🛍️ Retail Sales Data Analysis – White Paper

# 📘 Project Overview

This project aims to analyze retail transaction data to derive meaningful business insights. The analysis was conducted using Python, with popular libraries such as Pandas, Matplotlib, and Seaborn.

# 🎯 Objectives

- Analyze customer transactions to identify sales trends.

- Determine top-selling products and best-performing countries.

- Explore relationships between product quantity, pricing, and revenue.

- Understand monthly revenue trends to aid in business decision-making.

# 🧰 Tools & Technologies Used

- Python (Jupyter Notebook)

- Pandas

- Matplotlib

- Seaborn

- NumPy

# ⚙️ Steps Performed

1. \*\*Data Loading & Initial Exploration\*\*

- Loaded the dataset using Pandas.

- Inspected the structure and basic statistics of the data.

2. \*\*Data Cleaning\*\*

- Renamed the `Description` column to `Products`.

- Removed rows with missing `CustomerID` and `Products`.

- Converted `InvoiceDate` to datetime format and `CustomerID` to integer.

- Removed transactions with negative `Quantity` or `UnitPrice`.

3. \*\*Feature Engineering\*\*

- Created a new column `TotalAmount = Quantity × UnitPrice`.

- Extracted `Month` from `InvoiceDate` for time-based analysis.

4. \*\*Exploratory Data Analysis (EDA)\*\*

- Top Products by Revenue: Horizontal bar chart of top 10 products.

- Top Countries by Revenue: Vertical bar chart of countries contributing most to sales.

- Transaction Amount Distribution: Histogram with KDE curve.

- Quantity vs Unit Price: Scatter plot showing correlation.

- Monthly Revenue Trend: Line plot of revenue over time.

# 📊 Key Insights

- A small number of products account for the majority of total revenue.

- The UK dominates in sales volume, followed by a few key European markets.

- Most transactions are low in value, with occasional high-value spikes.

- Negative transactions (returns/cancellations) were effectively filtered.

- Revenue trends show monthly variation, indicating possible seasonality.

# ✅ Recommendations

- Focus marketing efforts on top revenue-generating products.

- Increase outreach in top countries showing strong revenue potential.

- Analyze underperforming products for potential removal or discounting.

- Use monthly trends for inventory planning and forecasting.

- Maintain consistent data cleaning to ensure quality analysis.

# 📁 Project Files

- `retail\_analysis.ipynb` – Main analysis notebook.

- `README.md` – Project documentation.

- `requirements.txt` – List of Python libraries used.

- `data.csv` – Source data file (if applicable).

# 📌 Final Note

This analysis provides valuable insights into retail sales data, helping businesses make data-driven decisions. The findings can be used to improve product strategies, expand in high-performing regions, and plan for seasonal sales effectively.