Marketing Insights for E-Commerce Company

Introduction

The dataset provided offers a comprehensive glimpse into the operational dynamics and consumer interactions within an e-commerce ecosystem. With over 52,000 entries, it encapsulates a rich array of attributes ranging from customer demographics to transactional details, presenting an invaluable opportunity to derive actionable insights into the behavior and preferences of the target market.

This project report endeavors to harness the power of data analytics to unearth compelling marketing insights for the e-commerce company under examination. By scrutinizing the dataset's dimensions, including customer demographics, transactional behaviors, product preferences, and marketing initiatives, we aim to construct a holistic understanding of the company's performance and its resonance within the competitive e-commerce landscape.

Data Overview and Preprocessing

The dataset comprises 21 columns, each offering valuable insights into various aspects of customer transactions within the e-commerce platform. Before delving into analysis, it is imperative to understand the structure of the dataset and undertake preprocessing steps to ensure its quality and suitability for analysis.

1. **Dataset Structure:** The dataset contains 52524 entries, each representing a unique customer transaction. The columns encompass diverse attributes, including customer demographics, transaction details, product information, and marketing metrics.

2. **Data Cleaning:** The dataset appears to be free of missing values, as indicated by the non-null counts across all columns. No outliers or anomalies are immediately

apparent; however, further exploratory analysis may reveal any potential inconsistencies.

3. **Feature Engineering:** A new feature, "Total Prices," has been derived, likely representing the total transaction value including product prices, delivery charges, taxes, and discounts. Additional feature engineering techniques may be explored to derive more informative variables or transform existing ones.

4. **Data Transformation:** The "Transaction_Date" column has been appropriately converted to datetime format, facilitating temporal analysis. Categorical variables such as "Gender," "Location," "Product_SKU," "Product_Description," "Product_Category," "Coupon_Status," and "Coupon_Code" may require encoding for numerical analysis.

5. **Integration of Multiple Datasets:** It is noted that the dataset is a result of combining data from five separate datasets. Steps undertaken to merge the datasets, including any data alignment, deduplication, or reconciliation processes, should be documented to ensure transparency and reproducibility.

Data Visualization:

1. **Top 20 Costliest Purchase IDs** : Visualizes the top 20 purchase IDs with the highest total prices, providing insights into significant transactions.
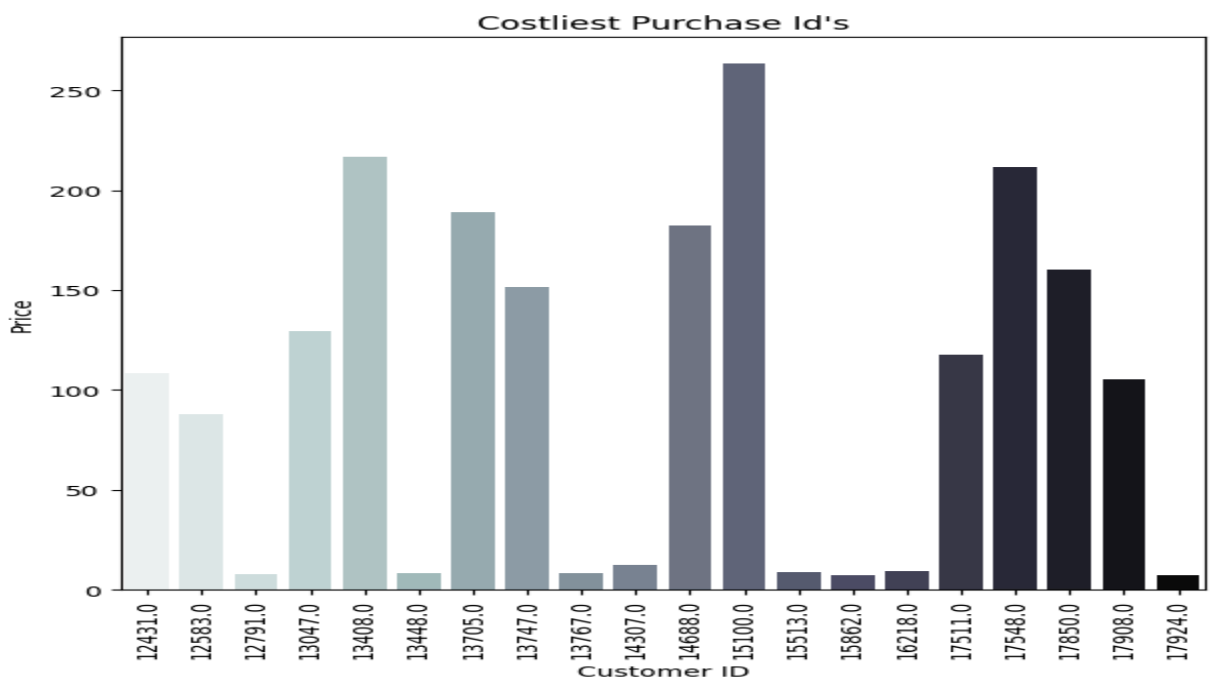
2. **Top 30 Popular Customer IDs with Purchase Count**: Presents the top 30 customer IDs with the highest purchase counts, indicating customer popularity and engagement.

