Advanced Data Visualization
Experiment - 1

Akshaya Karande
2021300056
BE COMPS A - BATCH H

Aim:

Analyze e-commerce dataset using various charts like bar graph, pie chart, time series charts and observe product wise and region wise sales.

Description of Dataset:

The dataset is a typical e-commerce dataset containing 541,909 entries and 8 columns.

1. Columns in the Dataset:

- InvoiceNo: A unique identifier for each transaction. It's an alphanumeric code (object type).
- StockCode: A unique code assigned to each distinct product in the inventory (object type).
- Description: A brief description of each product. There are some missing values in this column (object type).
- Quantity: The number of units of the product sold in each transaction. The values vary widely, indicating both high and low sales volumes (int64 type).
- InvoiceDate: The date and time when the transaction was made. It's in a string format (object type), which can be converted to a datetime format for time series analysis.
- UnitPrice: The price of one unit of the product. Some negative values may represent refunds or errors (float64 type).
- CustomerID: A unique identifier for each customer. This column has missing values, which could be due to anonymous purchases (float64 type).
- Country: The country where the customer is located. This will be useful for region-wise sales analysis (object type).

2. Dataset Characteristics:

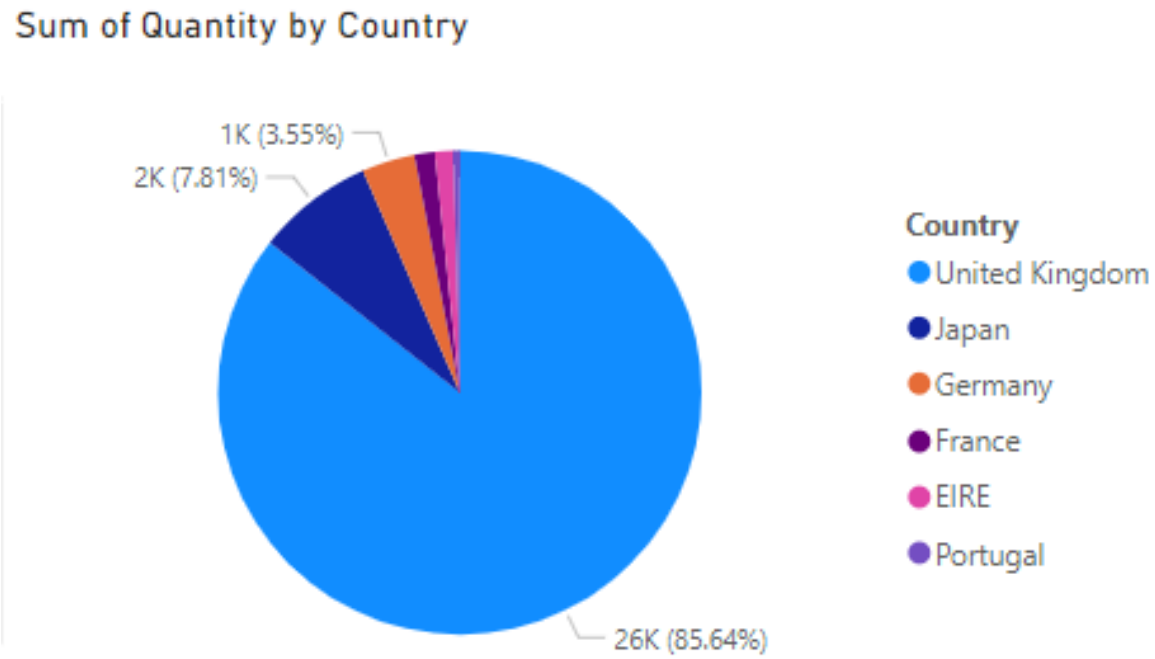
- The dataset contains sales data from various regions, with the majority of entries likely from the United Kingdom.
- The date range of the sales data is not yet determined but will be critical for performing time series analysis.
- Product-wise analysis can be conducted using the StockCode, Description, and Quantity columns.
- Region-wise analysis can be performed using the Country column.

