

Customer Segmentation using RFM Analysis



INTRODUCTION:

RFM:

Recency, Frequency, Monetary Value, or RFM, analysis is a sophisticated marketing tactic that gives companies a detailed picture of consumer behavior. Businesses can strategically adjust their marketing efforts to optimize engagement, retention, and revenue by exploring the frequency of transactions, the recentness of purchases, and the monetary value of customer spending.

RFM analysis has become an indispensable tool in modern marketing to meet the demand for individualized and focused strategies. In a time when customer-centric strategies are critical, knowing the subtleties of each customer's purchasing behavior is essential for long-term company expansion.

RFM Elements:

Recency (R): The 'R' in RFM, recency, measures how recently a customer has interacted with the business. It is calculating the amount of time that has passed since a customer's last purchase yields the recency score. When evaluating the temporal aspect of customer engagement, this metric is essential.

Frequency (F): The 'F' in RFM stands for frequency, which signifies the regularity of a customer's transactions. The frequency score gives information about how frequently a customer interacts with the brand by counting the number of purchases made within a given period.

Monetary value (M): The 'M' in RFM stands for monetary value, which measures a customer's financial contribution to the company. The customer's overall spending amount is used to calculate the monetary value score, which provides a metric to assess everyone's economic impact.

Advantages of RFM Analysis:

Segmentation: Businesses can create customized marketing campaigns for each category by using segmentation, which guarantees that promotions and messaging speak to the distinct tastes and habits of each group.

Customer Retention: Businesses can create retention strategies that improve customer loyalty, lower attrition rates, and create enduring relationships by identifying high-value customers.

Cross-Selling and Up-Selling: RFM Analysis reveals prospects for up-selling and cross-selling by figuring out what customers want and recommending related or better products.

Resource Allocation: Marketing efforts, funds, and resources should be directed toward market segments with the greatest potential for revenue generation to achieve efficient resource allocation.

TABLE OF CONTENTS

1. Introduction
2. Data Acquisition
3. Data Preprocessing
4. RFM Calculation
5. Exploratory Data Analysis
6. Customer Segmentation
7. Conclusion

DATASET

In this project assignment, we have worked with the eCommerce dataset provided (<https://www.kaggle.com/datasets/carrie1/ecommerce-data>) to create a Customer Segmentation model using the RFM (Recency, Frequency, Monetary) analysis method.

DATA PREPROCESSING

The data set contains 541909 rows and 8 columns before cleaning.
The 8 columns represent as follows:

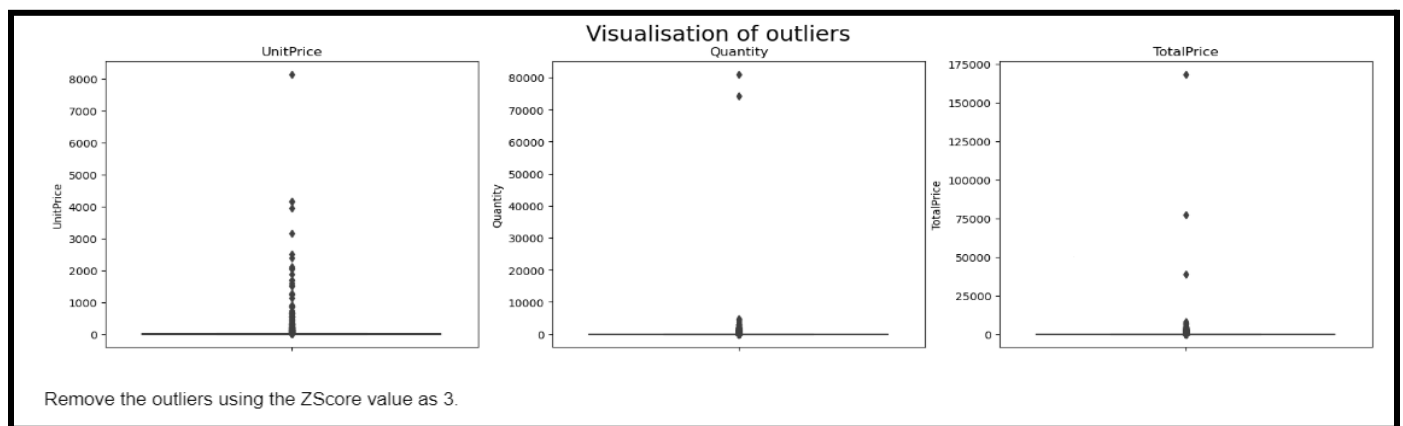
- **InvoiceNo**: Invoice number is a 6-digit integral number uniquely assigned to each transaction.
- **StockCode**: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
- **Description**: Product (item) name. Nominal.
- **Quantity**: The quantities of each product (item) per transaction. Numeric.
- **InvoiceDate**: Invoice Date and time. Numeric, the day and time when each transaction was generated.
- **UnitPrice**: Unit price. Numeric, Product price per unit in sterling.
- **CustomerID**: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
- **Country**: Country name. Nominal, the name of the country where each customer resides.

CLEANING THE DATASET

```
Check for null values
InvoiceNo      0
StockCode      0
Description    1454
Quantity       0
InvoiceDate    0
UnitPrice      0
CustomerID    135080
Country        0
dtype: int64
```

- There are missing values in Customer ID and Description. So it's better to remove that as it's difficult to replace with our assumptions.
- Check for the types for all the attributes and convert it to the proper format

- Looking at the data, we can see that there are some negative and zero numbers for quantity and unit pricing. As a result, we must disregard it in our analysis.
- There are stock codes which contain 'C' which states that the order is canceled/returned.
- Check for the duplicate rows in the data and drop it(there were 5192 duplicates found in our dataset)
- There are some inconsistencies with the Description as there are more than one description for one particular stock code. So we need to check the size and handle that appropriately
- Check for the outliers and remove it if it's necessary



RFM CALCULATION

Recency:

InvoiceDate	
CustomerID	
12347	1
12348	74
12349	18
12350	309
12352	35
...	...
18280	277
18281	180
18282	7
18283	3
18287	42

4333 rows × 1 columns

The recency analysis conducted on the provided dataset reveals valuable insights into customer purchasing behavior. The resulting DataFrame, 'recency_df,' succinctly captures the recency of each customer's last transaction, quantified in terms of the number of days since their latest purchase. For instance, customer 12347 made a purchase merely 1 day ago, indicating a high level of recent engagement, while customer 12350's recency value of 309 days suggests a longer gap between transactions. This information is crucial for understanding customer engagement patterns, tailoring marketing strategies, and identifying opportunities for targeted promotions or re-engagement efforts with customers who have been less active recently. The recency metric serves as a practical tool for customer segmentation and strategic decision-making in efforts to enhance overall customer satisfaction and maximize the effectiveness of marketing initiatives.

Frequency:

Frequency	
CustomerID	
12347	7
12348	4
12349	1
12350	1
12352	8
...	...
18280	1
18281	1
18282	2
18283	16
18287	3

4333 rows × 1 columns

The updated 'Frequency' analysis on the provided dataset, encapsulated in the DataFrame 'frequency_df,' illuminates distinct patterns of customer engagement. Each row signifies a unique customer, with the corresponding 'Frequency' column quantifying the number of their transactions. For example, customer 18283 exhibits a high frequency of 16, indicative of repeated interactions, while others like customer 12349 show a lower frequency of 1. This granularity in customer purchasing habits provides actionable insights for segmentation and targeted marketing efforts. Combined with the previously calculated recency metrics, this comprehensive view of customer behavior allows for the identification of various customer segments—ranging from highly engaged to less active. Such insights are invaluable for crafting personalized marketing strategies, optimizing customer retention initiatives, and ultimately enhancing the overall effectiveness of business decisions.

Monetary:

TotalExpense	
CustomerID	
12347	4310.00
12348	1797.24
12349	1757.55
12350	334.40
12352	2506.04
...	...
18280	180.60
18281	80.82
18282	178.05
18283	2045.53
18287	1837.28

