

**TOPIC: CUSTOMER SEGMENTATION USING RFM
ANALYSIS**

**SUBMITTED BY:
NEHA PATIL**

AIM OF THE PROJECT:

To perform RFM analysis on the dataset and segment the customers into distinct groups based on their RFM scores. These will provide us with insights which will be valuable for marketing and customer retention strategies.

TASK 1

DATA PREPROCESSING

```
import pandas as pd

df = pd.read_csv('data 2.csv', encoding='latin1')
```

[] df

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	12/9/2011 12:50	0.85	12680.0	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	12/9/2011 12:50	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	12/9/2011 12:50	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	12/9/2011 12:50	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	12/9/2011 12:50	4.95	12680.0	France

541909 rows x 8 columns

```
[ ] data_types = df.dtypes
print(data_types)
```


```
InvoiceNo      object
StockCode      object
Description    object
Quantity       int64
InvoiceDate    object
UnitPrice      float64
CustomerID     float64
Country        object
dtype: object
```

```
# Check for missing values
df.isnull().sum()
```

```
InvoiceNo      0
StockCode      0
Description    1454
Quantity       0
InvoiceDate    0
UnitPrice      0
CustomerID    135080
Country        0
dtype: int64
```


RFM CALCULATION

Calculated the Recency, Frequency and Monetary for each customer.

	CustomerID	Recency	Frequency	Monetary
0	12346.0	325	2	0.00
1	12347.0	1	7	4310.00
2	12348.0	74	4	1797.24
3	12349.0	18	1	1757.55
4	12350.0	309	1	334.40
...
4367	18280.0	277	1	180.60
4368	18281.0	180	1	80.82
4369	18282.0	7	3	176.60
4370	18283.0	3	16	2094.88
4371	18287.0	42	3	1837.28

RFM SEGMENTATION

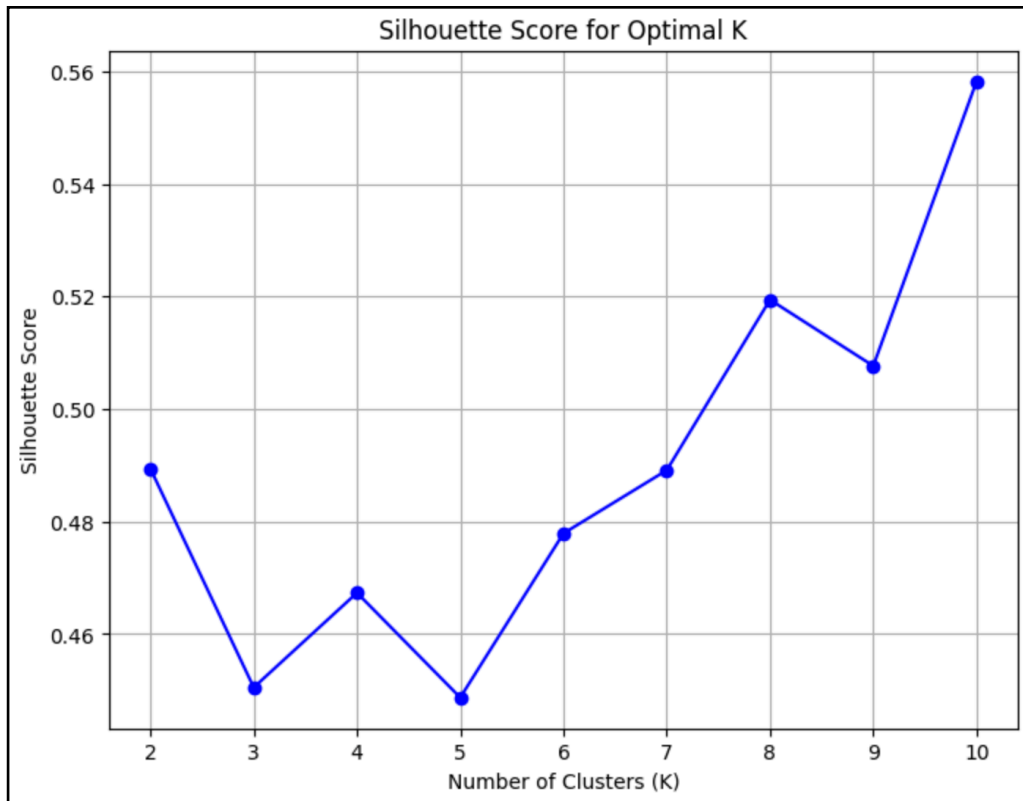
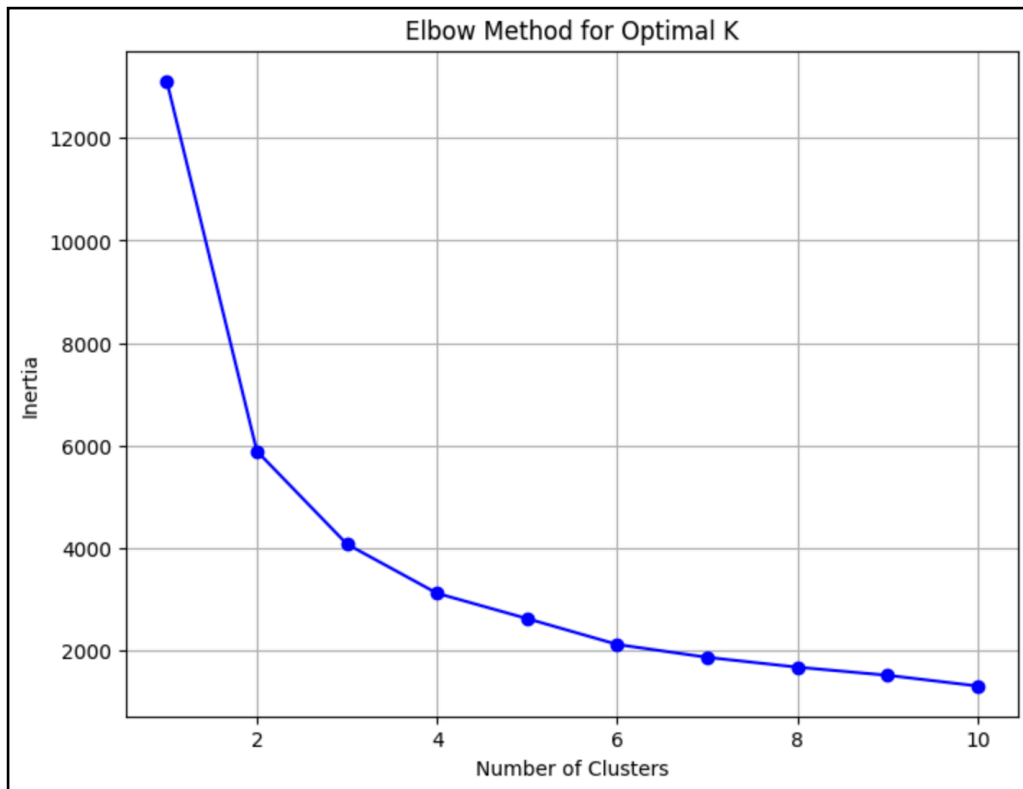
Assigned the RFM scores to each customer based on their quartiles. RFM score for each customer is assigned below.

