Exploratory Data Analysis: Insights from Online Retail Dataset

# Introduction:

This analysis dives into a real-world dataset from an online retail business based in the UK. The data spans a year’s worth of transactions and was sourced from the UCI Machine Learning Repository. My goal was to clean up the data, uncover trends, and pull-out meaningful insights—like which products sold best, who the top customers were, and how sales varied across countries and time.

# How I Did It (Methodology):

I used Python and a few popular libraries to get the job done:

* Pandas for data manipulation
* Matplotlib & Seaborn for visualizations
* NumPy for numerical operations

# The steps included:

1. Loading and understanding the structure of the dataset
2. Cleaning it up by removing duplicates and handling missing or invalid entries
3. Calculating a new column: *Total Revenue* (Quantity × Unit Price)
4. Aggregating data to analyse products, customers, and countries
5. Studying monthly sales patterns
6. Creating visual summaries like bar charts and line graphs

# What I Found (Key Insights):

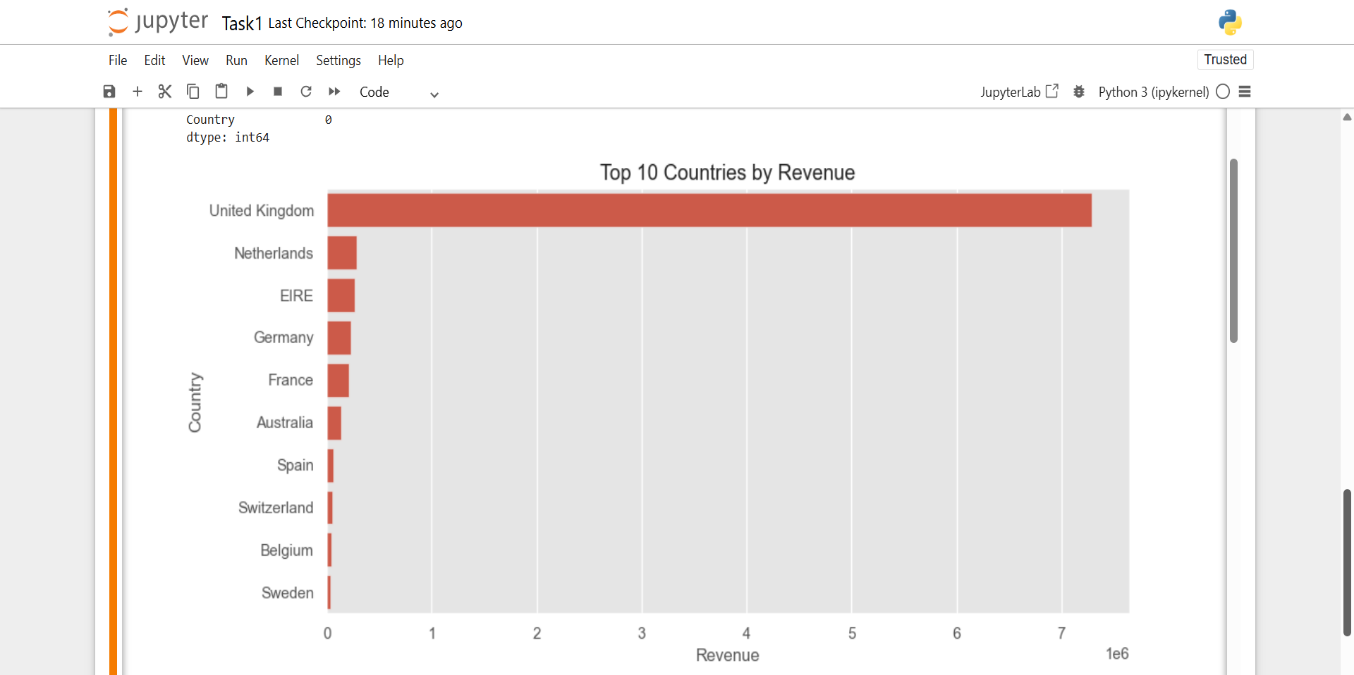
These are some of the standout insights:

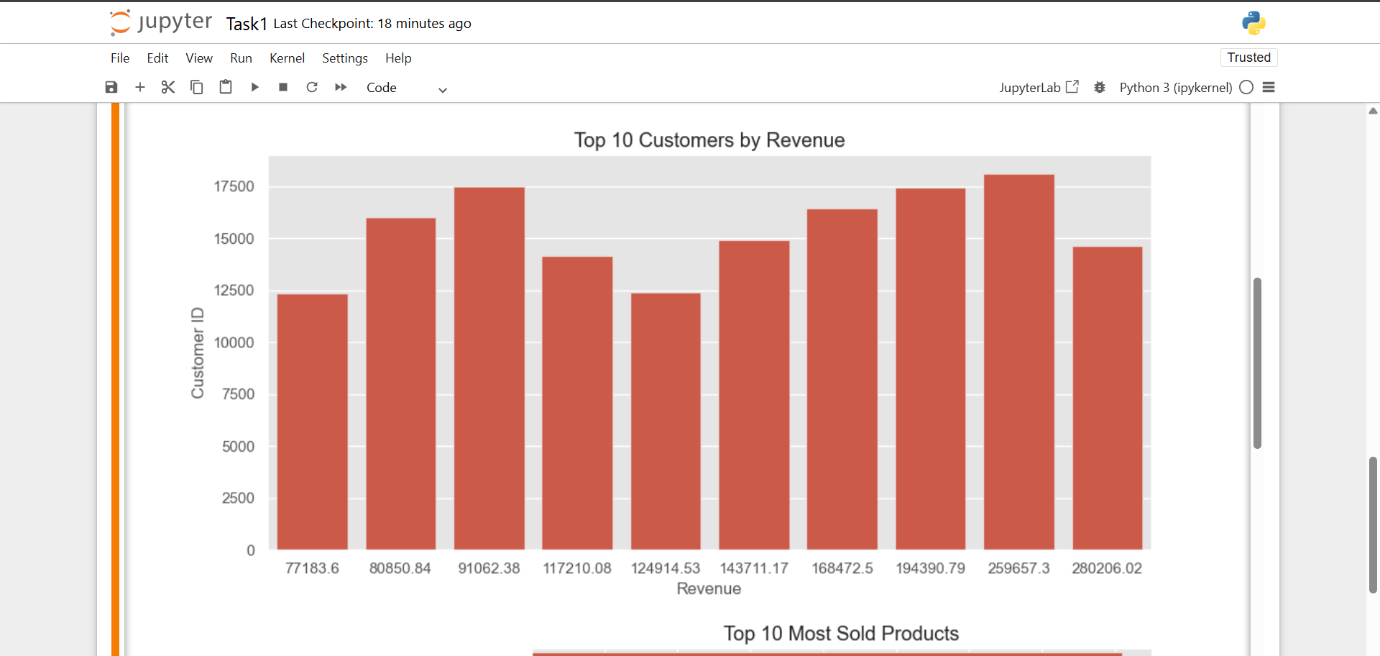
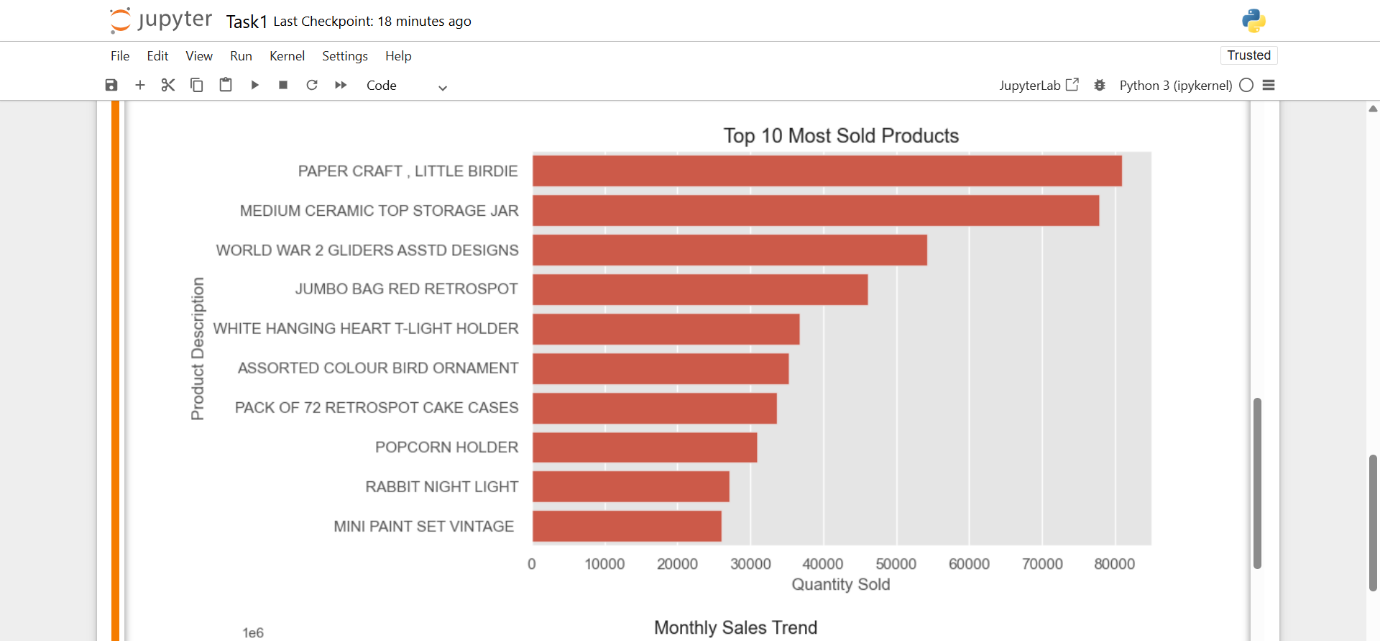
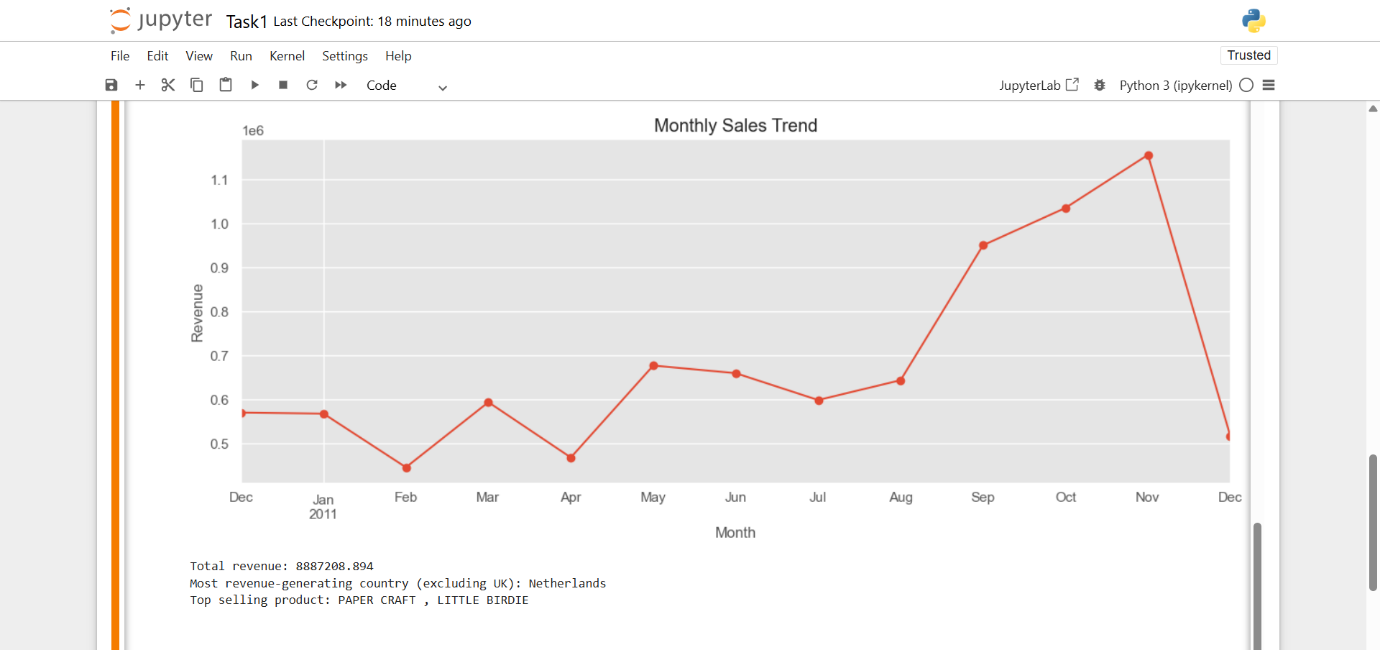
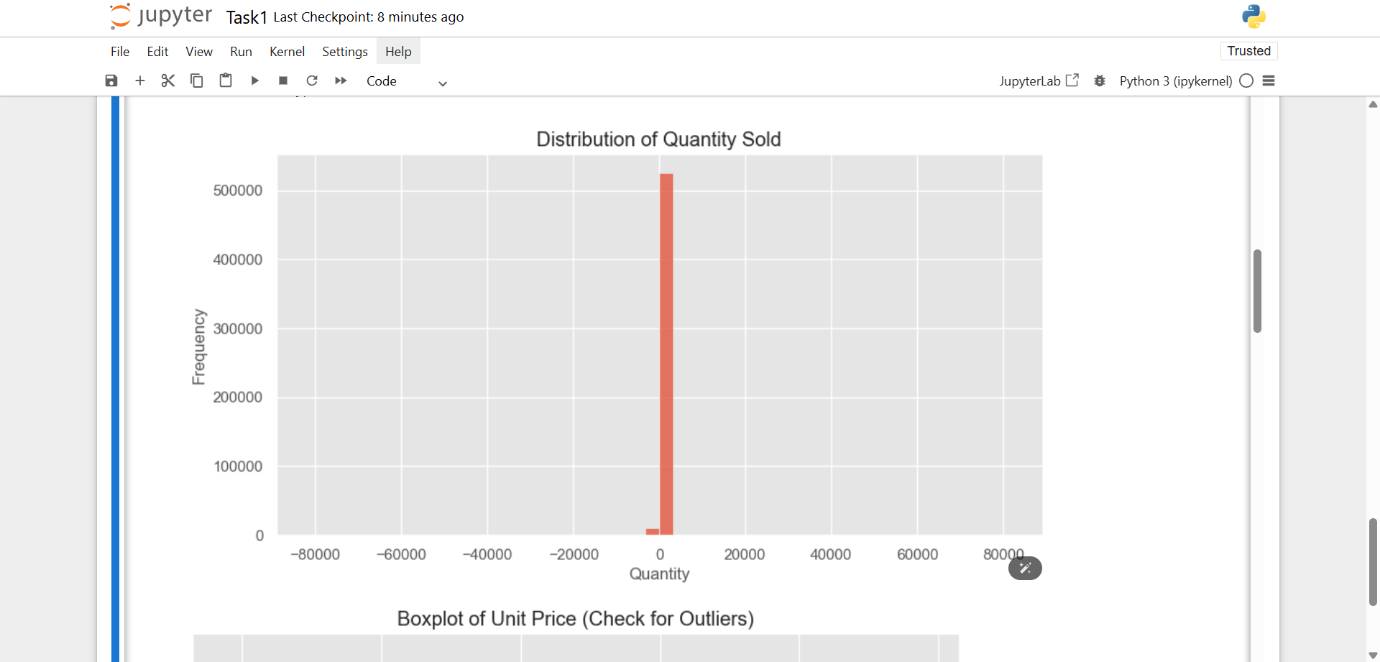