Top 30 Popular Customer ID with Purchase Count
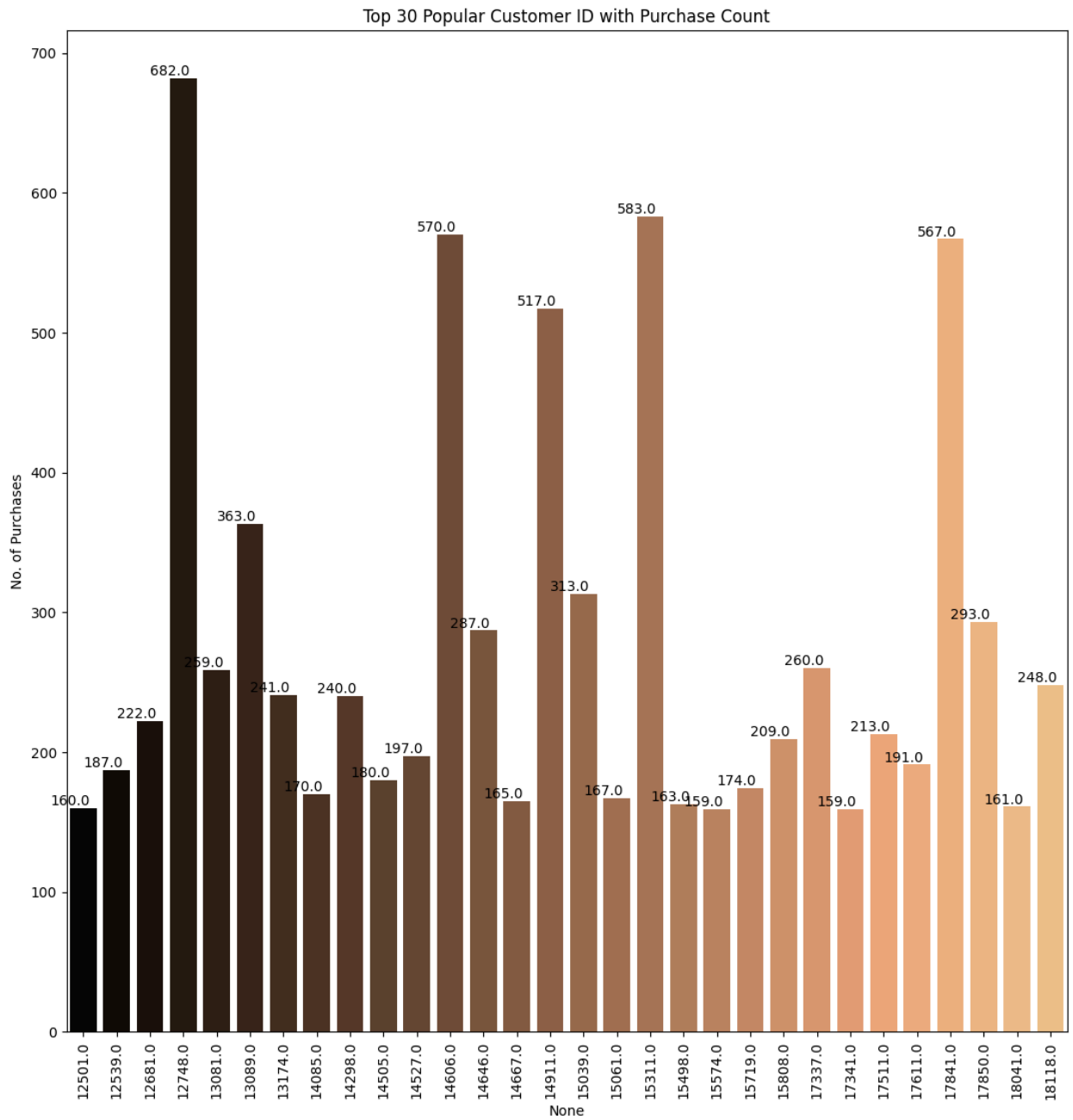
3. **Total Male-Female Purchase Comparison:** Illustrates the distribution of purchases between male and female customers through a pie chart, highlighting any gender-based disparities in spending habits.
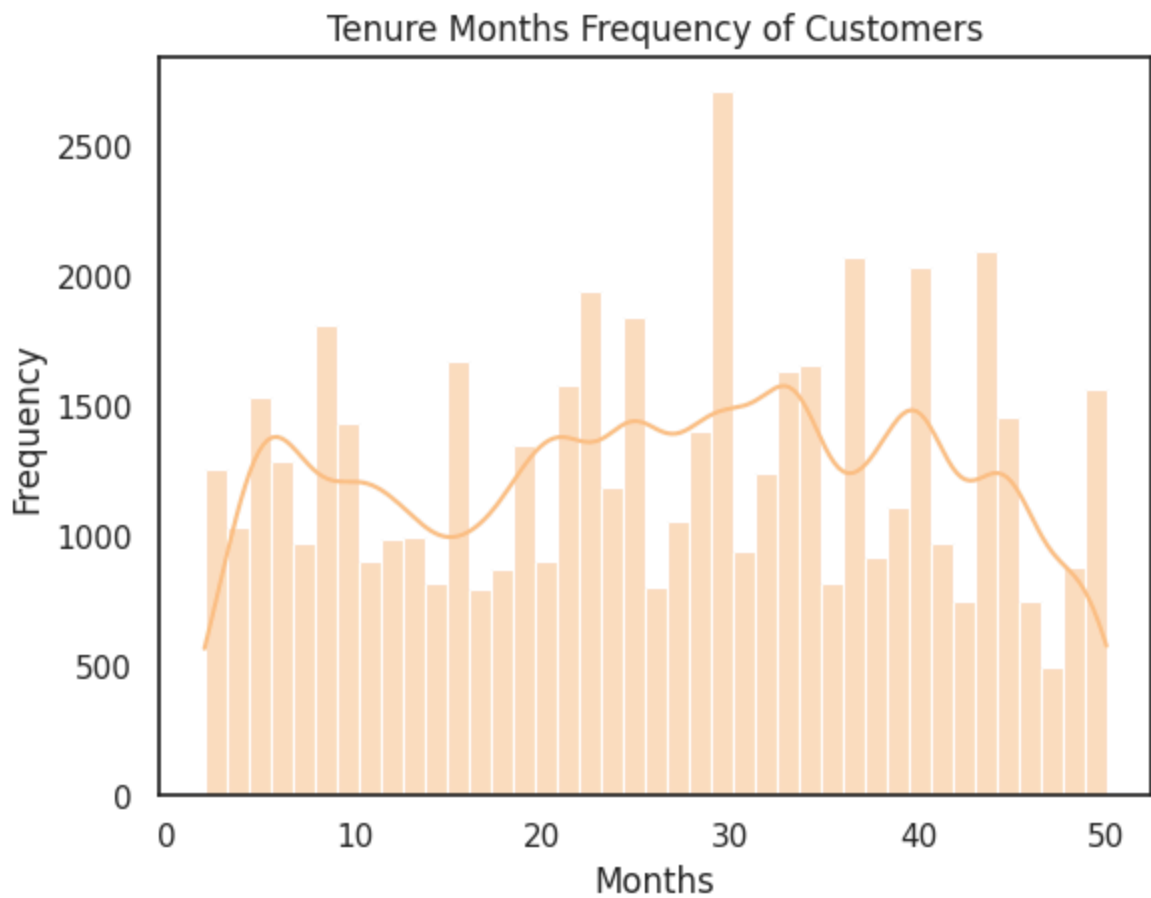