	CustomerID	Recency_Score	Frequency_Score	Monetary_Score	RFM_Score
0	12346.0	4	1	1	6
1	12347.0	1	3	4	8
2	12348.0	3	2	4	9
3	12349.0	2	1	4	7
4	12350.0	4	1	2	7
...
4367	18280.0	4	1	1	6
4368	18281.0	4	1	1	6
4369	18282.0	1	1	1	3
4370	18283.0	1	3	4	8
4371	18287.0	2	1	4	7

[4372 rows x 5 columns]

CUSTOMER SEGMENTATION

For clustering purpose, we first found out the K value, that is the number of clusters that the customers can be classified into to give optimal results.



RESULTS

```
----- Profile for Cluster 0 -----
Cluster Size: 1415
Recency:
- Mean: 1.5943462897526501
- Min: 1
- Max: 4

Frequency:
- Mean: 2.758303886925795
- Min: 2
- Max: 3

Monetary Value:
- Mean: 3.6367491166077737
- Min: 2
- Max: 4

----- Profile for Cluster 1 -----
Cluster Size: 1466
Recency:
- Mean: 3.6132332878581175
- Min: 3
- Max: 4

Frequency:
- Mean: 1.0661664392905865
- Min: 1
- Max: 3

Monetary Value:
- Mean: 1.465211459754434
- Min: 1
- Max: 2
```

```
----- Profile for Cluster 2 -----
Cluster Size: 786
Recency:
- Mean: 1.5623409669211197
- Min: 1
- Max: 2

Frequency:
- Mean: 1.0966921119592876
- Min: 1
- Max: 3

Monetary Value:
- Mean: 1.7162849872773538
- Min: 1
- Max: 3

----- Profile for Cluster 3 -----
Cluster Size: 705
Recency:
- Mean: 2.9602836879432624
- Min: 1
- Max: 4

Frequency:
- Mean: 1.252482269503546
- Min: 1
- Max: 2

Monetary Value:
- Mean: 3.2439716312056737
- Min: 2
```

CUSTOMER SEGMENTATION

High-Value Customers (Cluster 0: 1415)

Characteristics:

- **Recency:** Recently made purchases.
- **Frequency:** Regularly makes purchases.
- **Monetary Value:** High spending.

Recommendations:

1. **Exclusive Loyalty Program:** Launch an exclusive loyalty program for high-value customers with tiered rewards, such as early access to sales, exclusive products, and personalized offers.
2. **VIP Events:** Host VIP events or sales specifically for this segment to make them feel valued.
3. **Personalized Product Recommendations:** Leverage their purchase history to provide personalized product recommendations through email or on the website.
4. **Special Anniversary Offers:** Send special offers or discounts on the anniversary of their first purchase to encourage repeat business.

Mid-Value Customers (Cluster 1466)

Characteristics:

- **Recency:** Moderately recent purchases.
- **Frequency:** Makes purchases somewhat regularly.
- **Monetary Value:** Moderate spending.

Recommendations:

1. **Loyalty Program:** Implement a loyalty program offering rewards for consistent purchases to increase frequency.
2. **Limited-Time Promotions:** Create time-limited promotions or flash sales to stimulate more immediate purchases.
3. **Cross-Sell Campaigns:** Implement cross-sell campaigns to introduce customers to additional products based on their purchase history.
4. **Feedback Surveys:** Encourage feedback through surveys to understand their preferences and improve offerings.

Low-Value Customers (Cluster 2: 786)

Characteristics:

- **Recency:** Less recent purchases.
- **Frequency:** Infrequent purchases.
- **Monetary Value:** Low spending.

Recommendations:

1. **Reactivation Campaigns:** Run targeted reactivation campaigns with special discounts or promotions to encourage a return.

2. **Bundle Deals:** Introduce bundle deals or package discounts to increase the average transaction value.
3. **Customer Education:** Provide educational content to showcase the value of your products and encourage repeat purchases.
4. **Limited-Time Discounts:** Offer time-sensitive discounts to create a sense of urgency.

Dormant Customers (Cluster 3: 705)

Characteristics:

- **Recency:** No recent purchases.
- **Frequency:** Rarely makes purchases.
- **Monetary Value:** Little to no spending.

