**📊 E-commerce Sales Analysis Report**

**1. Project Overview**

This project analyzes the sales data of an e-commerce store to uncover insights such as monthly sales trends, best-selling products, and customer behavior.

The analysis is performed using **Python (pandas, matplotlib)**, and the data comes from a real e-commerce transaction dataset.

**2. Data Loading and Exploration**

**Dataset Information**

• **File:** ecommerce-data.csv

• **Total Rows:** 541,909

• **Columns: [ ]**

• InvoiceNo: Order number

• StockCode: Product code

• Description: Product name

• Quantity: Number of units sold

• InvoiceDate: Date of purchase

• UnitPrice: Price per unit

• CustomerID: Unique identifier for each customer

• Country: Country where the order was placed

**Initial Data Check**

import pandas as pd

df = pd.read\_csv("ecommerce-data.csv", encoding="ISO-8859-1")

print(df.info())

print(df.head())

**3. Monthly Sales Trend Analysis**

**Objective**

We want to understand the **monthly revenue trend** to identify seasonal patterns and potential sales peaks.

**Process**

1. Convert InvoiceDate to datetime format.

2. Extract YearMonth to group data by month.

3. Calculate monthly total sales using:

4. Plot the monthly sales trend using matplotlib.

**Code**

import matplotlib.pyplot as plt

df["InvoiceDate"] = pd.to\_datetime(df["InvoiceDate"])

df["YearMonth"] = df["InvoiceDate"].dt.to\_period("M")

df["TotalSales"] = df["Quantity"] \* df["UnitPrice"]

monthly\_sales = df.groupby("YearMonth")["TotalSales"].sum()

plt.figure(figsize=(12, 6))

monthly\_sales.plot(marker="o", linestyle="-", color="b")

plt.xlabel("Month")

plt.ylabel("Total Sales")

plt.title("Monthly Sales Trend")

plt.grid(True)

plt.show()

**Key Insight**

• Sales **increase significantly from September to November**, indicating potential holiday shopping effects (Black Friday, Christmas).

• **January and February have the lowest sales**, possibly due to post-holiday seasonality.

**4. Top 5 Best-Selling Products**

**Objective**

Identify the **top 5 products with the highest sales volume** to understand product popularity and inventory priorities.

**Process**

1. Group data by StockCode.

2. Calculate total quantity sold for each product.

3. Sort products by total quantity in descending order.

4. Plot the top 5 products using a bar chart.

**Code**

top\_products = df.groupby("StockCode")["Quantity"].sum()

top\_products = top\_products.sort\_values(ascending=False).head(5)

plt.figure(figsize=(10, 6))

top\_products.plot(kind="bar", color="skyblue")

plt.xlabel("Product (StockCode)")

plt.ylabel("Total Quantity Sold")

plt.title("Top 5 Best-Selling Products")

plt.xticks(rotation=45)

plt.grid(True)

plt.show()

**Key Insight**

• The top 5 products contribute significantly to total sales.

• Understanding which products drive sales can help **optimize inventory management and marketing campaigns**.

**5. Tools Used**

• **pandas**: Data manipulation and analysis

• **matplotlib**: Data visualization

• **Python**: Core programming language

**6. Summary of Findings**

**Analysis Step** **Key Takeaways**

Monthly Sales Trend Sales peak in November, indicating strong seasonal effect (holiday shopping)

Top 5 Products Concentrated sales in a small number of products, highlighting potential best-seller campaigns

A graph with blue line and numbers

AI-generated content may be incorrect.

**📊 Analysis of Most Returned Products**

**🎯 Objective**

Identify **which products were returned the most**.

• In this dataset, **returns are recorded as negative quantities** (e.g., Quantity = -5 means 5 units were returned).

• Understanding returns is important for **quality control, supplier negotiation, and customer satisfaction improvement**.

This analysis identifies the **Top 5 most returned products** based on total return quantities.

Returns can indicate potential issues such as:

• Product defects

• Misleading product descriptions

• Logistics and delivery damage

•Incorrect items shipped

**📜 Code**

*# Filter only returned products (Quantity < 0)*

returned\_products = df[df["Quantity"] < 0]

*# Group by product (StockCode) and sum absolute quantity to find total returns per product*

returned\_products\_summary = returned\_products.groupby("StockCode")["Quantity"].sum().abs()

*# Sort by highest number of returns*

top\_returned\_products = returned\_products\_summary.sort\_values(ascending=False).head(5)

*# Plot top 5 returned products*

plt.figure(figsize=(10, 6))

top\_returned\_products.plot(kind="bar", color="salmon")

plt.xlabel("Product (StockCode)")

plt.ylabel("Total Quantity Returned")

plt.title("Top 5 Most Returned Products")

plt.xticks(rotation=45)

plt.grid(True)

plt.show()

**📜 Process**

**Step 1: Filter Return Data**

We filtered transactions where Quantity is negative, indicating returns.

returned\_products = df[df["Quantity"] < 0]

**Step 2: Group and Aggregate**

We grouped by StockCode and calculated the **total absolute return quantity** per product.

returned\_products\_summary = returned\_products.groupby("StockCode")["Quantity"].sum().abs()

**Step 3: Identify Top 5**

We sorted the products by return quantity and selected the **Top 5**.

top\_returned\_products = returned\_products\_summary.sort\_values(ascending=False).head(5)

**Step 4: Visualization**

Finally, we visualized the result using a bar chart.

top\_returned\_products.plot(kind="bar", color="salmon")

**Key Findings**

**StockCode** **Total Returns**

23843 80,000+

23166 74,000+

23005 Around 30,000

84347 Around 10,000

23003 Around 10,000

• Products **23843 and 23166** stand out with exceptionally high return counts.