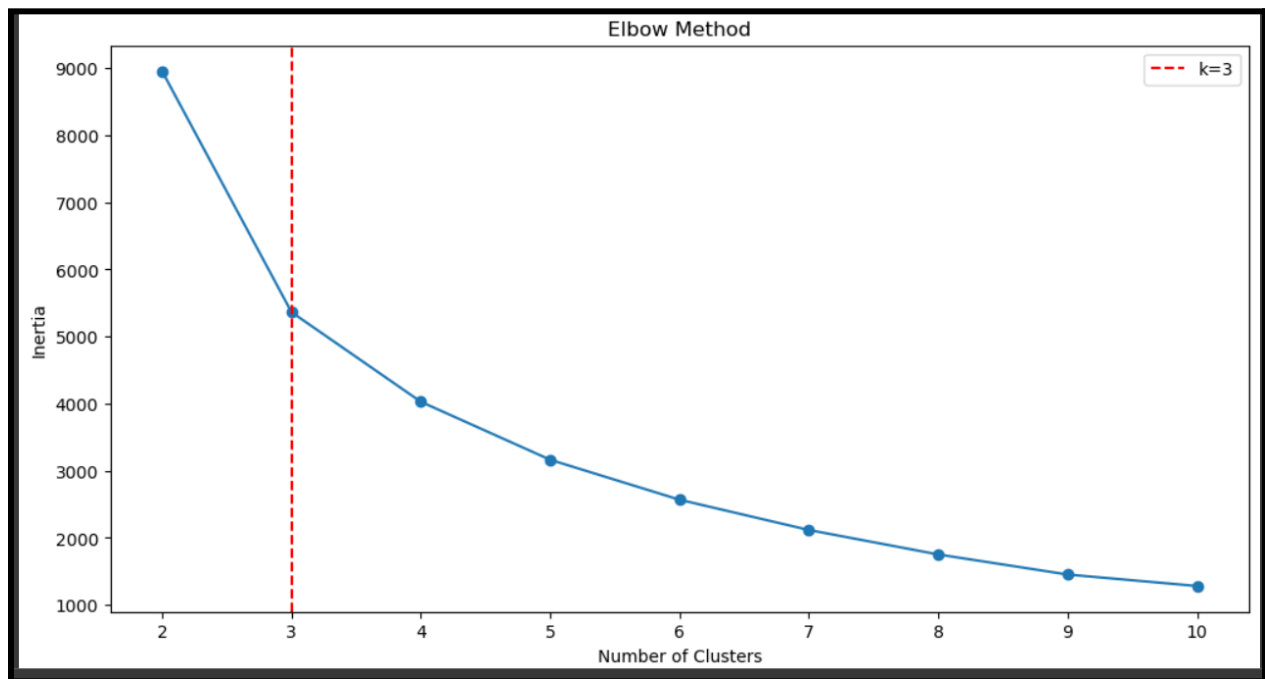
4333 rows × 1 columns

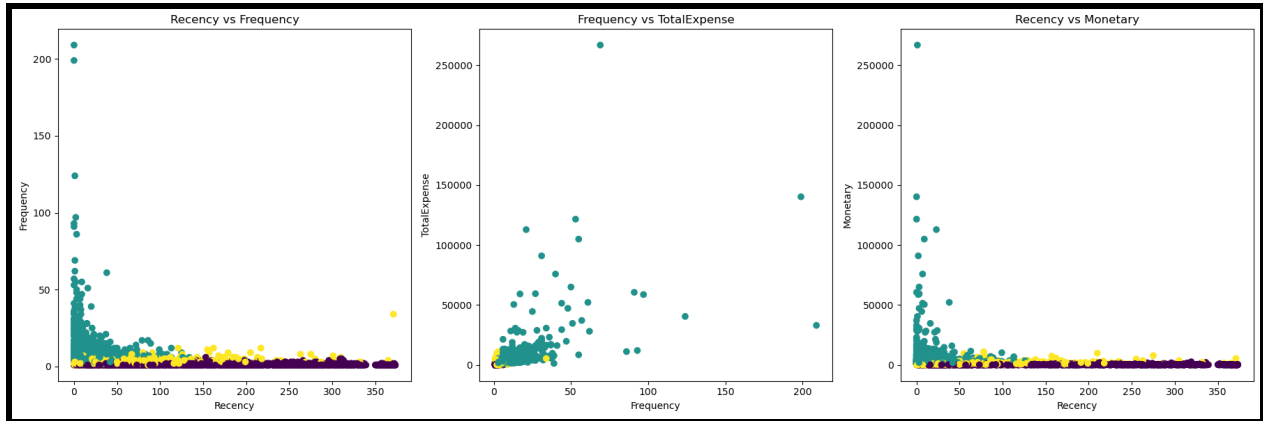
The 'TotalExpense' analysis conducted on the dataset, as encapsulated in the 'moneytary_df' DataFrame, offers a comprehensive understanding of each customer's financial contribution to the business. With 'CustomerID' as the index and 'TotalExpense' as the corresponding column, the DataFrame reveals the cumulative monetary expenditure for each customer. For instance, customer 12347 exhibits a substantial 'TotalExpense' of 4310.00, highlighting their significant financial impact on the business. Integrating this monetary dimension with previously calculated recency and frequency metrics provides a nuanced perspective on customer value. Businesses can leverage this information to identify high-value customers, tailor marketing strategies to retain and nurture their loyalty, and allocate resources effectively. Such insights empower businesses to make informed decisions that prioritize customer relationships based on both transactional behavior and financial significance, ultimately contributing to long-term success and profitability.

RFM SCORE:

	Recency	Frequency	TotalExpense	Recency_Quartile	Frequency_Quartile	Monetary_Quartile	RFM_Score
CustomerID							
12347	1	7	4310.00	4	4	4	12
12348	74	4	1797.24	2	3	4	9
12349	18	1	1757.55	3	1	4	8
12350	309	1	334.40	1	1	2	4
12352	35	8	2506.04	3	4	4	11
...
18280	277	1	180.60	1	1	1	3
18281	180	1	80.82	1	1	1	3
18282	7	2	178.05	4	2	1	7
18283	3	16	2045.53	4	4	4	12
18287	42	3	1837.28	3	3	4	10

4329 rows × 7 columns

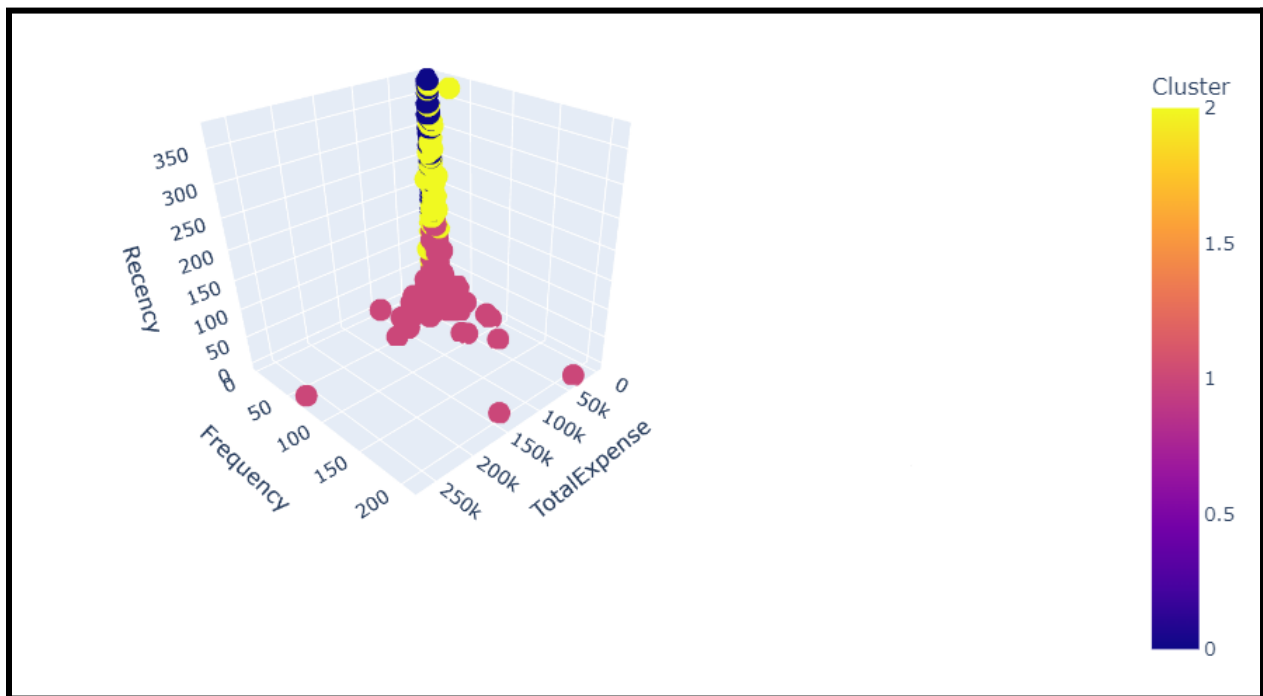




New customers have the lowest average order value. This may be because new customers are still learning about the business and its products or services.

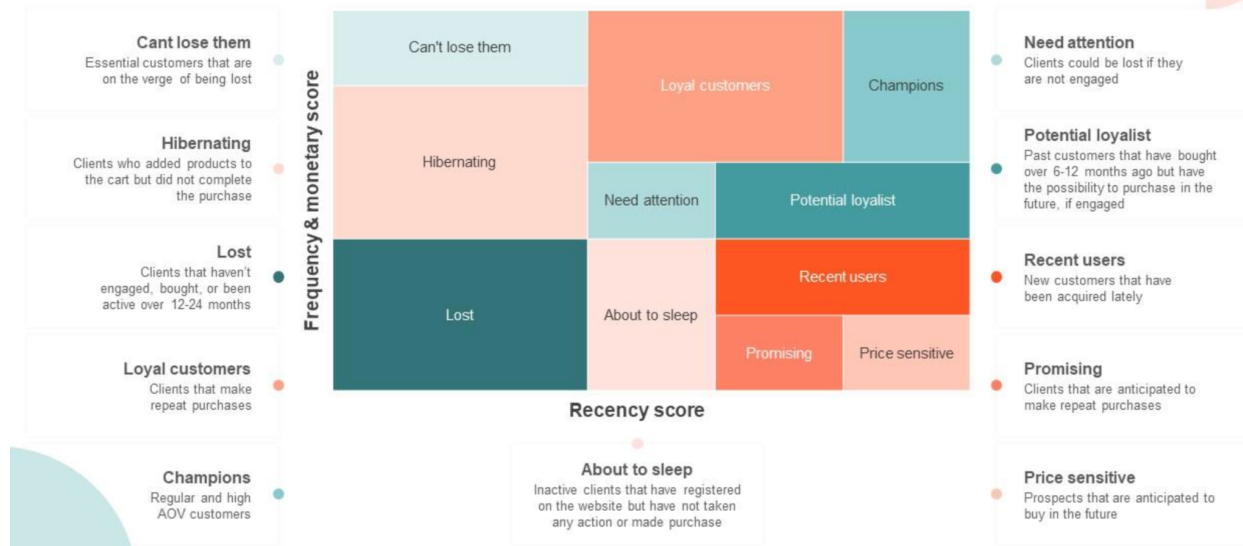
Repeat customers have a higher average order value than new customers. This may be because repeat customers are more familiar with the business and its products or services, and they are more likely to make repeat purchases.

Loyal customers have the highest average order value. This may be because loyal customers are the most satisfied with the business and its products or services, and they are more likely to spend more money with the business



RFM analysis matrix for customer segmentation

This slide highlights a matrix of RFM analysis for segmenting market customers to gain better understanding of target audience and develop effective marketing strategies. It classifies segments such as can't lose them, hibernating, lost, loyal customers, champions, need attention, potential loyalist, recent users, etc.



Let's break down the explanations for each category:

1. **Lost:** Customers who were active but haven't made a recent purchase. These individuals might need targeted reactivation campaigns to bring them back to the business.
2. **Hibernating:** Customers who have been inactive for an extended period. It's crucial to re-engage these customers through personalized offers or promotions to revive their interest.
3. **Can't Lose Them:** Highly valuable and consistently active customers. It's essential to maintain the satisfaction of these customers and ensure they continue to receive personalized services to retain their loyalty.
4. **About to Sleep:** Customers showing signs of reduced engagement. Businesses should implement strategies to rekindle their interest, such as exclusive offers or loyalty programs.

5. **Need Attention:** Customers who have been inactive for a longer period and require special efforts to regain their attention. Targeted campaigns or incentives may be effective in reactivating them.

6. **Loyal Customers:** Consistently active and loyal customers. These individuals form a strong customer base and should be rewarded with loyalty programs, exclusive offers, or personalized services to maintain their loyalty.

7. **Promising:** Customers who recently started engaging. Businesses should nurture these relationships with targeted communication and incentives to encourage further engagement and potential loyalty.

8. **Recent Users:** Customers who recently began engaging but haven't yet reached a high level of loyalty. It's an opportunity to solidify the relationship with personalized experiences and promotions.

9. **Potential Loyalist:** Customers showing potential for long-term loyalty. Businesses should focus on building stronger connections through personalized interactions and incentives to foster loyalty.