Male - Female Purchase Comparision
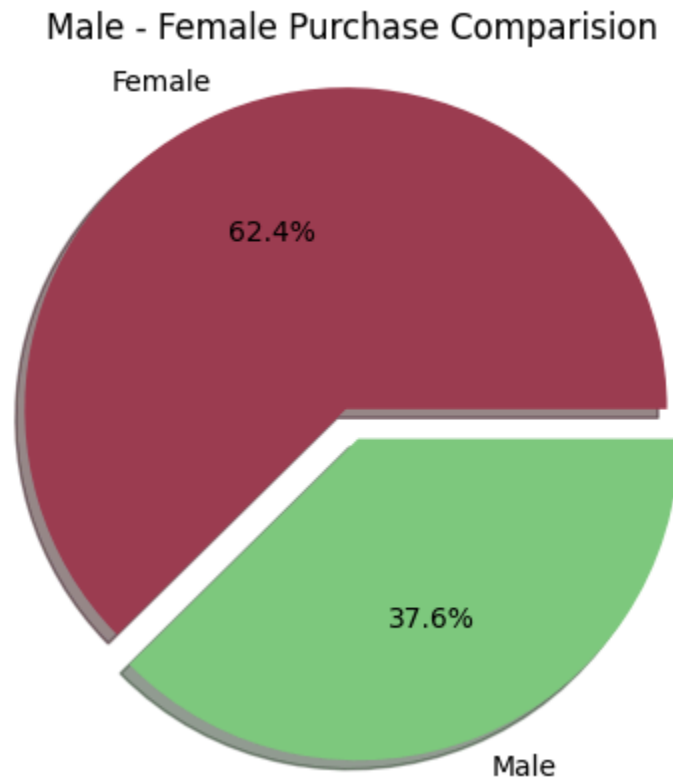
Females Ocuurence to purchase is more than males.

4. **Tenure Months Frequency of Customers**: Displays the frequency distribution of customer tenure months, shedding light on the distribution of customer loyalty and

retention.

## Tenure Months Frequency of Customers

The term "tenure" is often used in the context of customer relationship management to measure how long a customer has been associated with a business. In the case of a shopping dataset, "tenure months" might represent the number of months since a customer first made a purchase or started using a particular service

5. **Product Category Frequency Analysis:** Offers insights into the distribution of purchases across different product categories, aiding in identifying popular product segments.
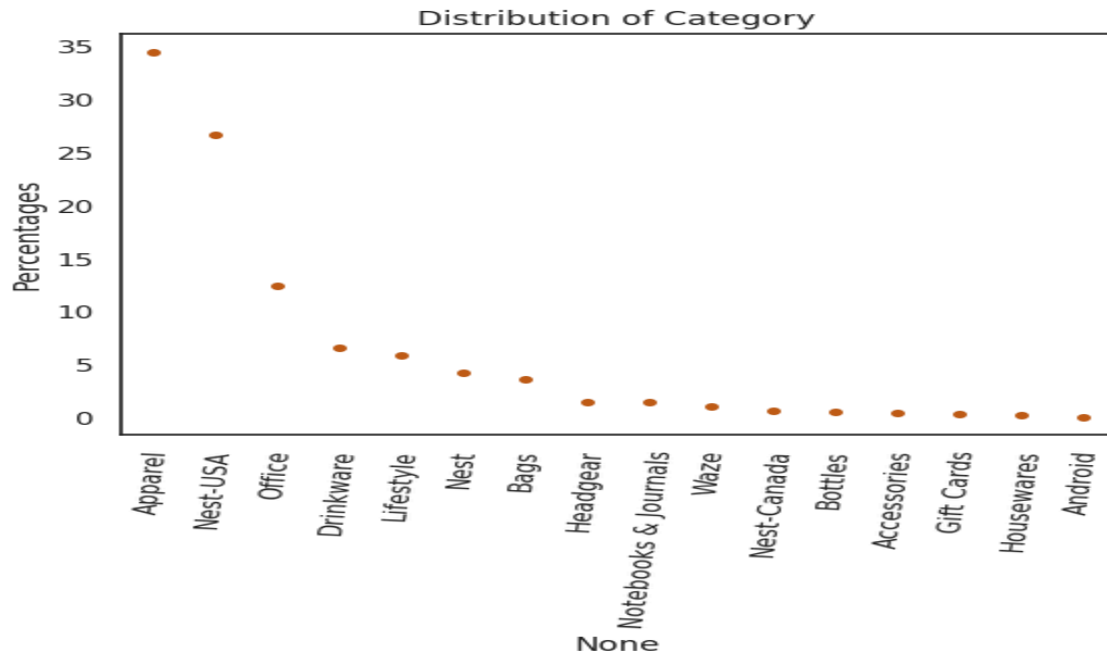
Apparel Category has high frequency

