

# E-COMMERCE CUSTOMER SEGMENTATION ANALYSIS

---



Presented By : M. Alwan Nur Ajie

# Table Of Content

1

Background  
Project

2

Data  
Understanding

3

Data Cleaning

3

EDA

4

Preprocessing

5

Modelling - K  
Means

6

Rekomendasi

# Project Background

The company is a UK-based e-commerce business that has been selling gifts and household items for adults and children through their website. The dataset used includes sales transactions over one year (December 2018 to December 2019). This company needs to identify potential customer segments to target their market more effectively and efficiently. This information will ensure that all future campaigns and marketing efforts are directed at customers who are most likely to make a purchase.

# Objective

- Identifying potential customer segments to help target the market more effectively.
- Can provide business recommendations that align with the characteristics of each customer segments.

# Data Understanding

The dataset has 536,350 rows and 8 columns, here's the description of each feature:

- TransactionNo : Unique number that define each transaction
- Date: The date when transaction was made.
- ProductNo: Unique character used to identify specific product.
- Product: product name.
- Price: Price of each product per unit in pound sterling (£).
- Quantity: Quantity of each product per transaction
- CustomerNo: Unique number that defines each customer.
- Country: Name of country where the customer resides

# MISSING VALUE & DUPLICATE VALUE

## Missing Value

There are 55 rows (0.01%) with missing values in the CustomerNo column.

## Duplicated Value

There are 5,200 duplicated values.

## Handling Missing Value

Drop the rows with missing values in the CustomerNo column.

## Handling Duplicated Value

Drop the rows with duplicated values.

# TREATMENT LANJUTAN

## Quantity & TransactionNo Columns

There are negative values in the quantity column and TransactionNo values that start with 'C', indicating canceled transactions. These values are invalid in transactions and will be dropped.

There are 8494 columns--> Drop

## Date Column

Change the data type of the date column to datetime.

## Create new feature: Revenue

The Revenue columns is obtained from the Price \* Quantity columns.



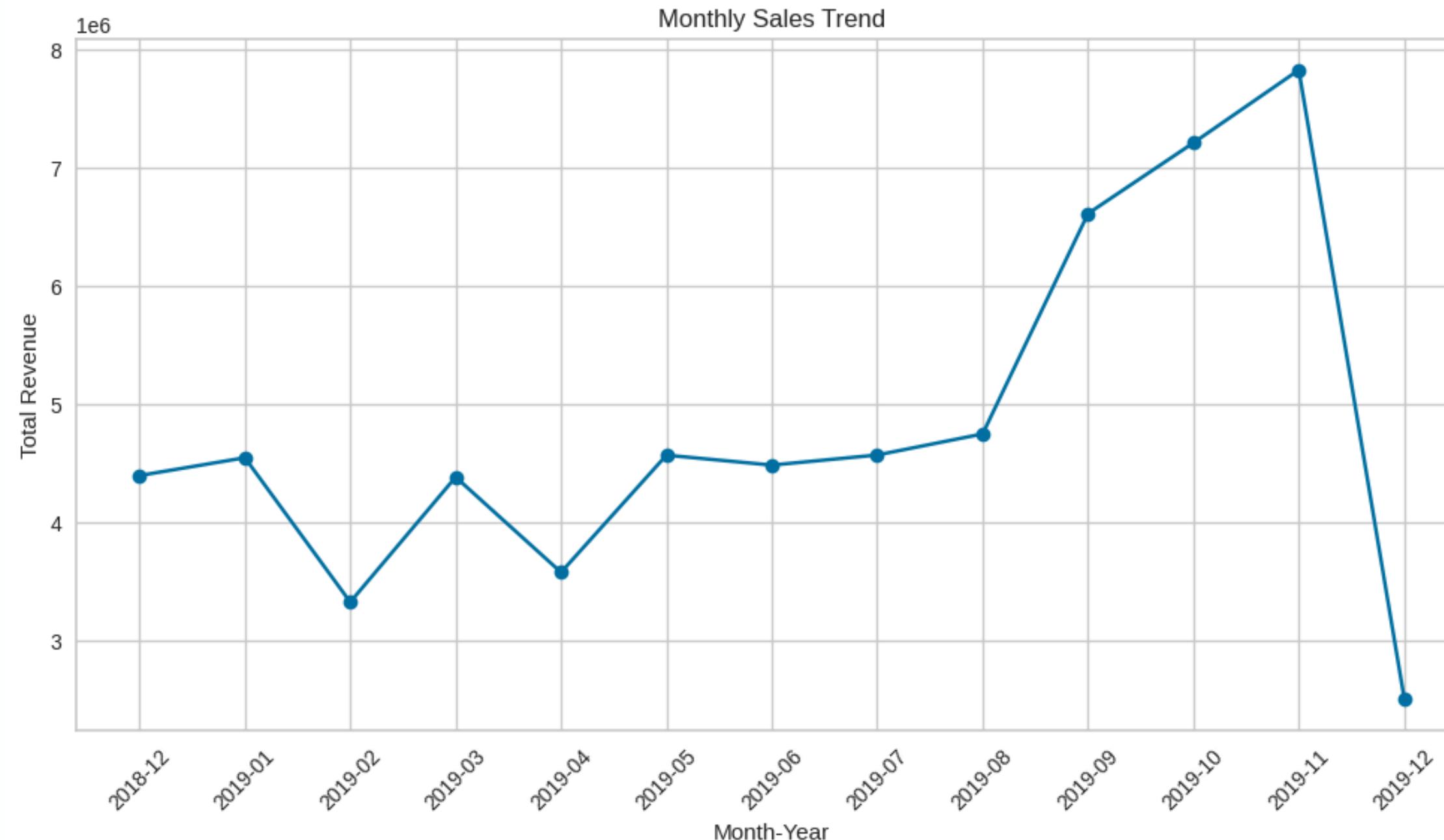
# EDA

For Better Visualization:

