

INFORME EXPLICATIVO DEL ANÁLISIS DEL PROYECTO

INTRODUCCIÓN

Este proyecto constituye el trabajo final del Máster en Data Analytics, cuyo objetivo principal es demostrar la capacidad para aplicar un flujo completo de análisis de datos: desde la extracción y limpieza, pasando por el análisis descriptivo y estadístico, hasta el diseño de visualizaciones y dashboards que permitan tomar decisiones informadas.

El análisis se centra en estudiar las ventas de un ecommerce internacional, vinculándolas con indicadores macroeconómicos como el PIB per cápita para encontrar patrones y relaciones significativas.

OBJETIVOS

- Desarrollar un fichero completo de análisis de datos con Python utilizando las librerías de Pandas, Numpy, Matplotlib y Seaborn.
- Construir un dashboard interactivo con PowerBI que permita explorar los resultados
- Identificar puntos claves que nos ayuden a establecer estrategias comerciales, como la relación entre el ticket medio y el PIB per capita

DATASETS

Se han utilizado dos conjuntos de datos principales provenientes de fuentes distintas:

1. Dataset de Ecommerce (OnlineRetail):

- Contiene detalles de facturación, como InvoiceNo, Quantity, UnitPrice, Country, InvoiceDate.
- Más de 500.000 registros iniciales.

2. Dataset de Indicadores Económicos (World Development Indicators- Kaggle):

- Se incluyen variables macroeconómicas como el PIB per cápita (NY.GDP.PCAP.CD)
- Datos estructurados por Country Name , Year , Month e indicadores económicos.

LIMPIEZA Y TRANSFORMACIÓN DE LOS DATOS

Dataset Ecommerce:

- Eliminamos valores nulos y registros como Quantity o UnitPrice negativos o nulos.
- Conversión de InvoiceDate a formato datetime y extracción de columnas derivadas como: Year, Month, Day , Weekday.
- Calculo de variables enriquecidas como:
 - TotalSales= Quantity * UnitPrice
 - IsWeekend, IsHighValue(tickets altos definidos por percentil 75)
 - DescriptionLength, Qty_Squared, Log_UnitPrice, Month_x_Day

Dataset Indicators:

- Filtramos únicamente por el indicador PIB per cápita (NY.GDP.PCAP.CD)
- Reestructuración del dataset (de wide a long con melt y luego pivot) para tener columnas como Country, Name, Year, GDP_per_capita

Unión de los dos dataset: ecommerce e indicators

- Normalización de nombres de país (strip y title())
- Unión de los dos dataset finales por country y year
- A partir de ahí se obtuvo un dataset de 401,604 filas y 22 columnas, cumpliendo con el requisito del proyecto.

ANÁLISIS DESCRIPTIVO Y ESTADÍSTICO

Se realizó un análisis exhaustivo del dataset final en Python:

- df_final.describe() para obtener estadístico como media, mediana, desviación estándar.
- df_final.corr() para revisar correlaciones entre variables numéricas , encontrando una baja correlación directa entre GDP_per_capita y TotalSales, pero una leve relación positiva con AvgTicket
- Histogramas y boxplots mostraron:
 - Distribuciones asimétricas con presencia de ticks extremadamente altos (outliers)
 - Variabilidad significativa del TotalSales según país

Este análisis permito confirmar la validez del dataset para construir le dashboard , destacando patrones estacionales y diferencias claras entre países

DASHBOARD INTERACTIVO EN POWERBI

El dashboard se diseño con el objetivo de facilitar la exploración dinámica de los datos, permitiendo al usuario filtrar por Country, year y tickets altos (IsHighValue)

Elementos principales:

- **KPI cards con:**
 - Ticket medio (AvgTicket): 373€
 - Ventas online (TotalSales): 8 M€
- Gráfico de dispersión: mostrando la relación entre AvgTicket y GDP_per_capita, con el tamaño representando TotalSales, permitiendo analizar si los países con mayor PIB tienden a gastar más por factura
- Grafico lineal con evolución mensual de TotalSales por país, mostrando picos altos en el último trimestre del año
- Tootips mostrando Country, Year, TotalSales al pasar el ratón.

RESULTADOS OBTENIDOS

Los países con mayor PIB per capita no siempre muestran un ticket medio proporcionalmente más alto, lo que nos da una idea de que hay otros factores locales (comerciales, culturales...) influyendo en el gasto medio.

Se observa una tendencia positiva moderada, es decir, los países con mayor PIB per cápita tienden a tener un ticket medio algo superior, aunque la relación no es lineal ni perfecta.

Se pueden identificar países con un ticket medio extraordinariamente alto (algunos superando los 3.000 €), pese a tener un PIB per cápita comparable a otros países con tickets mucho menores.

El país es el que concentra la mayor parte de las ventas totales es Reino Unido.

CONCLUSIONES

Este proyecto demuestra la capacidad para gestionar un ciclo completo de análisis de datos: desde la limpieza y transformación avanzada en Python hasta la construcción de un dashboard interactivo en Power BI, generando perspectivas accionables que pueden guiar decisiones estratégicas en un contexto real de negocio.