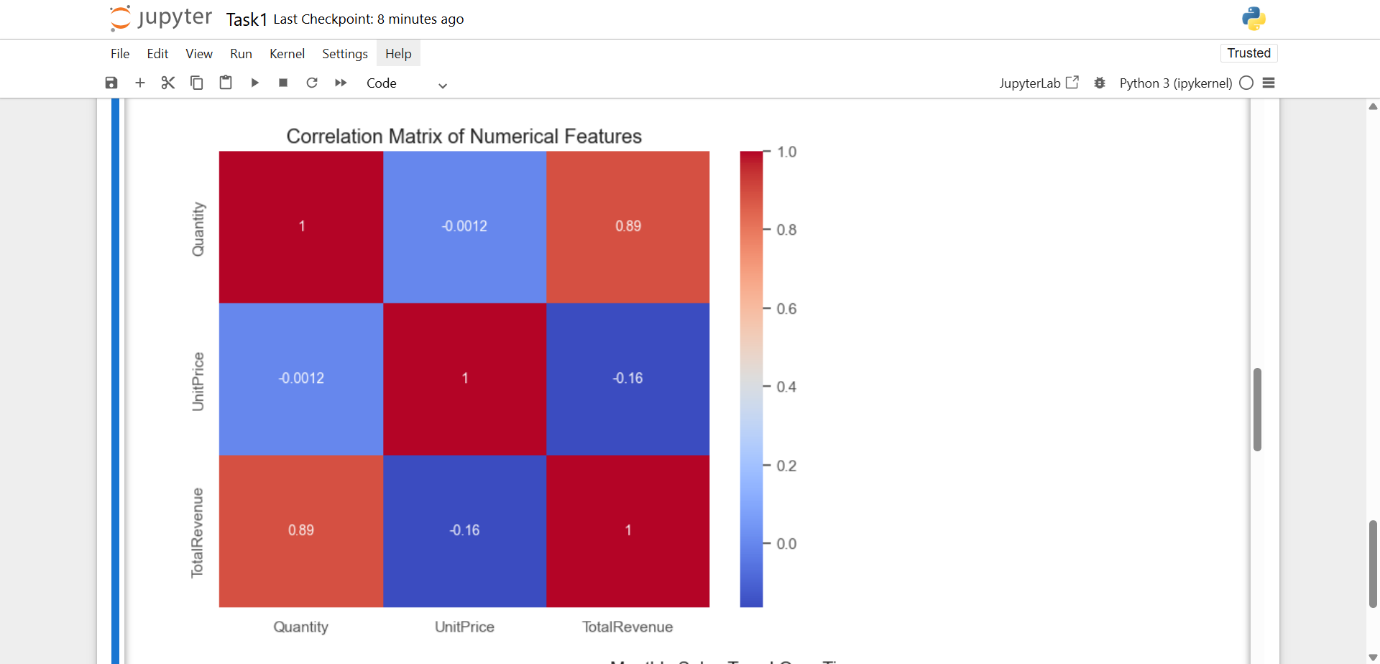
* The UK led in revenue, contributing the lion’s share of total sales.
* A single product stood out as the top seller by quantity—and it had a major impact on overall revenue.
* Top 10 customers accounted for a significant portion of the income, showing how important high-value shoppers are.
* Sales spiked during the holiday season, pointing to strong seasonal demand.
* Netherlands and Germany came in second and third after the UK in terms of international revenue.

