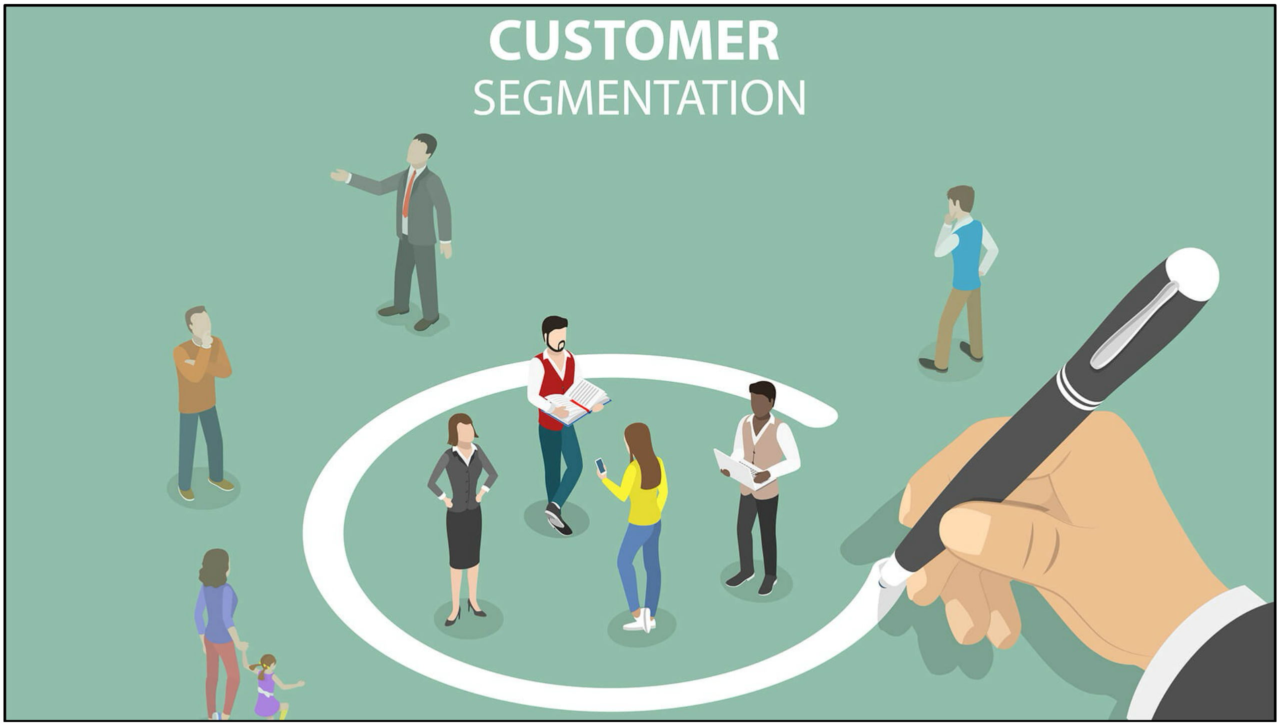


Customer Segmentation Analysis

IE6400 – Foundations Data Analytics Project Report 2

Group Number 09

Vikas Nijaguna Kagawad (002294765)

Deepikasai Mettu (002289270)

Kuchibhotla Anirudha Raghava Sarma (002207582)

Siddhant Pandit (002838340)

Sri Reshma Paramel (002821717)

Part 1: **Introduction and Research Questions**

In this project assignment, we will delve into the realm of e-commerce analytics using the "Online Retail" dataset, obtained from the UCI Machine Learning Repository. This dataset captures real transactions that occurred between December 2010 to 2011, for a different regions non-store online retail business.

The company specializes in offering unique all-occasion gifts, and a significant portion of its customer base comprises wholesalers. The dataset provides a comprehensive view of customer interactions with the online retail platform, encompassing a variety of transactions during the specified timeframe.

To unlock valuable insights and enhance strategic decision-making, we will employ the RFM (Recency, Frequency, Monetary) analysis method. RFM segmentation is a powerful technique widely used in businesses to categorize customers based on their recent purchasing behavior, purchase frequency, and monetary value.

Our overarching objective is to perform a rigorous RFM analysis on the dataset, leveraging the intrinsic patterns within customer transactions. By doing so, we aim to segment customers into distinct groups, each characterized by unique RFM scores. These segments will serve as a foundation for targeted marketing initiatives and tailored customer engagement strategies.

Through the lens of RFM analysis, we seek to unravel patterns that will not only enhance our understanding of customer behavior but also empower the formulation of effective marketing and customer retention strategies. As we embark on this journey through the eCommerce dataset, we anticipate gaining nuanced insights that will not only illuminate the dynamics of customer interactions but also inform data-driven decisions for optimizing the business's marketing and customer engagement efforts.

Introduction:
In this project, we delve into the realm of e-commerce analytics with a focus on customer segmentation using the RFM (Recency, Frequency, Monetary) analysis method. Our dataset, sourced from the UCI Machine Learning Repository, spans transactions from 2010 to 2011 for a UK-based non-store online retail business specializing in unique all-occasion gifts, with a substantial customer base of wholesalers.

1. **Objective**:
Our primary objective is to conduct a rigorous RFM analysis on the dataset, employing the intrinsic patterns within customer transactions. RFM segmentation is a powerful technique used by businesses to categorize customers based on their recent purchasing behavior, purchase frequency, and monetary value. By segmenting customers into distinct groups with unique RFM scores, we aim to unlock valuable insights for targeted marketing and customer retention strategies.

2. **Dataset Overview**:
The dataset provides a comprehensive snapshot of customer interactions, offering a wealth of transactional data. It serves as a valuable resource for understanding customer behavior, preferences, and engagement with the online retail platform.

3. **Project Scope**:
Our exploration extends beyond data preparation and RFM analysis. We will venture into the realm of exploratory data analysis (EDA), identifying patterns and trends within customer segments. By doing so, we aim to provide actionable insights that can guide marketing strategies, enhance customer experiences, and contribute to overall business growth.

4. **Methodology**:
Our journey involves meticulous data cleansing to ensure the dataset's integrity. Following this, the RFM analysis will be conducted, categorizing customers into segments based on recency, frequency, and monetary metrics. The subsequent EDA phase will unveil nuanced patterns, allowing for a deeper understanding of customer behavior.

5. **Expected Outcomes**:
By the conclusion of this project, we anticipate revealing customer segments with distinct behaviors and preferences. These insights will empower the business to tailor marketing strategies, optimize product offerings, and enhance customer engagement, ultimately fostering long-term customer relationships and business success. As we embark on this journey through the eCommerce dataset, the goal is to extract actionable intelligence that will not only benefit the business's bottom line but also elevate the overall customer experience within the e-commerce landscape.

Data Preprocessing:
This phase marks the initiation of our exploration and refinement process for the provided dataset. The dataset comprises actual transactions from a UK-based and registered non-store online retail company. We will scrutinize transactions that occurred between December 1, 2010, and December 9, 2011, to ensure the dataset's integrity and prepare it for a robust analysis.

1.1 **Overview of the Dataset**
Our dataset encompasses 541,909 entries with 8 columns, namely InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.

1.2 **Data Types**
Initial inspection revealed three primary data types: float64 (2 columns), int64 (1 column), and object (5 columns).

1.3 **Missing Values**
We identified missing values in the "Description" and "CustomerID" columns. Approximately 0.27% of "Description" entries and 24.93% of "CustomerID" entries were missing.

2. **Data Cleaning**

2.1 **Handling Missing Values**

To address the significant proportion (25%) of missing "CustomerID" entries, we opted to remove these rows, resulting in the elimination of 135,080 entries.

2.2 **Data Type Conversion**
In preparation for subsequent analysis, we converted the "InvoiceDate" column to a datetime format. Additionally, the data type of "CustomerID" was transformed into an object to maintain its categorical nature.

2.3 **Result**
The cleaned dataset is now devoid of missing values and features suitable data types. Specifically, "CustomerID" has been converted to an object, and "InvoiceDate" is now in datetime format. This preprocessing sets the stage for a more profound exploratory data analysis.

Recency, Frequency, Monetary

```
In [21]: #2.2 Recency
         recency_df = df[['CustomerID','InvoiceDate']]
         recent_date=max(recency_df.InvoiceDate)
         #Calculate recency for each invoice
         recent_date = recent_date + pd.DateOffset(days=1)
         recency_df['Diff'] = recent_date - recency_df.InvoiceDate
         recency_df

/var/folders/vw/d534bn6s4n93v5rfxxf3_my80000gn/T/ipykernel_2
A value is trying to be set on a copy of a slice from a Data
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.
rsus-a-copy
  recency_df['Diff'] = recent_date - recency_df.InvoiceDate
```

Out[21]:

	CustomerID	InvoiceDate	Diff
0	17850	2010-12-01 08:26:00	374 days 04:24:00
1	17850	2010-12-01 08:26:00	374 days 04:24:00
2	17850	2010-12-01 08:26:00	374 days 04:24:00
3	17850	2010-12-01 08:26:00	374 days 04:24:00
4	17850	2010-12-01 08:26:00	374 days 04:24:00
...	...	...	...
541904	12680	2011-12-09 12:50:00	1 days 00:00:00
541905	12680	2011-12-09 12:50:00	1 days 00:00:00
541906	12680	2011-12-09 12:50:00	1 days 00:00:00
541907	12680	2011-12-09 12:50:00	1 days 00:00:00
541908	12680	2011-12-09 12:50:00	1 days 00:00:00

401604 rows × 3 columns

Recency DataFrame:

Extracts 'CustomerID' and 'InvoiceDate' columns. Determines the most recent transaction date in the dataset. Computes the difference between the recent date and each customer's transaction date, storing it in the 'Diff' column.

```
#2.3 Frequency
frequency_df = df.groupby("CustomerID")["InvoiceNo"].count().reset_index()
frequency_df.columns=["CustomerID","Frequency"]
frequency_df
```

	CustomerID	Frequency
0	12346	2
1	12347	182
2	12348	31
3	12349	73
4	12350	17
...	...	...
4367	18280	10
4368	18281	7
4369	18282	13
4370	18283	721
4371	18287	70

4372 rows × 2 columns

Frequency DataFrame:
Groups the data by 'CustomerID'. Counts the number of unique invoices for each customer. Renames columns to 'CustomerID' and 'Frequency'.

```
#2.4 Monetary
monetary_df = df.groupby("CustomerID")["Total Price"].sum().reset_index()
monetary_df.columns = ["CustomerID","Monetary"]
monetary_df
```

	CustomerID	Monetary
0	12346	0.00
1	12347	4310.00
2	12348	1797.24
3	12349	1757.55
4	12350	334.40
...	...	...
4367	18280	180.60
4368	18281	80.82
4369	18282	176.60
4370	18283	2045.53
4371	18287	1837.28

4372 rows × 2 columns

Monetary DataFrame:
Groups the data by 'CustomerID'. Sums the 'Total Price' for each customer, representing their total monetary contribution. Renames columns to 'CustomerID' and 'Monetary'.

