Retail Sales Analysis

Project Overview

Project Title:
Retail Sale Analysis

**Tools and Technologies used:**
Python, Jupiter Notebook, MySQL, Matplotlib, Seaborn, and Excel.

Objective:
This project is designed to demonstrate SQL and Python skills and techniques typically used by data analysts to explore, clean, and analyse retail sales data. The project involves setting up a retail sales database, performing exploratory data analysis (EDA), and answering specific business questions through SQL queries and Excel. This project is ideal for those who are starting their journey in data analysis and want to build a solid foundation in SQL and Python.

Project Structure

1. **Set up a retail sales database:** Create a retail sale database with provided sales data transactions and sales data responses which are inserted into respective tables.
2. **Importing the dataset:** Using Pandas Library, import the sales data transactions, sales data responses and merge them to single dataset and understand the dataset to perform further analysis.
3. **Data Cleaning:** Identify and remove any records with missing or null values.
4. **Exploratory Data Analysis:** Perform basic exploratory data analysis to understand and get some insights in the dataset.
5. **Business Analysis**: Use SQL, Advanced Analytics and Excel to answer specific business questions and derive insights from the sales data.
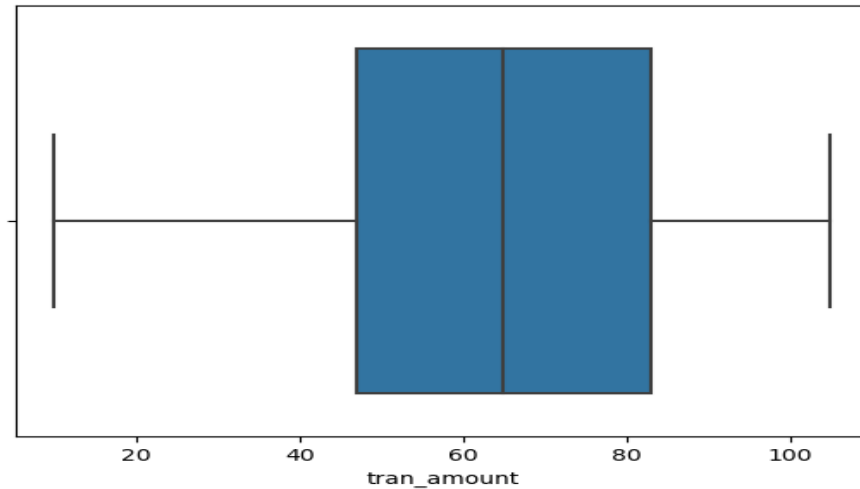
♦ The boxplot reveals a nearly symmetrical distribution of transaction amounts. The median is around 60, and the IQR spans from 50 to 80. No outliers are visible, suggesting a consistent range of transaction values.
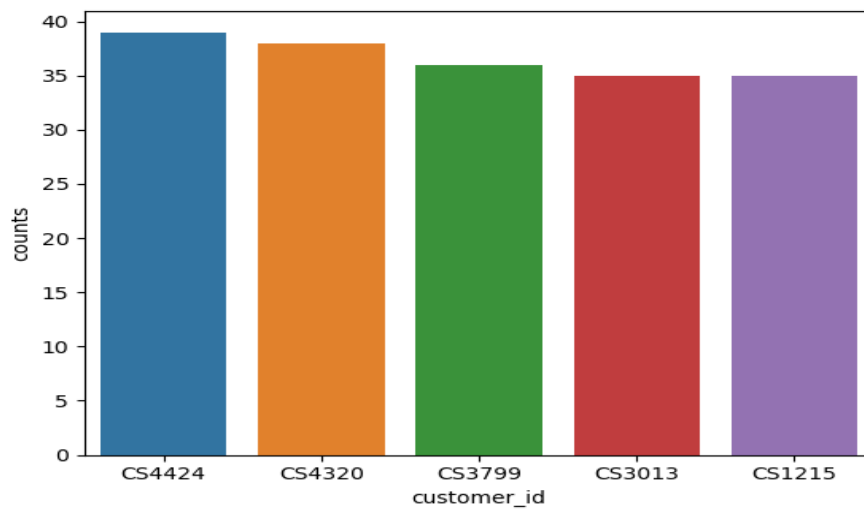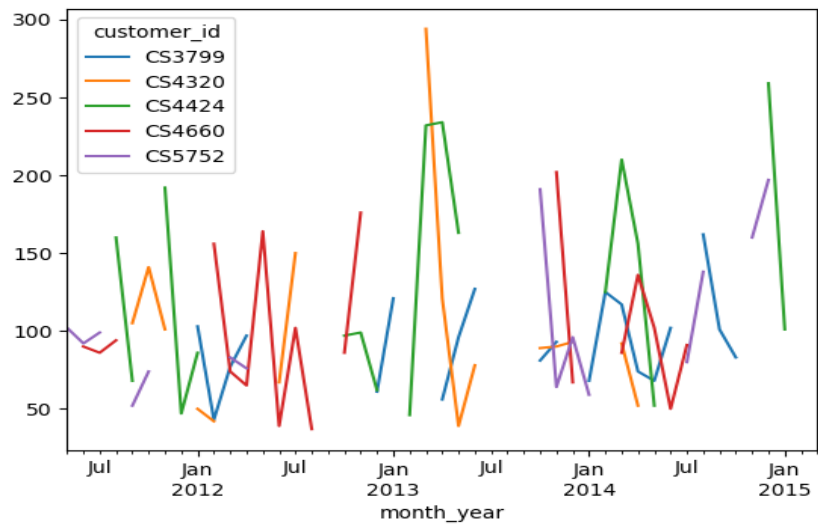

