

Experiment 1
Hardik Garg
2021300036
BE Comps A - BATCH H

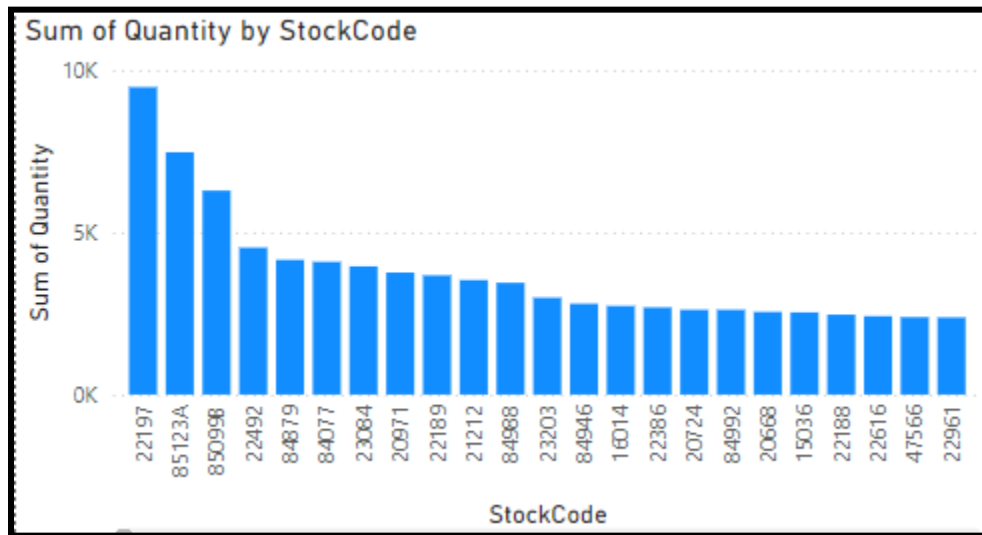
Aim: Analyze e commerce dataset using various charts like bar graph, pie chart, time series charts and observe product wise and region wise sales.

1. Columns in the Dataset:

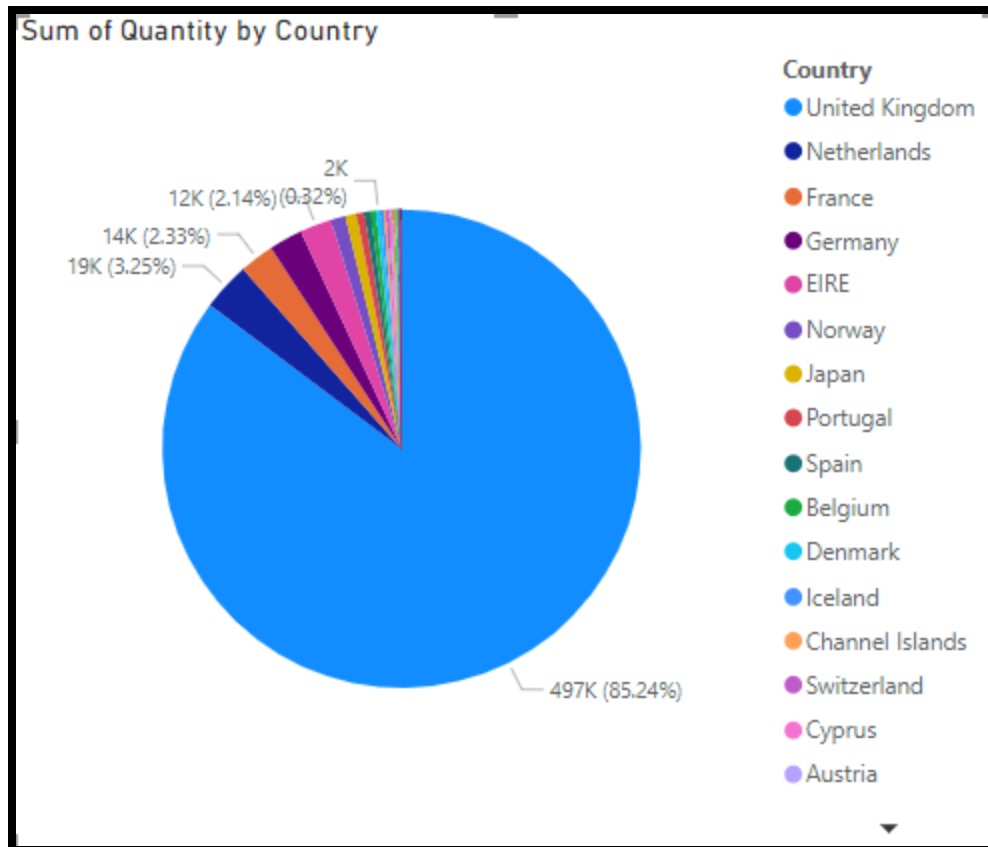
- **InvoiceNo:** A unique identifier for each transaction (alphanumeric, object type).
- **StockCode:** A unique code assigned to each distinct product in the inventory (object type).
- **Description:** A brief description of each product, with some missing values (object type).
- **Quantity:** The number of units of the product sold in each transaction, with values ranging widely, indicating varied sales volumes (int64 type).
- **InvoiceDate:** The date and time of each transaction, currently in string format (object type), which should be converted to a datetime format for time series analysis.
- **UnitPrice:** The price per unit of the product, with some negative values that may represent refunds or errors (float64 type).
- **CustomerID:** A unique identifier for each customer, with missing values that could be due to anonymous purchases (float64 type).
- **Country:** The country where the customer is located, useful for region-wise sales analysis (object type).