3. Potential Issues:

- Missing Data: The Description and CustomerID columns have missing values, which need to be handled during analysis.
- Negative Values: The Quantity and UnitPrice columns contain negative values, which may represent returns or errors and should be investigated.
- Data Format: The InvoiceDate is currently in a string format and needs to be converted to a datetime format for time series analysis.

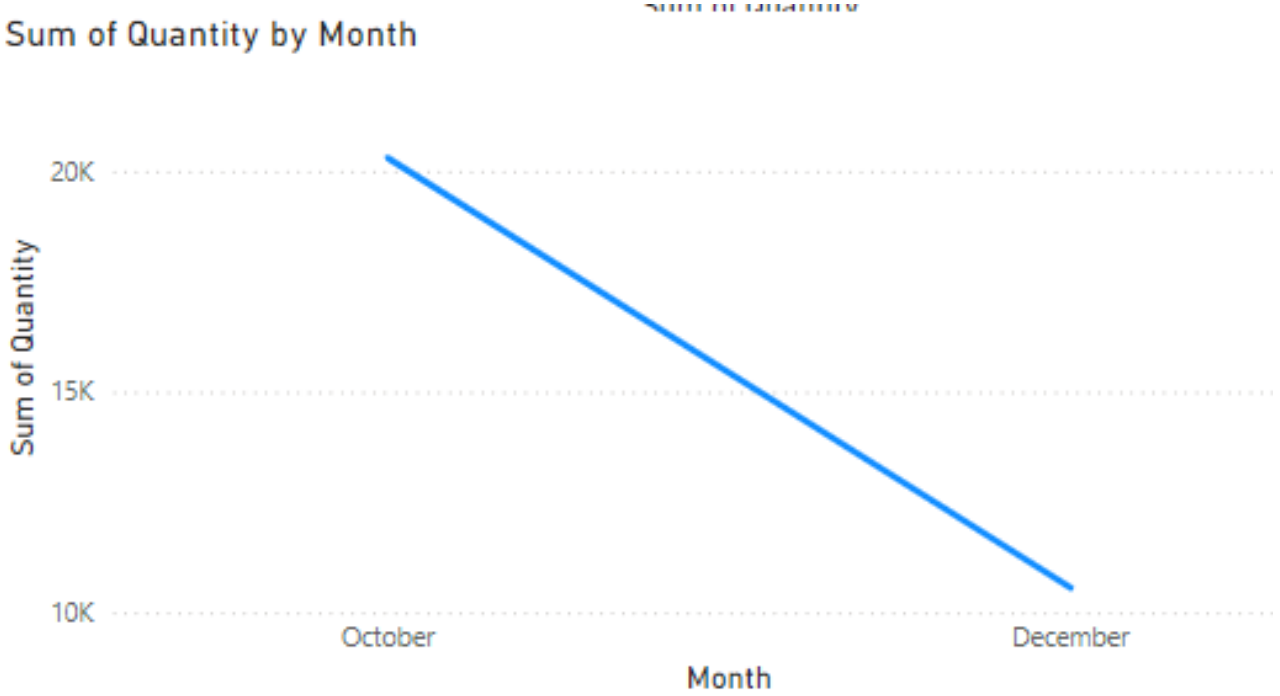
Charts and Analysis:

1) Pie Chart - Sum of Quantity by Country



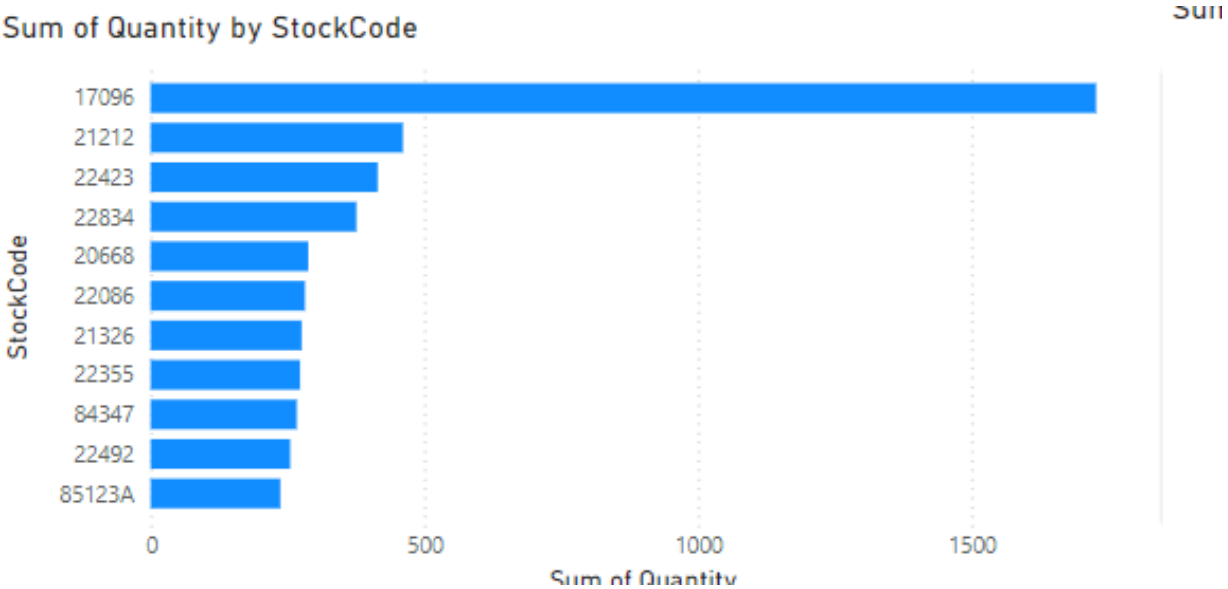
1. **United Kingdom (UK)** dominates the sales, accounting for 85.64% of the total quantity sold.
2. Other notable regions include **Japan (7.81%)**, **Germany (3.55%)**, **France**, **EIRE**, and **Portugal** with significantly smaller shares.
3. This indicates that the majority of sales are concentrated in the UK, with other regions contributing relatively smaller portions.

2) Time Series - Sum of Quantity by Month



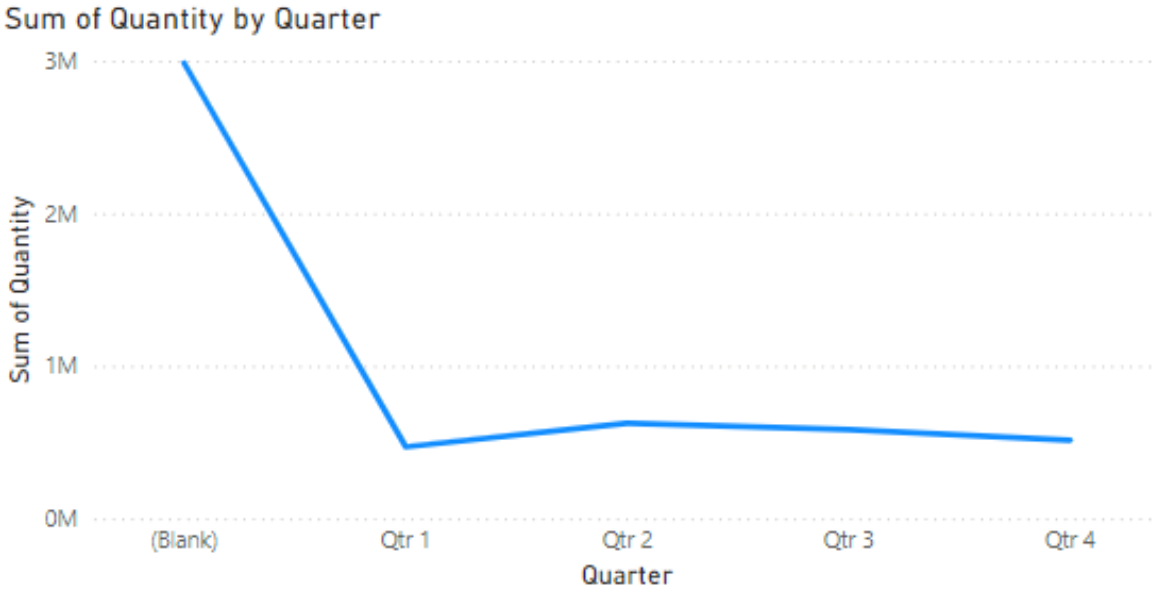
1. The chart shows a downward trend in the sum of quantities sold from **October** to **December**.
2. The sales quantities decrease sharply, indicating a potential seasonal drop in demand during this period.
3. The chart could suggest that there might be a need for marketing efforts or promotions in the months with declining sales.