♦ The bar chart shows the top 5 customers with the highest number of transactions. Customer CS4424 had the most transactions, followed by CS4320, CS3799, CS3013, and CS1215. The chart also shows that the number of transactions decreases as we move from CS4424 to CS1215.
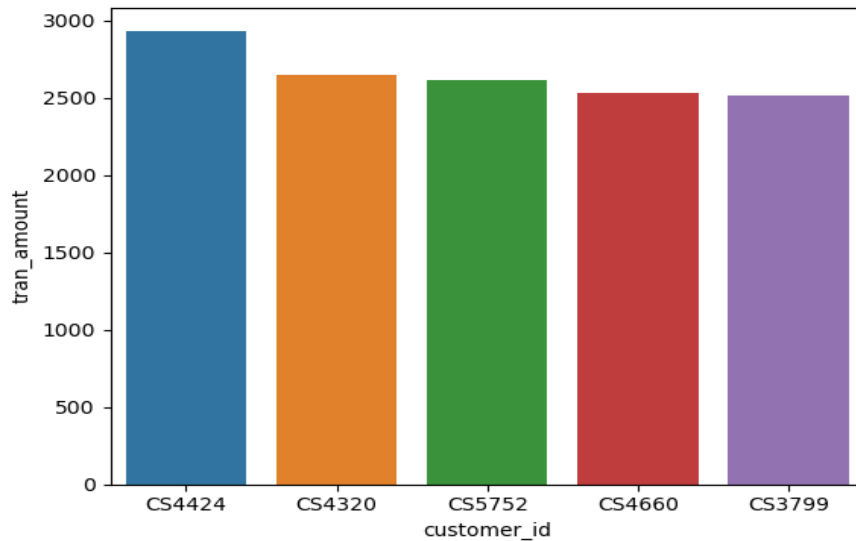
- ♦ The bar chart shows the top 5 customers with the highest transaction amounts. Customer CS4424 had the highest transaction amount, followed by CS4320, CS5752, CS4660, and CS3799. The chart also shows that the transaction amount decreases as we move from CS4424 to CS3799.
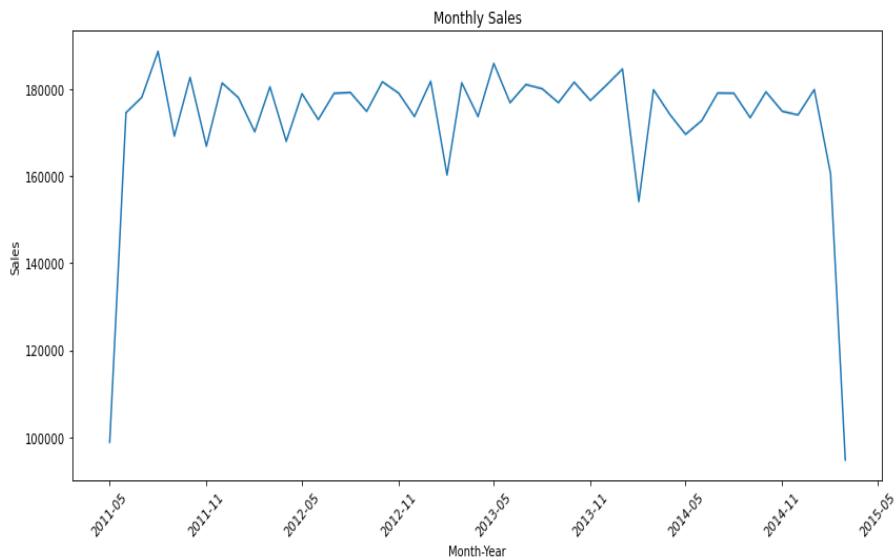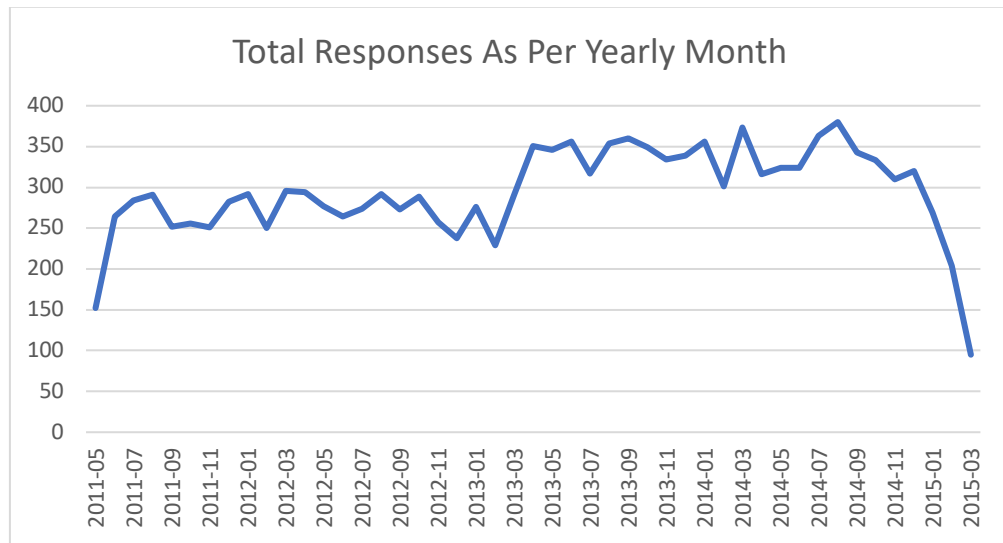

