**Taller: Repaso Corte 2 y Corte 3**

**Estudiante**

LAURA VALENTINA BUSTOS LUPACO

**PROGRAMA INGENIERÍA INDUSTRIAL**

**CIENCIA DE DATOS**

**Profesor**

JESUS ARIEL GONZALEZ BONILLA

25 Octubre 2025

**Introducción**

Se utilizó el conjunto de datos Online Retail, que incluye detalles sobre las transacciones de una tienda de regalos por Internet ubicada en el Reino Unido. La meta fundamental fue llevar a cabo un proceso integral de análisis de datos, que incluyó desde la carga, la limpieza y la transformación de los datos, hasta la elaboración de consultas SQL, la visualización gráfica y la integración con una API externa.

**Metodología**

A lo largo del taller, se siguió un procedimiento de trabajo bien definido que abarcó las siguientes fases:

* Extracción, Transformación y Carga (ETL): Los datos fueron descargados desde el repositorio UCI Machine Learning Repository y se llevó a cabo un proceso de limpieza para eliminar duplicados, valores ausentes y registros incompletos.
* Modelado y almacenamiento: Los datos fueron almacenados en una base de datos SQLite, lo que facilitó la realización de consultas SQL.
* Consultas SQL: Se calcularon métricas importantes, incluyendo el total de ventas por país, el promedio de ventas en cada transacción y la identificación de los clientes más valiosos procedentes del Reino Unido.
* Análisis temporal: Los datos se organizaron por mes y país para examinar las tendencias a lo largo del tiempo.
* Visualización: Se emplearon herramientas como Matplotlib y Seaborn para graficar las ventas según el país y el cambio de ventas a través del tiempo.
* Integración con API: Por último, se estableció una conexión con la API de divisas Frankfurter, que permite la conversión de valores de ventas a diferentes monedas para llevar a cabo un análisis comparativo a nivel internacional.

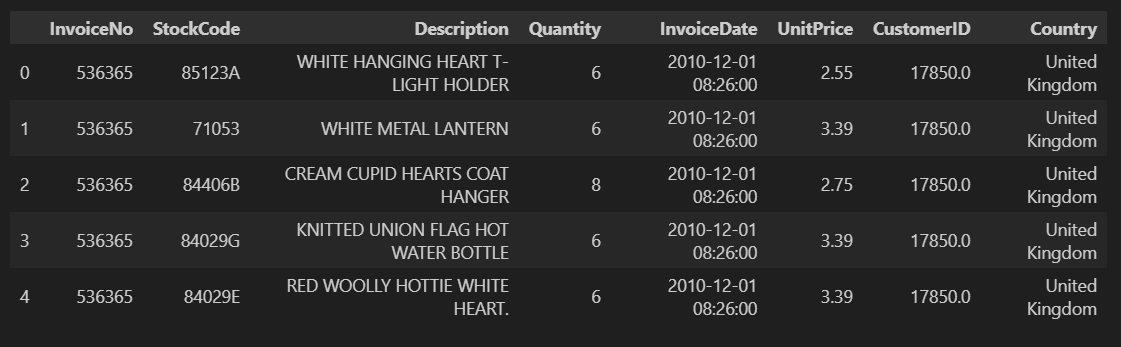
**Resultados**

Los principales resultados obtenidos fueron:

* El conjunto de datos original contenía 541,909 registros, de los cuales quedaron 401,604 registros tras la limpieza, eliminando un 25.89% de filas.
* El rango de fechas disponible abarca desde diciembre de 2010 hasta diciembre de 2011.
* Los países con mayores ventas fueron Reino Unido, Países Bajos, Alemania y Francia.
* El valor promedio por transacción fue superior en Reino Unido.
* Los clientes más valiosos pertenecen al Reino Unido, siendo los ID 18102, 17450 y 14911 los de mayor gasto total.

**Gráficos y Tablas**

Primera vista del DataFrame (las primeras filas de los datos cargados).

****

Resultado de df.info() (tipos de datos).

**Texto

El contenido generado por IA puede ser incorrecto.**

Resultado de los valores faltantes



Las métricas (porcentaje de eliminación y rango de fechas).