2. Analysis Goals:

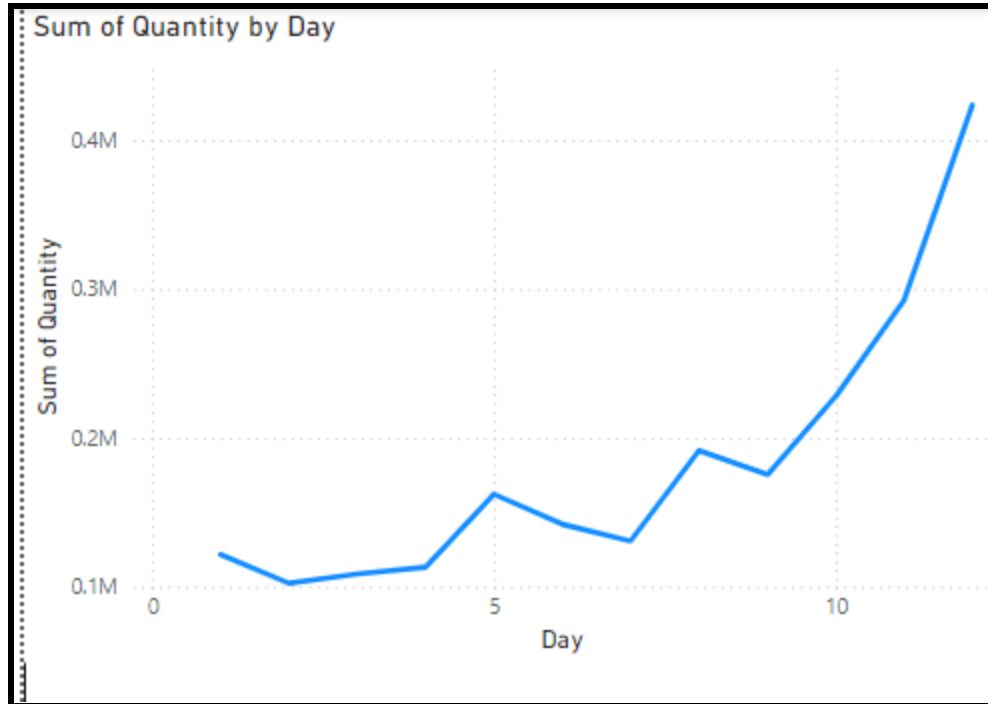
- **Product-Wise Sales Analysis:** Aggregate the Quantity column by StockCode to identify top-selling and low-selling products.
- **Region-Wise Sales Analysis:** Group data by Country to observe sales trends across different regions.
- **Time Series Analysis:** Convert InvoiceDate to datetime format and analyze sales trends over time by grouping data by time intervals (e.g., daily, monthly).
- **Visualization:** Use bar graphs for product-wise sales, pie charts for region-wise distribution, and line charts for time series trends.