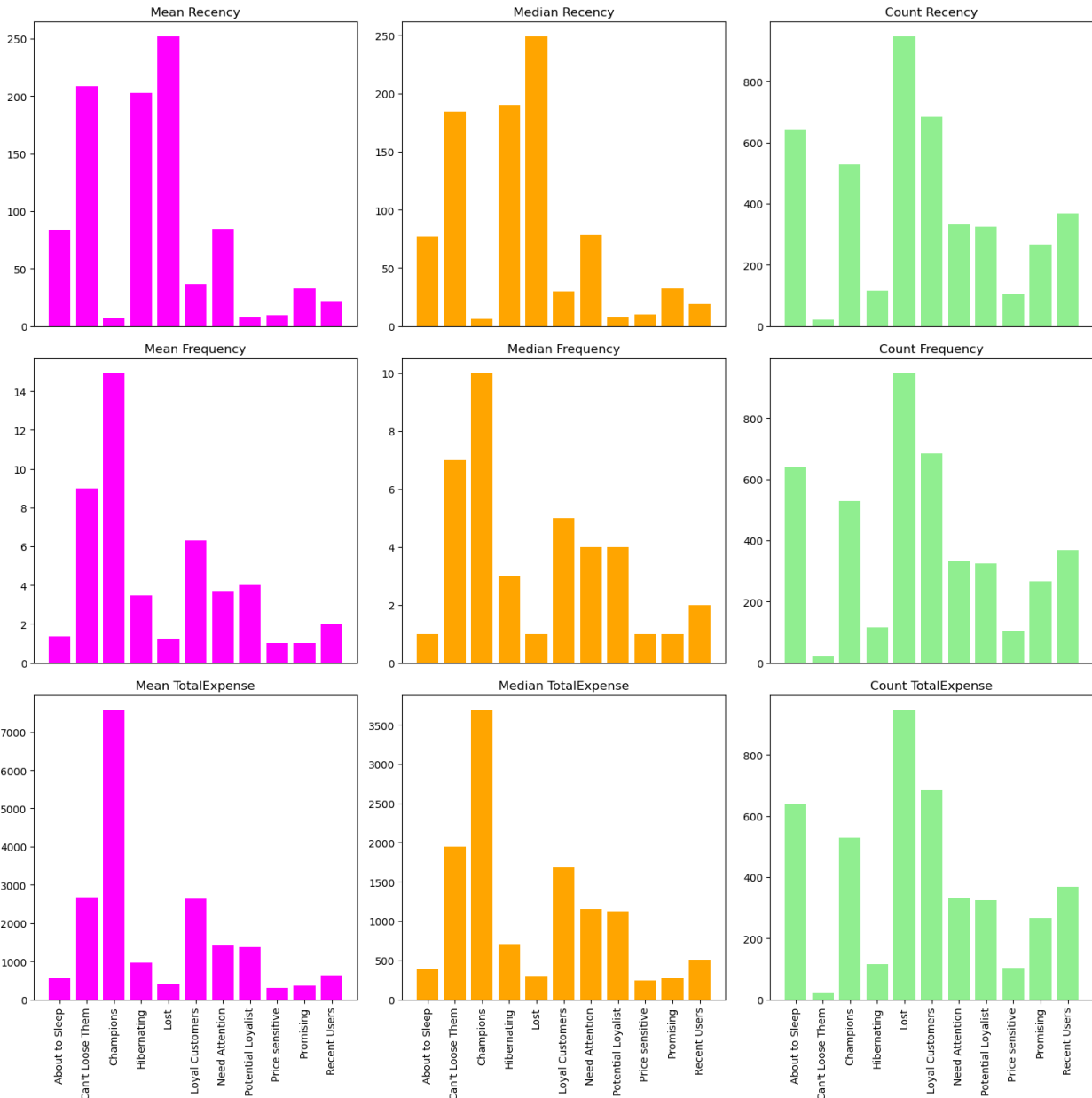
10. **Price Sensitive:** Customers sensitive to pricing. Offering discounts, special promotions, or value-added services may be effective in retaining these customers.

11. **Champions:** Highly valuable and loyal customers consistently making significant purchases. These customers should be recognized and rewarded to maintain their loyalty and potentially increase their lifetime value.

These categories help businesses tailor their marketing strategies to different customer segments, addressing the unique needs and behaviors of each group.

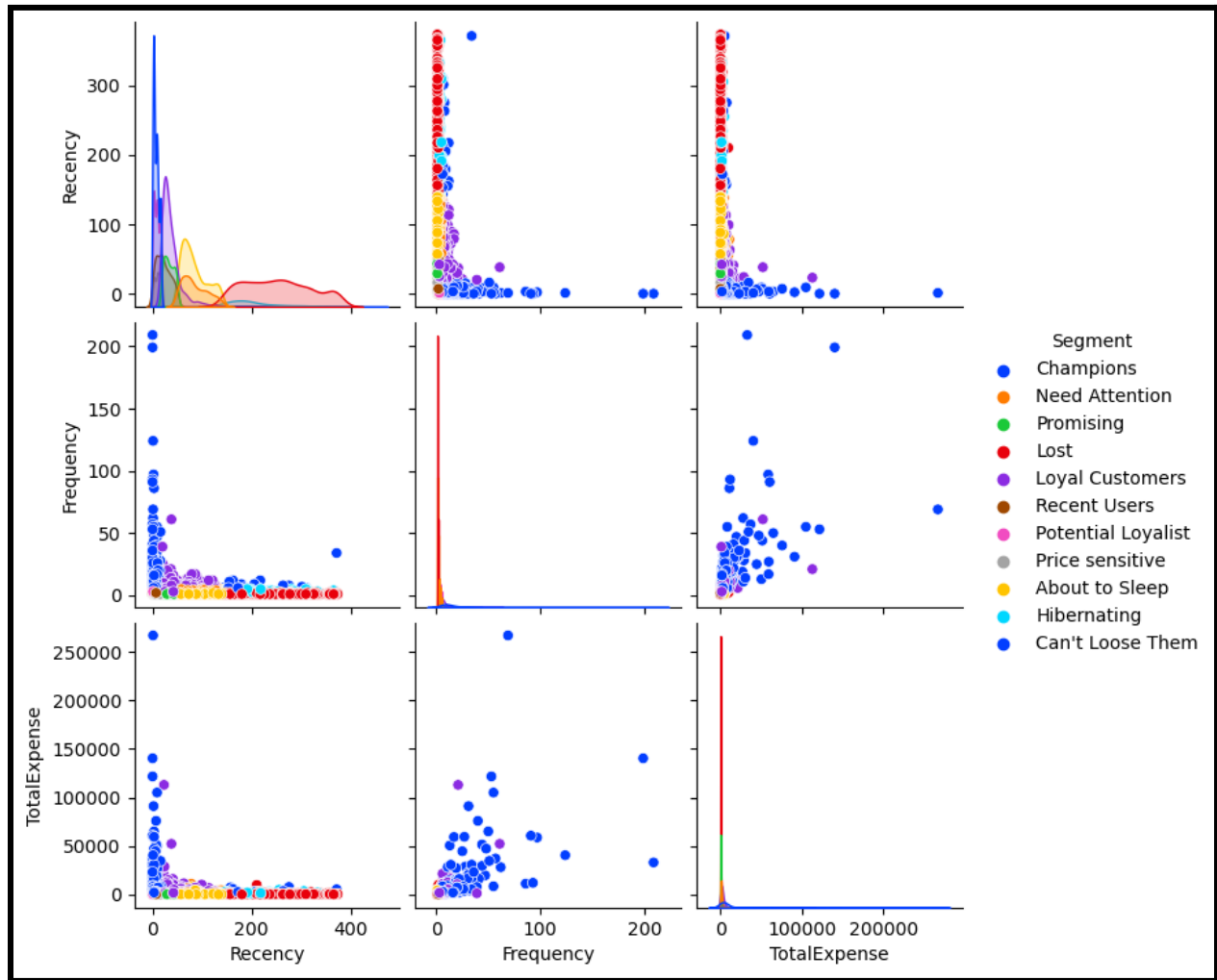
CUSTOMER SEGMENTATION SUMMARY:

	Recency			Frequency			TotalExpense		
	mean	median	count	mean	median	count	mean	median	count
Segment									
About to Sleep	83.609984	77.0	641	1.355694	1.0	641	556.226382	383.750	641
Can't Loose Them	208.800000	184.5	20	9.000000	7.0	20	2679.997000	1943.665	20
Champions	6.504726	6.0	529	14.930057	10.0	529	7579.942722	3692.280	529
Hibernating	202.556522	190.0	115	3.452174	3.0	115	962.171565	702.000	115
Lost	251.976769	249.0	947	1.243928	1.0	947	391.014109	291.050	947
Loyal Customers	36.582723	30.0	683	6.291362	5.0	683	2642.125814	1680.720	683
Need Attention	84.474320	78.0	331	3.685801	4.0	331	1402.156650	1149.020	331
Potential Loyalist	8.012308	8.0	325	3.987692	4.0	325	1378.555477	1126.370	325
Price sensitive	9.291262	10.0	103	1.000000	1.0	103	301.326699	243.550	103
Promising	32.973684	32.0	266	1.000000	1.0	266	359.300038	266.945	266
Recent Users	21.474255	19.0	369	2.000000	2.0	369	641.641382	511.350	369



Customers who purchased "About to Sleep" products around 83 days ago represent a potential for re-engagement with targeted techniques. Despite a reduction in engagement, the brief recency period has the ability to rekindle their interest. "Potential Loyalists" on the other hand, show promise with a short recency of 8 days and a propensity to engage more regularly, showing their potential long-term worth to the organization. Customers that are "Price Sensitive" have a great sensitivity to pricing and make frequent purchases every 1 day on average. Their value-driven behavior is reflected in a total spend of \$301, highlighting the significance of strategic pricing and value-based services. Similarly, "Promising" clients with a moderate recency of 32.97

days show higher engagement potential, making them a category worthwhile requiring deliberate nurturing. Finally, the "Lost" sector, which has been inactive for an average of 251.89 days, needs focused reactivation efforts, as seen by their low frequency and lower average total cost.