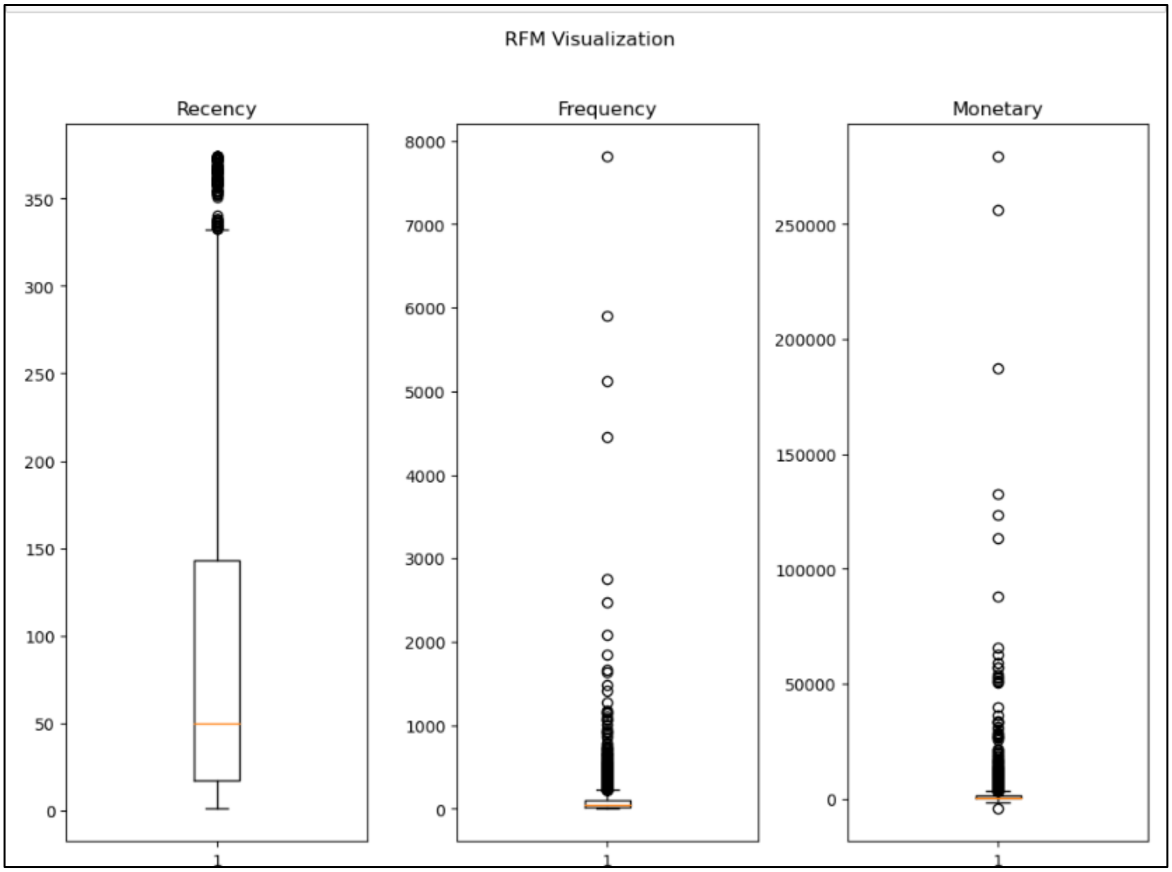
```python
#2.4 Monetary
monetary_df = df.groupby("CustomerID")["Total Price"].sum().reset_index()
monetary_df.columns = ["CustomerID","Monetary"]
monetary_df
```

	CustomerID	Monetary
0	12346	0.00
1	12347	4310.00
2	12348	1797.24
3	12349	1757.55
4	12350	334.40
...	...	...
4367	18280	180.60
4368	18281	80.82
4369	18282	176.60
4370	18283	2045.53
4371	18287	1837.28

4372 rows × 2 columns

Outlier Treatment:
Calculates the first quartile (Q1), third quartile (Q3), and interquartile range (IQR) for each RFM metric. Filters out data points beyond a certain range (1.5 times the IQR from Q1 and Q3) to mitigate the impact of outliers.

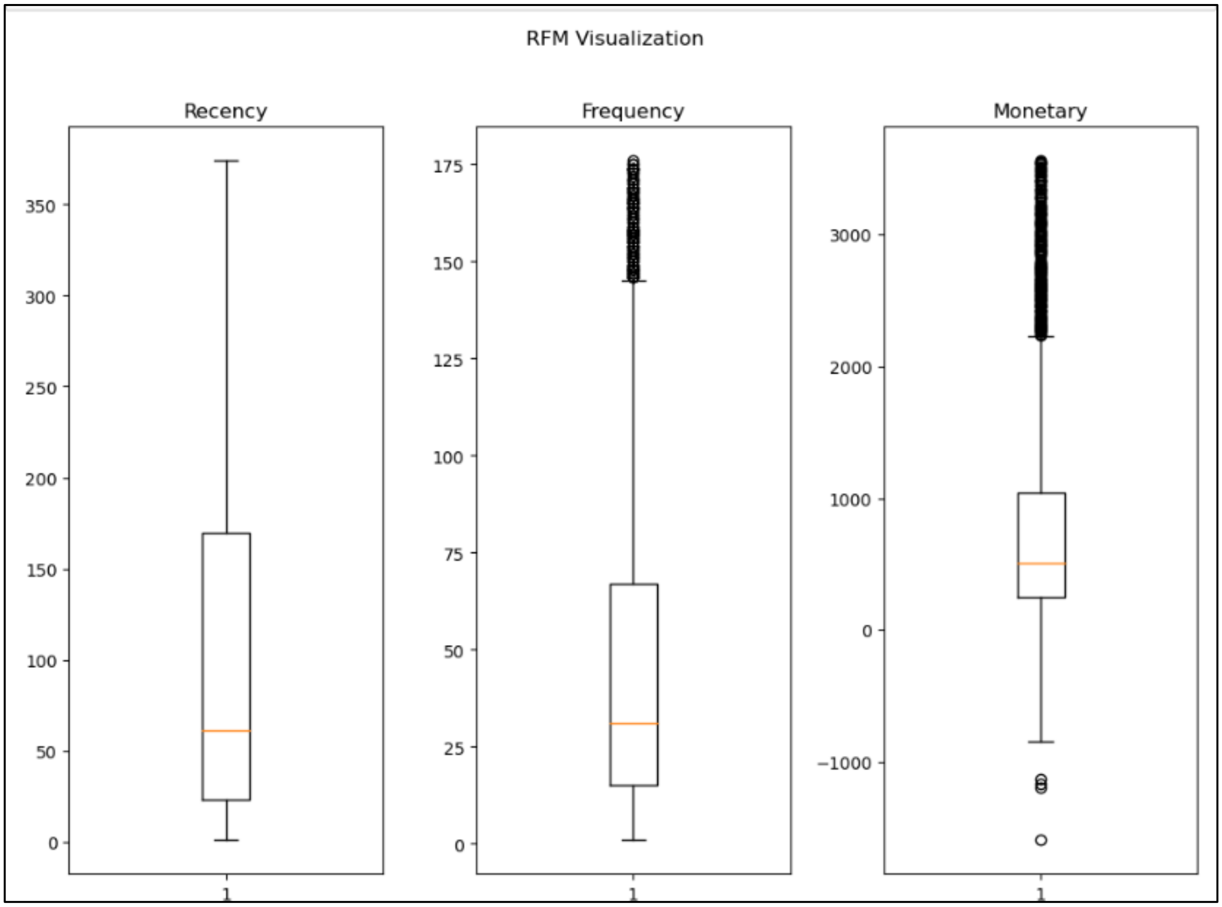
RFM Visualization before outlier treatment



In this phase, we aimed to visualize the distribution of key RFM (Recency, Frequency, Monetary) metrics before addressing outliers. The Matplotlib library was employed to create a comprehensive boxplot illustration.

1. Recency:
   - The first subplot showcases the distribution of Recency values, representing the number of days since the last transaction for each customer.
2. Frequency:
   - The second subplot illustrates the distribution of Frequency values, indicating the count of transactions for each customer.
3. Monetary:
   - The third subplot displays the distribution of Monetary values, representing the total monetary value of transactions for each customer.

RFM Visualization after Outlier Treatment:



Following the identification of outliers in the RFM metrics (Recency, Frequency, Monetary), a robust outlier treatment was applied to refine the dataset. This involved the utilization of the Interquartile Range (IQR) method to filter out extreme values.

1. Recency:
   - The first subplot of the updated visualization displays the distribution of Recency values after outlier treatment. Recency represents the number of days since the last transaction for each customer.
2. Frequency:
   - The second subplot depicts the distribution of Frequency values post-outlier treatment. Frequency indicates the count of transactions for each customer.
3. Monetary:
   - The third subplot exhibits the distribution of Monetary values after addressing outliers. Monetary reflects the total monetary value of transactions for each customer.

RFM Segmentation:

To derive meaningful insights from the RFM metrics (Recency, Frequency, Monetary), a segmentation approach was employed. Before clustering, all parameters were standardized using the Standard Scaler from scikit-learn. Standardization ensures that each metric contributes equally to the clustering process, preventing any metric from dominating the analysis.