- ♦ The line chart shows monthly sales over a period of several years. Sales fluctuate over time, with peaks and troughs. There are periods of high sales followed by periods of lower sales. Overall, there appears to be a slight downward trend in sales over the entire period.

♦ The line chart shows the monthly sales for the top 5 customers over a period of several years. The customers have different sales patterns with some customers showing more consistent sales and others having more fluctuating sales. Overall, there appears to be a slight downward trend in sales for most customers over the entire period.
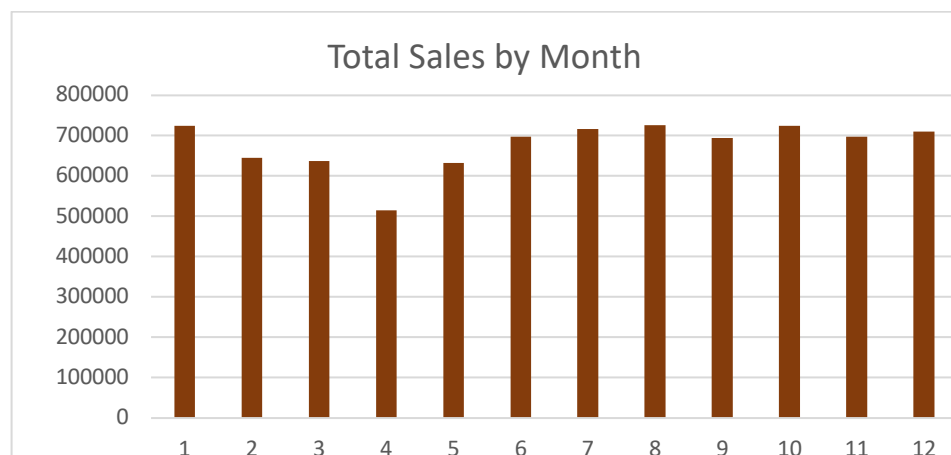

♦ The bar chart shows the total sales for each month of the year. Overall, sales are relatively consistent throughout the year, with some months having slightly higher sales than others. There is a significant drop in sales in month 4, and a slight increase in sales in month 12.

Total Responses As Per Yearly Month

- The line chart shows the total responses per month over a period of several years. There is a significant increase in responses in 2011, followed by a period of fluctuation with peaks and troughs. Overall, there appears to be a slight downward trend in responses over the entire period.