6. **Heatmap for Correlation:** Depicts the correlation matrix heatmap among numerical variables, revealing potential relationships between different attributes.
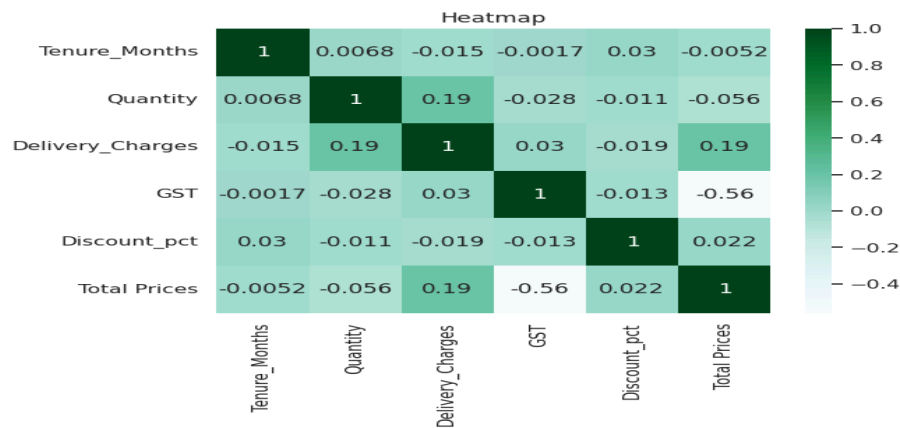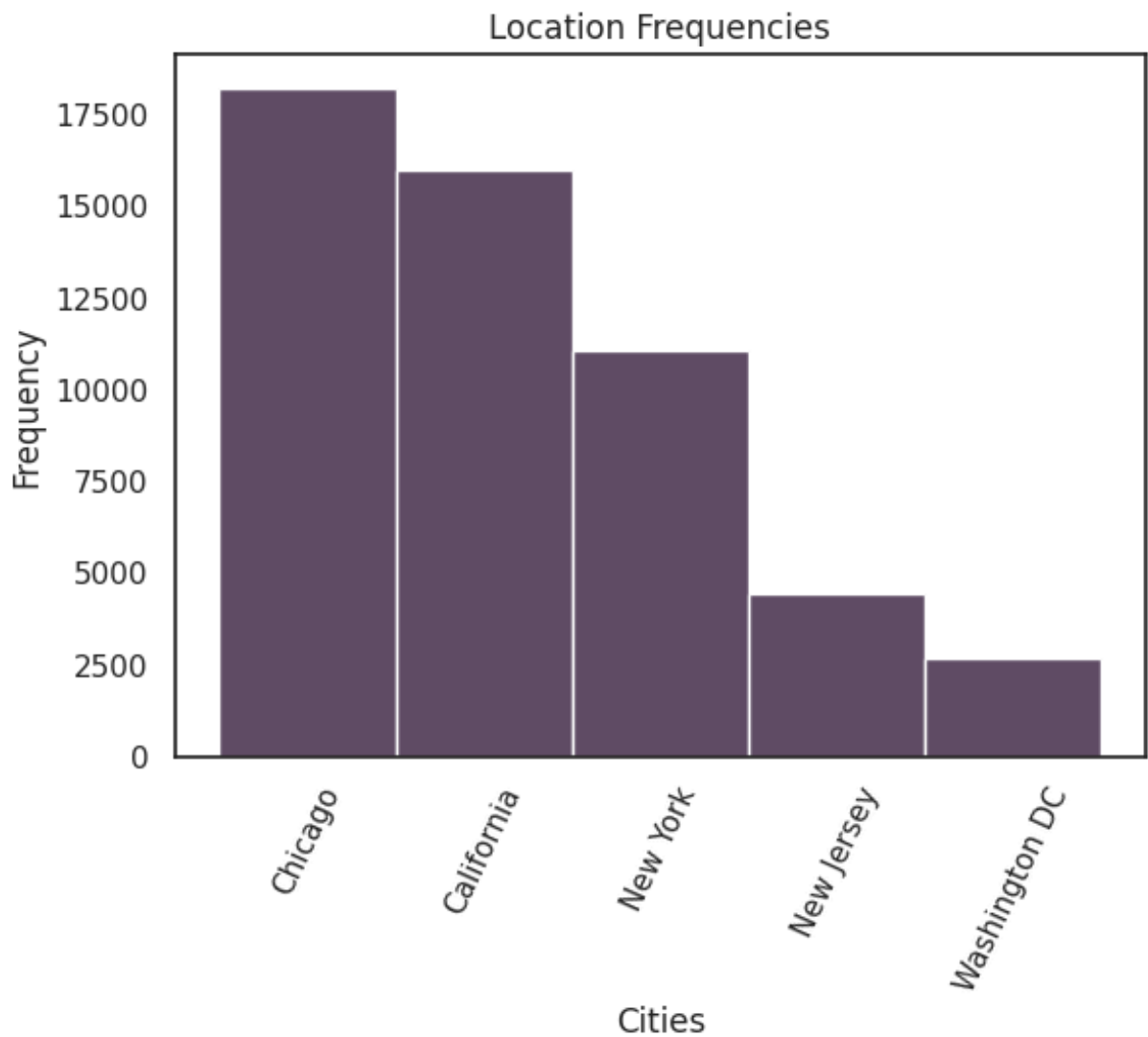
- **Shows the correlation of different features**

  1)Total Price and GST are negatively correlated

 **2)Other all features have less correlation with each other**

7. **Location Frequencies**: Presents the frequency distribution of customer locations, providing an overview of geographical engagement.