Texto

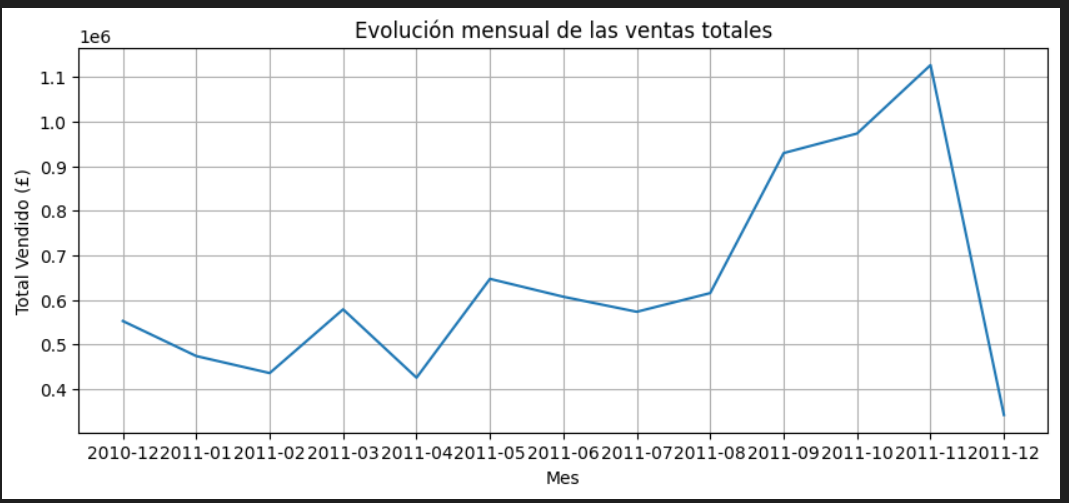
El contenido generado por IA puede ser incorrecto.

tabla top\_paises (Top 10 países con mayores ventas).

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

Evolución Mensual de las ventas totales



Paises con Mayores ventas totales

Gráfico

El contenido generado por IA puede ser incorrecto.

Evolución mensual de ventas en Reino Unido.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Ejemplo de conversión de ventas de USD → EUR

Texto

El contenido generado por IA puede ser incorrecto.

**Análisis e interpretación**

El estudio reveló que el Reino Unido agrupa la mayoría de las ventas, lo que se atribuye a que la tienda tiene su principal clientela en esa nación.

Las cifras de ventas evidencian una variación estacional, con incrementos notables en los meses previos a diciembre, lo que indica un crecimiento debido a la época navideña.

En cuanto a la clientela, se detectaron variaciones significativas en el valor medio por transacción, resaltando a ciertos compradores con altos volúmenes, quienes podrían ser considerados como clientes clave.

Estos hallazgos subrayan la relevancia de segmentar a los clientes y entender sus hábitos de compra para crear estrategias efectivas de fidelización.

**Conclusiones**

Concluyendo este taller vemos que podemos de implementar un proceso integral de análisis de datos utilizando Python. Cada fase, abarcando desde la importación y depuración de datos hasta su visualización y la interacción con una API, proporcionó insights significativos sobre el comportamiento del negocio. Se evidenció que herramientas como pandas, sqlite3, seaborn y requests son fundamentales para desarrollar soluciones de análisis que sean tanto reproducibles como escalables. La información obtenida se puede utilizar para mejorar los procesos de ventas, analizar el rendimiento de los mercados y potenciar la toma de decisiones fundamentadas en datos.

**Reflexión Final**

La implementación de procesos de ETL, el uso de consultas SQL, la integración de APIs y la visualización de datos forman un pilar esencial en la era actual de la ciencia de datos. Estas técnicas permiten la conversión de datos sin procesar en información valiosa, descubrir patrones y producir conocimiento útil para las organizaciones. La conexión de múltiples fuentes de datos, enriquece las capacidades analíticas y permite una comparación a nivel internacional. En resumen, la ciencia de datos promueve decisiones más precisas, estratégicas y fundamentadas, mejorando así la competitividad y la efectividad en contextos empresariales reales.