<https://finalprojectcustsegment.streamlit.app>



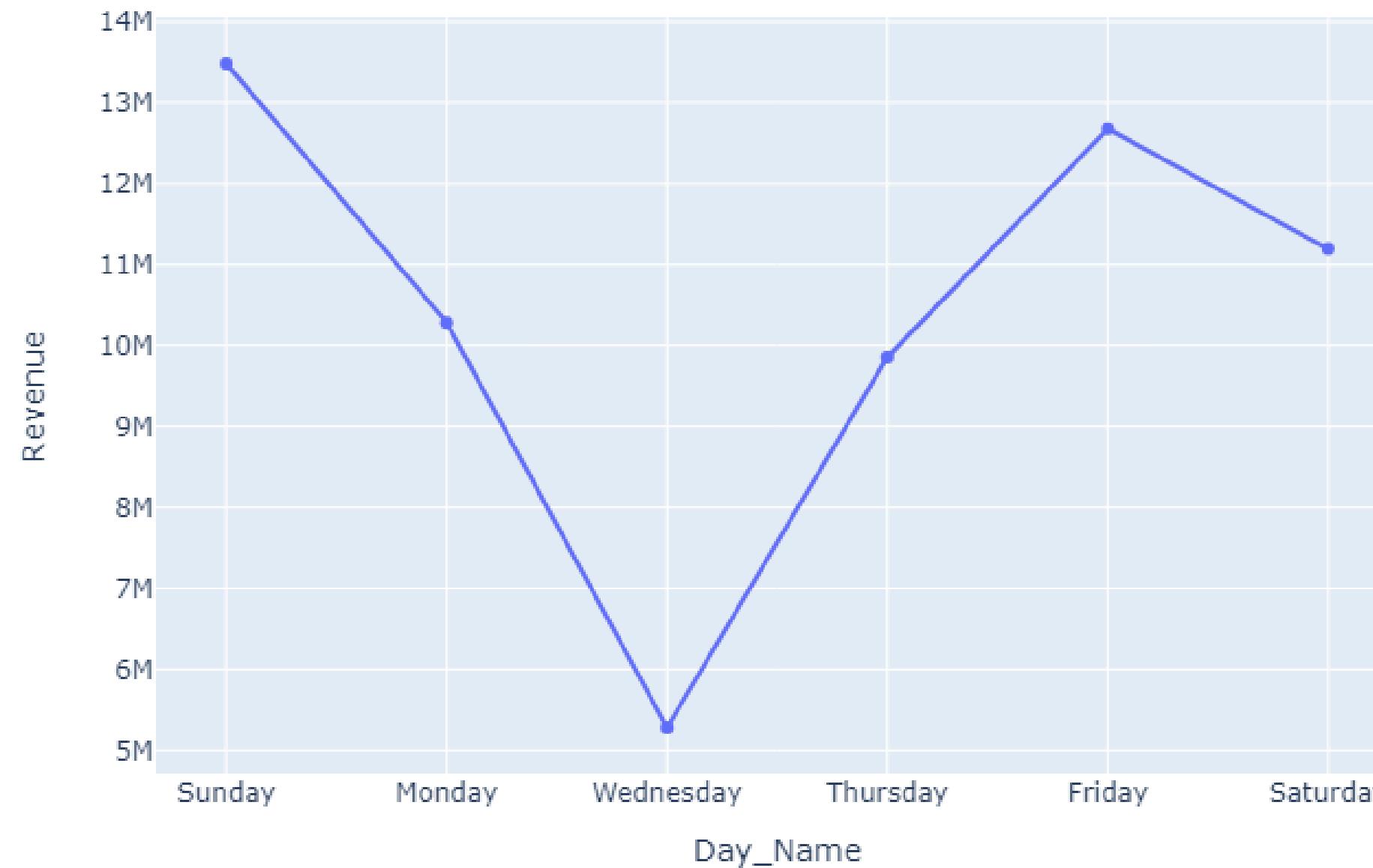
# WHAT ARE THE SALES TRENDS BY MONTH?



Fluctuating trend from December 2018 to May 2019. Sales remained stagnant from May to August. Afterward, there was an increase, reaching its peak in November.

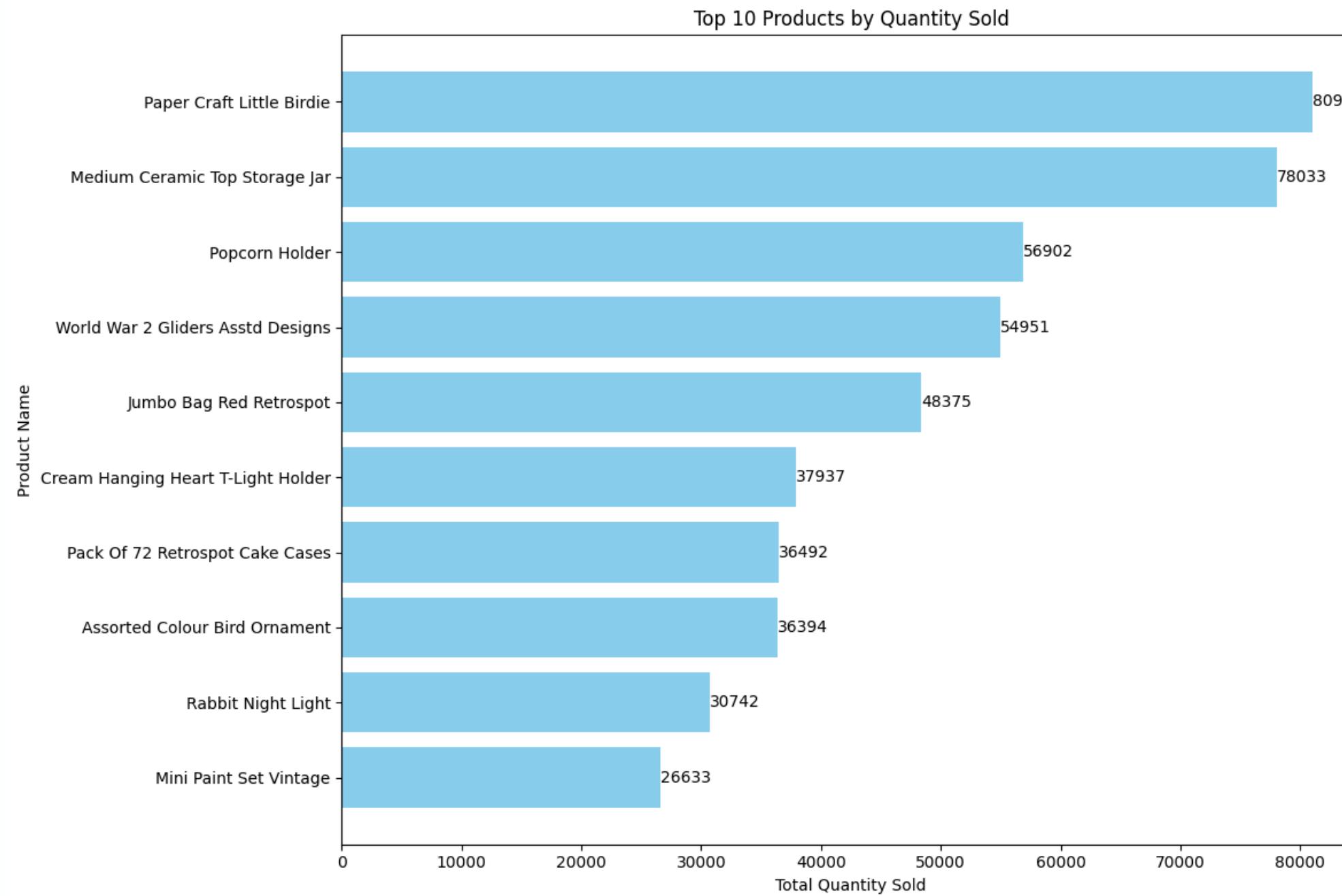
# WHAT ARE THE SALES TRENDS BY DAY?