Recommendations:

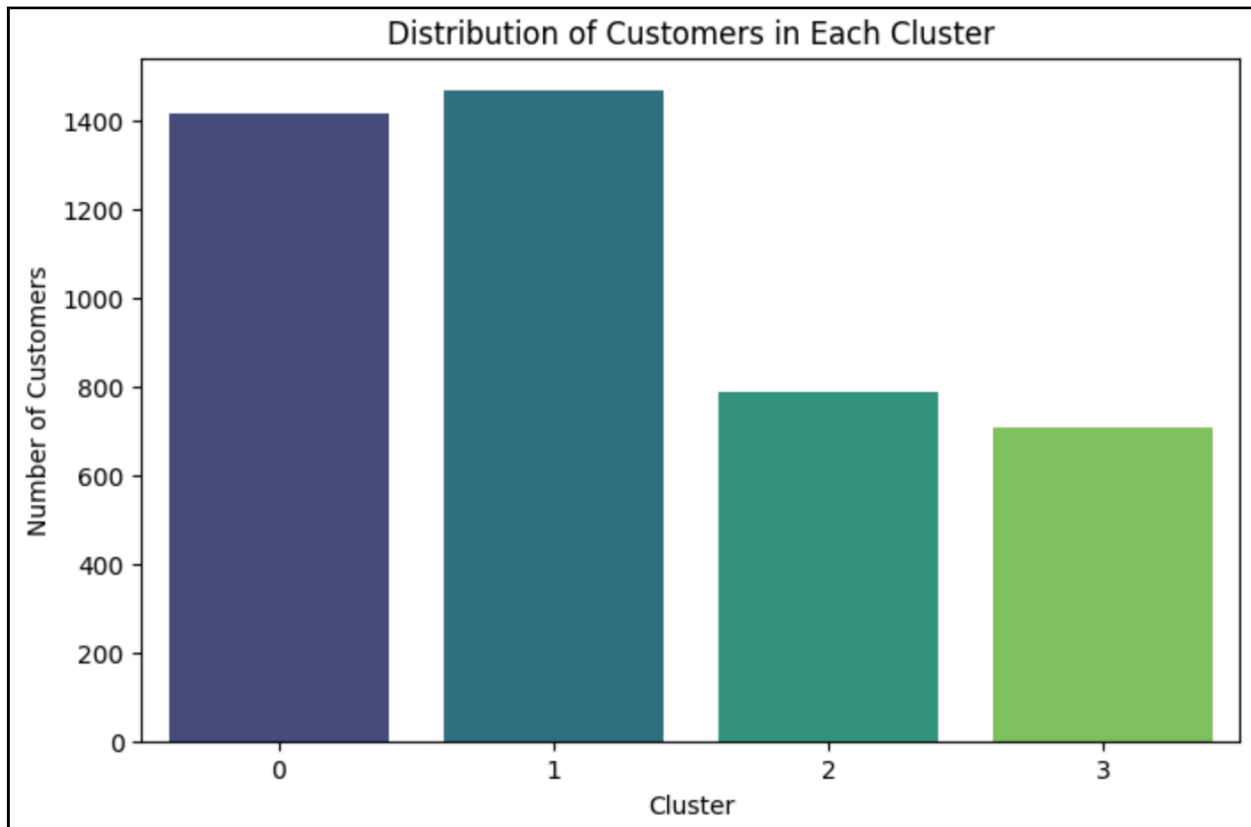
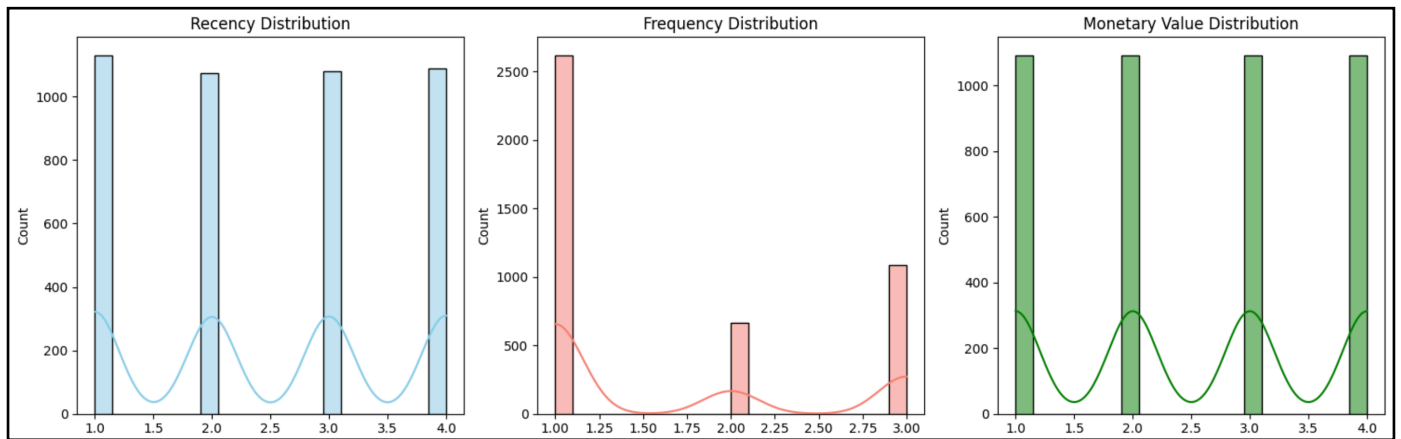
1. **Win-Back Campaigns:** Launch win-back campaigns with compelling offers to re-engage dormant customers.
2. **Reactivation Incentives:** Provide additional incentives, such as free shipping or exclusive discounts, for their first purchase upon returning.
3. **Feedback Collection:** Conduct surveys or feedback campaigns to understand the reasons for dormancy and address any concerns.
4. **Personalized Apology Offers:** Consider personalized apology offers for dormant customers to express appreciation for their return.