Location Frequencies

Chicago -1st, California -2nd, New York -3rd, New Jersey -4th, Washington DC -5th in Purchase Frequency

8. Male-Female Price Spent: Compares the total price spent by male and female customers, highlighting any gender-based differences in purchasing power.

Price Spend

Males
Price :1260583

38.2%

61.8%

Females
Price :2043492
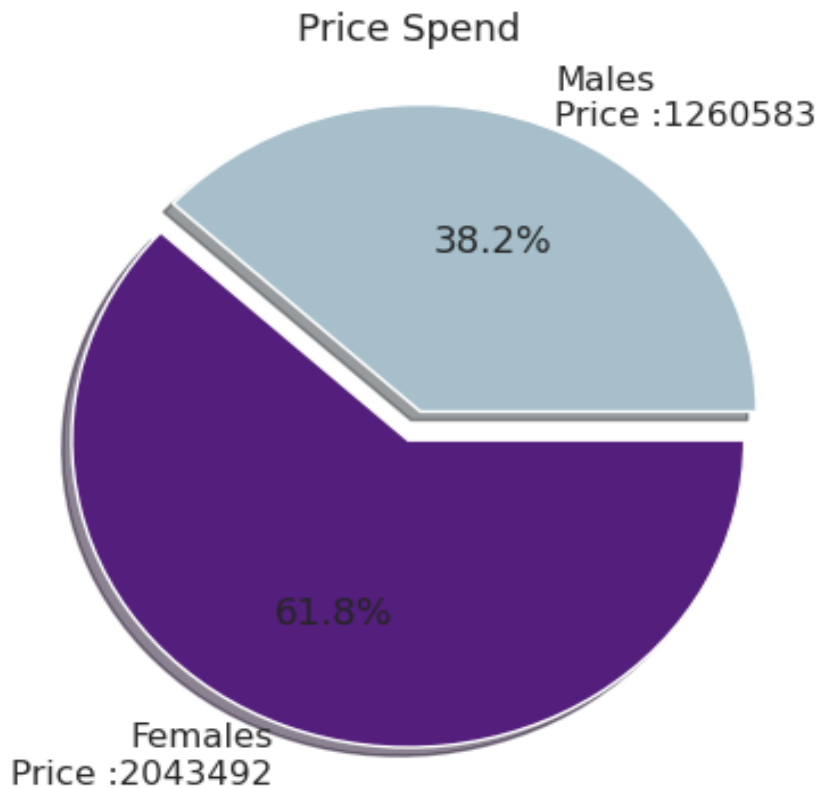
- **Female spend nearly 2 times higher than Males**
- **Similarly females occurrence to shop is high by our previous bar plot**

9. Cities with Price Spent: Shows the distribution of price spent across different cities, enabling geographical analysis of purchasing behavior.

Cities with Price Spent

- **The Occurence table and this plots gives same result**
- **Thus, the most frequent cities spents high amount in total**

10. Delivery Charges vs. Average Price: Explores the relationship between delivery charges and average product price, uncovering insights into pricing strategies and delivery costs.


Delivery Charges Vs Avg_Price
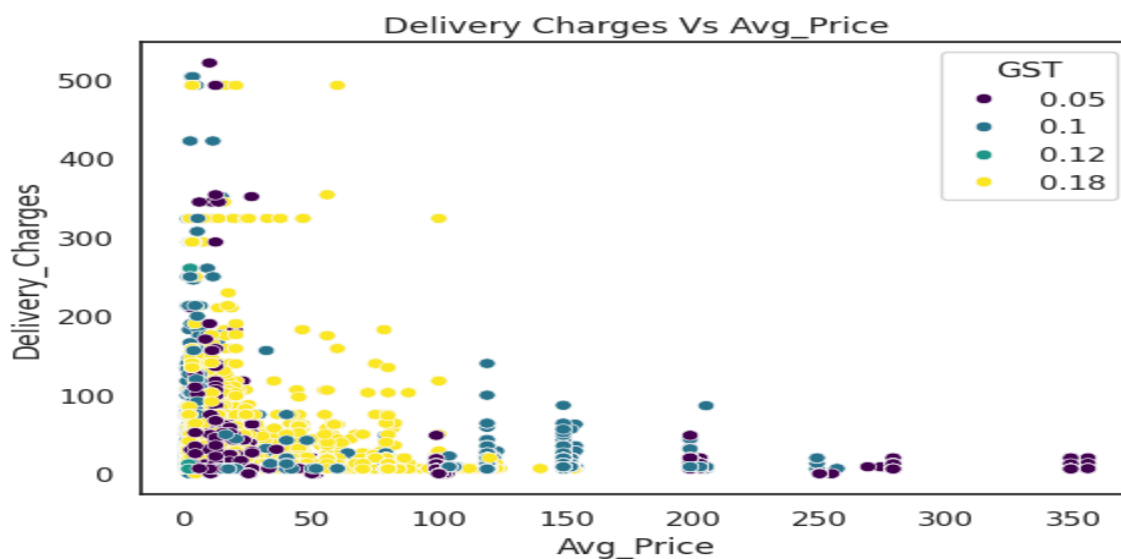
- **Low GST Product has high delivery Charges.Nice tactic in profit.Isn't it?**

Time series Analysis

Time series analysis is a statistical technique that deals with time-ordered data points. It involves studying the patterns and trends in the data to make predictions or gain insights into the underlying processes. Time series data consists of observations or measurements taken at different points in time, and the goal of time series analysis is often to understand and model the structure that generates the data.