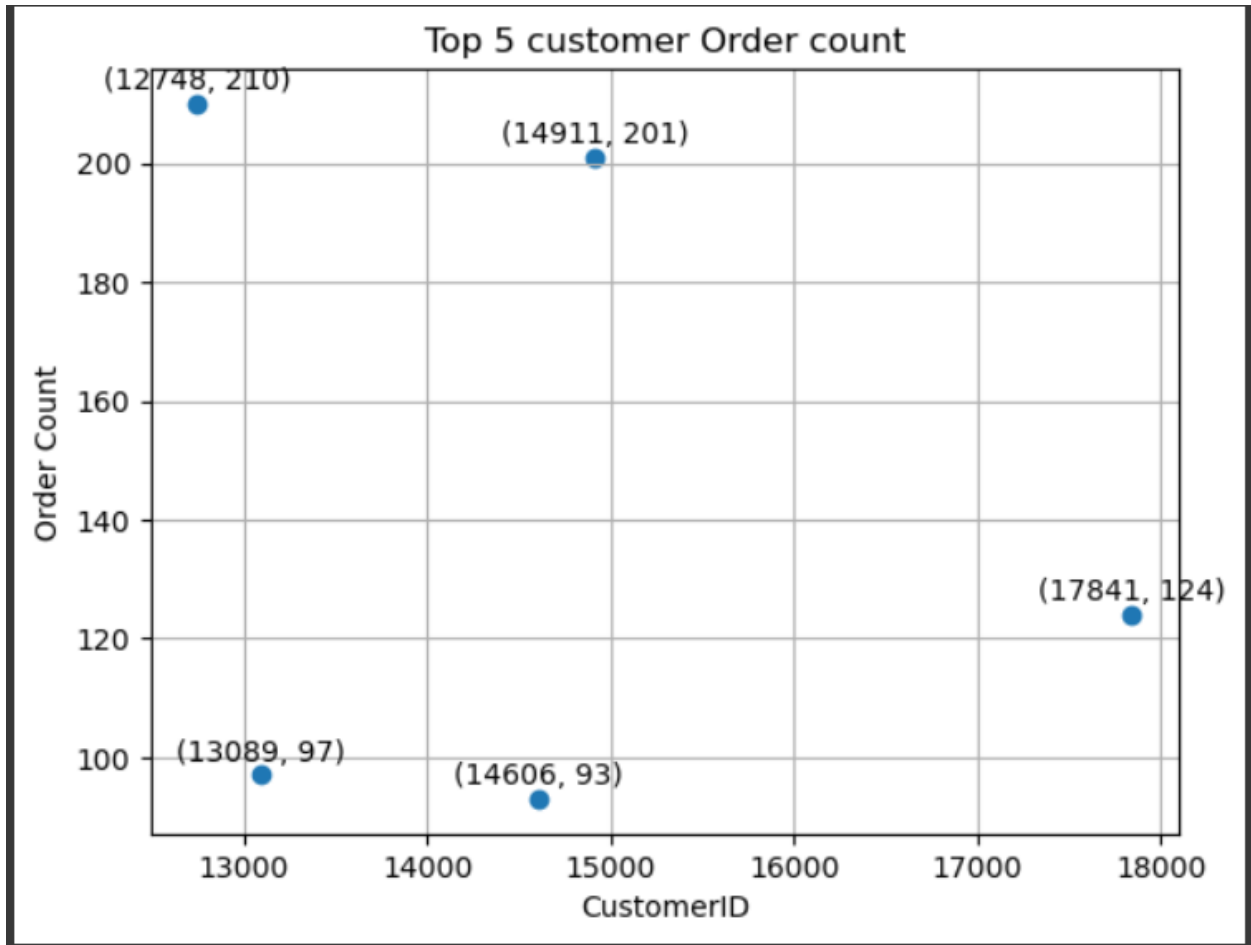


MARKETING RECOMMENDATION:

Re-engage lost customers with targeted promotions, enticing recent purchases. Revive hibernating customers through personalized offers and time-limited promotions. Maintain loyalty for consistently active customers with exclusive incentives. Retain "About to Sleep" customers by implementing exclusive offers or loyalty programs. Reactivate longer inactive customers with strategic campaigns and tailored incentives. Reinforce loyalty for consistently engaged customers through rewards and exclusive offers. Nurture promising customers with personalized communication and special promotions. Solidify relationships with recent users through personalized experiences and incentives. Foster potential loyalists with stronger connections via personalized interactions. Retain price-sensitive customers with discounts and tailored cost-saving opportunities. Recognize and reward champions with exclusive perks to increase their lifetime value.

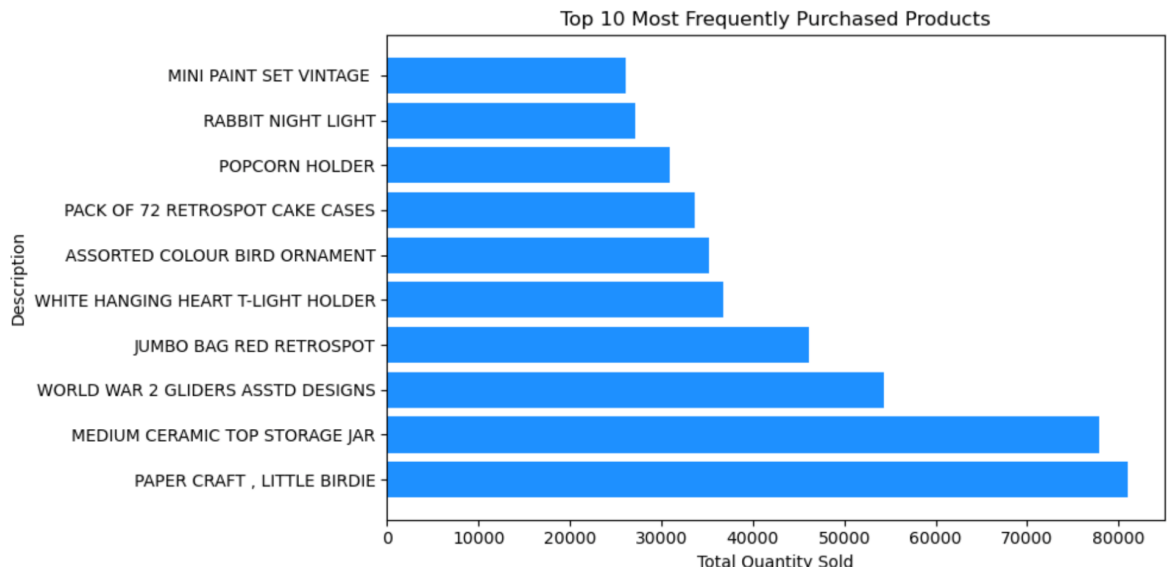
EXPLORATORY DATA ANALYSIS

CUSTOMER ANALYSIS:



The scatter plot illustrates the unique invoice counts for the top 5 customers. CustomerID 12748 stands out with the highest count of 209, suggesting a consistently high level of transactional activity and significant engagement with the business. CustomerID 14911 follows closely with 199 unique invoices, indicating another major contributor to the business. CustomerID 17841, 13089, and 14606 exhibit progressively lower unique invoice counts, suggesting varying levels of engagement. Analyzing these patterns provides insights into customer behavior and can guide strategic decisions, such as targeted marketing or customer retention efforts, to maximize business impact and enhance customer relationships.

PRODUCT ANALYSIS :



Based on total sales quantity, the purpose of this report is to visually represent the top 10 most frequently purchased products. The top 10 products are shown in the horizontal bar plot below according to their total sales quantity. Every bar is a product, and the length of the bar indicates the total number of units sold. The bar graph shows the highest products purchased in an increasing manner. Based on this the “Paper Craft, Little Birdie” is the most frequently purchased product.

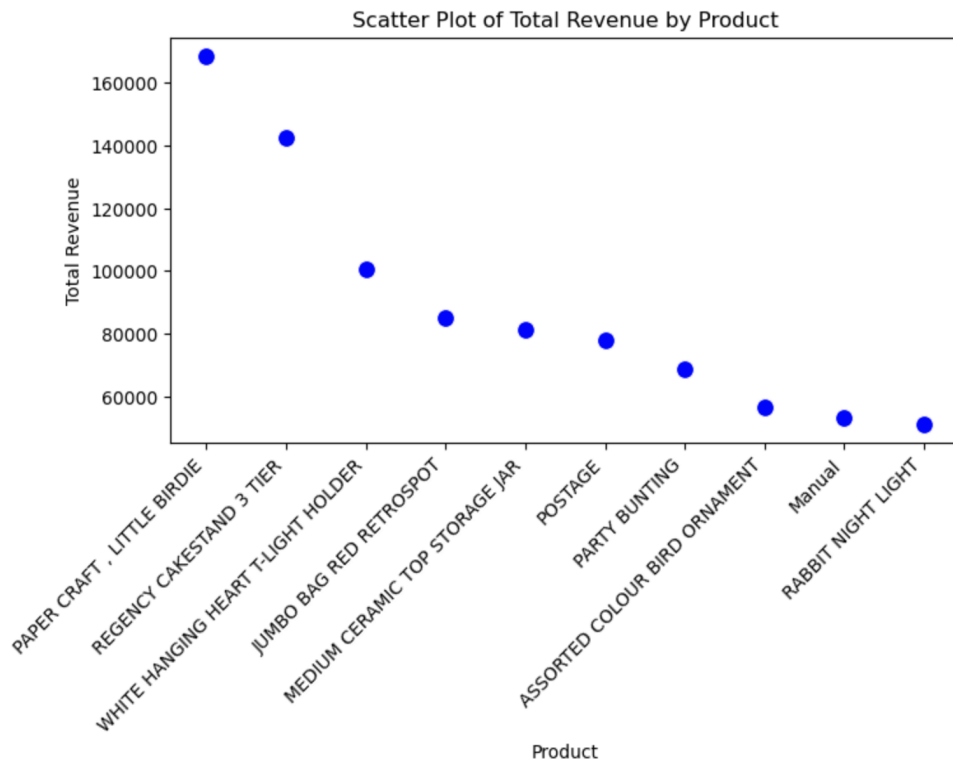
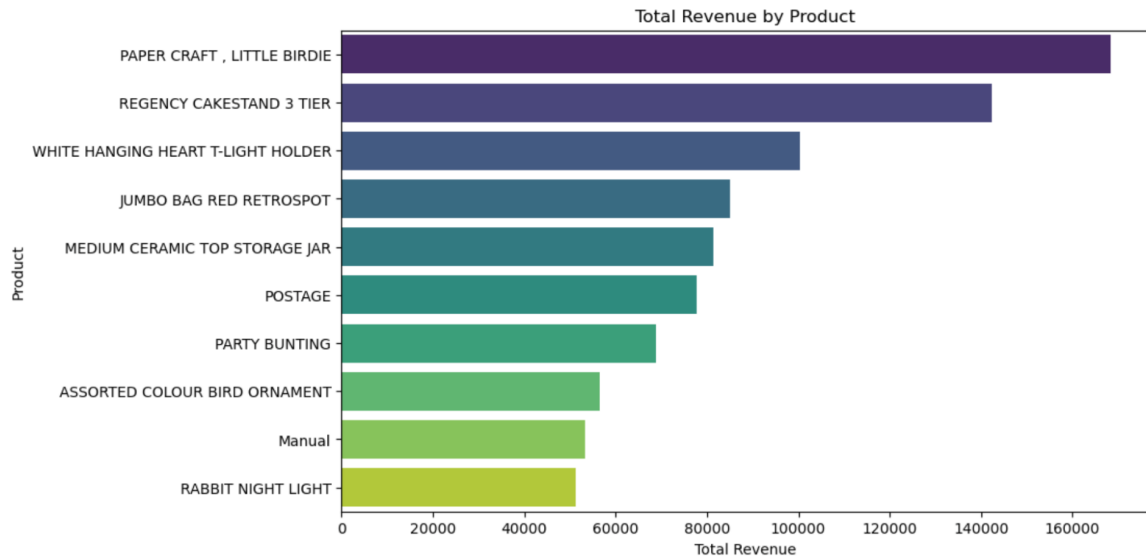
Top 10 average:

	Description	UnitPrice
1036	DOTCOM POSTAGE	744.147500
2395	PICNIC BASKET WICKER 60 PIECES	649.500000
2130	Manual	174.650211
2885	RUSTIC SEVENTEEN DRAWER SIDEBORD	158.076923
2769	REGENCY MIRROR WITH SHUTTERS	156.428571
3555	VINTAGE BLUE KITCHEN CABINET	146.750000
3604	VINTAGE RED KITCHEN CABINET	143.421053
689	CHEST NATURAL WOOD 20 DRAWERS	118.076923
1932	LOVE SEAT ANTIQUE WHITE METAL	114.024390
3600	VINTAGE POST OFFICE CABINET	66.360000