Sales Trend Over the Day



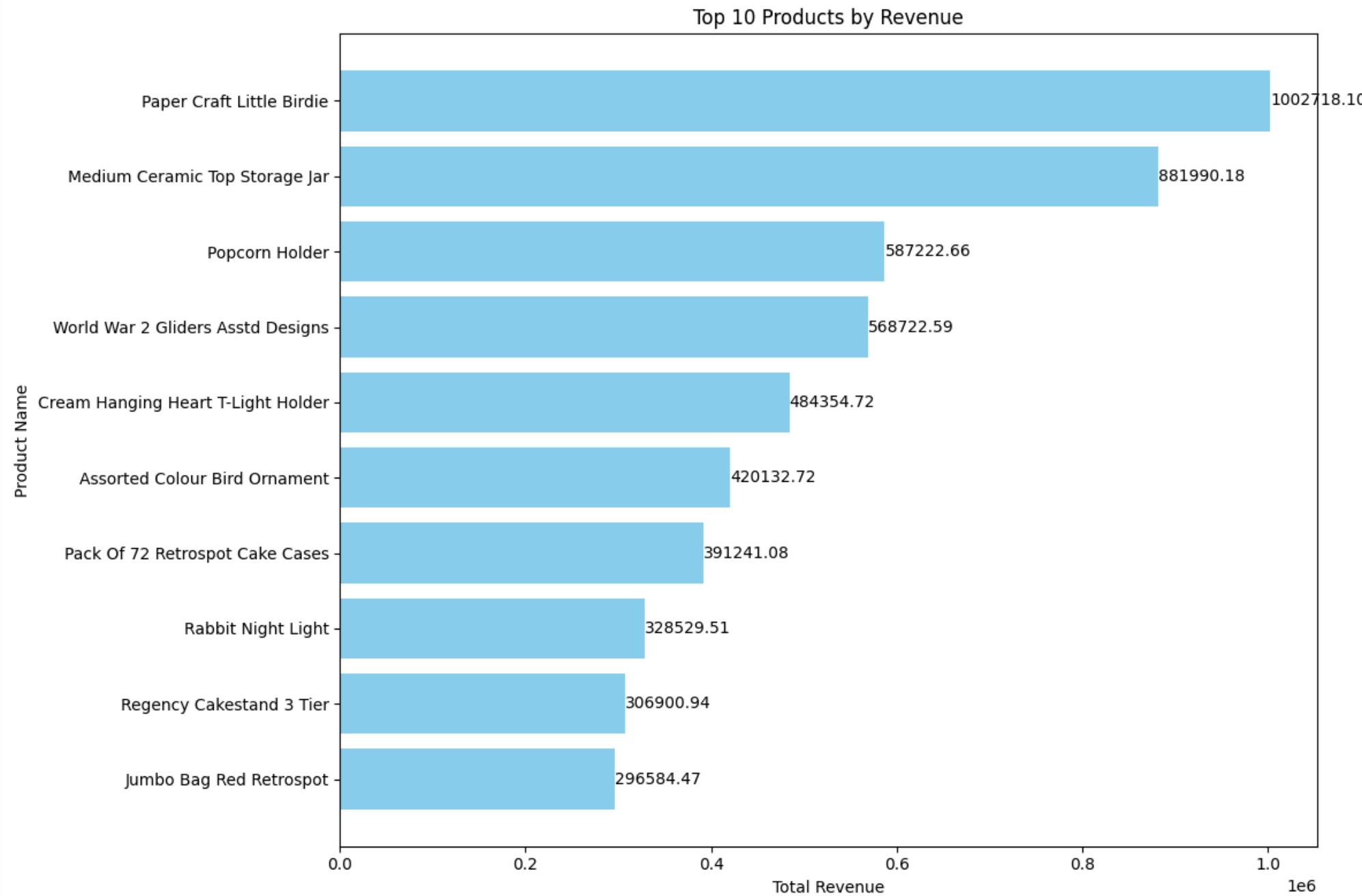
Customers tend to make purchases on weekends.  
The store is closed on Tuesdays.

# WHAT ARE THE TOP 10 PRODUCTS FREQUENTLY BOUGHT BY CUSTOMERS?



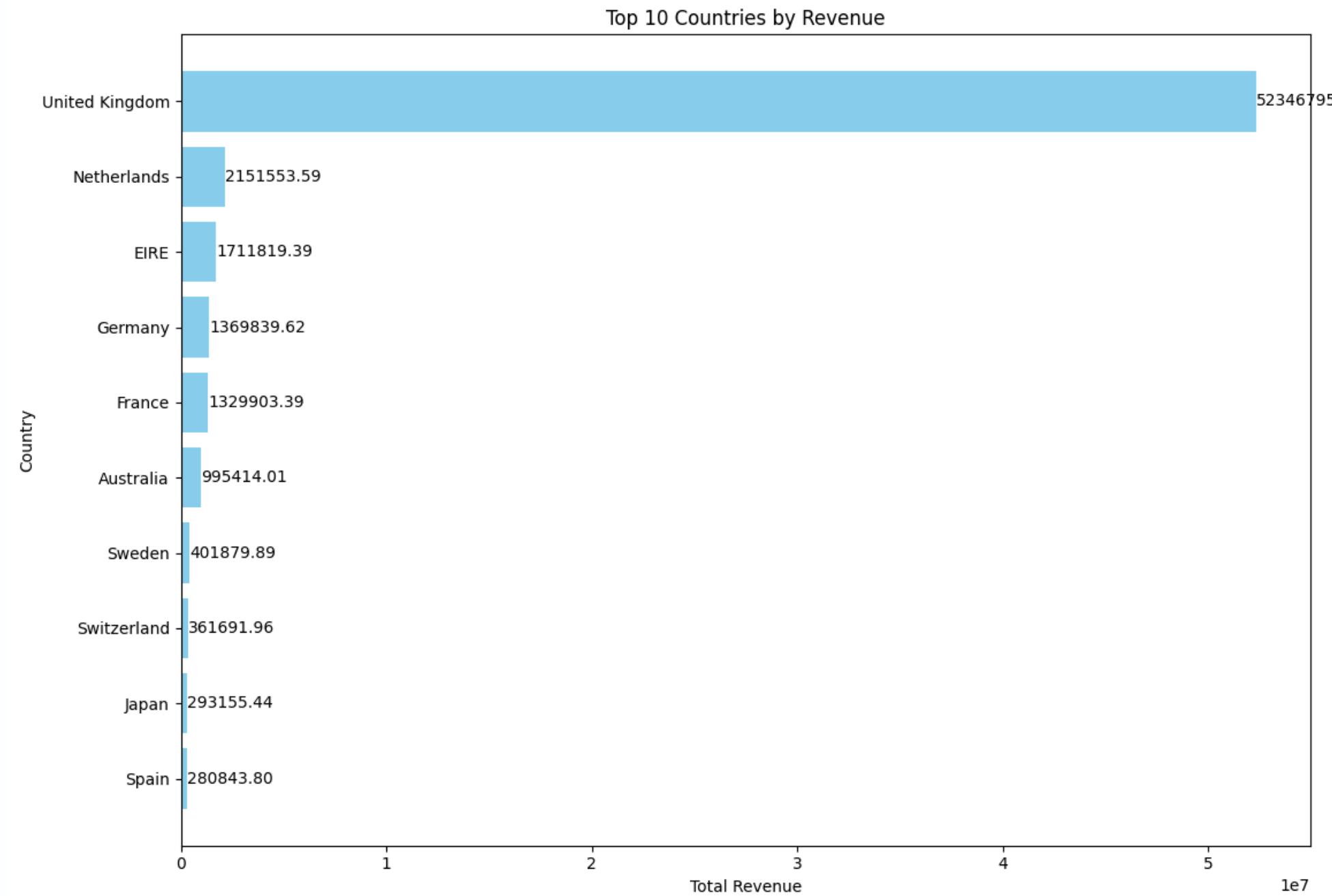
Paper Craft Little Birdie is the top-selling product with a total of 80,995 units sold. The frequently bought products include various categories such as paper crafts, ceramic storage, home decorations, and children's items like toys.

# WHAT ARE THE TOP 10 PRODUCTS THAT GENERATE THE HIGHEST REVENUE?



The top 4 products generating the highest revenue are also the top 4 most frequently bought items by customers.

# WHAT ARE THE TOP 10 COUNTRIES THAT GENERATE THE HIGHEST REVENUE?



The UK is the country with the highest revenue, totaling £52,346,795.60. This makes sense since the store is based in the UK.

# EDA Recommendations

- Conduct post-holiday promotions to reduce the decline in sales, such as year-end discounts or clearance sales.
- Holding special promotions or discounts on Wednesdays can help boost sales on days with lower shopping activity.
- Focus marketing campaigns and promotions on these best-selling products to further increase sales. For instance, offer special discounts and ensure that these products are always in stock.
- Countries outside the UK, such as the Netherlands, Ireland, and Germany, show significant revenue. Expanding operations and marketing campaigns in these countries can increase additional revenue.