1. **Total spend with month**

Total Spend With Month
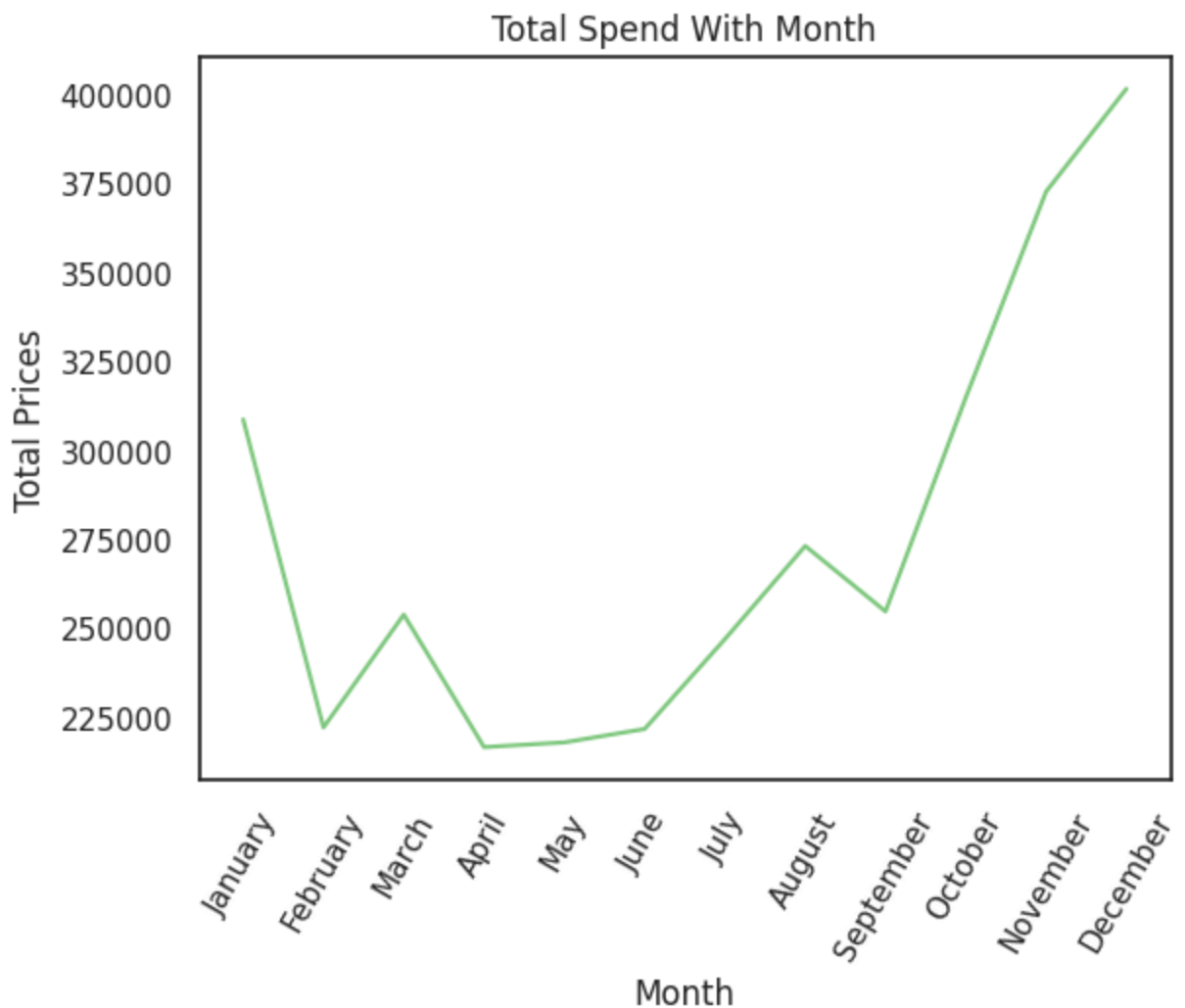
2. Online spend
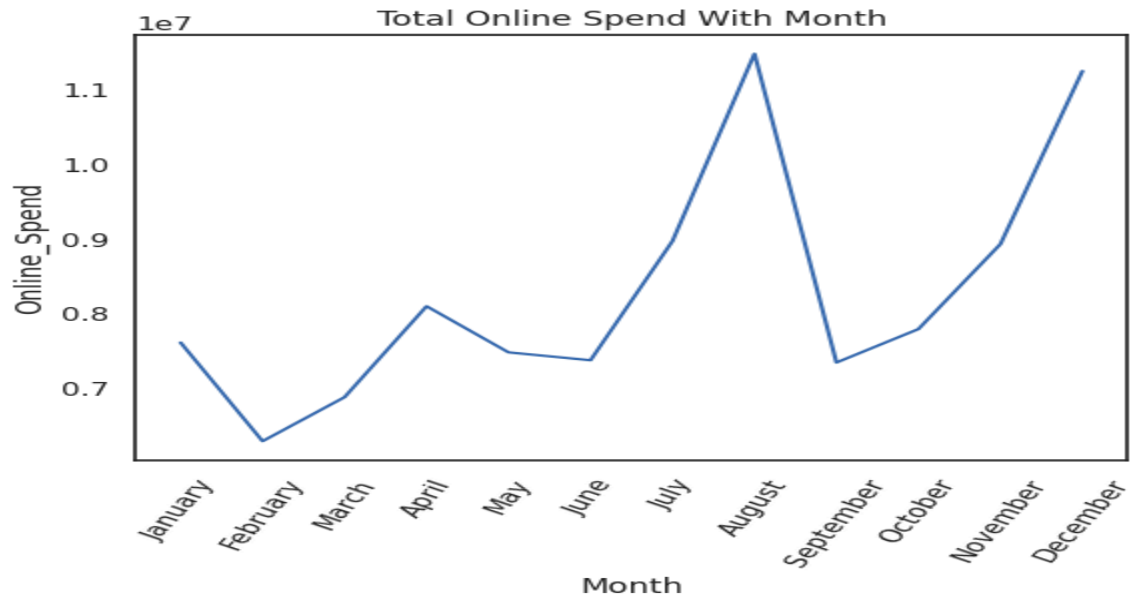
3. Offline spend