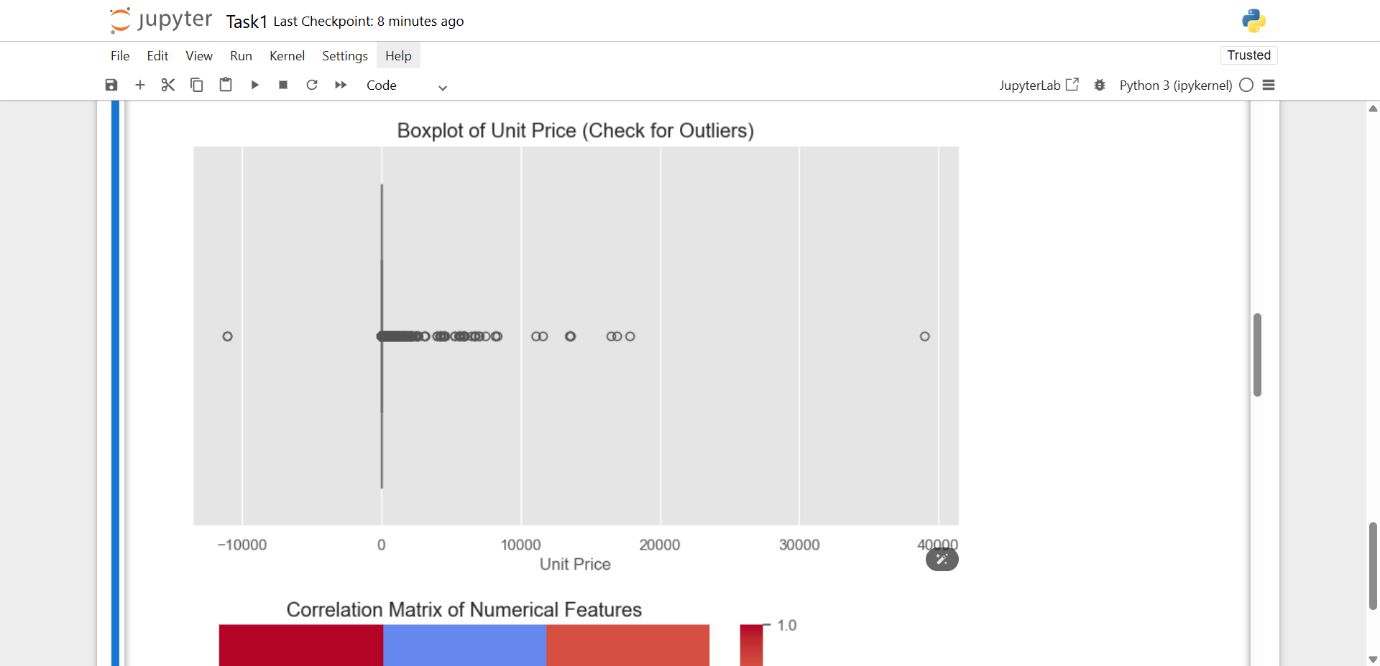
# Visual Highlights:

To help make sense of the numbers, I created several visualizations:

* Bar chart showing the top 10 countries by revenue:



* Chart highlighting the top-spending customers:
* Visualization of the most-sold products:
* Line graph tracking monthly sales trends over the year:
* Correlation Matrix(Heatmap), Histograms(Boxplot, Count plot): 



# Conclusion:

The EDA helped me better understand customer behaviour and sales performance:

* Cleaning the data was crucial—it removed thousands of problematic records and improved accuracy.
* Seasonal trends suggest opportunities to boost sales through targeted promotions during peak times.
* Focusing inventory and marketing efforts on top-performing products could drive more revenue.
* Offering loyalty programs or special offers for high-spending customers might be a smart move for retention and growth.