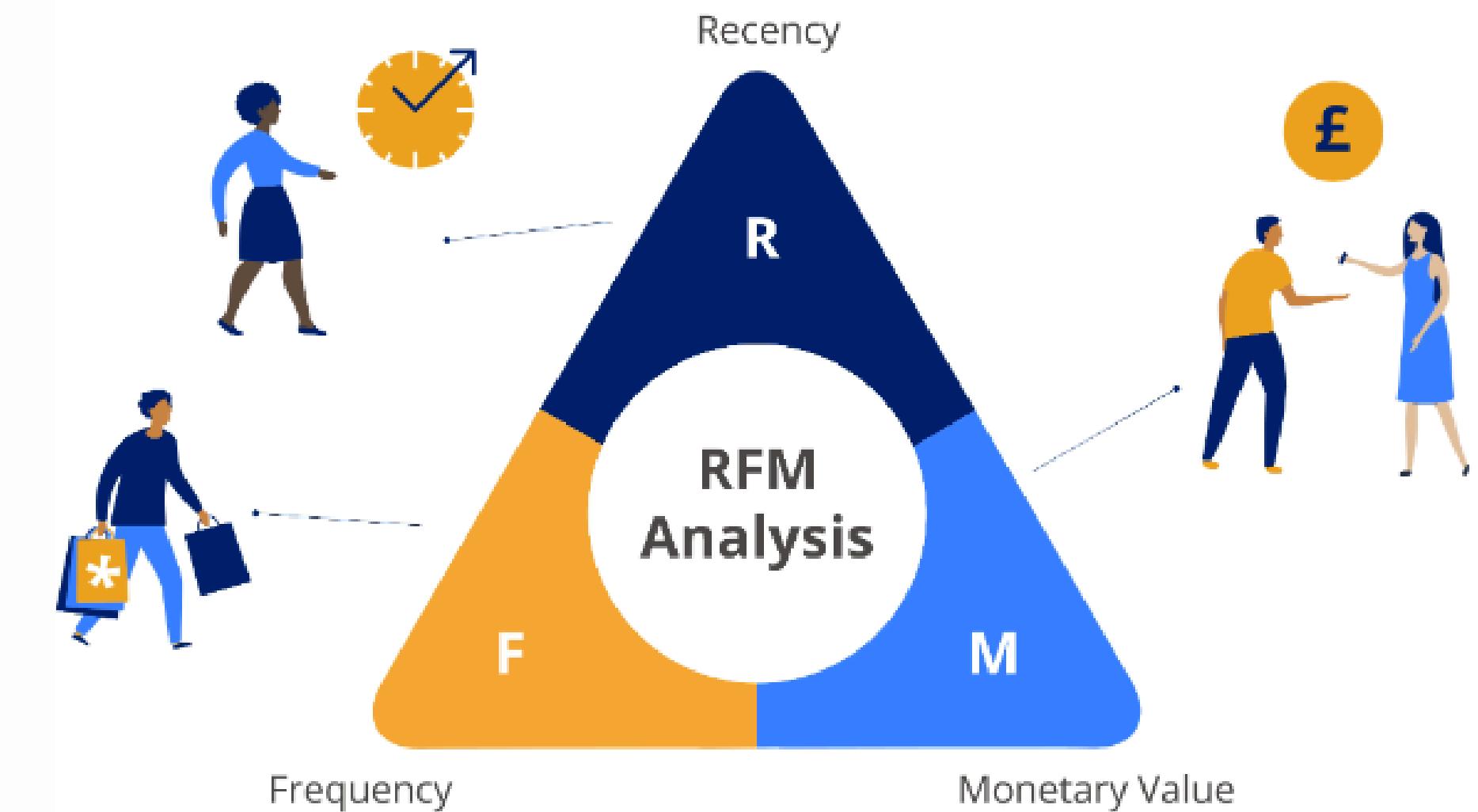
oooo

# PREPROCESSING



# RFM

Preprocessing is done to facilitate RFM analysis, which uses three key metrics: Recency, Frequency, and Monetary value. The current dataset is in long format, where a single customer (especially CustomerNo) can appear more than once in each row. I will group the data by each customer's CustomerNo beforehand.



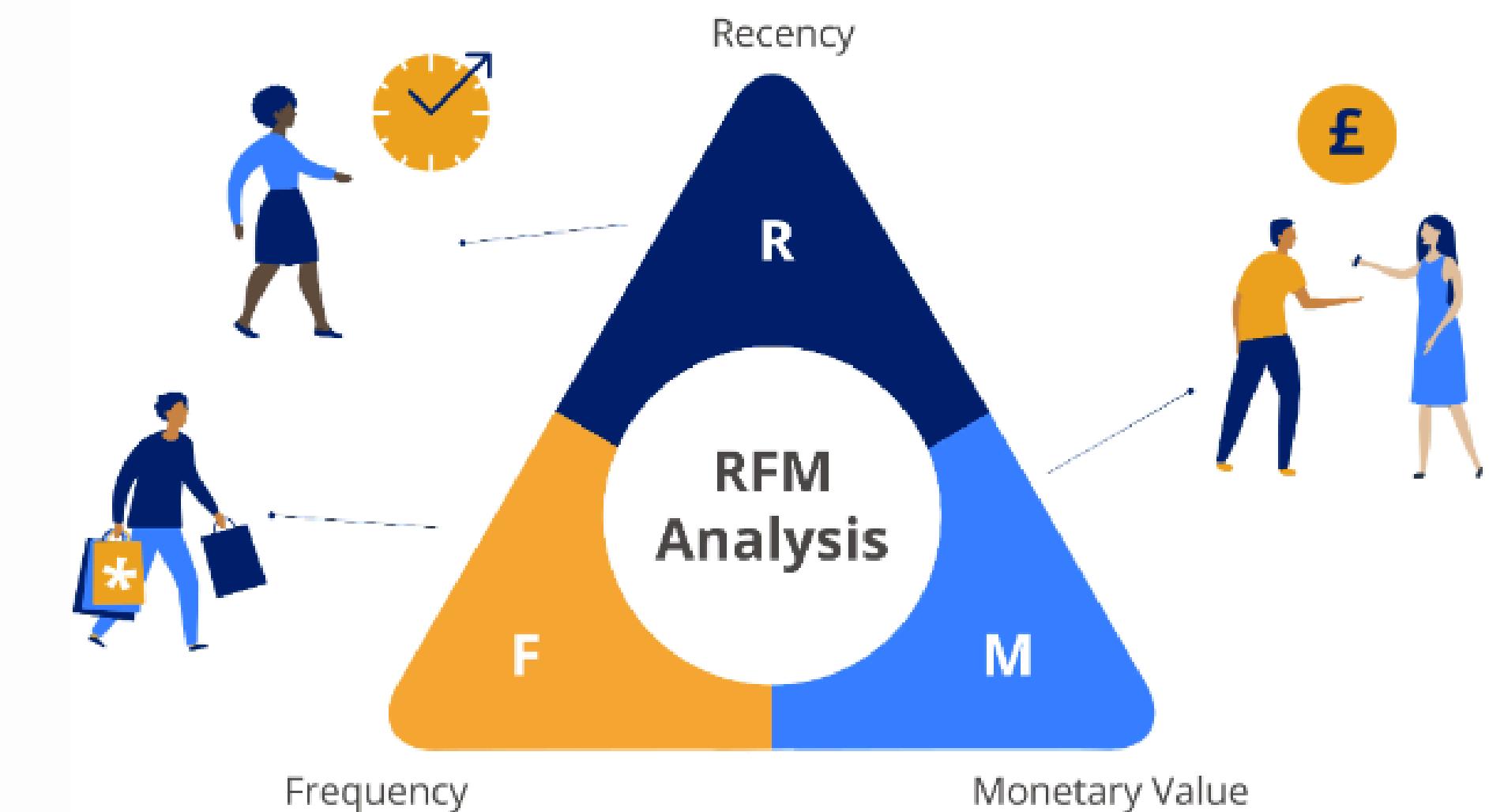
# Recency

The most recent order time of a customer.

