Online Retail EDA & Customer Insights Report

1. Project Overview

This project focuses on performing Exploratory Data Analysis (EDA) for an online retail business. The goal is to understand sales performance, customer purchasing behavior, and seasonal patterns that can help the company make better business decisions.

2. Business Problem

A small e-commerce company wants to analyze its past sales data to gain insights into how different factors influence revenue. The business seeks to identify top-selling products, seasonal and weekly sales trends, and customer purchasing patterns. These insights will support inventory planning, marketing, and customer relationship strategies.

3. Objectives

- Clean and preprocess the raw dataset.
- Perform detailed Exploratory Data Analysis (EDA).
- Identify top-selling products, seasonal patterns, and customer behavior.
- Segment customers based on purchasing activity.
- Provide business insights and recommendations for growth.

4. Data Requirements

The dataset contains transactional-level information, including order details, products, customers, and sales amounts.

Feature	Description
InvoiceNo	Unique transaction ID
StockCode	Product code
Description	Product name
Quantity	Quantity of products purchased
Price	Unit price of the product
InvoiceDate	Date of the transaction
CustomerID	Unique customer identifier
Country	Country where the order was placed

Dataset Source: Kaggle – Online Retail II Dataset (UCI)
Total Records: 541,909 transactions

5. Data Collection

The dataset was downloaded from Kaggle, which provides real transactional data from a UK-based online retail store. It contains purchase information for the years 2009–2011, covering multiple products and countries.

6. Data Validation

- Verified data types using df.info() and df.dtypes.
- Checked for duplicate records using df.duplicated().sum().
- Identified missing values using df.isnull().sum().
- Confirmed that InvoiceDate is in correct datetime format.

7. Data Cleaning and Preparation

Cleaning Steps:

- Removed duplicate entries.
- Dropped rows with missing Customer ID for accurate customer analysis.

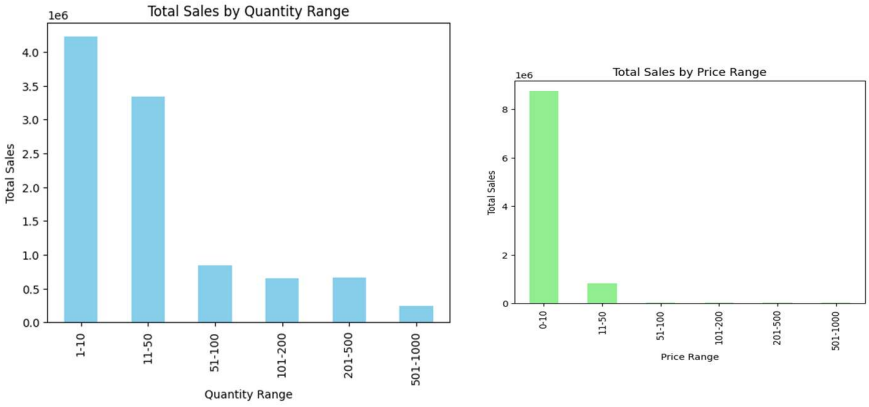
Feature Engineering:

- Created TotalSales = Quantity × Price.
- Extracted Month, Year, and Weekday from InvoiceDate.
- Retained outliers for realistic representation of large orders.

8. Exploratory Data Analysis (EDA)

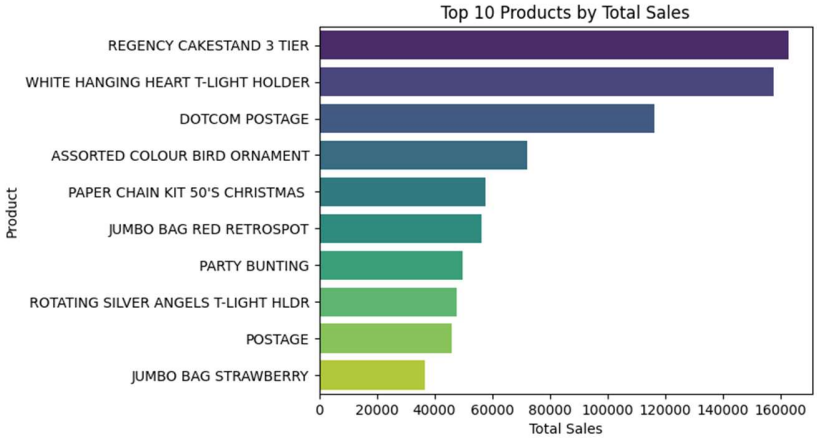
- Sales by Quantity and Price:

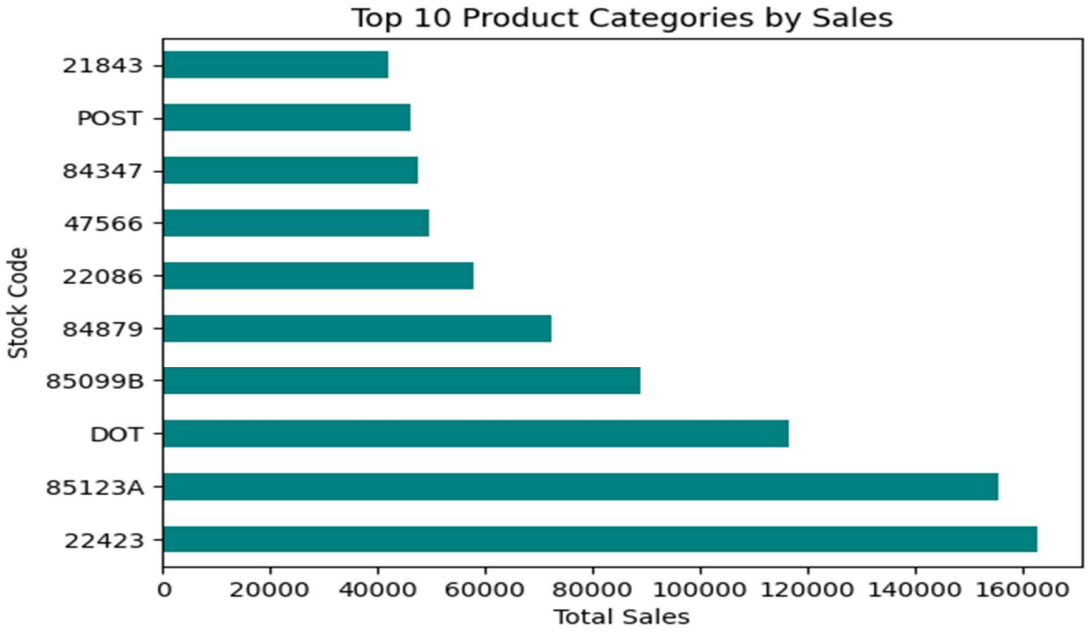
Most sales come from moderate Quantity and Price ranges. Very large orders are rare and often represent bulk purchases or returns.



- Top Products and Categories:

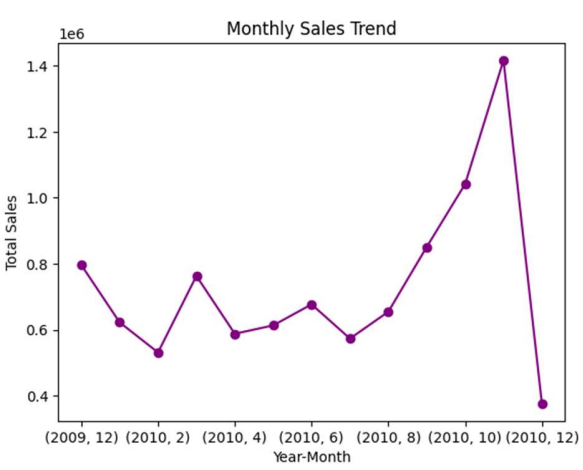
A few products and categories generate the majority of sales. These should be prioritized for restocking and promotional offer.

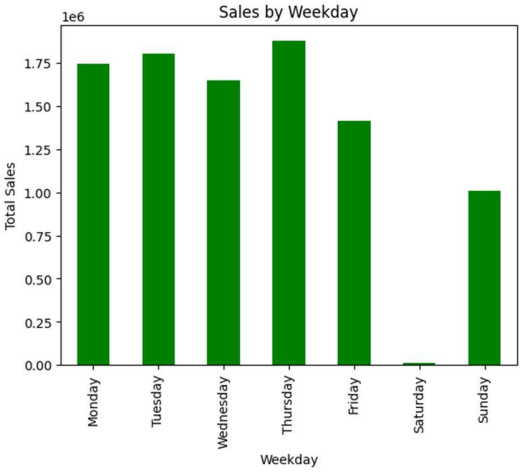




- Sales Trends:

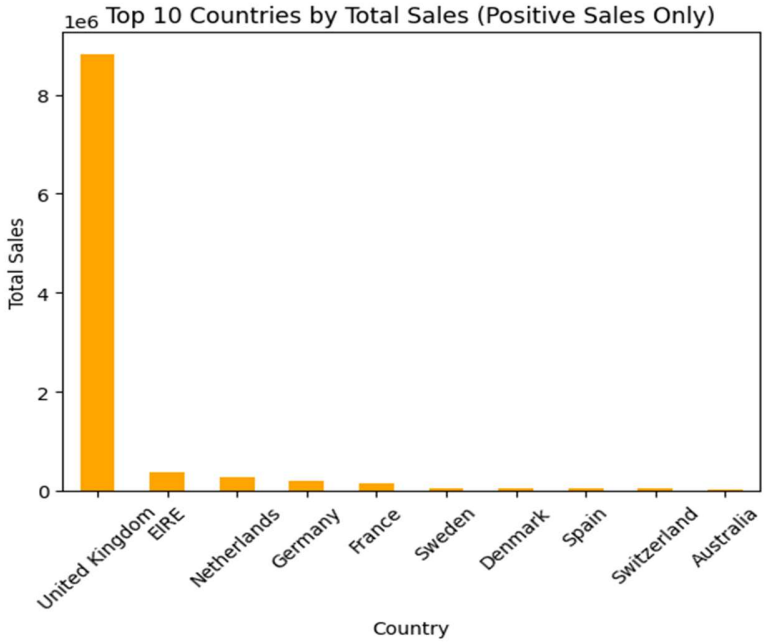
Monthly analysis shows peak sales in specific months. Weekly trends indicate higher sales on certain weekdays. Such patterns are useful for planning marketing campaigns.





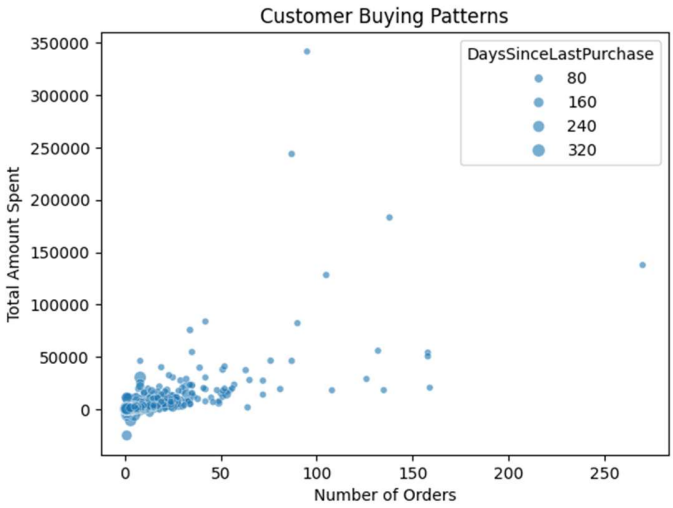
- Sales by Country:

The top 10 countries contribute most to total revenue. The UK leads in total transactions and sales volume.



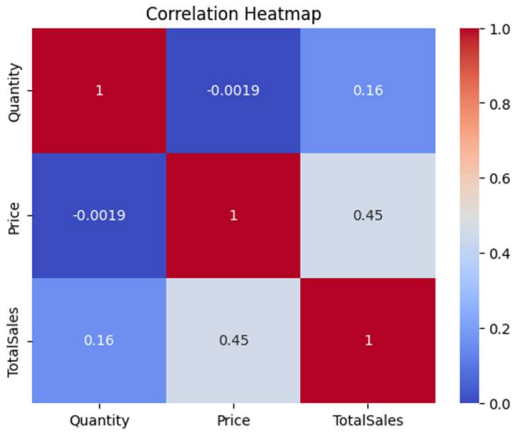
- Customer Segmentation (Purchase Behavior):

Some customers purchase frequently and spend significantly more. Others buy occasionally, indicating opportunities for targeted marketing or loyalty programs.



- Correlation Analysis:

Quantity and Price have a strong influence on Total Sales. Positive correlation shows that higher quantity and higher price both increase revenue.



9. Analysis and Insights

- Majority of sales are driven by moderately priced products.
- Certain products and categories consistently outperform others.
- Seasonal and weekly trends reveal predictable demand cycles.
- A small percentage of loyal customers contribute to a large share of total revenue.
- Quantity and Price directly impact overall sales performance.

10. Business Recommendations

- Focus on top-selling products and countries to maintain steady revenue.
- Offer discounts or loyalty rewards to frequent high-spending customers.
- Use sales trend insights for inventory and promotional planning.

- Maintain balanced pricing strategies to attract both small and bulk buyers.

11. Conclusion

This analysis helped uncover clear patterns in sales and customer behavior. The insights can guide marketing decisions, stock management, and customer engagement strategies. By leveraging these findings, the business can improve profitability, strengthen customer relationships, and plan more effectively for future demand.

12. Tools and Technologies Used

- Python: Data cleaning and analysis
- Pandas & NumPy: Data manipulation
- Matplotlib & Seaborn: Visualization
- Google Colab: Development environment

13. References

- Kaggle – Online Retail II Dataset (UCI)
- UCI Machine Learning Repository
- Python, Pandas, and Seaborn Official Documentation