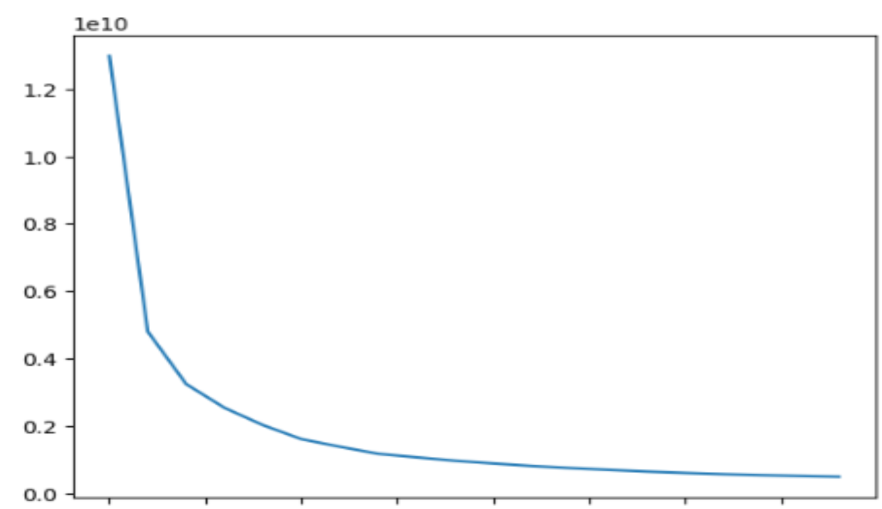
```
#RFM Segmentation
# standardise all parameters
from sklearn.preprocessing import StandardScaler
standard_scaler = StandardScaler()
standard_scaler.fit_transform(rfm_df)
```

```
array([[-1.72558978,  2.13727356, -1.07413759, -1.04176195],
       [-1.7244252 , -0.28178068, -0.37022813,  1.39975609],
       [-1.72384291, -0.82148999,  0.64922695,  1.34583794],
       ...,
       [ 1.73029055,  0.73981194, -0.95277389, -0.93196944],
       [ 1.73087284, -0.92750432, -0.80713745, -0.80185404],
       [ 1.73378428, -0.590186  ,  0.57640873,  1.45414971]])
```

Elbow Method for Optimal Cluster Selection:

Determining the optimal number of clusters is crucial for effective segmentation. The Elbow Method was employed to find the point where the rate of decrease in the sum of squared distances (SSD) sharply changes, indicating an appropriate number of clusters. A range of clusters, from 1 to 20, was evaluated, and the results were plotted to visualize the elbow curve. The optimal number of clusters is typically identified at the "elbow" of the curve, where adding more clusters provides diminishing returns in terms of reducing SSD.

```
#Plot elbow curve
plt.plot(ssd)
plt.show()
```



Analysis of clusters formed:
A new DataFrame (RFM_km) is created by combining the original RFM data with the cluster labels assigned by the K-Means algorithm. This DataFrame now includes a 'Cluster' column indicating which cluster each customer belongs to.
The mean values of monetary, frequency, and recency are calculated for each cluster, providing insights into the average behavior of customers within each cluster.
The results for each cluster are combined into a summary DataFrame (km_clusters), where each row represents a cluster and columns include the cluster identifier along with the mean values for amount, frequency, and recency.

```
# analysis of clusters formed
rfm_df.index = pd.RangeIndex(len(rfm_df.index))
RFM_km = pd.concat([rfm_df, pd.Series(model_clus3.labels_)], axis=1)
RFM_km.columns = ['CustomerID', 'Recency', 'Frequency', 'Monetary', 'Cluster']
km_clusters_amount = pd.DataFrame(RFM_km.groupby(["Cluster"]).Monetary.mean())
km_clusters_frequency = pd.DataFrame(RFM_km.groupby(["Cluster"]).Frequency.mean())
km_clusters_recency = pd.DataFrame(RFM_km.groupby(["Cluster"]).Recency.mean())
```

```
km_clusters = pd.concat([pd.Series([0,1,2]), km_clusters_amount, km_clusters_frequency, km_clusters_recency], axis=1)
km_clusters.columns = ["Cluster", "Amount_mean", "Frequency_mean", "Recency_mean"]
km_clusters.head()
```

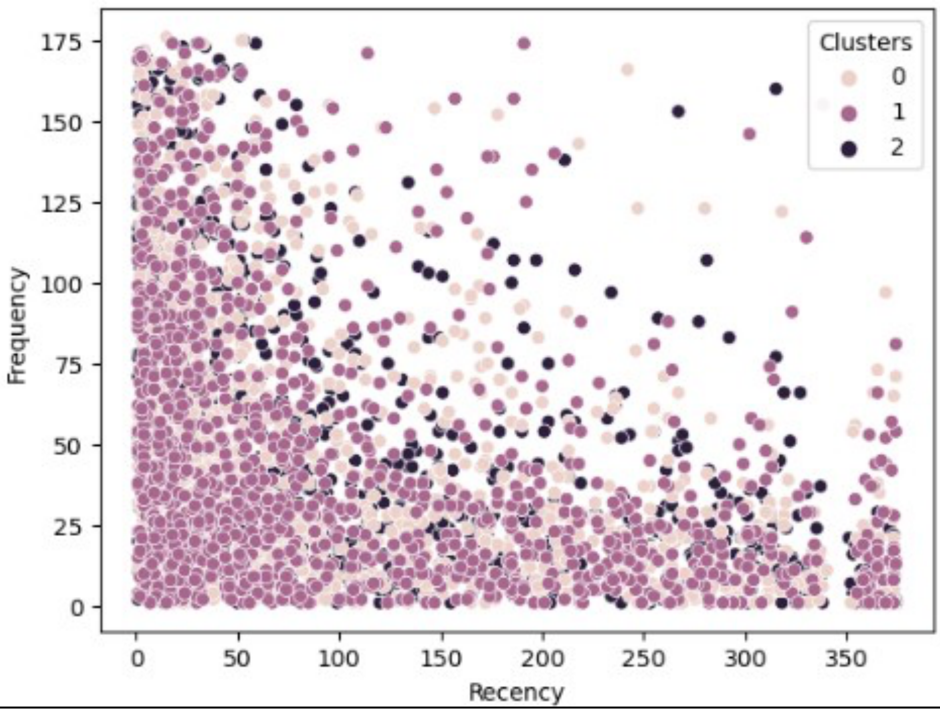
	Cluster	Amount_mean	Frequency_mean	Recency_mean
0	0	744.279510	46.068438	105.710145
1	1	688.262498	45.983727	104.403580
2	2	868.104331	46.707555	102.585703

Adding clusters to rfm dataframe

Scatter Plot - Recency vs Frequency:
This scatter plot visualizes the relationship between two RFM (Recency, Frequency, Monetary) parameters, specifically Recency and Frequency. Each point on the plot represents a customer, and the points are color-coded based on their assigned clusters. This visualization helps in identifying patterns and groupings within the dataset.
X-axis (Recency): Represents how recently a customer made a purchase. Lower values indicate more recent purchases.
Y-axis (Frequency): Represents the frequency of purchases made by each customer.
Hue (Clusters): Color-codes the points according to the clusters assigned by the K-Means algorithm.
Each color represents a distinct cluster. This plot provides insights into how customers are distributed based on their recency and frequency of purchases, making it easier to observe the segmentation achieved by the clustering algorithm.

```
# Recency vs Frequency
sns.scatterplot(x='Recency',y='Frequency',data=rfm_df,hue='Clusters')
```

```
<Axes: xlabel='Recency', ylabel='Frequency'>
```



Scatter Plot - Recency vs Monetary:

This scatter plot illustrates the relationship between Recency and Monetary parameters for each customer. Similar to the previous plot, points are color-coded based on their assigned clusters, allowing for the observation of patterns and distinctions within the dataset.

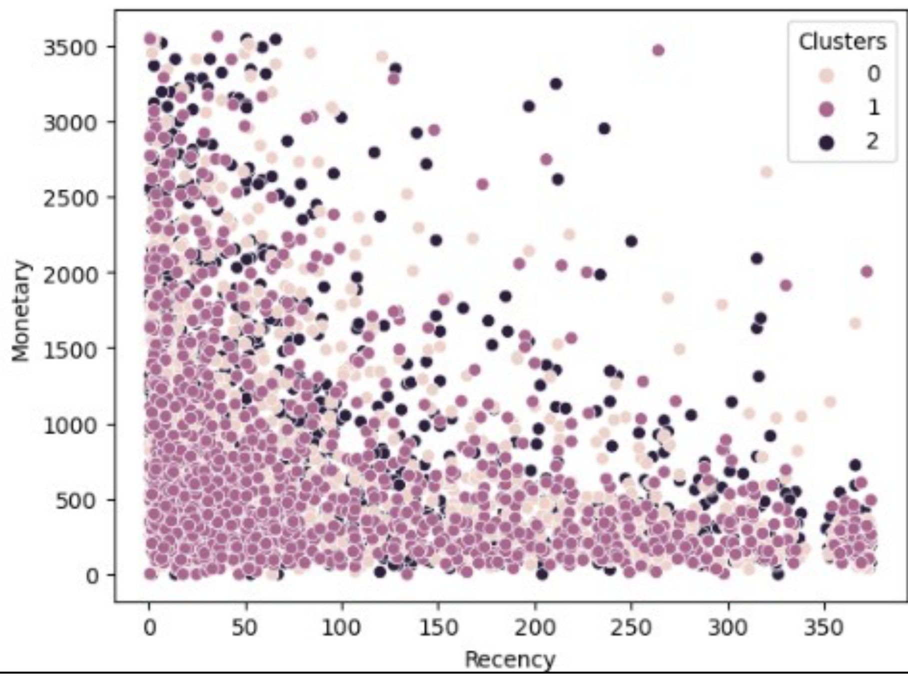
X-axis (Recency): Represents how recently a customer made a purchase. Lower values indicate more recent purchases.

Y-axis (Monetary): Represents the monetary value of purchases made by each customer.