recent\_date: 2019-12-10 00:00:00

	CustomerNo	LastPurchaseDate	Recency
0	12004	2019-04-26	228
1	12006	2019-05-05	219
2	12008	2019-03-08	277
3	12013	2018-12-15	360
4	12024	2019-06-16	177
...	...	...	...
4713	18280	2019-03-07	278
4714	18281	2019-06-12	181
4715	18282	2019-12-02	8
4716	18283	2019-12-06	4
4717	18287	2019-10-28	43

4718 rows × 3 columns

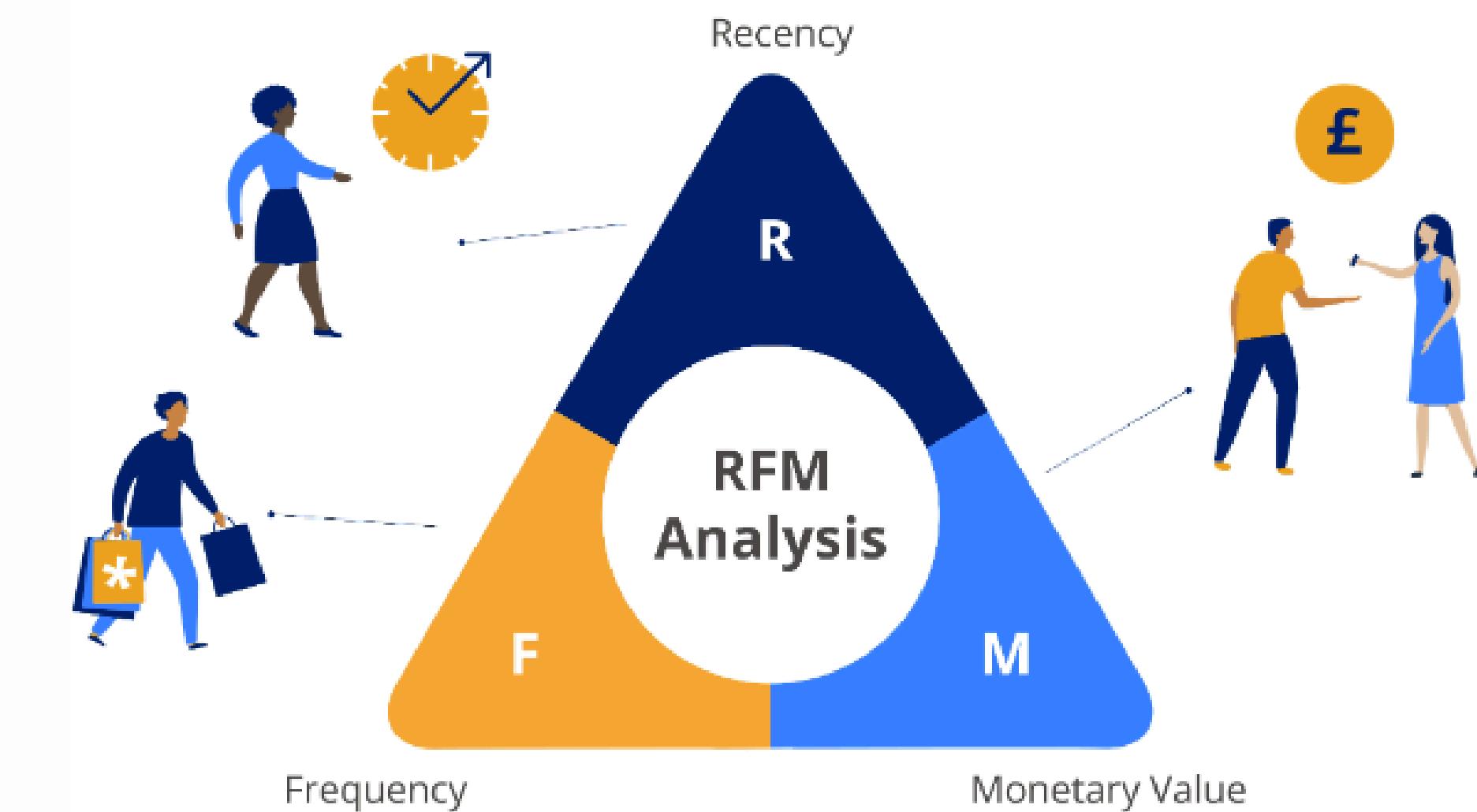


# Frequency

How frequent a customers return to make a purchase.

CustomerNo	Frequency
408	12748
2085	14911
4384	17841
667	13089
2406	15311
...	...
2339	15226
2345	15234
2347	15236
2357	15247
2359	15250