3) Bar Chart - Sum of Quantity by StockCode



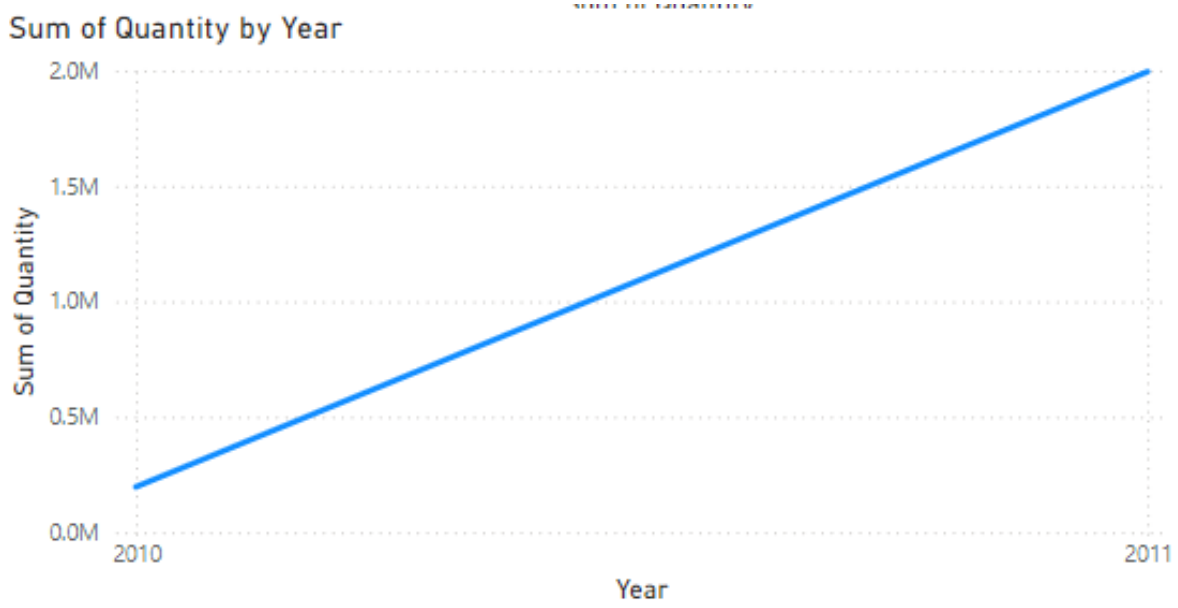
1. Majority of StockCodes have quantities in the range of 0-500 and they might not have a strong market presence.
2. StockCode 17096 has the highest quantity.