• These products require urgent quality control review and further investigation into potential product or delivery issues.

• Frequent returns directly impact **customer satisfaction** and **operational costs**.

**Next Step**

• Compare the return rates of these products across different countries to see if certain markets are more affected.

•Analyze product reviews and customer complaints associated with these items for deeper insight.

A graph with red squares

AI-generated content may be incorrect.

**📊 Comparison of Sales by Country**

**📄 Section 4: Sales by Country Analysis**

**Objective**

The objective of this analysis is to identify the **Top Countries by Total Sales** and understand the overall **geographic distribution of revenue**.

**Process**

1. **Data Preparation**

Total Sales was calculated using:

2. **Group and Aggregate**

We grouped the data by Country and calculated the **total sales per country**.

3. **Sorting**

Countries were sorted in descending order of total sales to easily identify the top markets.

4. **Visualization**

A bar chart was used to visualize the distribution of total sales across countries.

**Key Findings**

|  |  |
| --- | --- |
| **Country** | **Observation** |
| United Kingdom | Dominates total sales, far ahead of all other countries |
| Other EU countries | Small contribution to total sales |
| Non-EU countries | Minimal contribution |

• This indicates that the store’s main customer base is located in the United Kingdom.

• The store has **limited penetration in non-UK markets**, especially outside Europe.

• This presents a potential **global expansion opportunity** if localized marketing strategies are applied.

**Next Steps**

• Investigate which product categories are most popular in the UK.

• Analyze **return rates by country** to see if certain regions have more complaints or quality issues.

• Consider launching localized marketing campaigns for underperforming countries to boost international sales.

**📊 코드 정리 (완성본)**

*# Total Sales*

df["TotalSales"] = df["Quantity"] \* df["UnitPrice"]

*# Total sales in each country*

sales\_by\_country = df.groupby("Country")["TotalSales"].sum().sort\_values(ascending=False)

*# visualization*

plt.figure(figsize=(12, 6))

sales\_by\_country.plot(kind="bar", color="skyblue")

plt.title("Total Sales by Country")

plt.ylabel("Total Sales")

plt.xticks(rotation=45)

plt.grid(True)

plt.show()

A graph with lines and text

AI-generated content may be incorrect.

**📊 Final E-commerce Data Analysis Report (LaTeX)**

\documentclass{article}

\usepackage{graphicx} % Package for including images

\usepackage{amsmath} % Package for mathematical equations

\begin{document}

\title{E-commerce Data Analysis Report}

\author{Yelin}

\date{\today}

\maketitle

\section{Introduction}

This report presents an analysis of e-commerce data, focusing on key performance indicators such as monthly sales trends, most returned products, sales distribution by country, and the Average Order Value (AOV).

\section{1. Monthly Sales Trend}

\subsection{Analysis}

The total sales for each month were calculated using:

\[

\text{TotalSales} = \text{Quantity} \times \text{UnitPrice}

\]

The monthly trend is visualized in the figure below.

\begin{figure}[h]

\centering

\includegraphics[width=0.8\textwidth]{Figure\_1.png}

\caption{Monthly Sales Trend}

\end{figure}

\section{2. Top 5 Most Returned Products}

\subsection{Analysis}

The top 5 most returned products were identified by filtering transactions with negative quantities and summing the absolute values of the returns for each product. The results are shown in the bar chart below.

\begin{figure}[h]

\centering

\includegraphics[width=0.8\textwidth]{Figure\_2.png}

\caption{Top 5 Most Returned Products}

\end{figure}

\section{3. Sales by Country}

\subsection{Analysis}

Sales were aggregated by country to identify the regions contributing the most revenue. The results are presented in the figure below.

\begin{figure}[h]

\centering

\includegraphics[width=0.8\textwidth]{Figure\_3.png}

\caption{Total Sales by Country}

\end{figure}

\section{4. Average Order Value (AOV)}

\subsection{Calculation}

The Average Order Value (AOV) was computed as follows:

\[

\text{AOV} = \frac{\sum \text{TotalSales}}{\text{Total Orders}}

\]

where:

\begin{itemize}

\item \(\sum \text{TotalSales}\) represents the sum of all order values.

\item \(\text{Total Orders}\) is the number of unique invoices (transactions).

\end{itemize}

\subsection{Result}

The calculated AOV for this dataset is:

\[

\textbf{AOV} = 376.36

\]

\section{Conclusion}

This analysis provides insights into sales trends, product returns, and regional sales performance. The findings can assist in optimizing business strategies and improving operational efficiency.

\end{document}