MARKETING RECOMMENDATION

Tips for the business for tailoring marketing strategies:

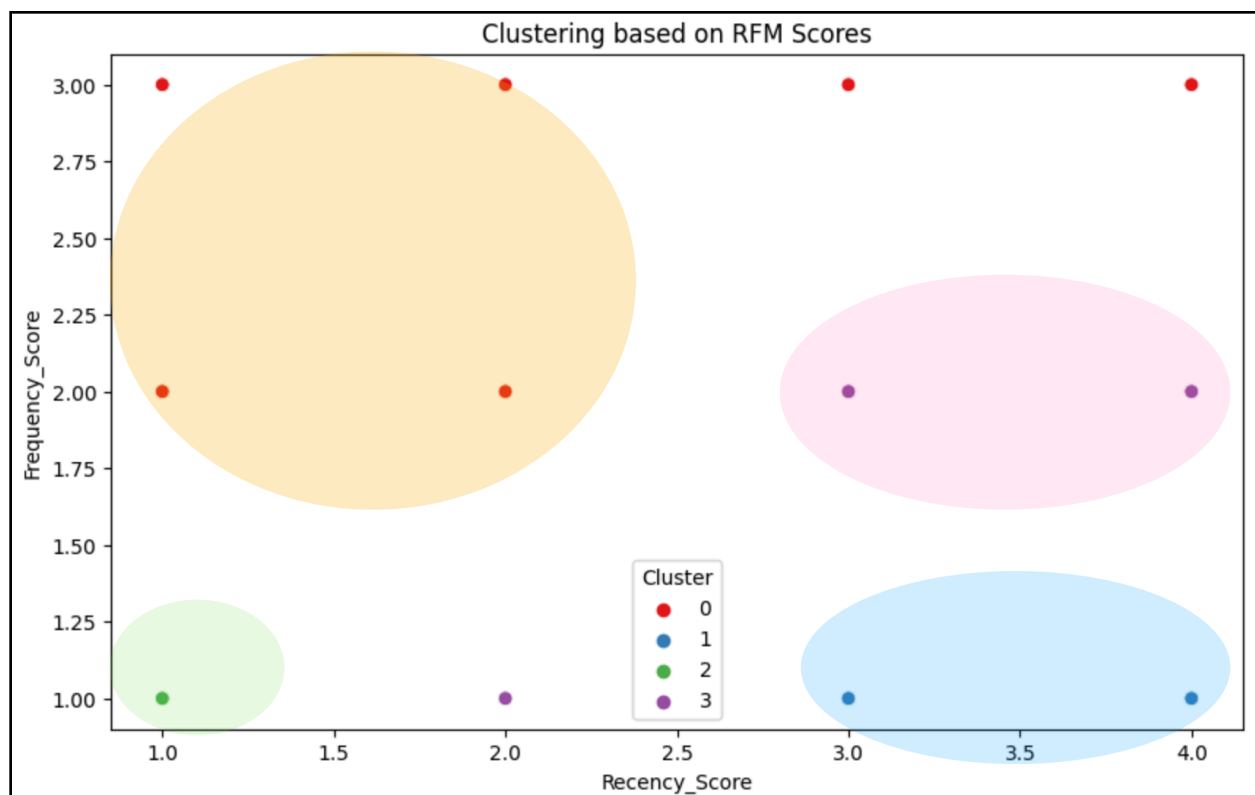
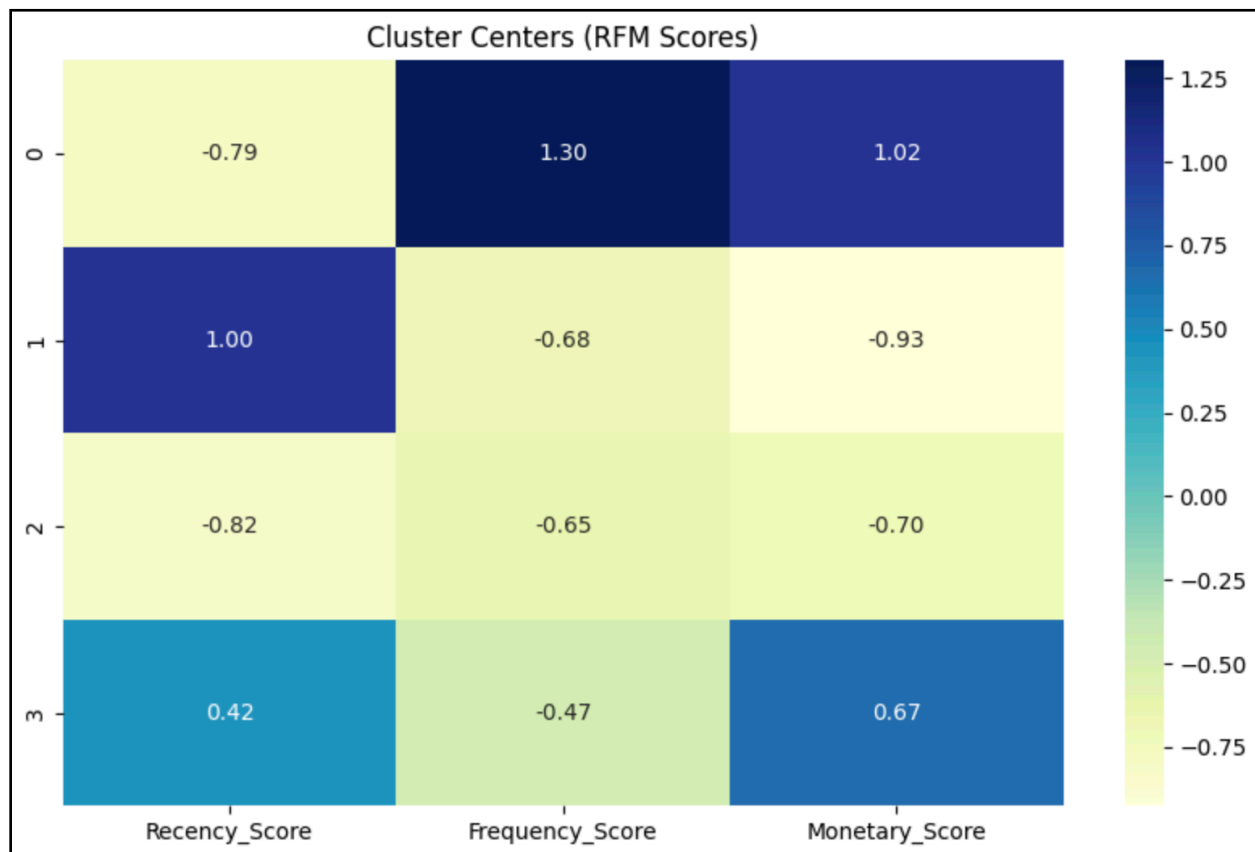
1. **Segment-Specific Communication:** Tailor marketing messages and channels based on the preferences of each segment.
2. **Social Media Engagement:** Engage with high and mid-value customers on social media platforms, showcasing new products or exclusive content.
3. **Customer Feedback Loop:** Encourage feedback from all segments to continually refine marketing strategies.
4. **Monitor and Iterate:** Regularly monitor the performance of these strategies and iterate based on customer responses and market dynamics.