This visualization helps in understanding how customers are distributed concerning both recency and the monetary value of their purchases.

```python
# Recency vs Monetary
sns.scatterplot(x='Recency',y=rfm_df[rfm_df['Monetary']>=0]['Monetary'],data=rfm_df,hue='Clusters')
```

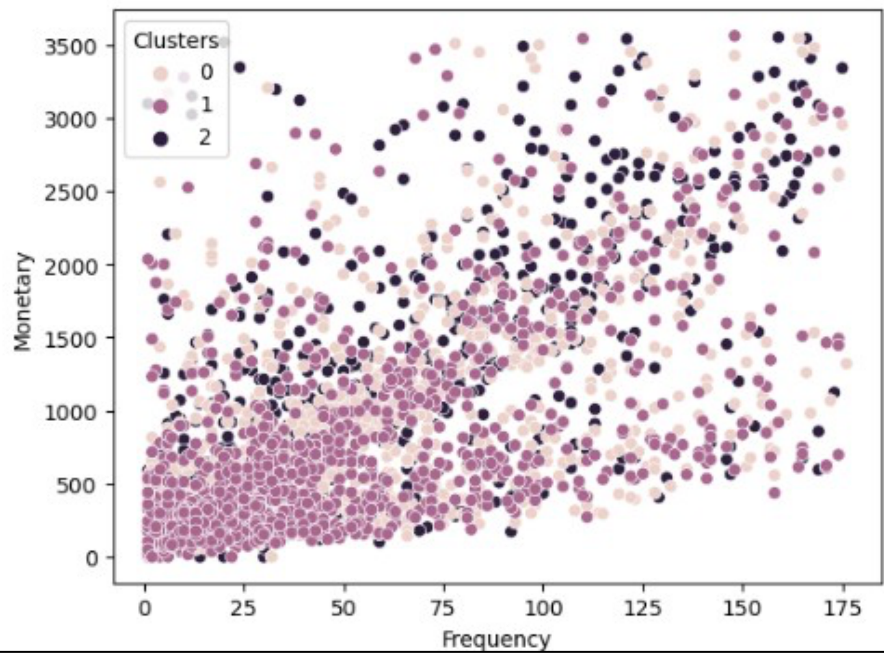
```
<Axes: xlabel='Recency', ylabel='Monetary'>
```



Scatter Plot - Frequency vs Monetary:
This scatter plot explores the relationship between Frequency and Monetary parameters. Points are color-coded by their assigned clusters, allowing for the identification of groups based on purchase frequency and monetary spending.
X-axis (Frequency): Represents the frequency of purchases made by each customer.
Y-axis (Monetary): Represents the monetary value of purchases made by each customer.
This visualization provides insights into how customer segments differ in terms of both purchase frequency and monetary value, aiding in the analysis of distinct clusters formed by the K-Means algorithm.

```python
# Frequency vs Monetary
sns.scatterplot(x='Frequency',y=rfm_df[rfm_df['Monetary']>=0]['Monetary'],data=rfm_df,hue='Clusters')
```

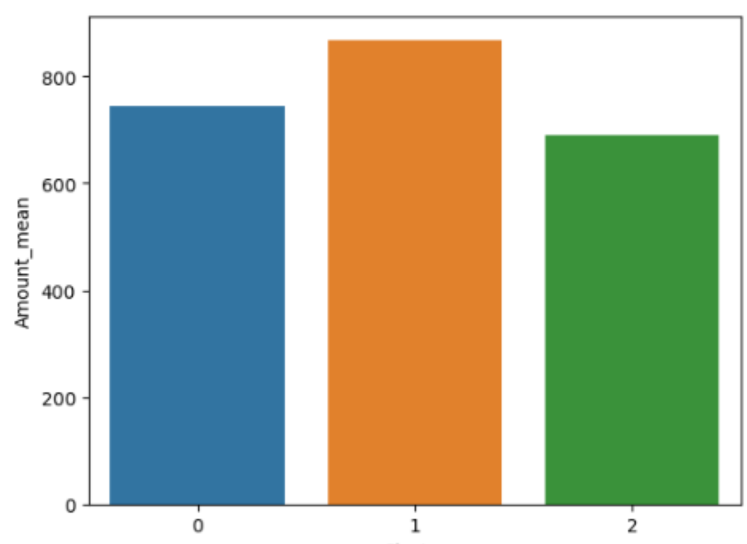
```
<Axes: xlabel='Frequency', ylabel='Monetary'>
```



Bar Plot - Cluster vs Average Monetary Amount:

```python
sns.barplot(data=km_clusters, x='Cluster', y='Amount_mean')
```

```
<Axes: xlabel='Cluster', ylabel='Amount_mean'>
```

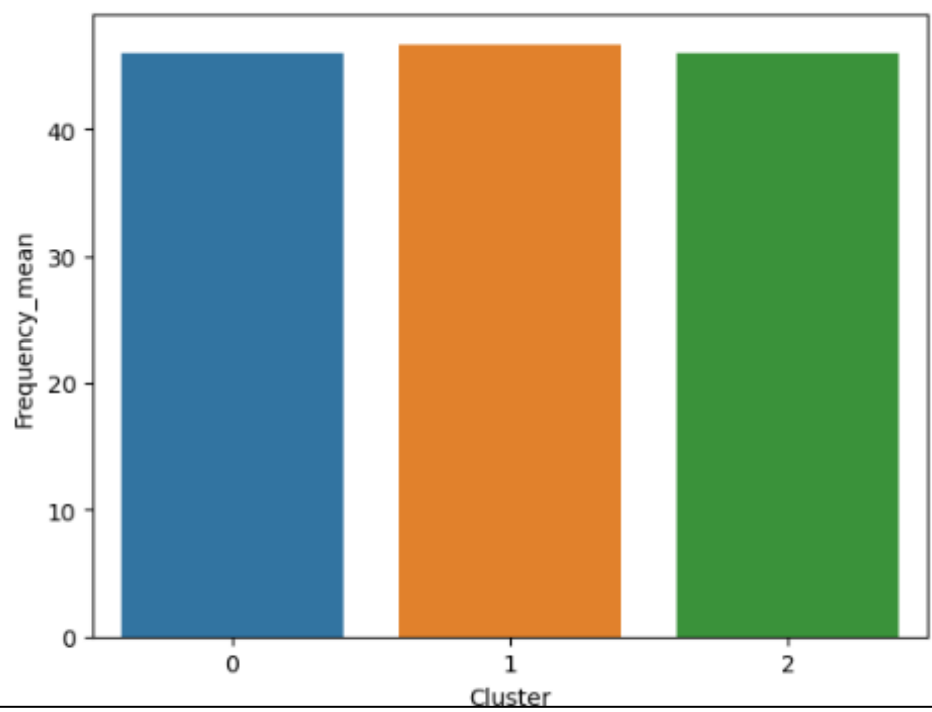


This bar plot displays the average monetary amount for each cluster. Each bar represents a different cluster, and the height of the bar corresponds to the mean monetary value of purchases for customers within that cluster.
X-axis (Cluster): Represents the clusters formed by the K-Means algorithm.
Y-axis (Amount_mean): Represents the average monetary amount of purchases made by customers within each cluster.
Analyzing this plot allows you to understand the variations in average spending across different clusters. It provides insights into how distinct customer segments, identified by the clustering algorithm, differ in terms of their average monetary contributions.

Bar Plot - Cluster vs Average Purchase Frequency:

```
sns.barplot(data=km_clusters, x='Cluster', y='Frequency_mean')
```

```
<Axes: xlabel='Cluster', ylabel='Frequency_mean'>
```

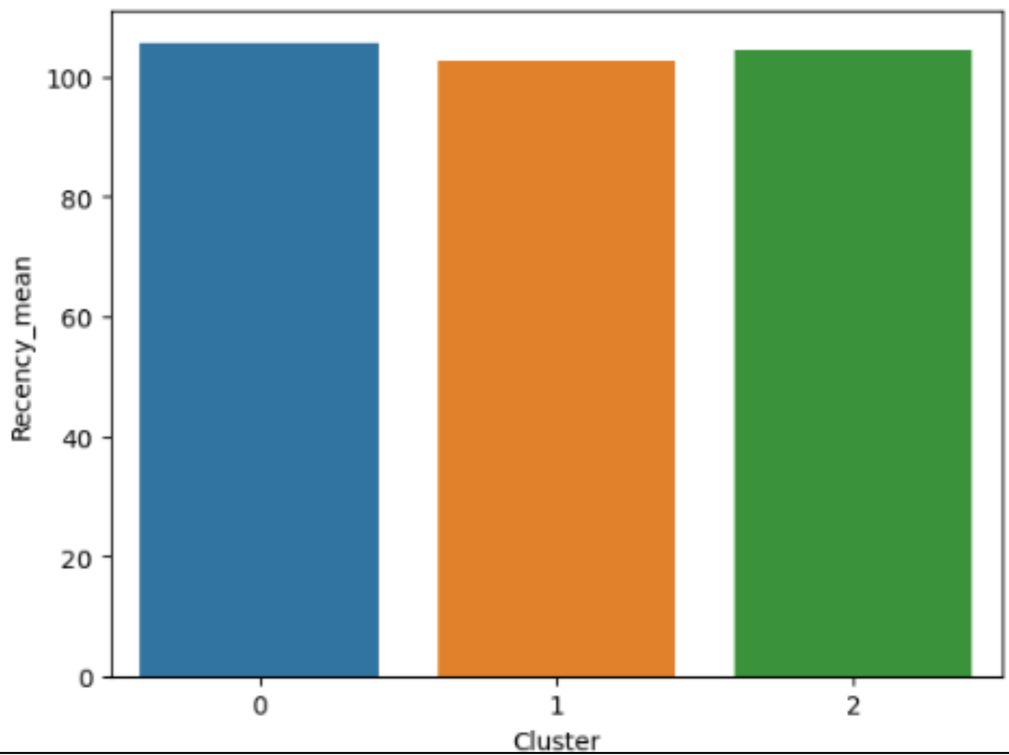


This bar plot illustrates the average purchase frequency for each cluster. Each bar represents a different cluster, and the height of the bar indicates the mean frequency of purchases for customers within that cluster.
X-axis (Cluster): Represents the clusters formed by the K-Means algorithm.
Y-axis (Frequency_mean): Represents the average frequency of purchases made by customers within each cluster.