4718 rows × 2 columns

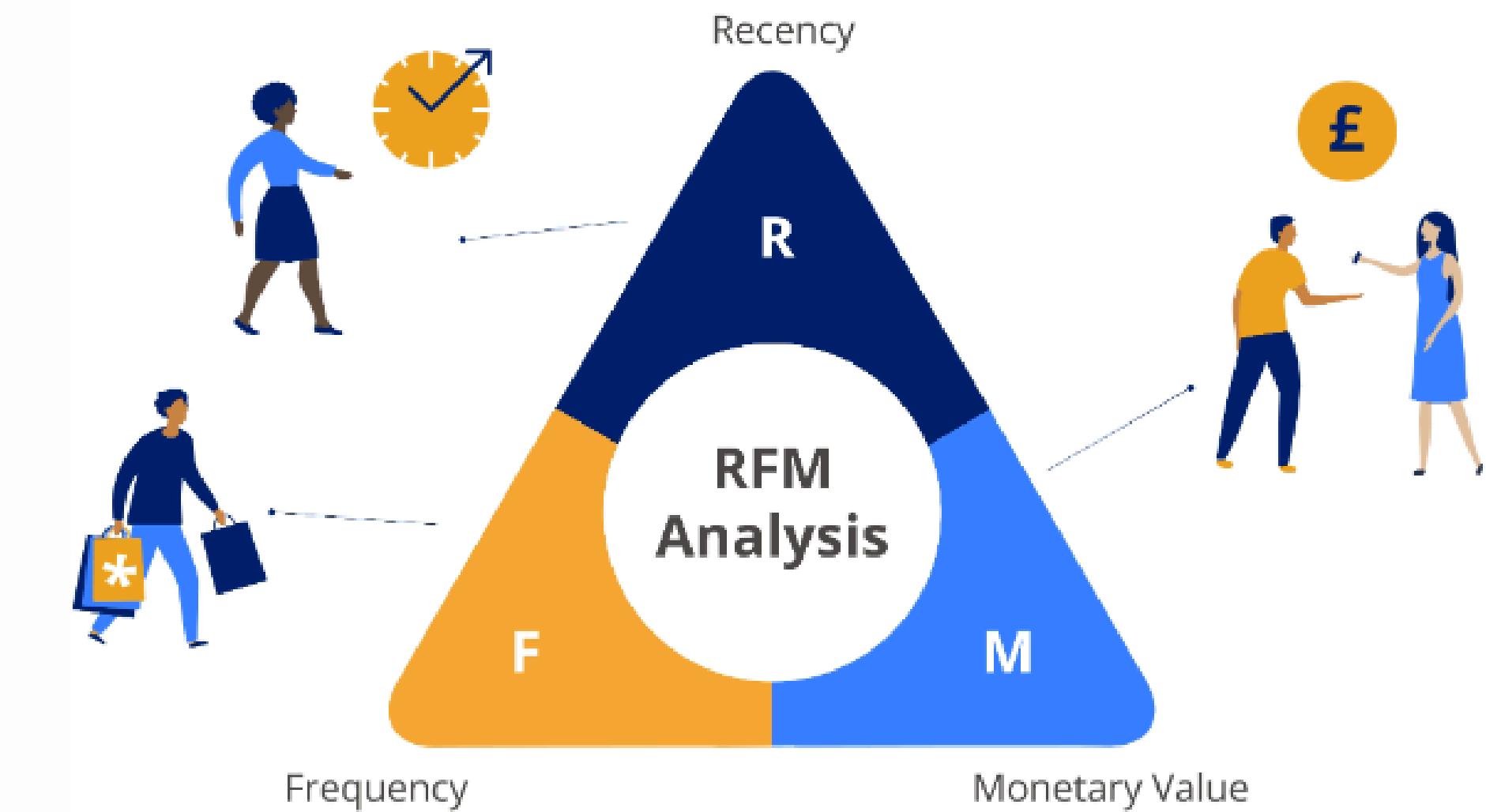


# Monetary

The total amount of money spent by the customer.

CustomerNo	Monetary
1880	14646 2112282.03
3302	16446 1002741.57
2085	14911 914204.19
126	12415 900545.54
4581	18102 897137.36
...	...
61	12309 12.86
1716	14435 12.38
453	12810 11.98
1208	13775 11.53
3693	16937 5.97