VISUALIZATIONS



- 0 - stands for the High valued customers
- 1- stand for the Mid value customers
- 2- stands for the low value customers
- 3 stands for the dormant customers

According to the above bar graph, we infer that most number of people belong to the mid value range(1500) followed by the people belonging to the high value range(1400).



RESULT OF ANALYSIS

1. What is the size of the dataset in terms of the number of rows and columns?

- The dataset contains 4372 rows and 8 columns.

2. Can you provide a brief description of each column in the dataset?

- **Description** of each column: InvoiceNo: Invoice number for each transaction. StockCode: Code identifying the product. Description: Description of the product. Quantity: Quantity of the product purchased. InvoiceDate: Date and time of the transaction. UnitPrice: Price per unit of the product. CustomerID: ID of the customer. Country: Country where the transaction took place.

1.3 What is the time period covered by this dataset?

- According to our analysis, the following time period has been covered.

Earliest Date: 2010-12-01 08:26:00

Latest Date: 2011-12-09 12:50:00

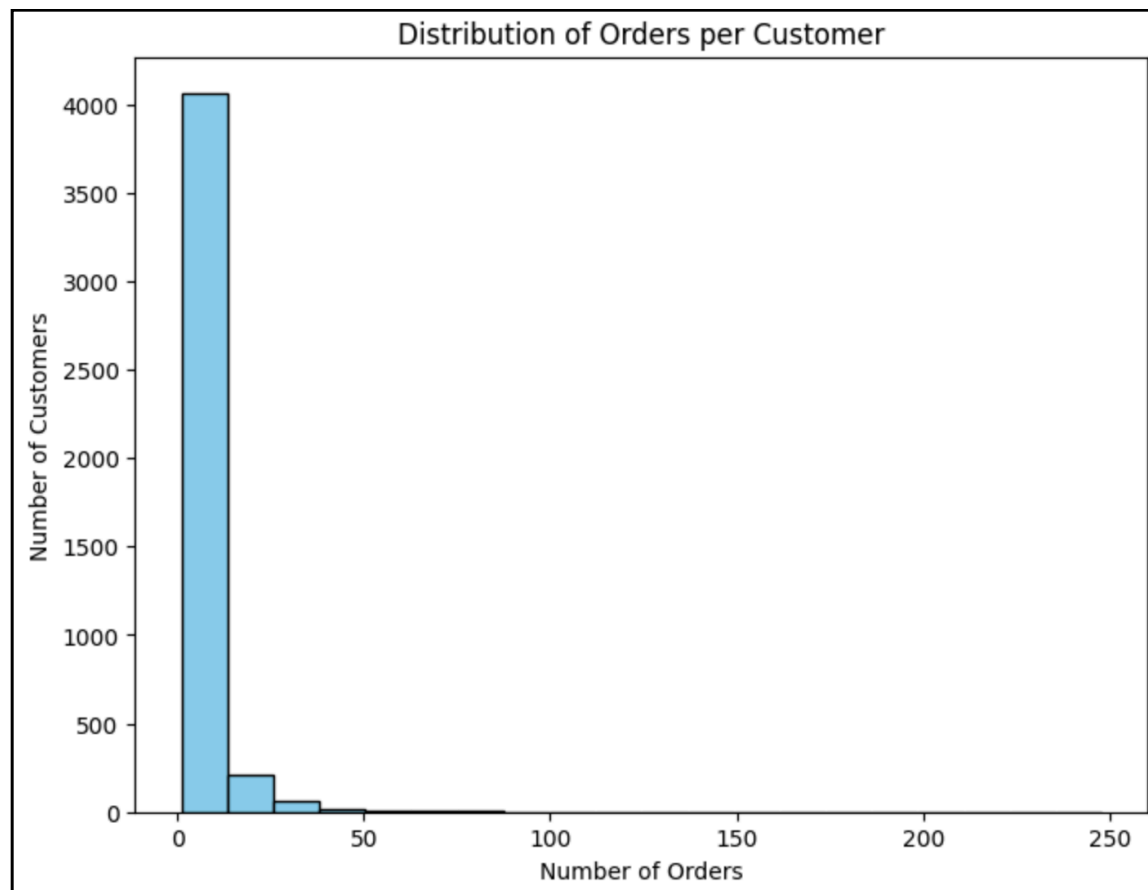
Time Period Covered: 373 days 04:24:00

2.1 how many unique cx are there?