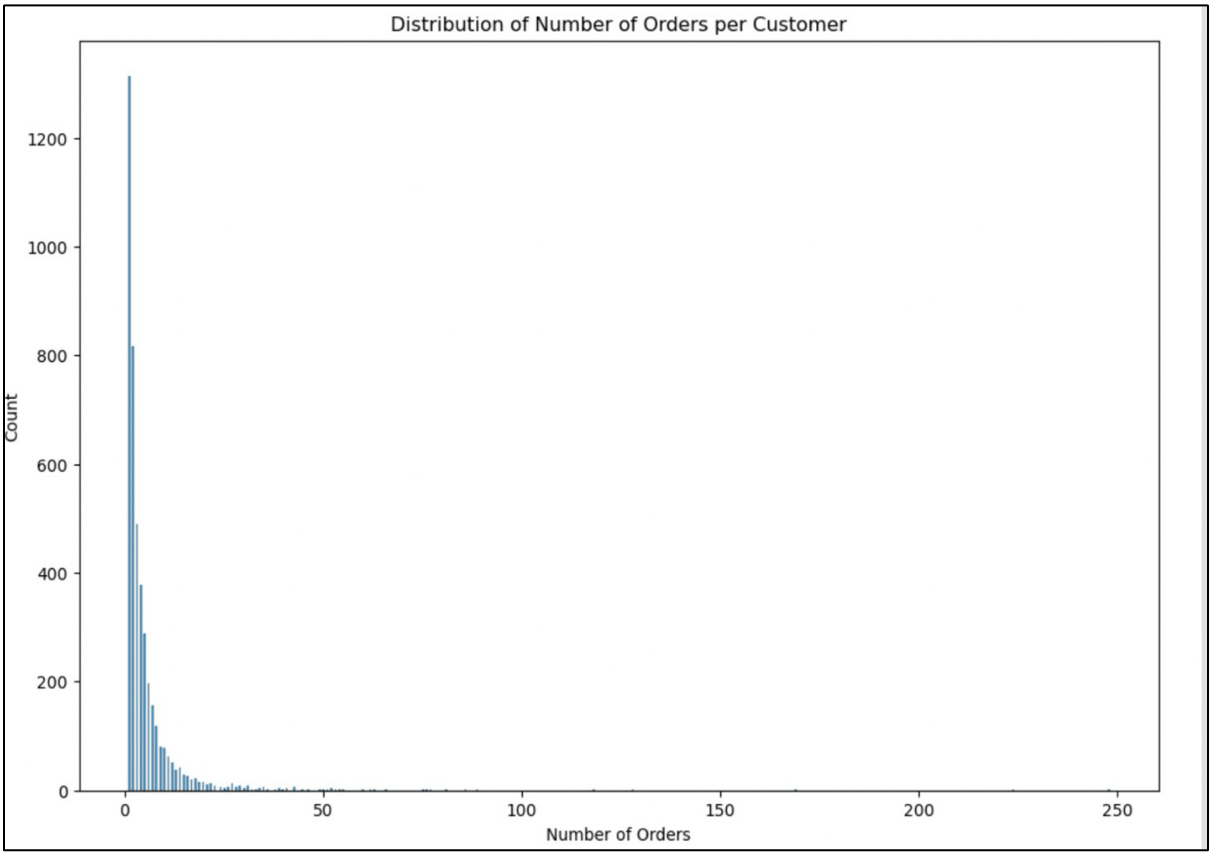
Examining this plot helps in understanding the variations in purchasing frequency across different clusters. It provides insights into how distinct customer segments, identified by the clustering algorithm, differ in terms of their average purchase frequency.
Bar Plot - Cluster vs Average Recency:

```
sns.barplot(data=km_clusters, x='Cluster', y='Recency_mean')
```

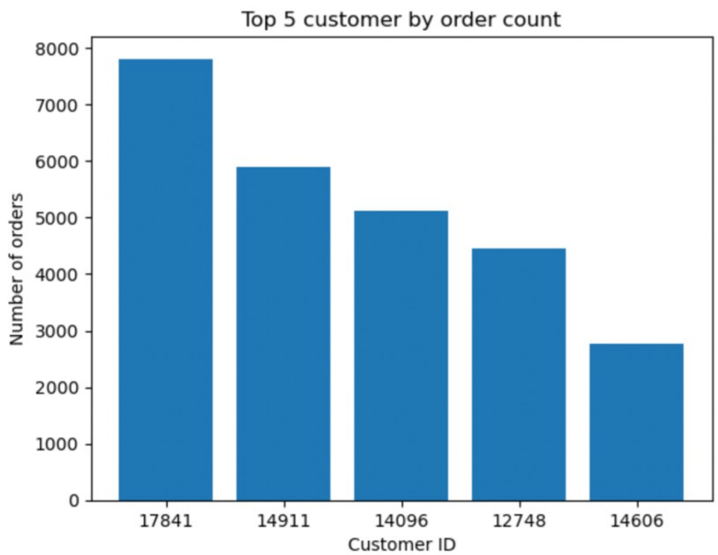
```
<Axes: xlabel='Cluster', ylabel='Recency_mean'>
```



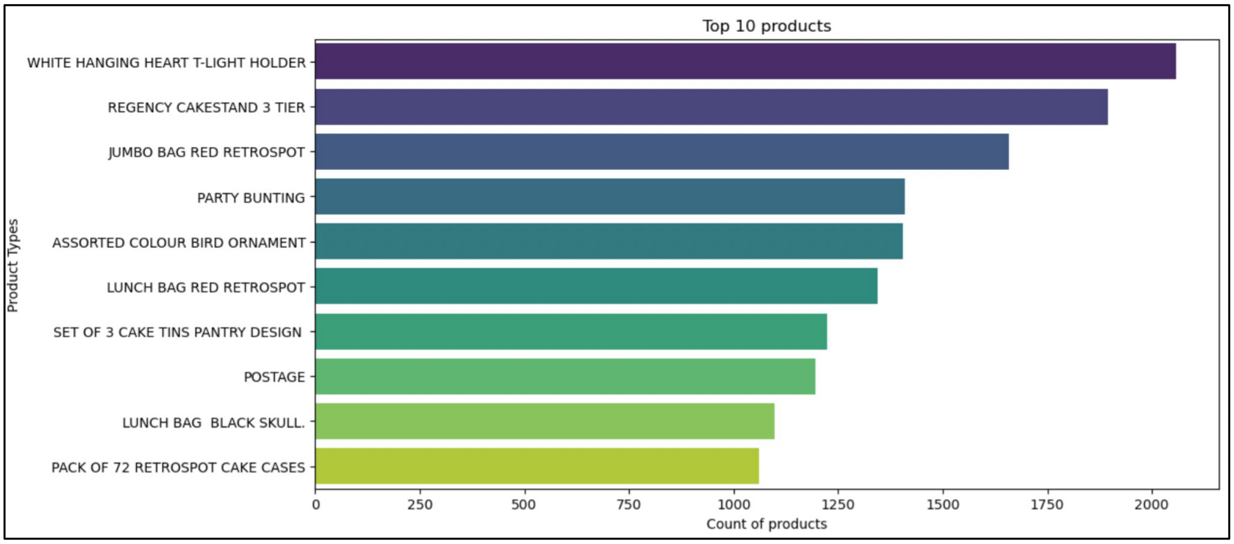
This bar plot illustrates the average recency (time since the last purchase) for each cluster. Each bar represents a different cluster, and the height of the bar indicates the mean recency for customers within that cluster.
X-axis (Cluster): Represents the clusters formed by the K-Means algorithm.
Y-axis (Recency_mean): Represents the average recency (time since the last purchase) for customers within each cluster.
Examining this plot helps in understanding how recently customers in each cluster have made purchases. It provides insights into the temporal behavior of different customer segments identified by the clustering algorithm.
Lower values on the Y-axis indicate that customers in that cluster have made more recent purchases on average.



The histogram graph visually illustrates how orders are distributed among customers. A significant portion of customers, the majority, falls within the 0 to 50 order range, indicating a concentration of transactions in this bracket. Beyond this range, there is a distinct decrease, with noticeably fewer orders extending up to 250. This pattern underscores the concentration of customer activity in the lower order count range, prompting further exploration into the factors influencing purchasing patterns and potential strategies to encourage increased order frequency.



The graph shows how orders are spread out among the top five customers, highlighting some interesting patterns in their buying habits. Customer ID 17841 comes out on top with more than 7500 orders, indicating they're a significant and regular customer. Next is Customer ID 1491 with around 6000 orders, showing they also play a big role in the customer base. In the third spot is Customer ID 14096, with almost 5000 orders, showcasing a solid level of engagement. The last two in the top five are Customer IDs 12748 and 14606, each with their own order counts, although not as high as the top three. So, it seems like a few customers are making a lot of orders compared to the rest. This analysis helps us understand the uneven distribution of orders among the top customers and provides insights for planning strategies and managing customer relationships.



The presented graph delineates the top 10 most frequently purchased products, offering insights into consumer preferences. Notably, the White Hanging Heart T-Light Holder emerges as the most sought-after item, boasting an impressive order count exceeding 2000. Following closely is the Regency Cake Stand 3-Tier, garnering substantial popularity with over 1750 orders. The Jumbo Bag Red Retro spot secures a notable position in the ranking, with a robust order count surpassing 1500. This analysis sheds light on the consumer trends, enabling a strategic understanding of the products that resonate most with the clientele.