4) Time Series Chart - Sum of Quantity by Quarter



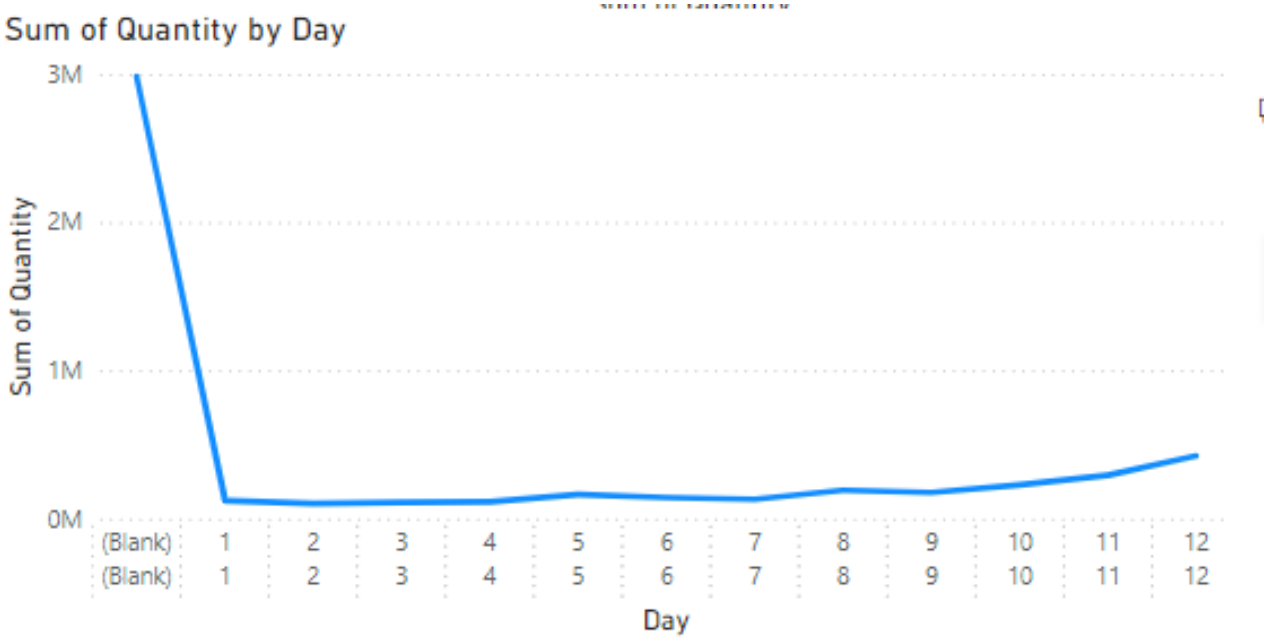
1. The quarterly analysis reflects a similar trend to the monthly chart, with Q1 showing a substantial peak in sales.
2. After Q1, there's a noticeable decline in sales quantities, which then stabilizes across Q2, Q3, and Q4.
3. This trend reinforces the seasonality observed in the monthly analysis, suggesting that most sales occur early in the year.

5) Time Series Chart - Sum of Quantity by Year



1. This chart shows a year-over-year increase in the total quantity of products sold.
2. The consistent upward trend suggests overall growth in sales from 2010 to 2011, which could indicate a successful period for the business, potentially due to expanding market reach, increasing customer base, or effective sales strategies.

6) Time Series Chart - Sum of Quantity by Day



1. This daily analysis shows a significant peak at the beginning of the period, after which the sales drop sharply and then remain relatively steady.
2. The initial peak could correspond to a specific event, such as a major sales promotion or the launch of a new product, followed by a period of more consistent daily sales.

Conclusion:

1. The data demonstrates a clear seasonal pattern, with significant sales occurring at the beginning of the year (January and Q1), followed by a stabilization in sales quantities.
2. Yearly growth in sales quantity indicates positive business performance, potentially reflecting successful strategies or market conditions during the period.
3. The day-by-day and month-by-month analyses highlight how short-term events can have substantial impacts on sales, which is crucial for understanding consumer behavior and planning future promotions.