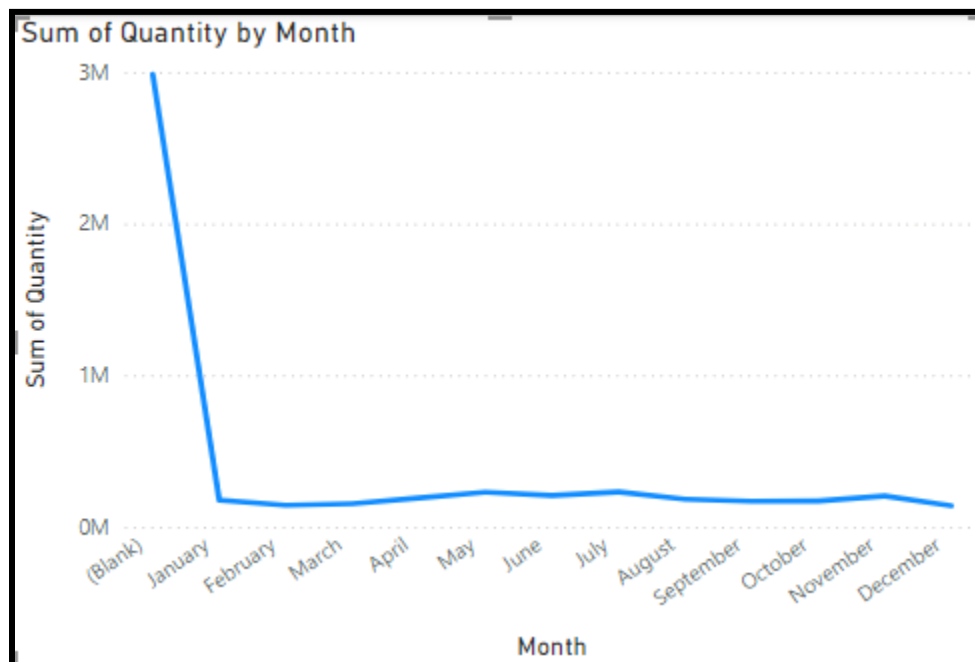
1. The product with stockCode 22197 has the highest quantity sold, nearly reaching 10k units. It significantly outperforms others in terms of sales in volume.
2. Majority of the stockcodes listed in the graph have quantities in the 4000 to 5000 range. These products are likely consistent sellers but may not have as strong of a market presence as the top 3.



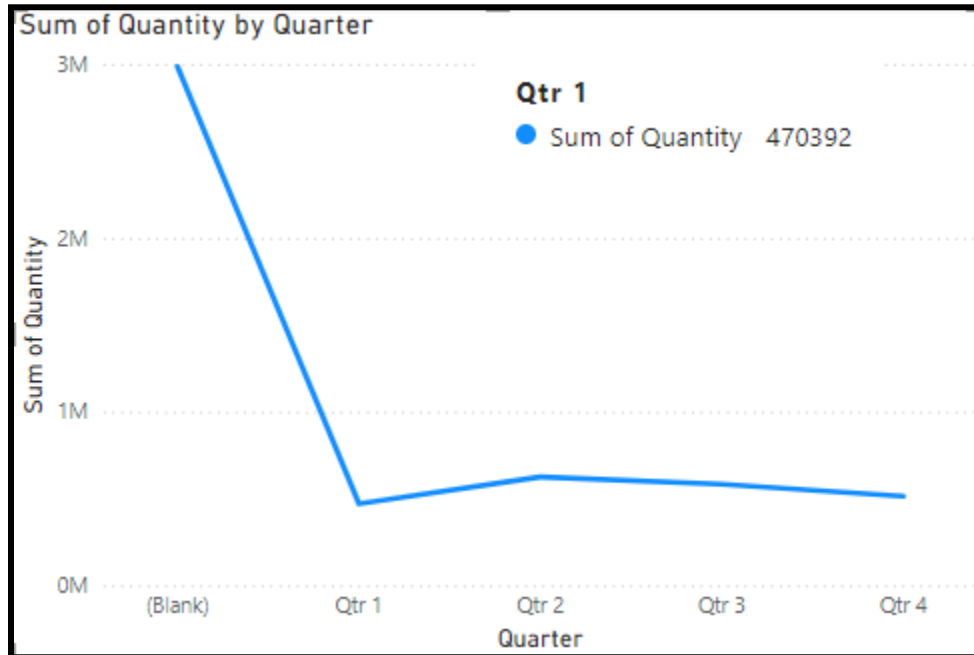
1. **Dominance of United Kingdom:** The United Kingdom (UK) overwhelmingly dominates the sales quantity, accounting for 85.24% of the total quantity sold. This indicates that the majority of transactions or sales come from the UK, making it the primary market for this dataset.
2. **Minor Contributions from Other Countries:** Other countries such as the Netherlands, France, Germany, and EIRE have smaller contributions. The quantities sold in these countries range between 2,000 to 19,000 units, with each country contributing between 0.32% to 3.25% of the total sales quantity. These countries represent secondary markets.