Least 10 average:

	Description	UnitPrice
969	DISCO BALL CHRISTMAS DECORATION	0.118478
490	BLUE STONES ON WIRE FOR CANDLE	0.102941
3800	WRAP BAD HAIR DAY	0.100000
2401	PIECE OF CAMO STATIONERY SET	0.096774
2606	PORCELAIN BUDAH INCENSE HOLDER	0.095263
1903	LETTER SHAPE PENCIL SHARPENER	0.094634
1671	HOUSE SHAPE PENCIL SHARPENER	0.090851
633	CARTOON PENCIL SHARPENERS	0.073200
2599	POPART WOODEN PENCILS ASST	0.043478
2275	PADS TO MATCH ALL CUSHIONS	0.000750

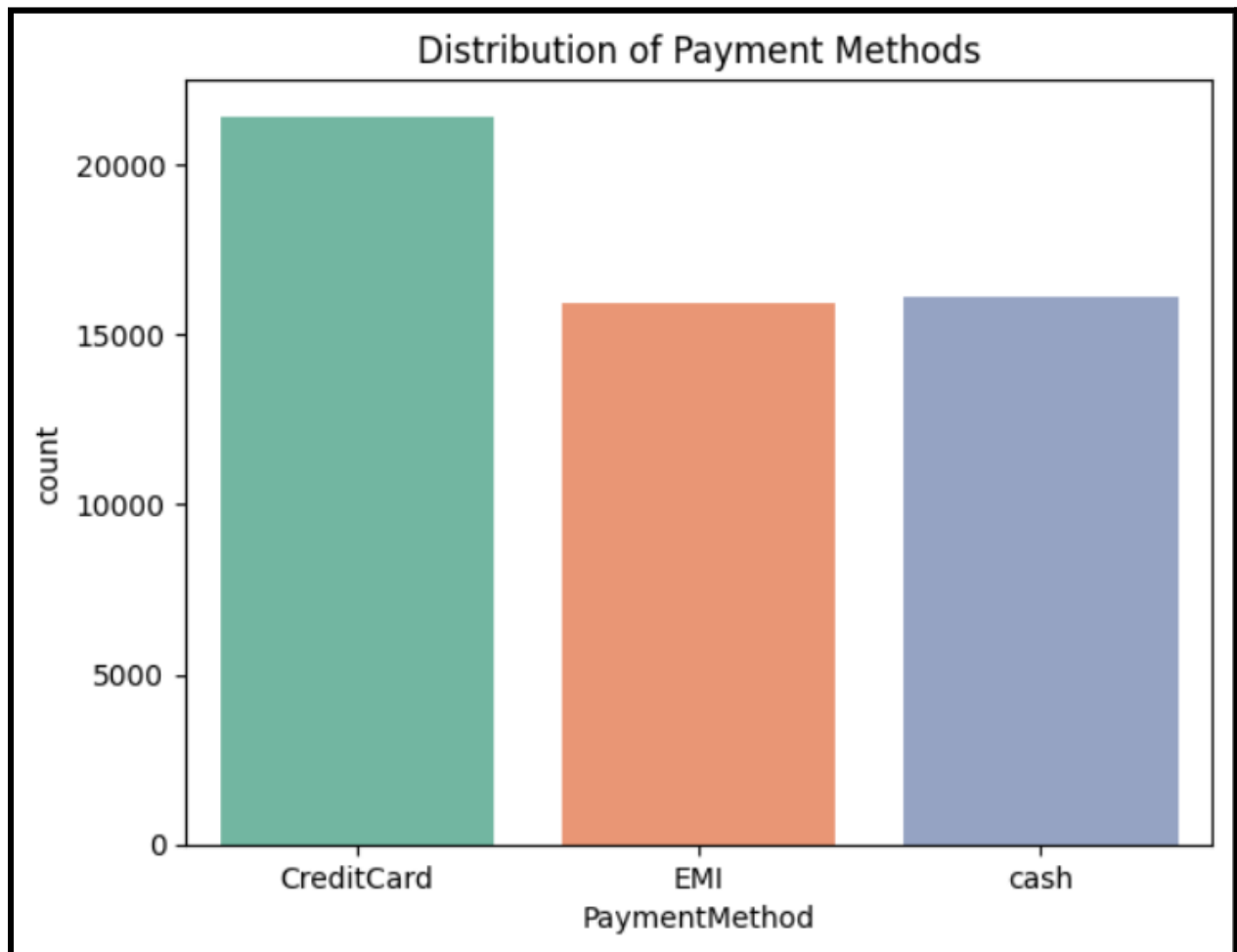
The table shows the top 10 and least 10 average price of products in the dataset. From the table we can infer that the top most average price of the product is “Dotcom Postage” which has a value of around 744.1475 and the least average price of the product is “Pads To Match All Cushions” which has a value of around 0.0007. This table clearly gives an idea about the average price of products in the dataset.



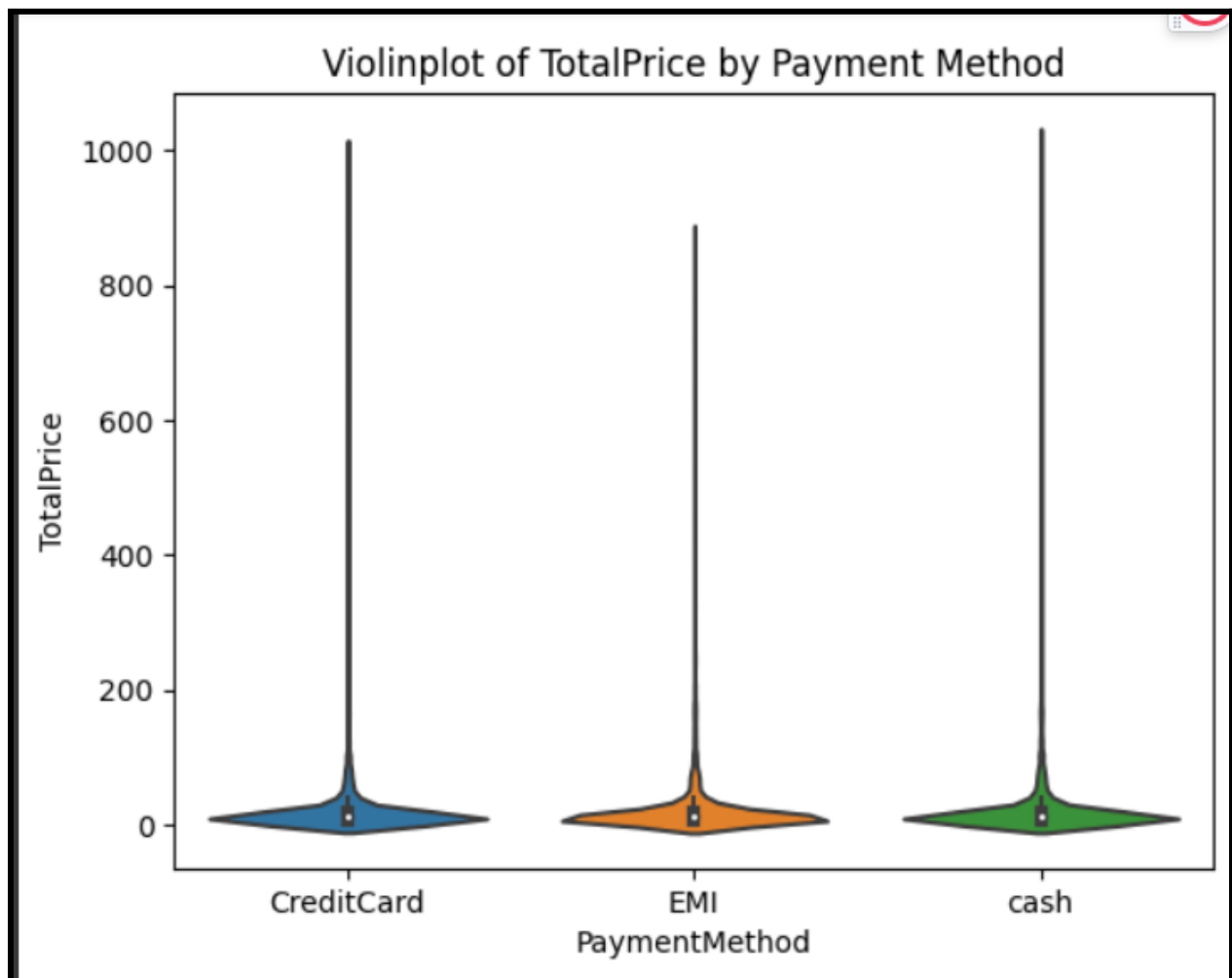
The above bar graph and scatter plot conveys about the total revenue generated by each product. This graphic helps with business planning by giving a brief but informative summary of the various products' respective revenue contributions. It also helps with

decision-making. Both the plots give an inference about the top ten revenue generated by each product. The figure shows the revenue generated in a decreasing order. From the report we can infer that the highest revenue generated is “Paper Craft, Little Birdie” which has a value above 160000.

PAYMENT ANALYSIS :



Payment method used the most is credit card. The mean total price is slightly higher for EMI payments compared to credit card payments, while the standard deviation is similar for both methods. Cash payments tend to have a slightly lower median and standard deviation compared to credit card and EMI payments.