- Number of unique customers: 4372

2.2 What is the distribution of the number of orders per customer?

2.3 Group orders by 'CustomerID' and count the number of unique orders per customer



2.3 The following depicts the top 5 customers with the most number of purchases:

```
Top 5 Customers with the Most Purchases by Order Count:
CustomerID
14911.0      248
12748.0      224
17841.0      169
14606.0      128
13089.0      118
Name: InvoiceNo, dtype: int64
```

3.1 10 most frequently purchased products are as follows:

```
Top 10 Most Frequently Purchased Products:
      StockCode  TotalQuantity
0  WORLD WAR 2 GLIDERS ASSTD DESIGNS    53215
1          JUMBO BAG RED RETROSPOT    45066
2    ASSORTED COLOUR BIRD ORNAMENT    35314
3  WHITE HANGING HEART T-LIGHT HOLDER    34147
4    PACK OF 72 RETROSPOT CAKE CASES    33409
5          POPCORN HOLDER    30504
6          RABBIT NIGHT LIGHT    27094
7      MINI PAINT SET VINTAGE    25880
8    PACK OF 12 LONDON TISSUES    25321
9  PACK OF 60 PINK PAISLEY CAKE CASES    24163
```

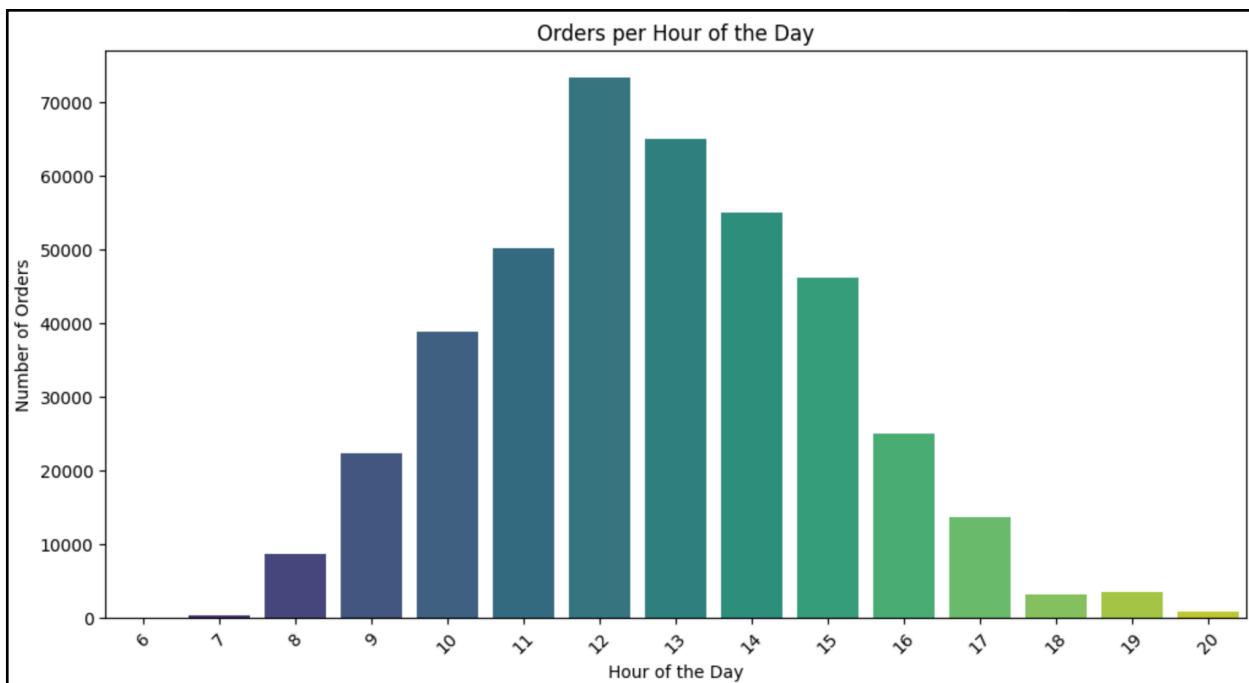
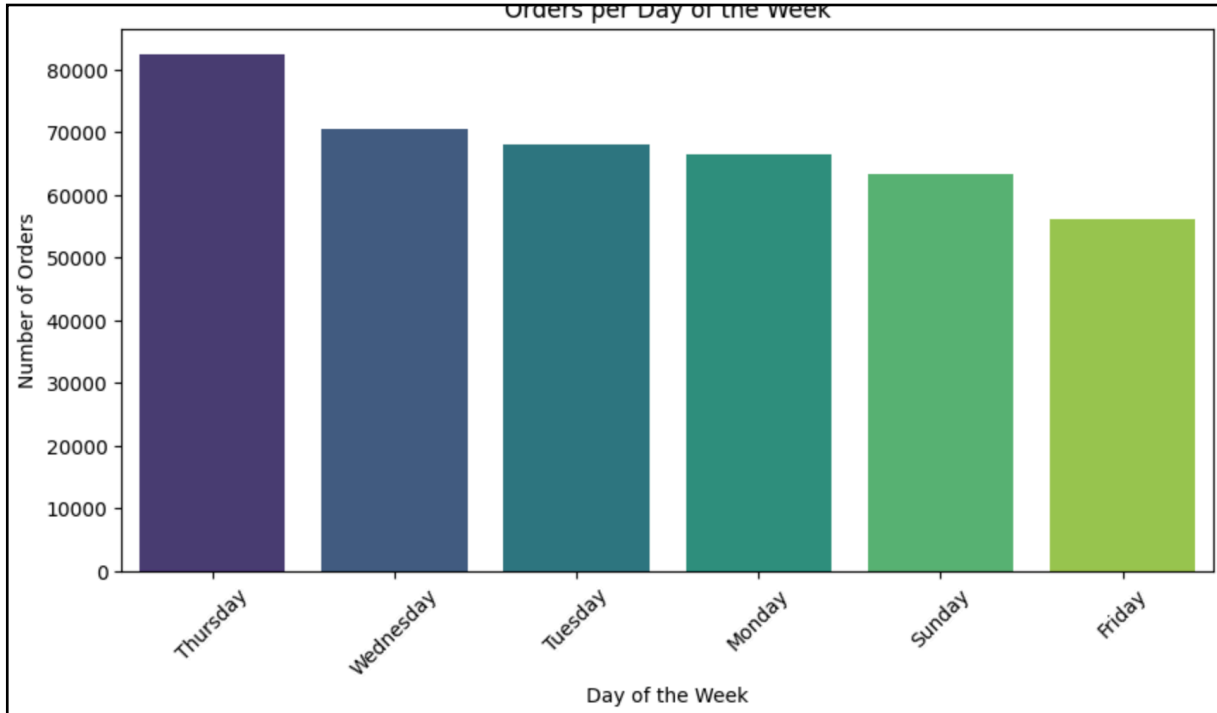
3.2 What is the average price of products in the database

```
Average price of products: 3.460471018536043
```

3.3 Can you find out which product category generates the hi

```
Description
REGENCY CAKESTAND 3 TIER    132870.4
Name: TotalPrice, dtype: float64
```

4.1 Is there a specific day of the week or time of day when most orders are place



- Most number of orders are placed on Thursdays.
- Most orders are placed at around 12pm.