5. Total Delivery charges per month
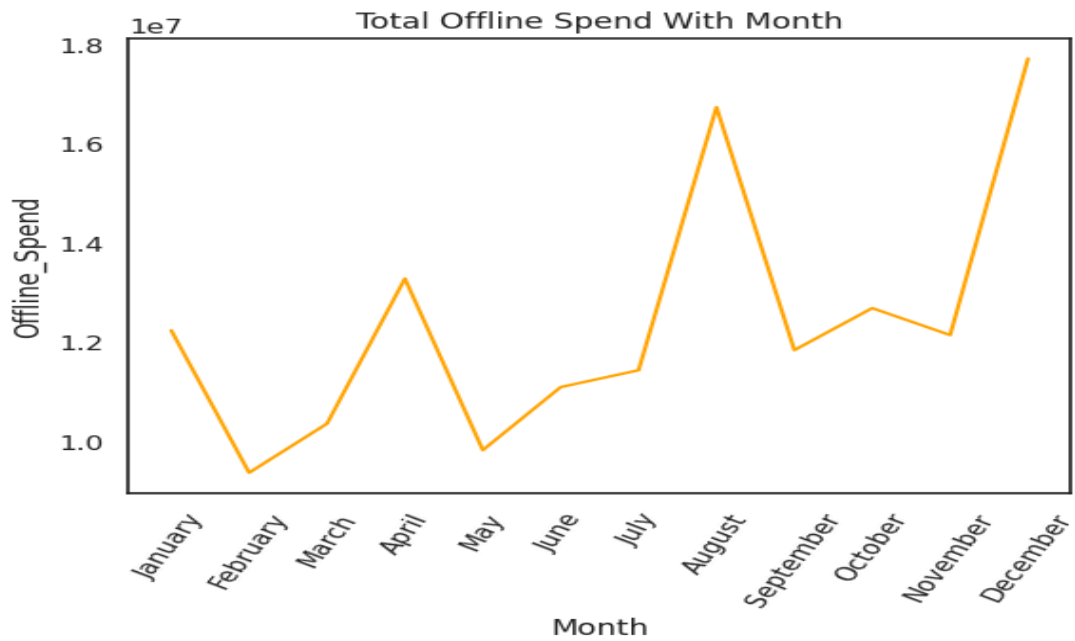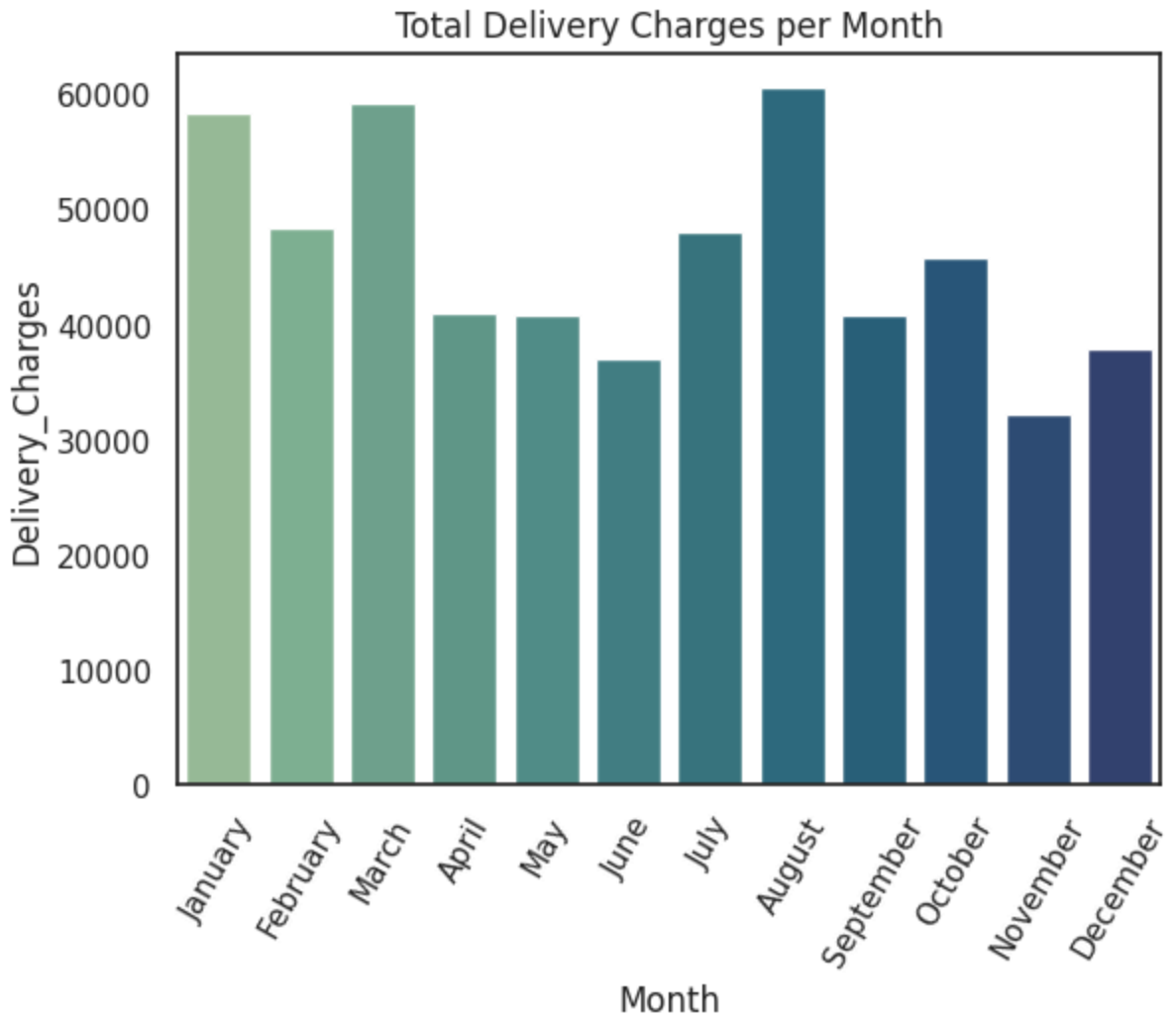

Total Delivery Charges per Month

- **In all three types (Total,Online,Offline) of Spends, the peak is at August and December Month**
- **Delivery Charges are high during the first 3 months of year and also at August month**