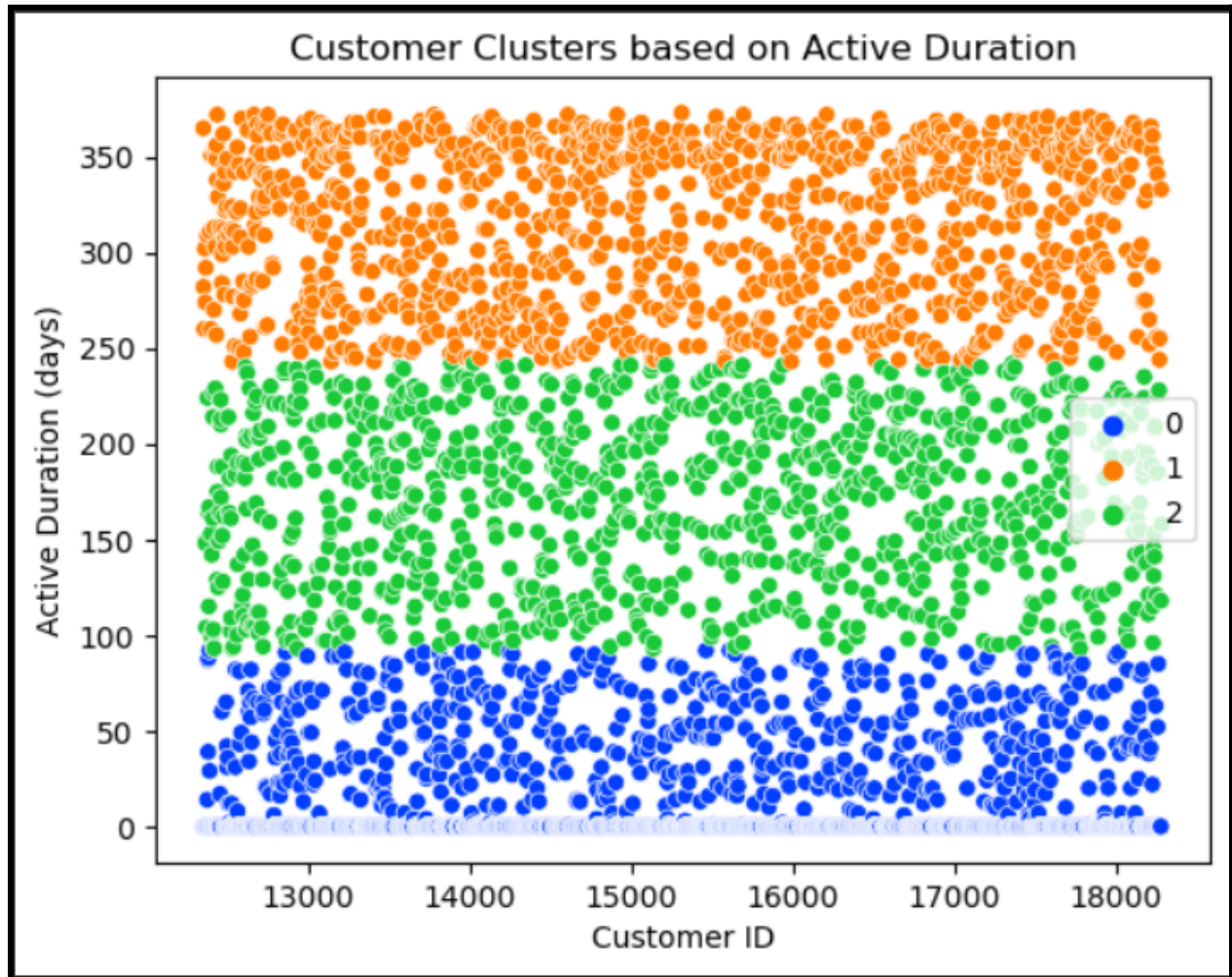
The median credit card payment method is high compared to that of the cash and EMI. IQR range is more in Credit card and then cash and at last for the EMI.

This shows customers prefer using Credit card and Cash payment more than that of the EMI option available to them.

The total price for all three payment methods is relatively widely distributed. This means that there is a wide range of total prices for each payment method.

There is also a significant amount of outliers present in the graph which shows that payment made could have been higher than expected during seasonal sales.

CUSTOMER BEHAVIOR:



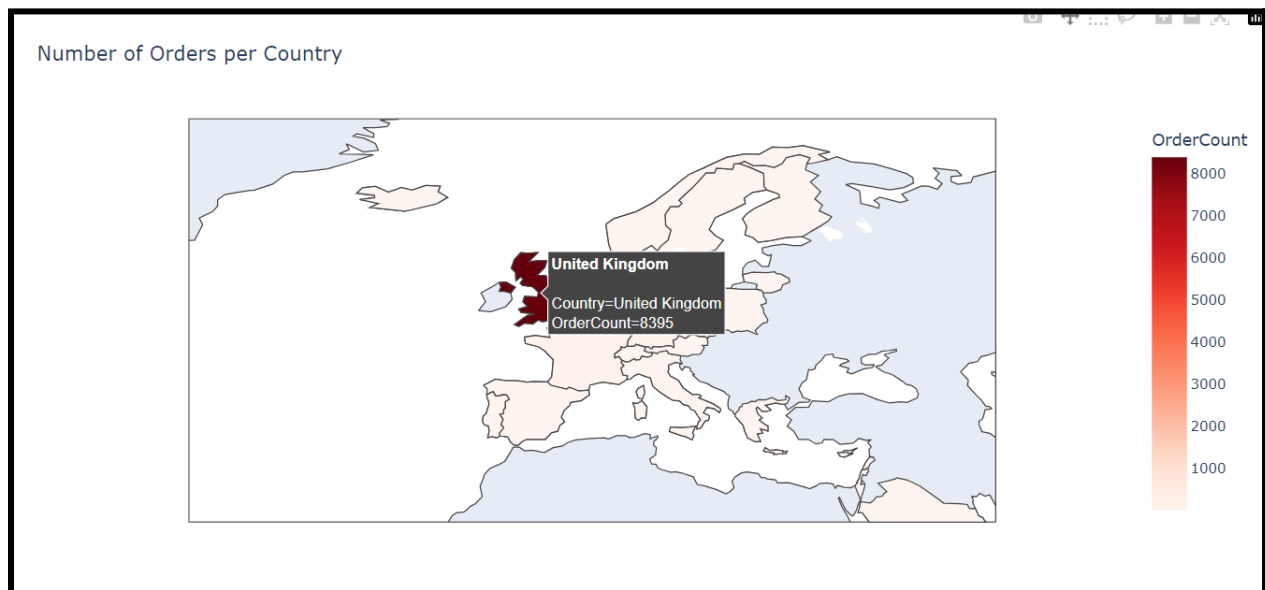
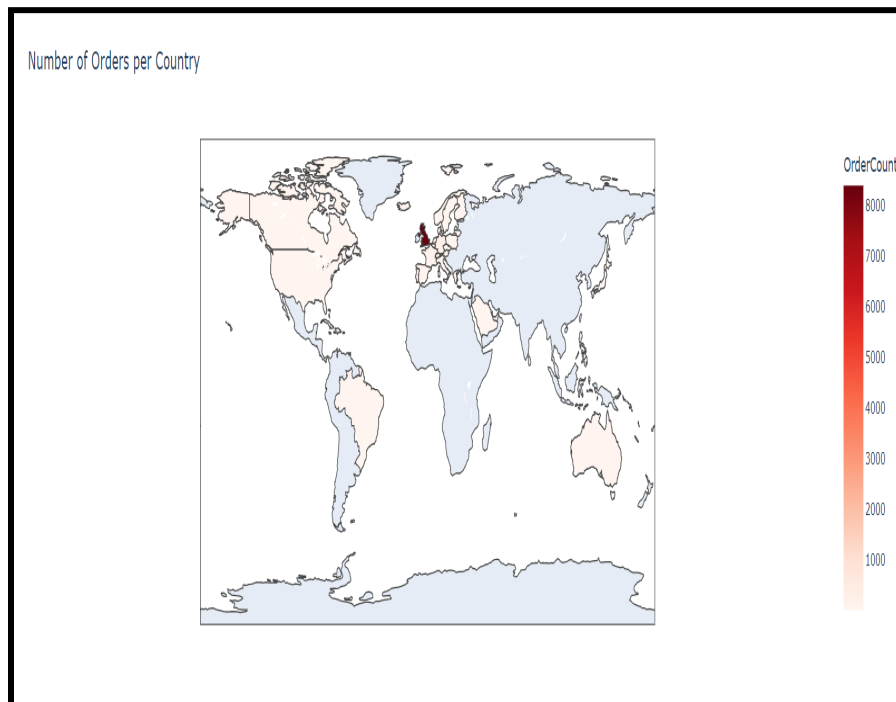
These clients are divided into three groups:

Cluster 0:- New clients, who began purchasing things only a few days ago.

Cluster 1:- Prospective Customers, They are continually purchasing items and have the potential to become loyal customers.

Cluster 2:- Loyal Customers, who buy items from the business on a regular basis and stay for a long time.

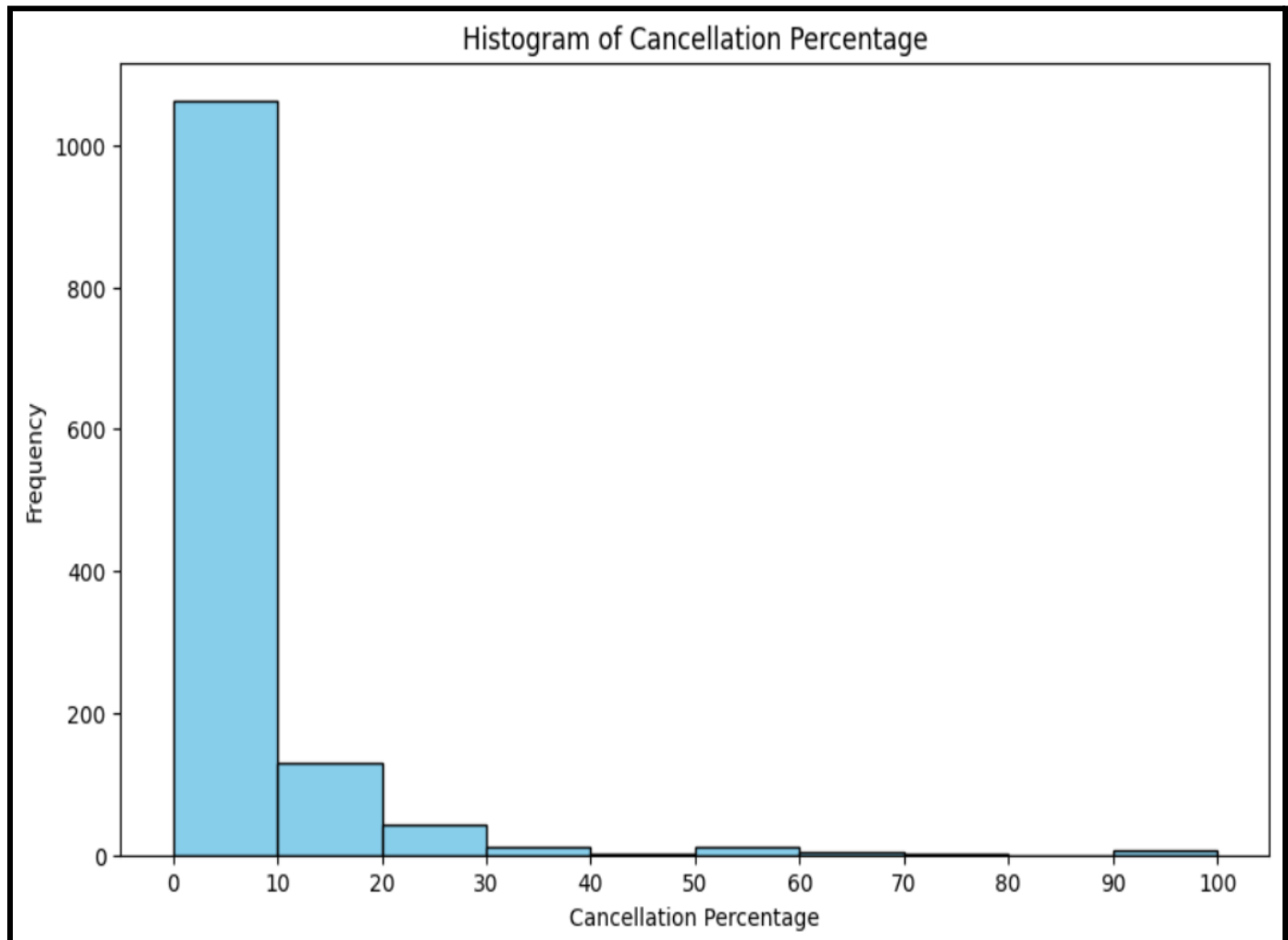
GEOGRAPHICAL ANALYSIS:



The choropleth map generated using Plotly Express provides a visually intuitive representation of the distribution of order counts across different countries. The map vividly illustrates the disparities in order volumes, with varying shades of red indicating the magnitude of orders in each country. Unsurprisingly, the United Kingdom stands out prominently with a deep color, signifying a significantly higher number of orders compared to other countries. This spatial visualization offers a quick and insightful

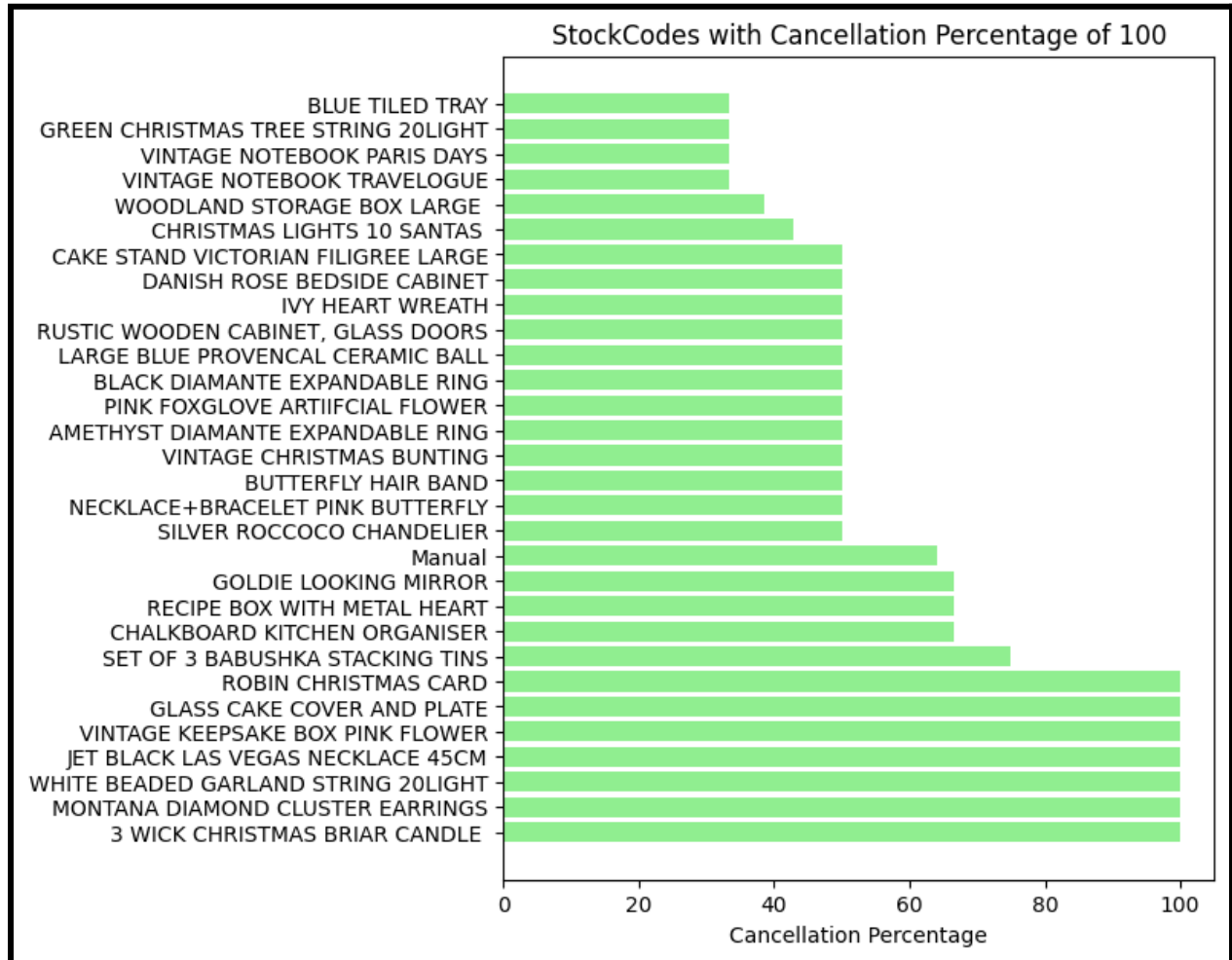
overview of the geographical distribution of business activity, aiding in identifying key markets and potential areas for strategic focus. The interactive nature of the choropleth map allows for a dynamic exploration of order counts across countries, enhancing the interpretability of the data

RETURN AND REFUNDS:



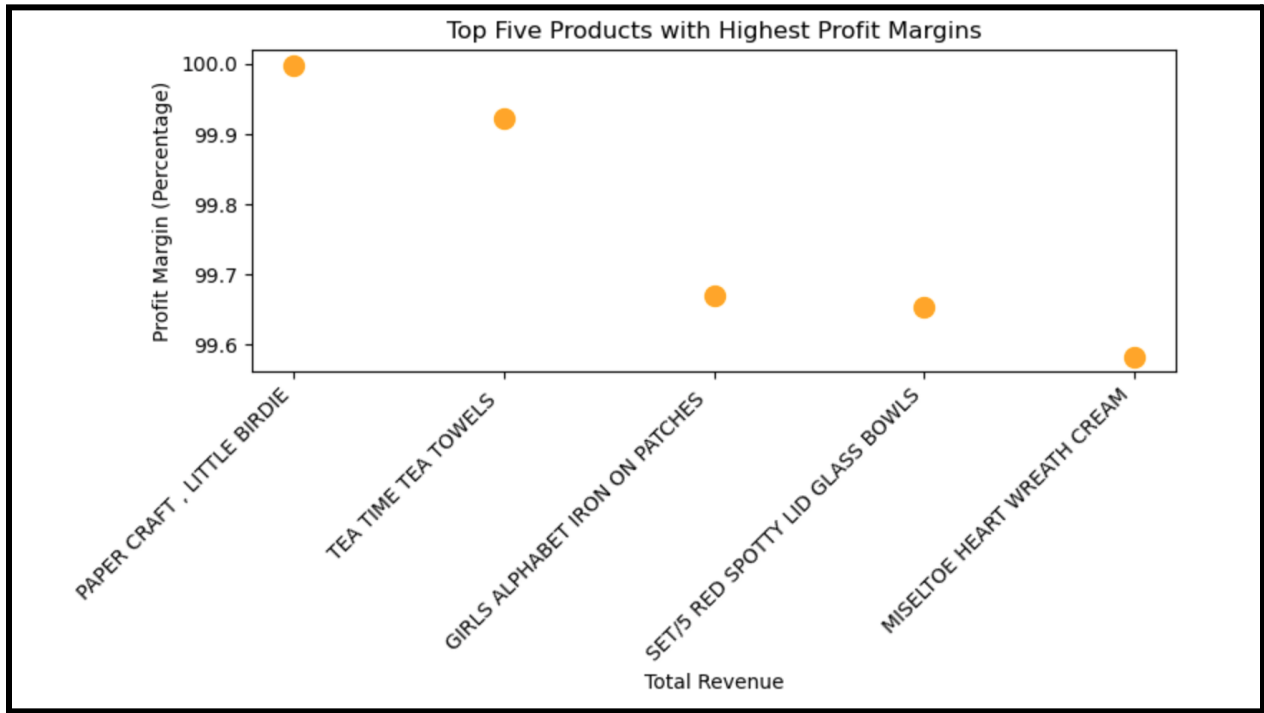
For 0-10 percent of canceled products we can infer a 1000 people returning the product ,for 10-20 % we can find less than 200 people returning the product as we go further the frequency of cancellation is less but there is subsequent rise in the mid in 50-60%.All the canceled the products are assumed to returned and refunded.

The most canceled product is a 3 wick Christmas Briar candle and least returned one is Blue tiled Tray.



PROFITABILITY ANALYSIS:

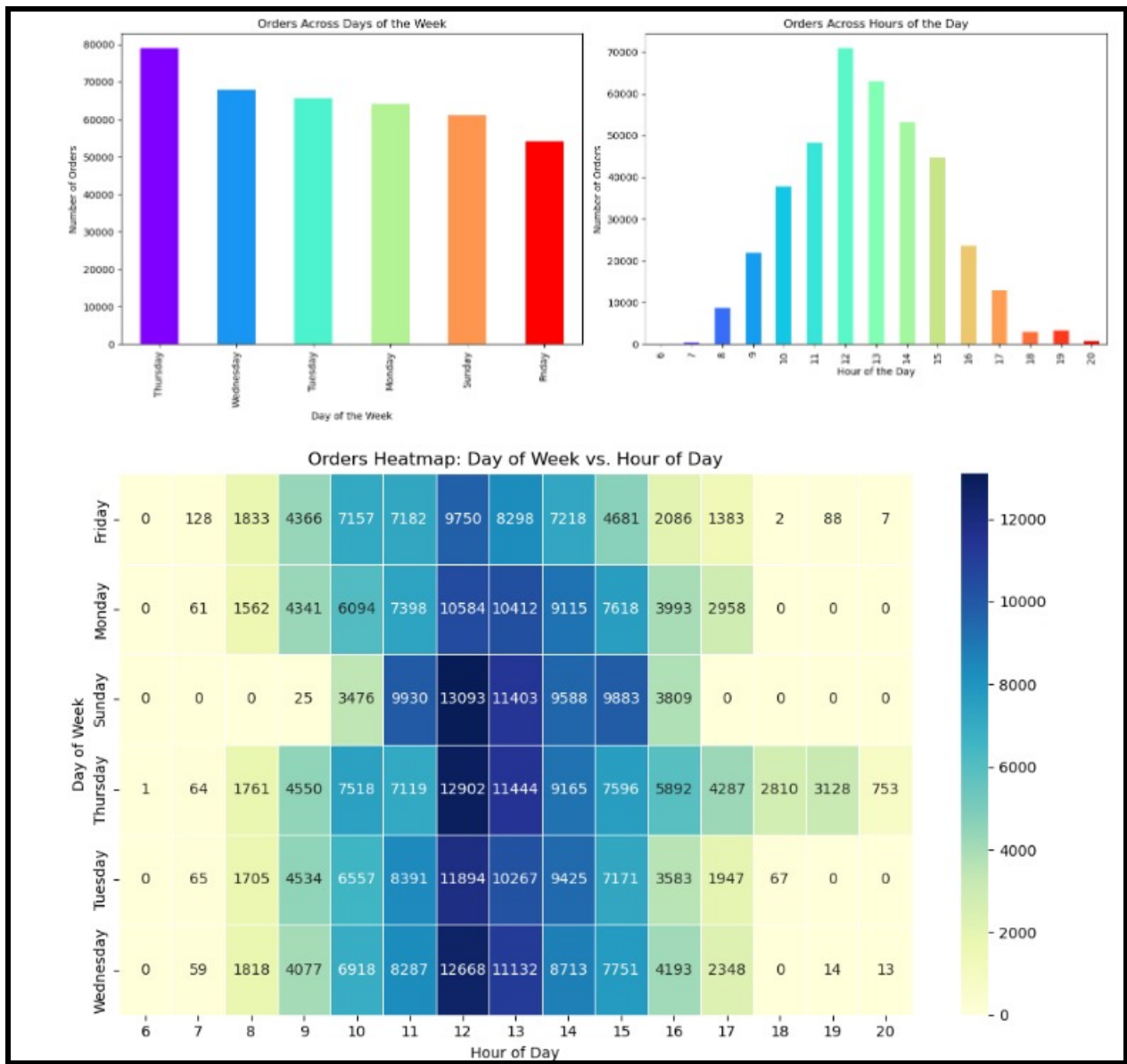
	Description	ProfitMargin
2319	PAPER CRAFT , LITTLE BIRDIE	99.998765
3423	TEA TIME TEA TOWELS	99.923077
1398	GIRLS ALPHABET IRON ON PATCHES	99.670894
3135	SET/5 RED SPOTTY LID GLASS BOWLS	99.652778
2082	MISELTOE HEART WREATH CREAM	99.583333