4.3 Seasonal Trends in Sales

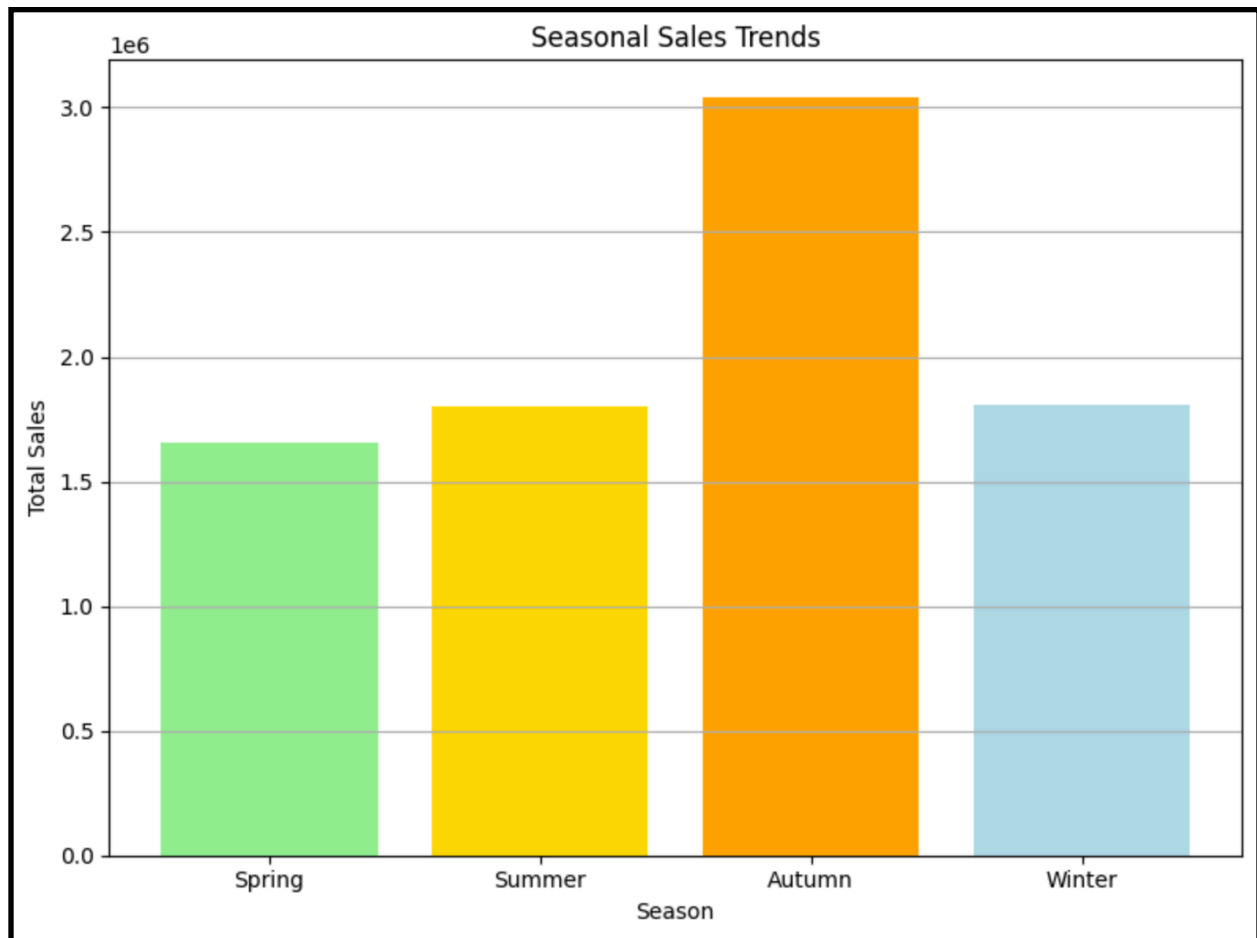
Extracting month from 'InvoiceDate'

```
Month
1      475074.380
2      436546.150
3      579964.610
4      426047.851
5      648251.080
6      608013.160
7      574238.481
8      616368.000
9      931440.372
10     974603.590
11    1132407.740
12     897110.400
Name: TotalPrice, dtype: float64
```

We mapped the months to the following seasons :
Winter, Spring, Summer, Autumn.

```
# Mapping months to seasons
seasons = {
    1: 'Winter', 2: 'Winter', 3: 'Spring',
    4: 'Spring', 5: 'Spring', 6: 'Summer',
    7: 'Summer', 8: 'Summer', 9: 'Autumn',
    10: 'Autumn', 11: 'Autumn', 12: 'Winter'
}

# Assigning seasons based on months
df['Season'] = df['Month'].map(seasons)
```



We analyzed that the maximum number of sales take place in the **autumn season** followed by summer, winter and spring season, respectively. This could be because at the ending of autumn, we see the holiday season take place resulting in a lot of time for people to shop and the Black Friday discounts that take place during this time.