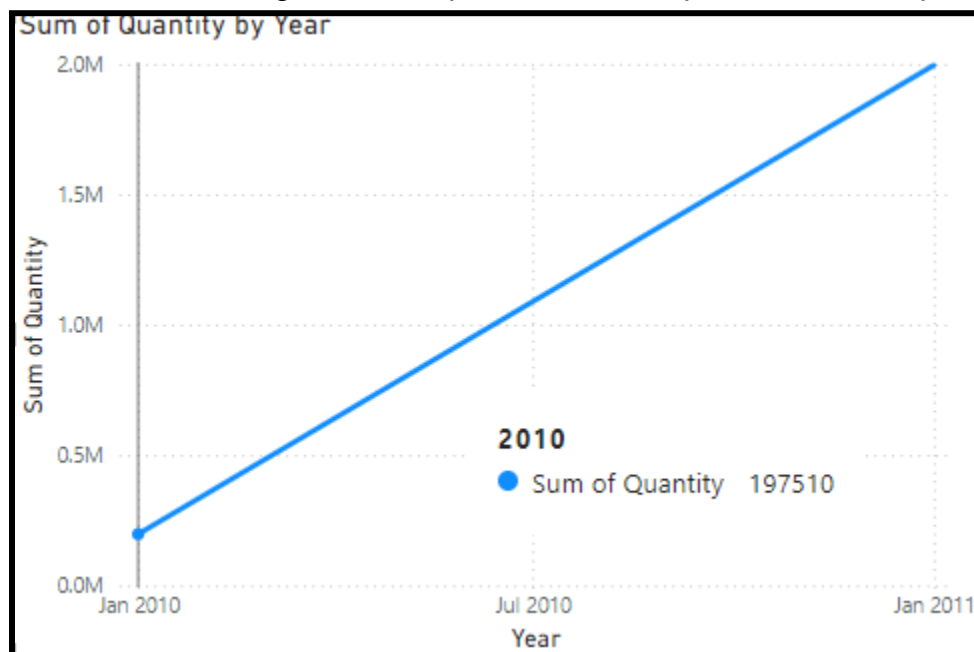
1. There is a growth of total quantity of products sold day by day and a significant increase by 4 times.
2. There is an occasional dip in the quantity of products sold which might be due to market fluctuations.



1. There is a significant drop after the initial peak in January then there is a very low number of products sold, which might indicate that people might be buying on new year's more or there might be a flash sale occurring at the start of the year.



1. There is a significant drop after the initial peak in the first quarter of the year.



There is a steady increase in the number of products sold year by year and the ecommerce site must be getting popular.

Conclusion:

- **Primary Market Focus:** The dataset is heavily skewed towards the UK market, which should be the primary focus for any targeted marketing or sales strategies.
- **Volatile Sales Patterns:** The line chart highlights the importance of understanding the reasons behind the fluctuating sales quantities, which could involve deeper analysis of seasonal trends, product launches, or economic conditions during the periods in question.