For each product in a DataFrame, the supplied code performs an extensive analysis of revenue, profit, and profit margins. The code determines the top five products with the

highest average profit margins by calculating each product's average profit margin and sorting the results in descending order. The profit margins of the top five products are displayed in a scatter plot created by the provided code, which also provides a visual depiction of the individual product contributions and the associated profit percentages. The products are shown on the horizontal axis, and the percentage profit margins are shown on the vertical axis. A clear picture of the distribution and fluctuation of profit margins among the best-performing products can be obtained from the scatter plot.

TIME ANALYSIS :



Most Common Day of the Week: Thursday

The highest number of orders occurs on Thursdays. This could indicate that Thursdays are particularly popular for shopping or that there might be specific promotions or marketing strategies implemented on that day. Most Common Hour of the Day: 12 (Noon)

The peak ordering hour is at noon.

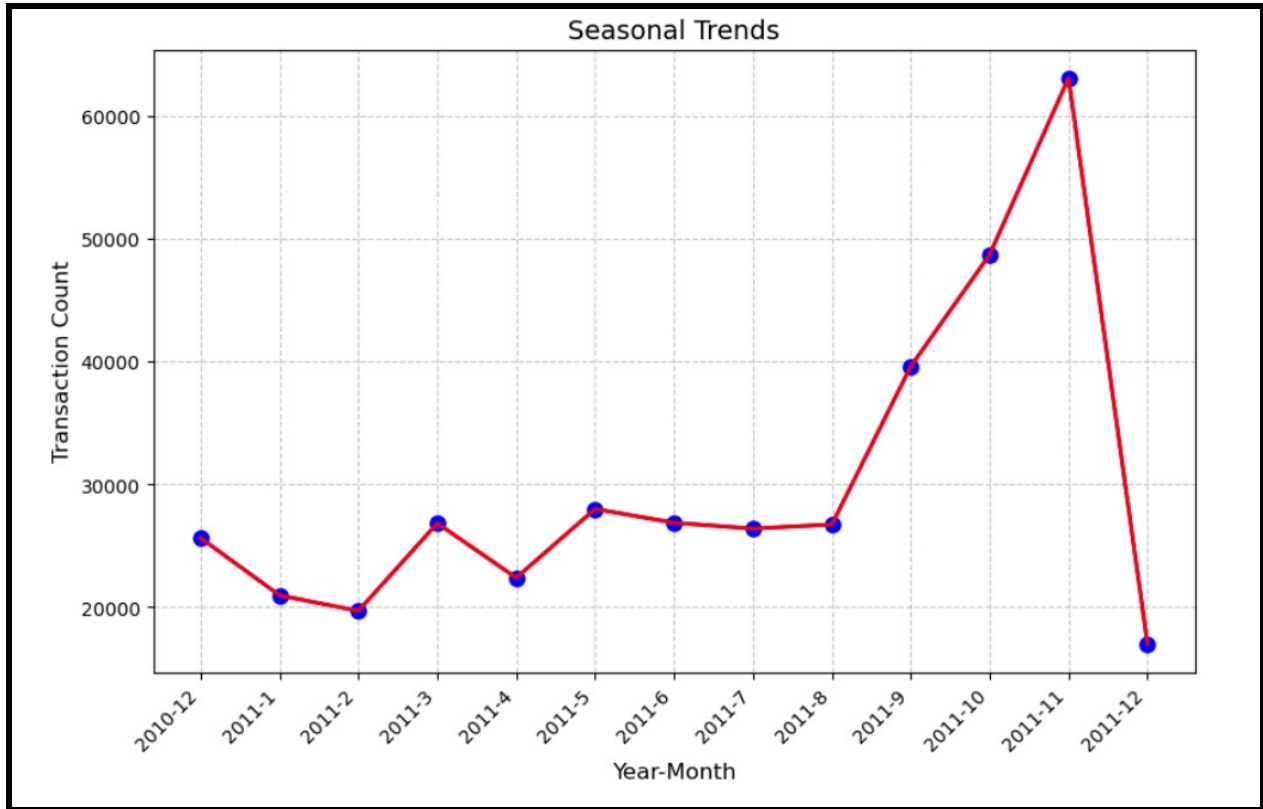
This could be associated with lunch breaks, where people might take the opportunity to browse and make purchases. Alternatively, it could be influenced by specific events, promotions, or activities happening around that time.

Customer Purchase Patterns:

The average processing time of 35 hours suggests that, on average, customers make repeat purchases within a relatively short time frame. This could indicate that the products offered are frequently used or consumed, prompting customers to replenish their supplies quickly. Opportunities for Targeted Marketing:

With a relatively short average processing time, there may be opportunities for targeted marketing strategies. For example, sending personalized promotions or recommendations shortly after a customer's previous purchase could encourage them to make additional orders. Efficient Inventory Management:

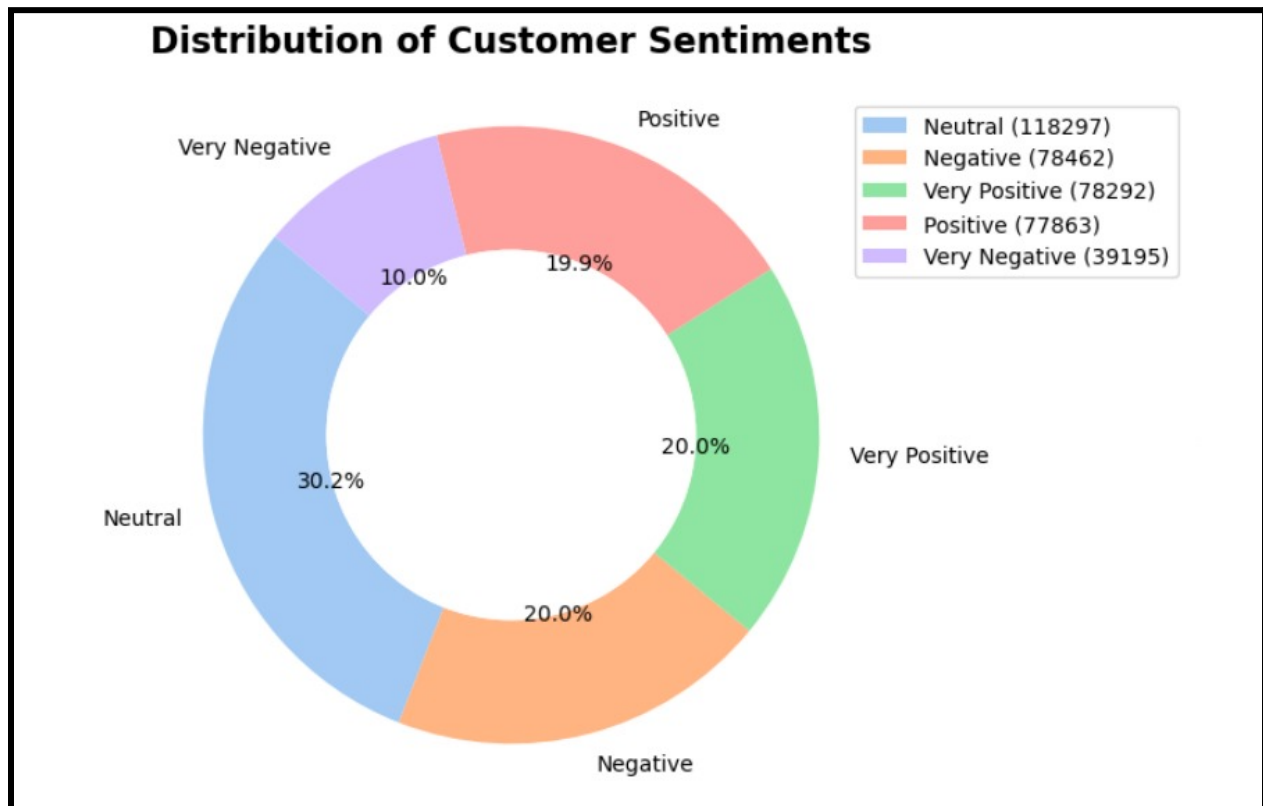
Understanding the average time between consecutive orders helps in optimizing inventory management. Businesses can adjust stock levels and reorder points to align with the expected frequency of customer purchases.



Yearly Shift in Transaction Volumes: The dataset highlights a significant shift in transaction counts from 2010 to 2011, indicating potential changes in customer behavior or business strategies. Analyzing the factors contributing to this transition can unveil key insights into the evolving dynamics of the business landscape during this period.

Identifying Seasonal Peaks: Specific months, notably November and October in 2011, showcase pronounced peaks in transaction volumes, suggesting strong seasonal influences. Understanding and leveraging these seasonal trends enables businesses to optimize resource allocation, inventory management, and marketing strategies to capitalize on peak

CUSTOMER SATISFACTION:



Customers are generally satisfied with the product. The fact that the majority of customer sentiments are positive suggests that customers are generally satisfied with the product.

There is room for improvement. The fact that 20% of customer sentiments are neutral or negative suggests that there is room for improvement in customer satisfaction.

The business should focus on retaining and marketing to its most satisfied customers. The fact that the majority of customer sentiments are positive suggests that the business should focus on retaining and marketing to its most satisfied customers.

The business should investigate the root causes of customer dissatisfaction. The fact that 20% of customer sentiments are neutral or negative suggests that the business should investigate the root causes of customer dissatisfaction and take steps to address them.