5.1 Top 5 countries with the highest number of orders

Country	
Australia	1986.627101
Austria	534.437895
Bahrain	274.200000
Belgium	343.789580
Brazil	1143.600000

5.2 Correlation between the country of the customer and the average order

Average order value for each country:

Country	
Australia	1986.627101
Austria	534.437895
Bahrain	274.200000
Belgium	343.789580
Brazil	1143.600000
Canada	611.063333
Channel Islands	608.675455
Cyprus	647.314500
Czech Republic	141.544000
Denmark	893.720952
EIRE	784.593166
European Community	258.350000
Finland	465.140417
France	429.504017
Germany	367.658723
Greece	785.086667
Iceland	615.714286
Israel	1165.708333
Italy	307.100182
Japan	1262.165000
Lebanon	1693.880000
Lithuania	415.265000
Malta	250.547000
Netherlands	2818.431089
Norway	879.086500
Poland	300.547500
Portugal	415.140143
RSA	1002.310000
Saudi Arabia	65.585000
Singapore	912.039000
Spain	521.662667
Sweden	795.563261
Switzerland	785.061972
USA	247.274286
United Arab Emirates	634.093333
United Kingdom	340.830609
Unspecified	333.383750

dtype: float64

Correlation between the number of orders and average order value: -0.335022638668918

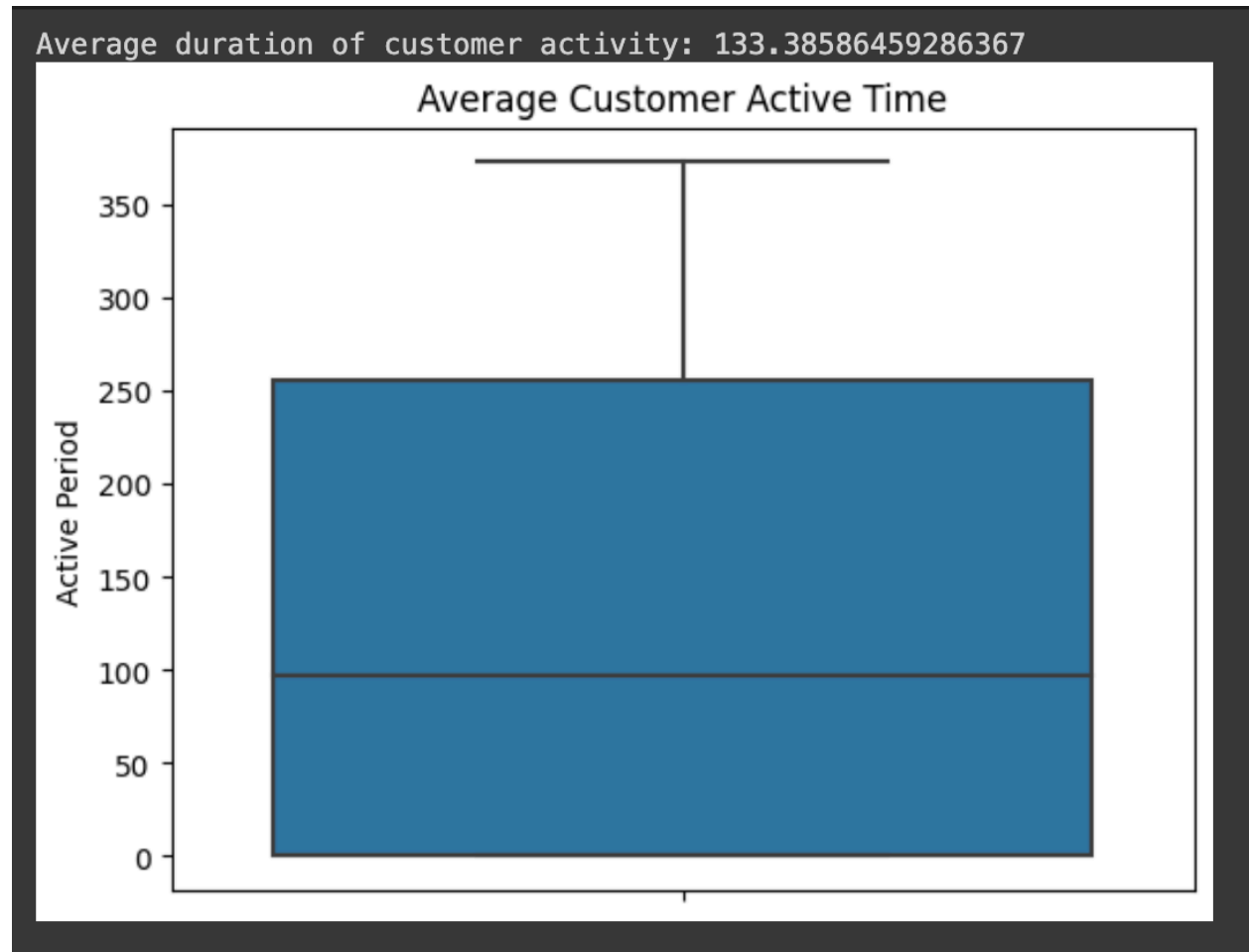
There is negative correlation between customer and average order i.e -0.33

6.

Payment Analysis

Data is not sufficient for doing the payment analysis

7. Customer Behavior



- On an average, Customers remain active for 133 days between their first and last purchase.

7.b Customer Segments distribution

	Recency	Frequency	Segment
12346.0	325	2	Low Activity
12347.0	1	182	High Activity
12348.0	74	31	Low Activity
12349.0	18	73	High Activity
12350.0	309	17	Low Activity
12352.0	35	95	High Activity
12353.0	203	4	Low Activity
12354.0	231	58	Low Activity
12355.0	213	13	Low Activity
12356.0	22	59	High Activity

8. Returns and Refunds

Data is not sufficient for doing the profitability analysis

9. Profitability Analysis

Data is not sufficient for doing the profitability analysis

10. Customer Satisfaction

Data is not sufficient for doing the analysis