gestión de inventario

- Mejora del 15% en eficiencia operativa por centro