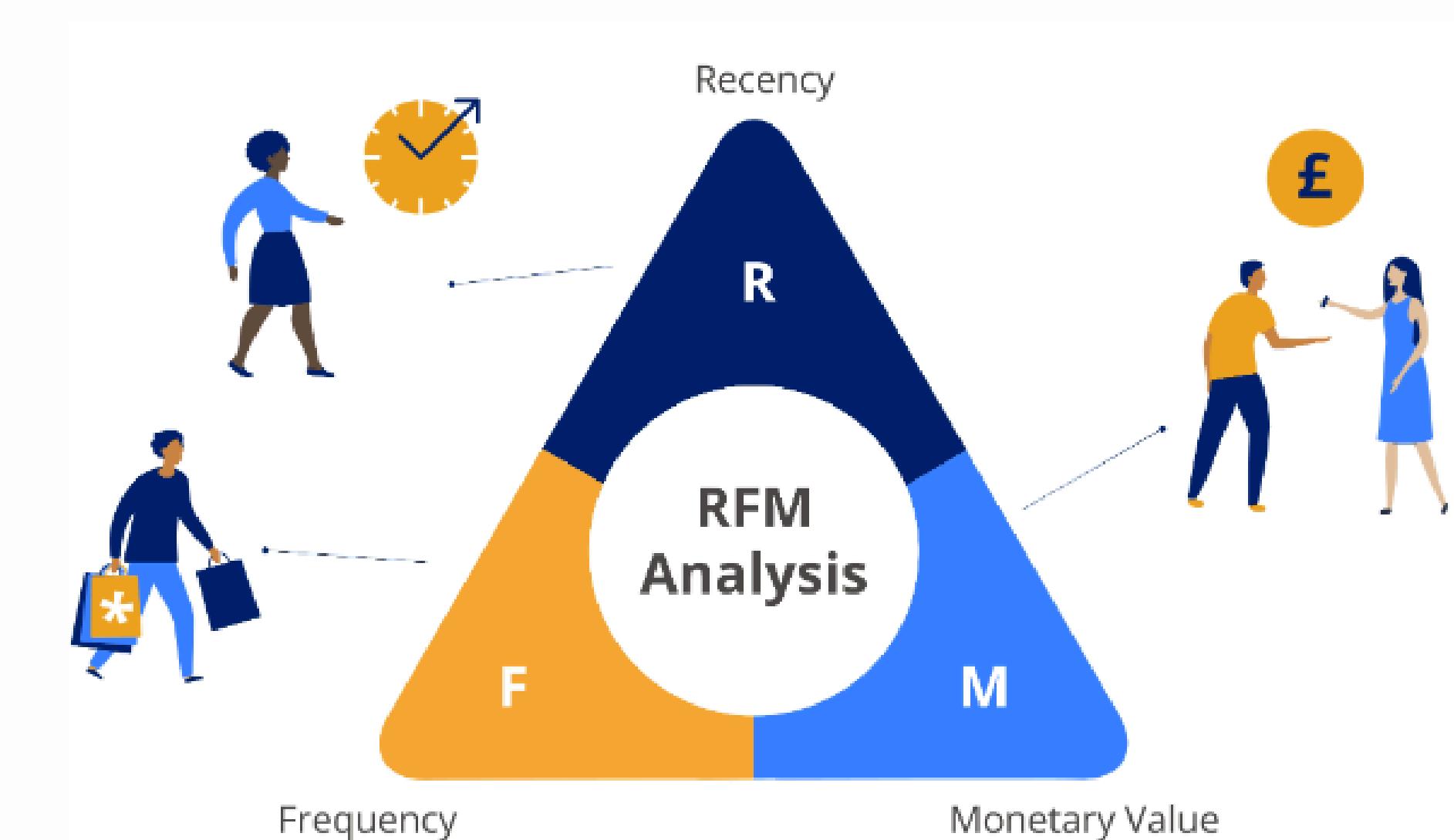
4718 rows × 2 columns



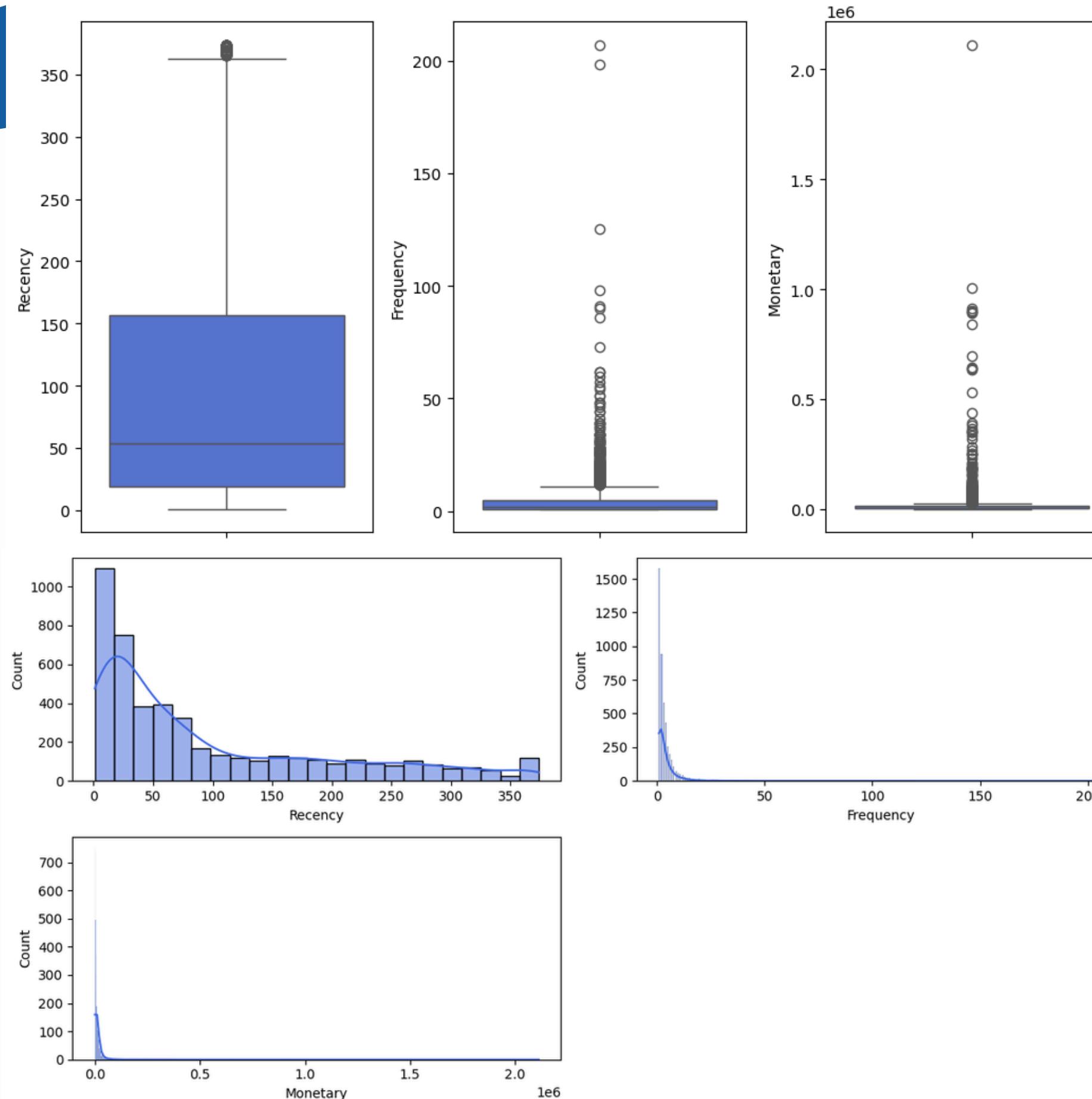
# Merging RFM into a DataFrame.

	CustomerNo	Recency	Frequency	Monetary
0	12004	228	1	1509.60
1	12006	219	1	24.76
2	12008	277	1	5689.57
3	12013	360	1	69.96
4	12024	177	1	149.52
...	...	...	...	...
4713	18280	278	1	623.26
4714	18281	181	1	576.58
4715	18282	8	2	1044.86
4716	18283	4	16	11773.90
4717	18287	43	3	18139.56

4718 rows × 4 columns



# Quick Brief RFM Distribution



- On average, customers made their last purchase within the past 50 days, while some customers haven't made a purchase in a long time.
- Some customers make large purchases.
- A significant portion of revenue comes from a small number of high-value customers.

# SCALING

Scaling is performed to prevent variables with a wide range from affecting distance calculations, for example in Monetary.

Scaling is done using a standard scaler.

	Recency	Frequency	Monetary
0	1.287572	-0.431944	-0.217001
1	1.199118	-0.431944	-0.244313
2	1.769156	-0.431944	-0.140113
3	2.584901	-0.431944	-0.243482
4	0.786331	-0.431944	-0.242018