Out[48]:

	Product	Average Unit Price	Total Revenue
0	4 PURPLE FLOCK DINNER CANDLES	2.312162	265.66
1	50'S CHRISTMAS GIFT BAG LARGE	1.248091	2269.75
2	DOLLY GIRL BEAKER	1.243796	2745.75
3	I LOVE LONDON MINI BACKPACK	4.138406	1449.85
4	I LOVE LONDON MINI RUCKSACK	4.150000	4.15
...	...	...	...
3891	ZINC T-LIGHT HOLDER STARS SMALL	0.836888	3843.46
3892	ZINC TOP 2 DOOR WOODEN SHELF	16.768182	92.75
3893	ZINC WILLIE WINKIE CANDLE STICK	0.872344	2165.90
3894	ZINC WIRE KITCHEN ORGANISER	7.175000	156.80
3895	ZINC WIRE SWEETHEART LETTER TRAY	3.454000	253.24

3896 rows × 3 columns

In [49]:
```
product_df.describe()
```

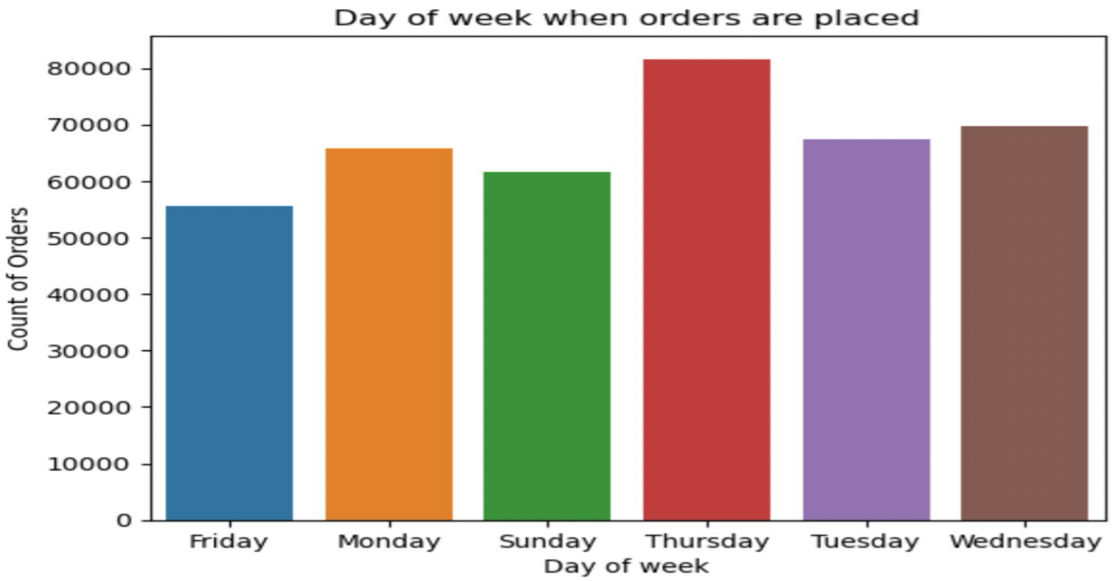
Out[49]:

	Average Unit Price	Total Revenue
count	3896.000000	3896.000000
mean	3.993164	2124.876649
std	19.668337	5403.630535
min	0.000750	-58745.460000
25%	1.019722	107.070000
50%	1.950000	558.355000
75%	3.886952	1930.900000
max	744.147500	132567.700000

The tables present a comprehensive overview of product-related financial metrics. The first table details specific products, their average unit prices, and the total revenue generated. Notably, the product "ZINC T-LIGHT HOLDER STARS SMALL" stands out with the highest total revenue of $3843.46, while "50'S CHRISTMAS GIFT BAG LARGE" and "DOLLY GIRL BEAKER" follow closely with revenues of $2269.75 and $2745.75, respectively. The second table provides a summary statistical analysis of the average unit price and total revenue across all products. The average unit price ranges from a minimum of $0.000750 to a maximum of $744.147500, with a mean of $3.993164. The total revenue exhibits significant variability, with a standard deviation of $5403.630535, highlighting the diversity in revenue generation across the product portfolio. It is important to note that while most products contribute positively to revenue, outliers such as the minimum total revenue of -$58745.46 warrant further investigation, as they may indicate anomalies or errors in the data collection process. The product

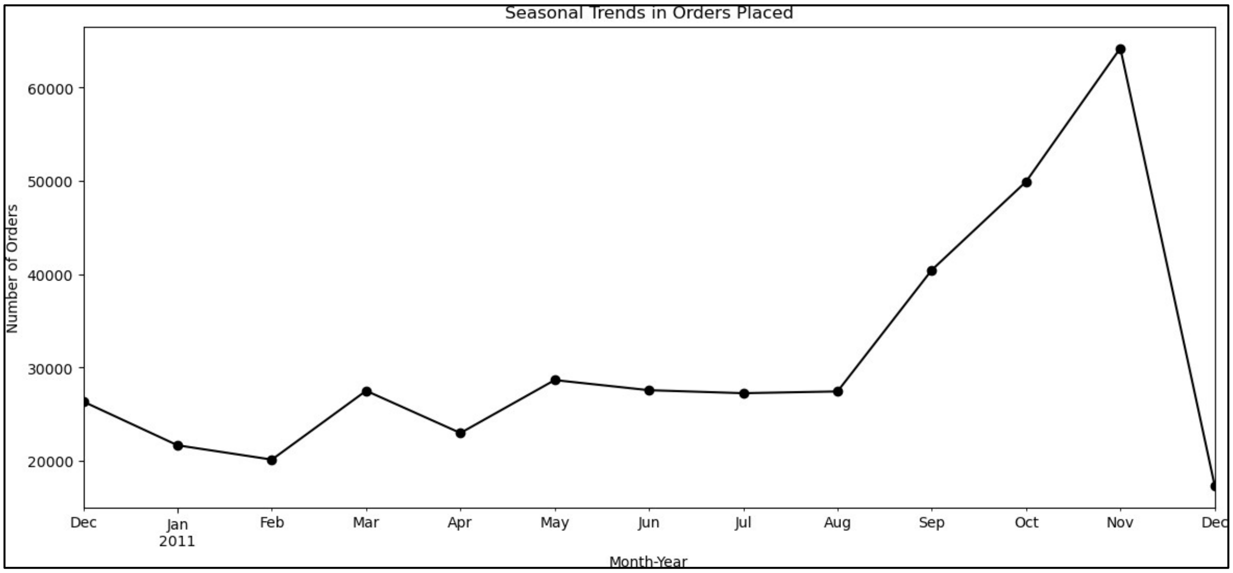
"REGENCY CAKESTAND 3 TIER" has an average unit price of $12.43, generating a total revenue of $132,567.70.

	Day	Number of Sales per day of week
0	Friday	55512
1	Monday	65715
2	Sunday	61673
3	Thursday	81575
4	Tuesday	67376
5	Wednesday	69753



The table displays the distribution of sales across different days of the week. Thursday emerges as the peak sales day, recording the highest number of sales at 81,575.
Following closely, Monday, Wednesday, and Tuesday also demonstrate robust sales figures, with 65,715, 69,753, and 67,376 sales, respectively. Sunday and Friday maintain strong sales performance, with 61,673 and 55,512 sales, respectively. This analysis provides valuable insights into the weekly sales patterns, helping to identify peak and off-peak days. Understanding the variations in sales across different days of the week can inform strategic decisions related to marketing, inventory management, and resource allocation to optimize overall business performance.

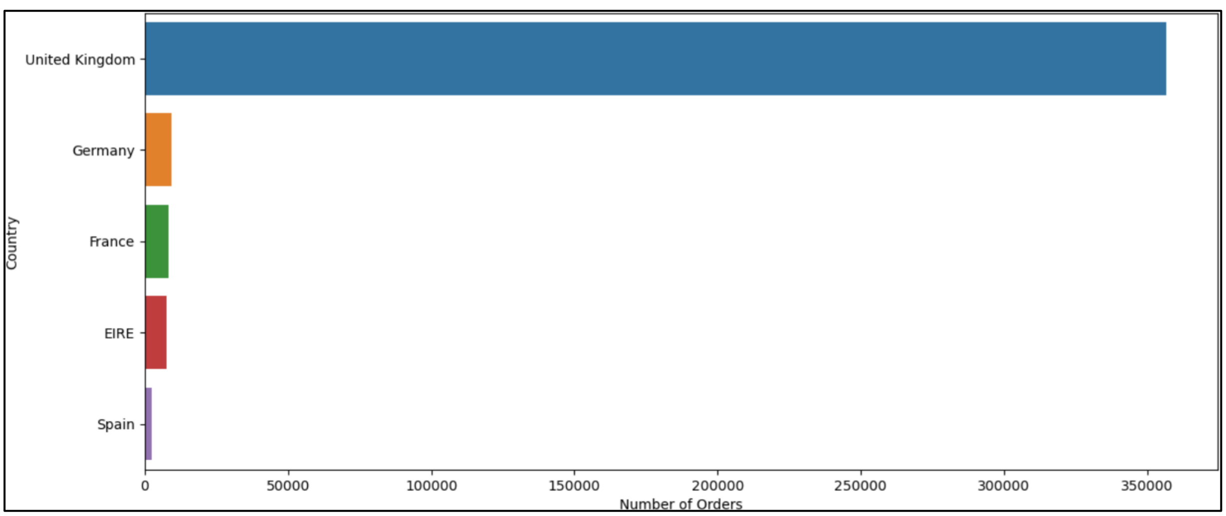
The graph depicting the time-of-day order placement reveals discernible trends in customer behavior over the course of the day. The day kicks off with a modest order rate, slightly above zero at 6 am, steadily increasing to around 10,000 at 8 am, marking the initial influx of orders in the morning. Subsequently, there is a notable surge, exceeding 20,000 orders at 9 am and peaking at over 40,000 at 10 am, suggesting heightened customer activity during mid-morning. The momentum continues to climb, reaching its zenith at noon with over 70,000 orders, representing the day's peak. Post-noon, there is a gradual decline in orders, with approximately 65,000 at 1 pm, 62,000 at 2 pm, and around 50,000 at 3 pm. The decline accelerates in the late afternoon, with orders dropping to 25,000 at 4 pm and 15,000 at 5 pm. Subsequently, after 6 pm, there is a noticeable decrease, with orders consistently falling below 10,000. This analysis illuminates the temporal dynamics of order placement, indicating that customers exhibit peak engagement during the morning and early afternoon hours, gradually tapering off in the evening. Understanding these patterns is essential for optimizing resource allocation, marketing strategies, and operational planning to align with customer preferences and enhance overall efficiency throughout the day.