Monetary value for each frequency
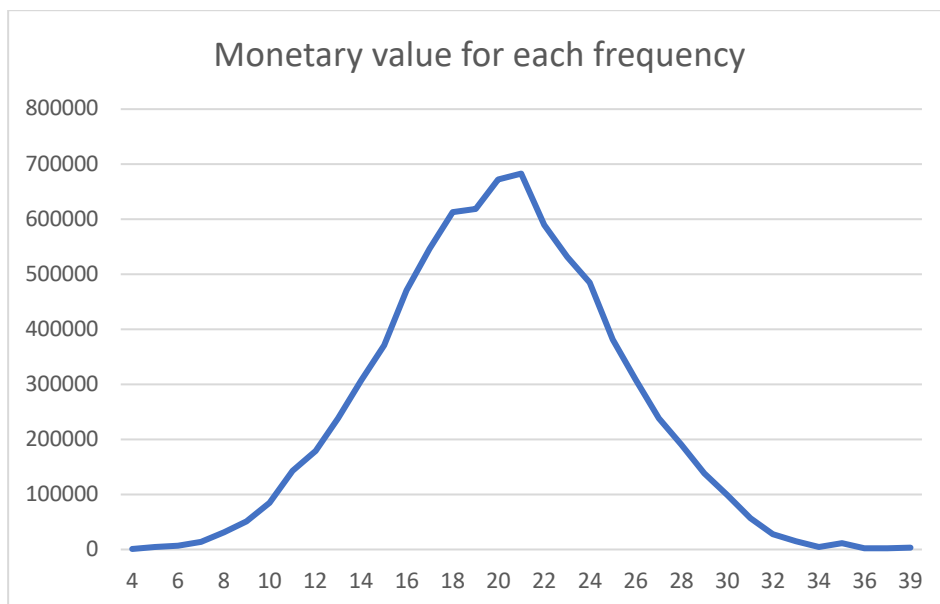
- The line chart shows the monetary value for each frequency. The monetary value increases steadily up to frequency 22, then decreases steadily. The highest monetary value is at frequency 22, and the lowest is at frequencies 4 and 39.

Monetary value for each category

- Gold-P1
- Premium-P0
- Silver-P2
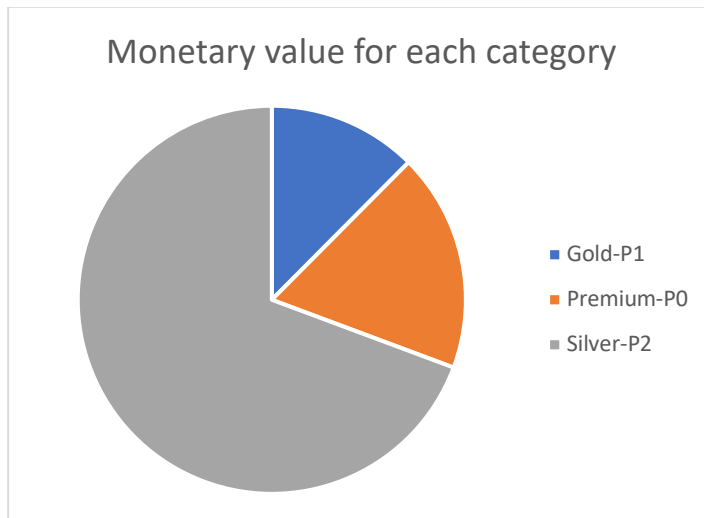
♦ The pie chart shows the distribution of monetary value across three categories: Gold-P1, Premium-PO, and Silver-P2. The majority of the value is in the Silver-P2 category, followed by Gold-P1 and Premium-PO. The Silver-P2 category holds the largest share, accounting for approximately 60% of the total monetary value.

Findings

- **Customer Insights**: The analysis identifies the top-spending customers.
- **Sales Trends**: Monthly analysis shows variations in sales, helping identify peak seasons.
- **High-Value Transactions**: Several transactions had a total sale amount greater than 1000, indicating premium purchases.
- **Cohort Segmentation:** Based on the frequency of transactions, recent transactions and monetary value, we would segment the customers into different sales category.

Reports

- **Sales Summary**: A detailed report summarizing total sales, customer demographics.
- **Trend Analysis**: Insights into sales trends across different months and year.
- **Customer Insights**: Reports on top customers and unique customer count.
- **Category Summary**: A detailed report summarizing a customer into different sales category based on recent, frequency and monetary value.

Conclusion

- This project serves as a comprehensive introduction to SQL, Python and Excel for data analysts, covering database setup, data cleaning, exploratory data analysis, and business-driven SQL queries. The findings from this project can help drive business decisions by understanding sales patterns, customer behaviour.