Key Findings and Insights

1. **Customer Engagement and Purchase Behavior:** The visualization of the top 30 popular customer IDs reveals significant variations in purchase frequency, indicating varying levels of customer engagement and loyalty. Analysis of tenure months frequency suggests a diverse customer base with varying levels of tenure, highlighting opportunities for targeted retention strategies.

2. **Product Preferences and Category Analysis**: The distribution of product category frequency showcases varying levels of demand across different product segments, with certain categories garnering more attention than others. The heatmap for correlation uncovers potential relationships between product attributes, offering insights into cross-category purchase patterns and preferences.

3. **Gender-based Spending Patterns:** The comparison of male-female purchase expenditure indicates comparable spending levels between genders, suggesting a balanced customer base in terms of purchasing power.

4. **Geographical Insights and Regional Variations:** Location frequencies illustrate the distribution of customers across different cities, with potential regional variations in market penetration and customer engagement. Analysis of cities with price spent reveals notable disparities in spending levels across geographical regions, suggesting opportunities for targeted marketing campaigns.

5. **Marketing Effectiveness and Pricing Strategies:** Delivery charges vs. average price scatterplot offers insights into the relationship between pricing strategies and associated delivery costs, guiding decisions on pricing optimization and delivery fee structures.

6. **Temporal Trends and Seasonality Effects:** Time series analysis of total spend, online spend, offline spend, and delivery charges per month unveils temporal trends and seasonality effects, informing seasonal marketing campaigns and inventory management strategies.

Recommendations and Actionable Insights

1. **Customer Engagement and Retention Strategies:** Implement a tiered loyalty program to incentivize repeat purchases and foster long-term customer loyalty. Tailor rewards based on customer segments and purchase behaviors identified through segmentation analysis. Launch personalized email campaigns leveraging customer data to provide relevant product recommendations and exclusive offers, enhancing engagement and driving conversions.

2. **Product and Inventory Management Recommendations:** Conduct a thorough review of product performance across different categories. Allocate resources to promote high-demand products and consider discontinuing low-performing ones to optimize inventory turnover. Introduce product bundles or curated collections based on cross-category purchase patterns identified in the correlation analysis, encouraging customers to explore complementary items and increase basket size.

3. **Marketing and Promotional Campaign Optimization:** Allocate marketing budgets based on the analysis of campaign effectiveness and ROI. Focus resources on channels and campaigns that yield the highest conversion rates and customer acquisition efficiency. Leverage dynamic pricing strategies to adjust prices in real-time based on demand and competitive dynamics, maximizing revenue while remaining competitive in the market.

4. **Geographical Expansion and Market Penetration Strategies:** Identify regions with high growth potential based on geographical engagement insights. Develop targeted marketing campaigns and localized offerings to penetrate new markets and capitalize on emerging opportunities. Invest in localized customer support and logistics infrastructure to enhance the shopping experience for customers in new regions, ensuring seamless order fulfillment and timely delivery.

5. **Operational Efficiency Enhancements:** Streamline the checkout process to minimize cart abandonment rates and improve conversion rates. Implement features such as guest checkout, one-click ordering, and seamless payment options to enhance user experience. Invest in logistics optimization technologies, such as route planning software and warehouse management systems, to streamline order fulfillment processes and reduce delivery lead times.

6. **Continuous Monitoring and Evaluation:** Establish a robust analytics framework to track key performance indicators (KPIs) and monitor the impact of implemented strategies. Regularly review and analyze data to identify emerging trends and opportunities for optimization. Foster a culture of data-driven decision-making within the organization by providing training and resources to employees across departments. Encourage cross-functional collaboration to leverage insights and drive continuous improvement initiatives.

Conclusion

In conclusion, the data analysis and visualization have provided valuable insights into the operational dynamics and customer interactions within the e-commerce platform. By exploring various aspects of customer behavior, product preferences, and marketing effectiveness, several key findings have emerged, offering actionable insights for strategic decision-making and business optimization.

1. **Customer Engagement and Purchase Behavior:** The analysis of customer engagement and purchase behavior revealed diverse patterns in customer tenure and transaction frequency. This suggests opportunities to implement targeted retention strategies to nurture existing customer relationships and drive repeat purchases.
2. **Product Performance and Category Analysis:** Insights into product performance and category analysis highlighted the varying popularity of different product segments. By optimizing product assortments and inventory management strategies, the company can capitalize on high-demand products and streamline operations.
3. **Marketing Effectiveness and Geographical Insights:** The evaluation of marketing effectiveness and geographical insights shed light on regional variations in customer behavior and engagement. Tailoring marketing campaigns and offerings to specific geographical regions can enhance market penetration and customer acquisition.

4. **Operational Efficiency Enhancements:** The analysis of delivery charges and pricing strategies underscored the importance of balancing customer expectations with operational costs. Implementing streamlined checkout processes and optimizing logistics infrastructure can improve overall operational efficiency and customer satisfaction.

5. **Time Series Analysis and Seasonality Effects:** Time series analysis revealed temporal trends and seasonality effects in total spend, online spend, offline spend, and delivery charges per month. Leveraging these insights can inform seasonal marketing campaigns and inventory management strategies to capitalize on peak demand periods.