The line graph shows the seasonal trends in orders placed over a period of one year, from January 2011 to December 2012. The number of orders placed shows a clear seasonal pattern, with peaks in May and December and lows in February and September. The highest peak in orders placed occurs in November, which is likely due to the holiday season. Many people do their holiday shopping in December, and this is reflected in the increase in orders placed during this month. The second-highest peak in orders placed occurs in May, which is likely due to the start of the summer season. People often purchase new clothes, sporting goods, and other items for the summer during this month. The lowest point in orders placed occurs in February, which is likely due to the cold weather and the fact that many people are recovering from the holiday season financially. The second-lowest point in orders placed occurs in September, which is likely due to the start of the back-to-school season. People often have less disposable income during this time of year, and they are more likely to be focused on spending money on school supplies and other necessities. Businesses should be aware of the seasonal trends in orders placed when planning their inventory levels and marketing campaigns. By understanding when customers are most likely to place orders, businesses can ensure that they have enough inventory on hand to meet demand and that they are targeting their marketing campaigns to the right audience at the right time.

	Country	Number of Orders
35	United Kingdom	356728
14	Germany	9480
13	France	8475
10	EIRE	7475
30	Spain	2528

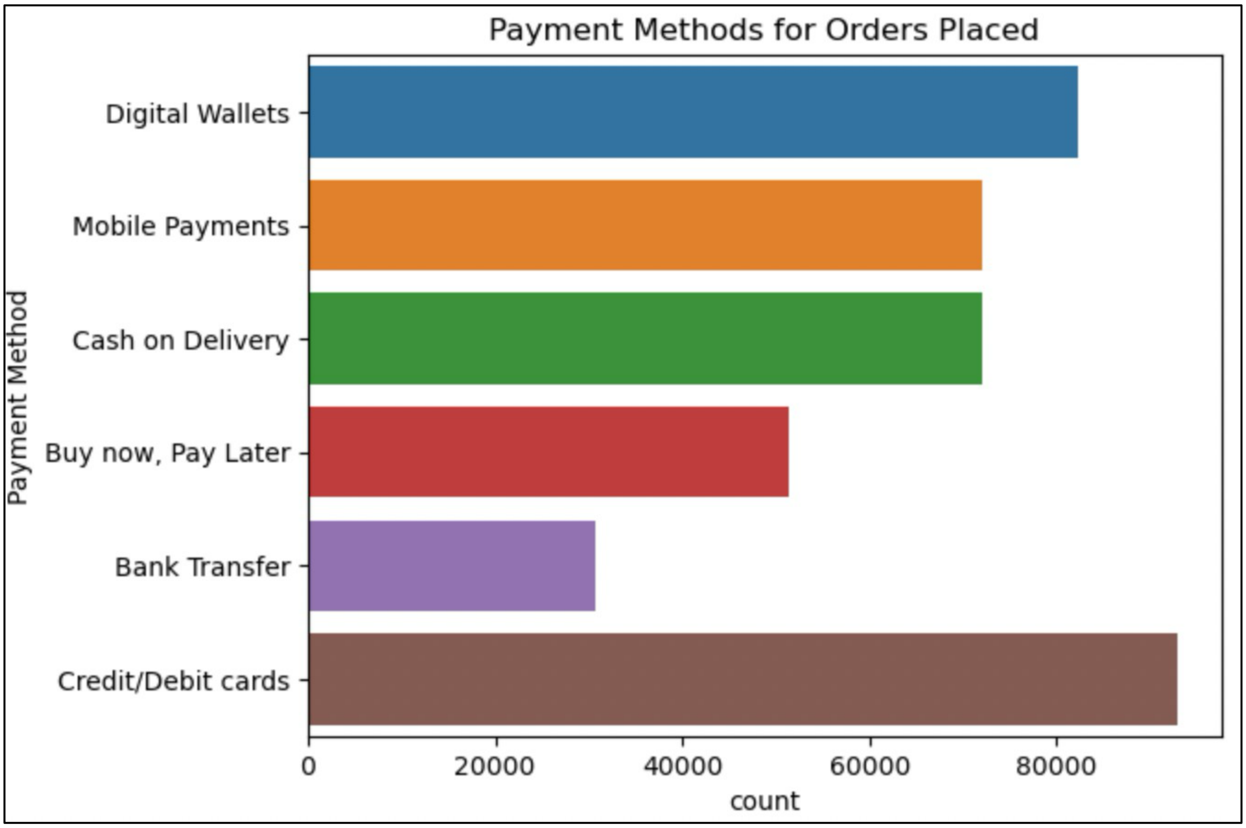
The presented table offers valuable insights into the distribution of orders among different countries. The United Kingdom stands out prominently, accounting for a significant number of orders at 356,728. In contrast, Germany, France, and EIRE follow with order volumes of 9,480, 8,475, and 7,475, respectively, indicating substantial yet comparatively lower order frequencies. Spain concludes the list with 2,528 orders. This analysis highlights the geographic dispersion of order activity, emphasizing the United Kingdom's pivotal role as the primary market. Recognizing these variations in order volumes across countries is crucial for tailoring effective marketing strategies, optimizing logistics, and refining customer engagement approaches to meet the diverse needs of the international customer base.



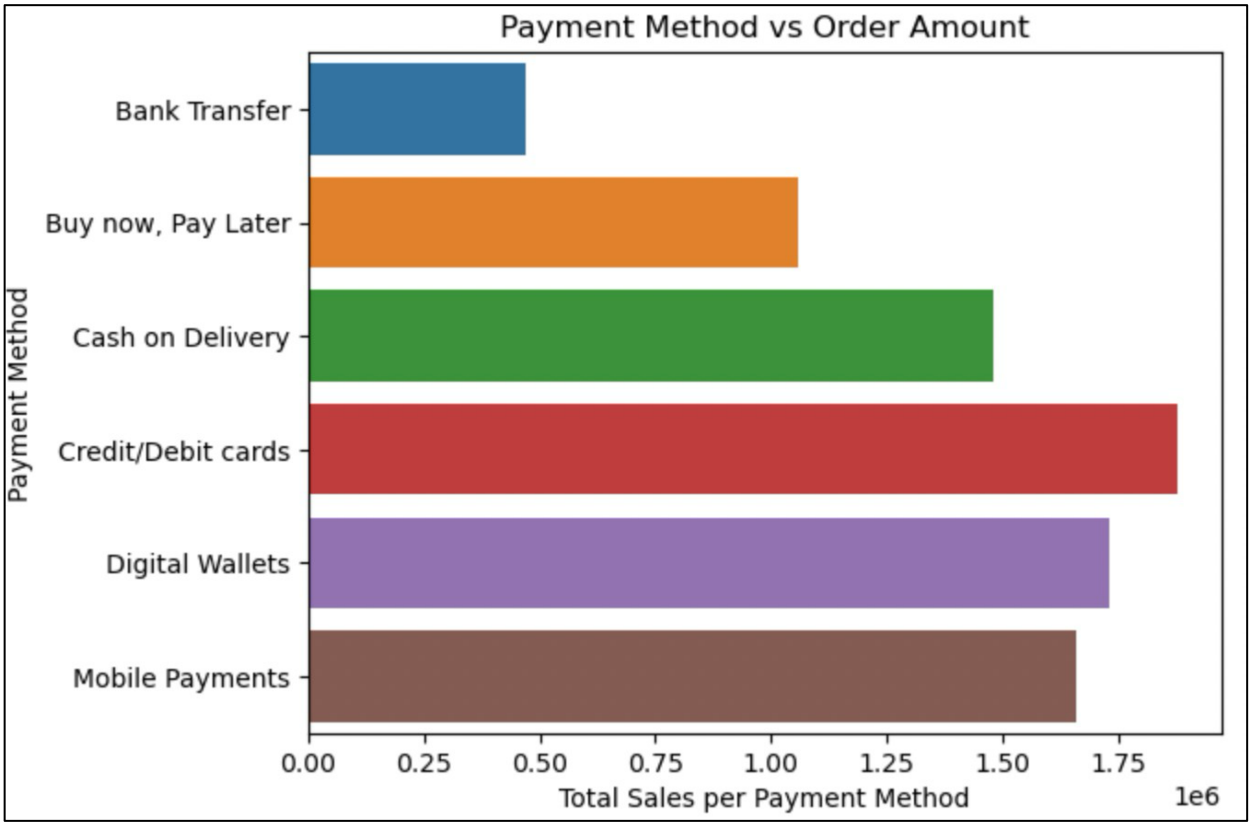
The line graph depicting 2023 order volumes in the United Kingdom, Germany, France, Spain, and Ireland illustrates the dominance of the UK in online shopping, surpassing the order frequencies of the other four countries. This underscores the UK's status as the largest online shopping market in Europe, propelled by high smartphone and mobile device penetration, facilitating widespread online transactions. Following the UK, Germany holds the second-largest market share, trailed by France and Spain, all showcasing substantial but comparatively lower order frequencies. The growth of online shopping in Europe is fueled by factors such as convenience, affordability, product availability, and the popularity of mobile shopping. Businesses eyeing European expansion must consider these trends and tailor their strategies based on market size, mobile device penetration, and cultural influences on shopping behavior.

	Country	Average Country Sales
23	Netherlands	120.059696
0	Australia	108.910787
19	Japan	98.716816
31	Sweden	79.360976
9	Denmark	48.247147
21	Lithuania	47.458857
29	Singapore	39.827031
20	Lebanon	37.641778
4	Brazil	35.737500
10	EIRE	33.445054
24	Norway	32.378877
15	Greece	32.263836
2	Bahrain	32.258824
12	Finland	32.124806
32	Switzerland	29.696004
17	Israel	28.293117
34	United Arab Emirates	27.974706
6	Channel Islands	26.520991
1	Austria	25.322494
5	Canada	24.280662
16	Iceland	23.681319
8	Czech Republic	23.590667
14	Germany	23.365978
13	France	23.200714
30	Spain	21.659822
11	European Community	21.176230
25	Poland	21.152903
7	Cyprus	21.045434
18	Italy	21.034259
3	Belgium	19.773301
22	Malta	19.728110
26	Portugal	19.711598
35	United Kingdom	18.914008
27	RSA	17.281207
28	Saudi Arabia	13.117000
36	Unspecified	11.040539
33	USA	5.948179

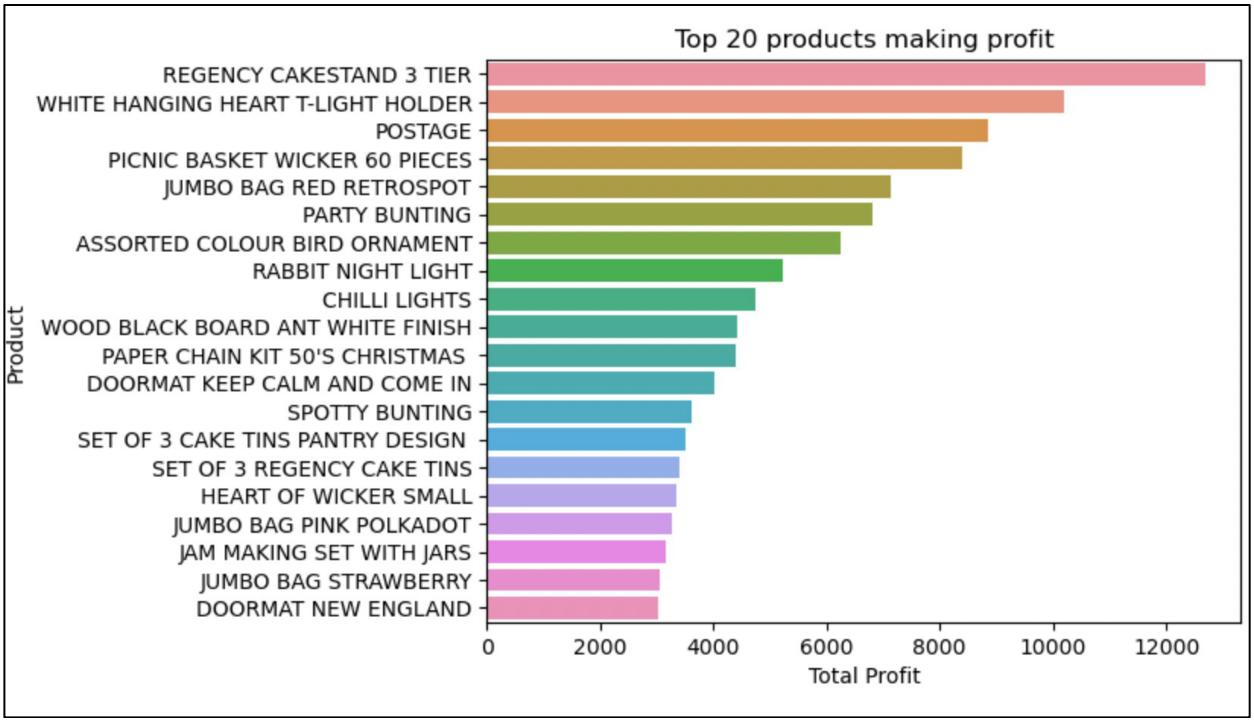
The table provides an insightful analysis of average sales per country, revealing distinct patterns in consumer behavior across different regions. The Netherlands emerges as the top-performing country with an average country sales value of 120.06, followed closely by Australia and Japan with average sales of 108.91 and 98.72, respectively. This analysis indicates varying purchasing power and consumer engagement levels among different countries. Interestingly, the United Kingdom, while a substantial market, exhibits a comparatively lower average country sales value of 18.91, suggesting a higher transaction frequency with lower individual order values. Understanding these variations is crucial for businesses aiming to tailor their marketing strategies and customer engagement initiatives to cater to diverse consumer behaviors in each country. For instance, in countries with higher average sales, strategies can focus on premium offerings, while in countries with lower average sales, volume-driven approaches may be more effective. This nuanced understanding of customer behavior enables businesses to optimize their market-specific strategies and enhance their overall international performance.



The breakdown of payment methods for placed orders reveals distinct customer preferences. Notably, both digital wallets and Credit/Debit cards emerge as highly favored options, each accounting for over 80,000 orders, indicating widespread adoption and user convenience. Mobile Payments and Cash on Delivery share a similar order count, hovering around 70,000, suggesting their comparable popularity among customers. Buy Now Pay Later follows closely with more than 50,000 orders, indicating a notable interest in deferred payment options. Bank Transfer records a lower order count, approximately 30,000, reflecting a less favored but still utilized payment method. The significant prevalence of digital wallets and credit/debit cards underscores a clear customer inclination towards seamless and digital transaction modes. Recognizing these preferences is pivotal for businesses, emphasizing the need to provide a diverse range of payment options to cater to varied customer choices and enhance overall satisfaction and order completion rates.

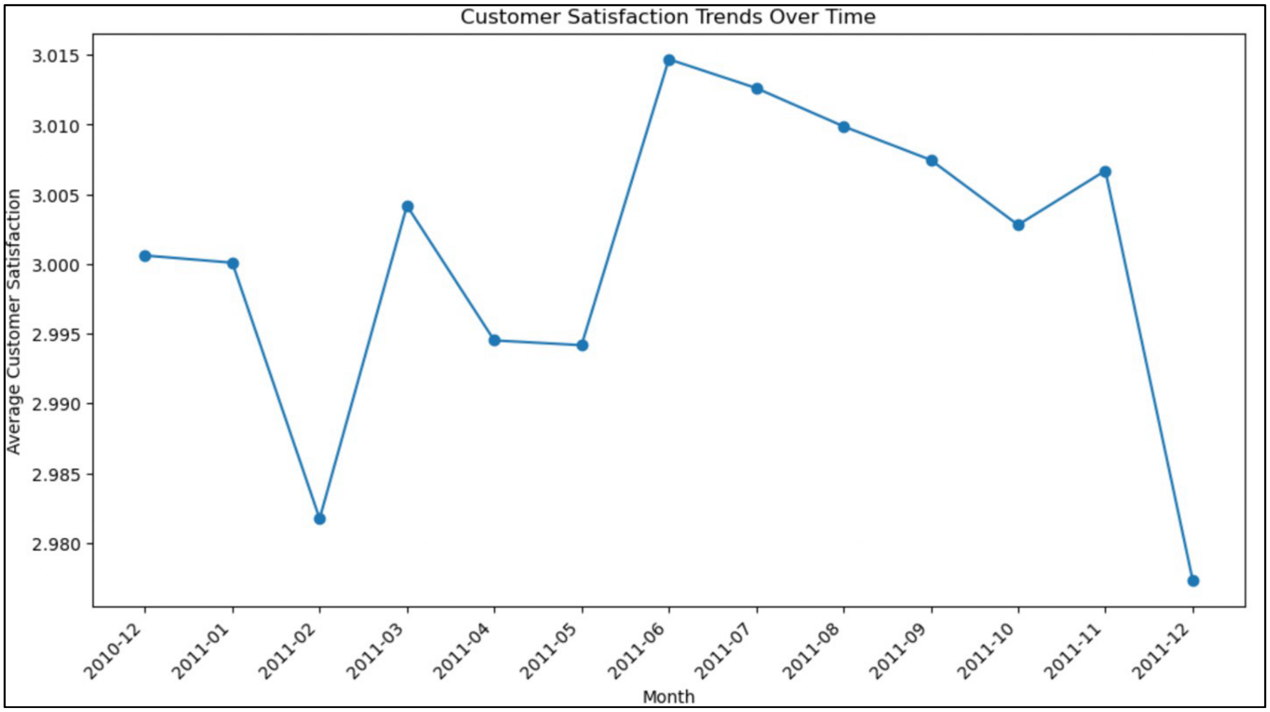


The graph illustrating payment methods against order amounts yields insightful observations into consumer spending patterns. Notably, Credit/Debit cards stand out as the predominant payment method, generating over 2 million in order amounts, indicating a substantial reliance on these cards for higher-value transactions. Digital Wallets and Mobile Payments closely trail, both contributing significant order amounts at approximately 1.75 million and 1.65 million, respectively. Despite being a popular method with over 1.5 million orders, Cash on Delivery exhibits a slightly lower order amount, hinting at a preference for this method in lower-value transactions. The Buy Now Pay Later method records order amounts exceeding 1 million, underscoring the attractiveness of deferred payment options. Bank Transfer, with an order amount of around 0.5 million, reflects a moderate contribution to overall transaction values. This analysis emphasizes the varied financial preferences of customers, with different payment methods influencing the scale of order amounts. Businesses can leverage these insights to optimize their payment offerings and align them with customer expectations and transaction values.



The analysis of the top five products with the highest profit margins provides valuable insights into the company's revenue generation. The product "Regency Cakestand 3 Tier" stands out with an average unit price of $12.43, contributing to a total revenue of $132,567.70 and an impressive total profit of $12,687.98. Following closely, the "White Hanging Heart T-Light Holder" demonstrates a solid profit margin, with an average unit price of $2.89, total revenue of $93,767.80, and a total profit of $10,196.75. Notably, the product "Postage" achieves a remarkable profit margin, given its higher average unit price of $37.89, leading to a total profit of $8,844.65 on a total revenue of $66,710.24.

The "Picnic Basket Wicker 60 Pieces" and "Jumbo Bag Red Retrospot" also showcase commendable profit margins, contributing to the company's overall profit percentage of approximately 10%. This analysis aids in strategic decision-making, allowing the company to focus on product lines with higher profitability and optimize its product portfolio for enhanced financial performance.



The analysis of the line graph depicting average customer satisfaction trends from December 2010 to December 2011 reveals a generally positive trajectory over the year. Despite slight dips in satisfaction observed in February, April, and October, the overall trend showcases an upward movement. Notably, the highest customer satisfaction score of 3.015 was attained in June 2011, reflecting a peak in positive customer experiences. Conversely, the lowest satisfaction score of 2.980 occurred in February 2011, potentially influenced by the aftermath of the holiday season and cold weather, impacting customers' financial recovery. Understanding these fluctuations is pivotal for businesses, allowing them to pinpoint specific periods of higher and lower customer satisfaction, thereby informing targeted strategies for improvement during critical times. The overall positive trend suggests that efforts to enhance customer satisfaction have been effective, but continuous monitoring and adjustments may be needed to maintain and further improve customer experiences.

Thank You!