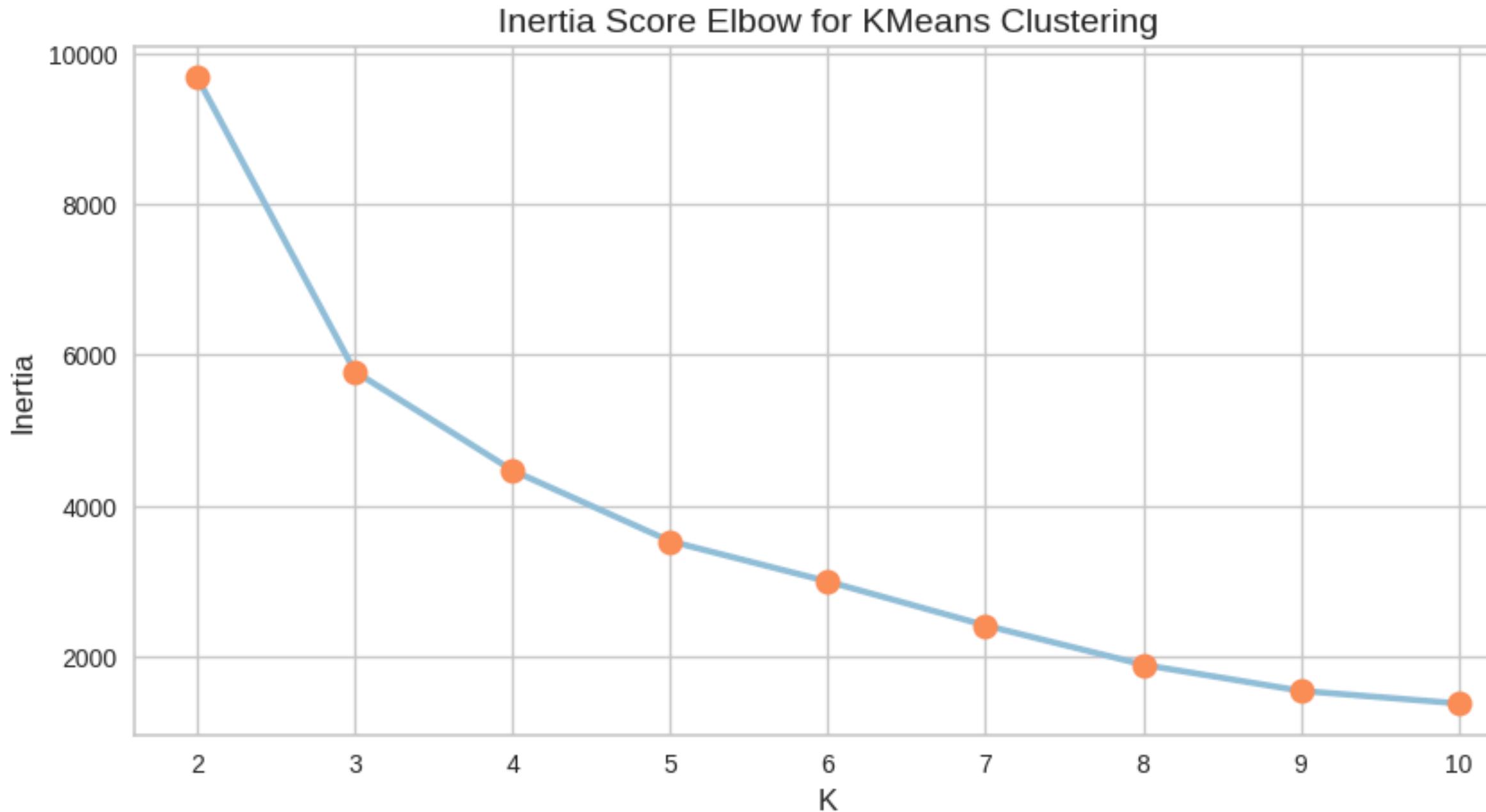


# MODELLING



# FINDING OPTIMAL K VALUE

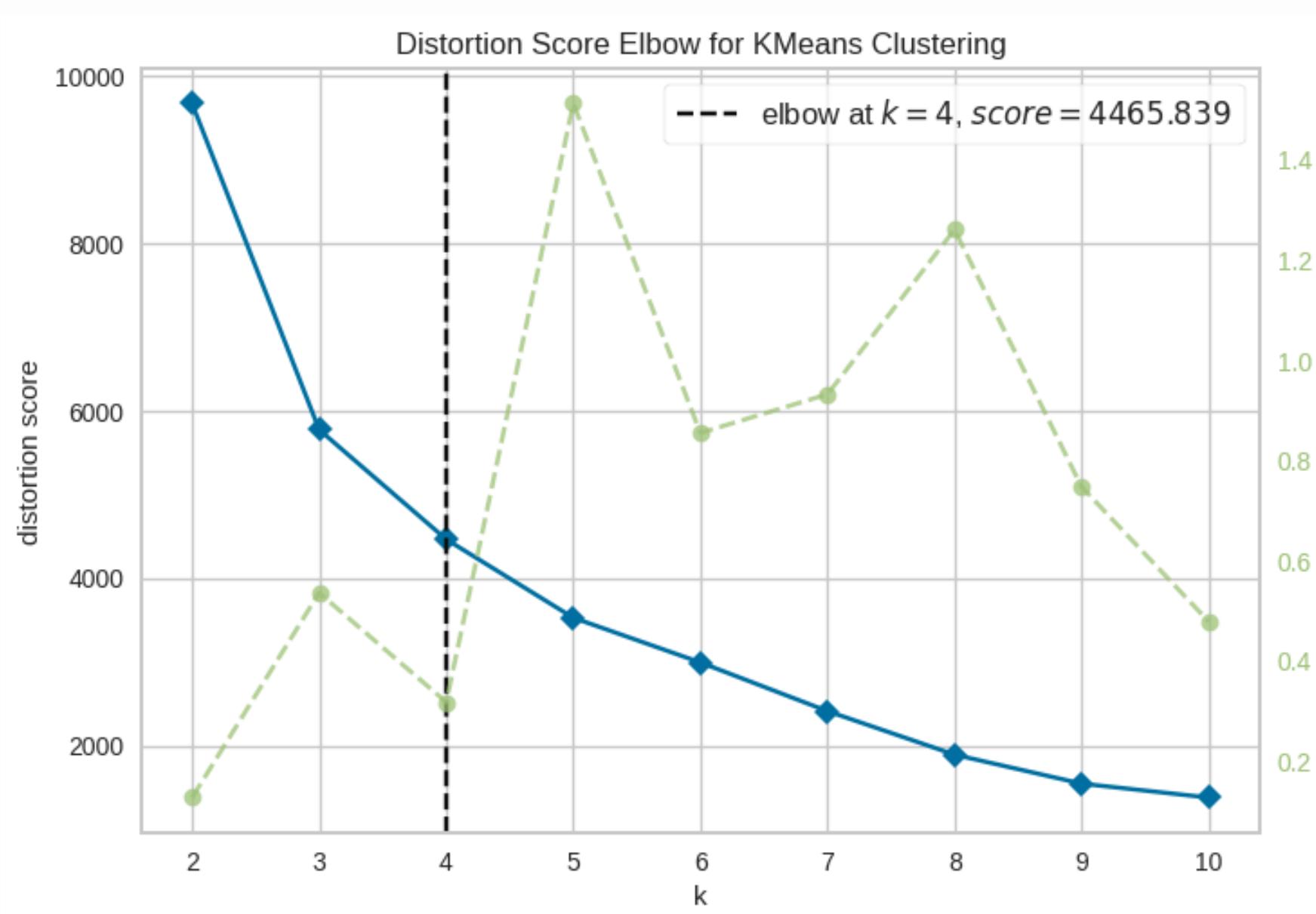
## ELBOW METHOD



Based on the generated graph, there is no clear sharp drop or significant bend, making it difficult to determine the optimal value of k with certainty.

# FINDING OPTIMAL K VALUE

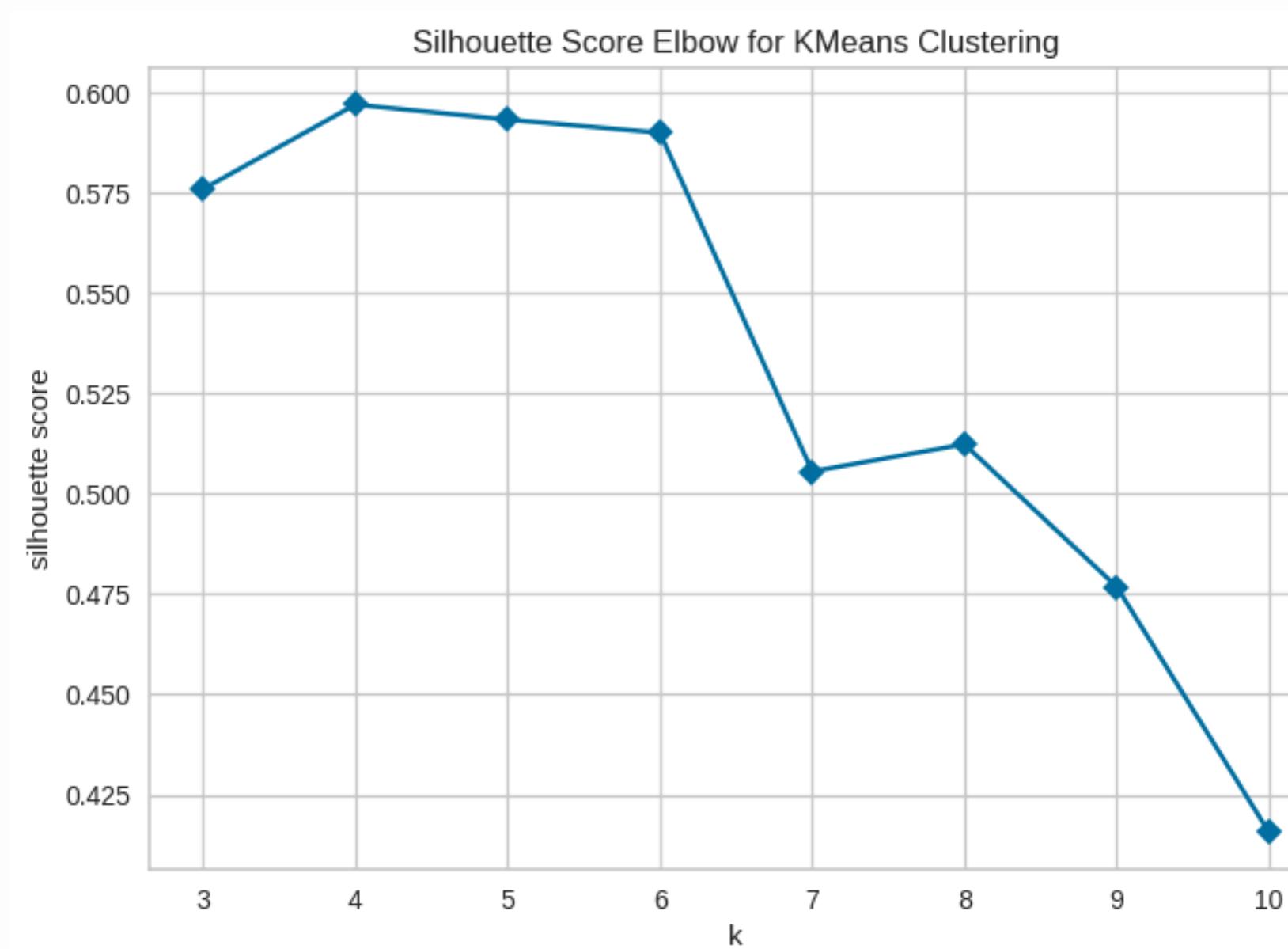
## ELBOW METHOD



Based on the graph, with the distortion parameter, the optimal value of  $k$  is 4.

# FINDING OPTIMAL K VALUE

## Silhouette Score



Based on the silhouette graph, the optimal value of  $k = 4$ .

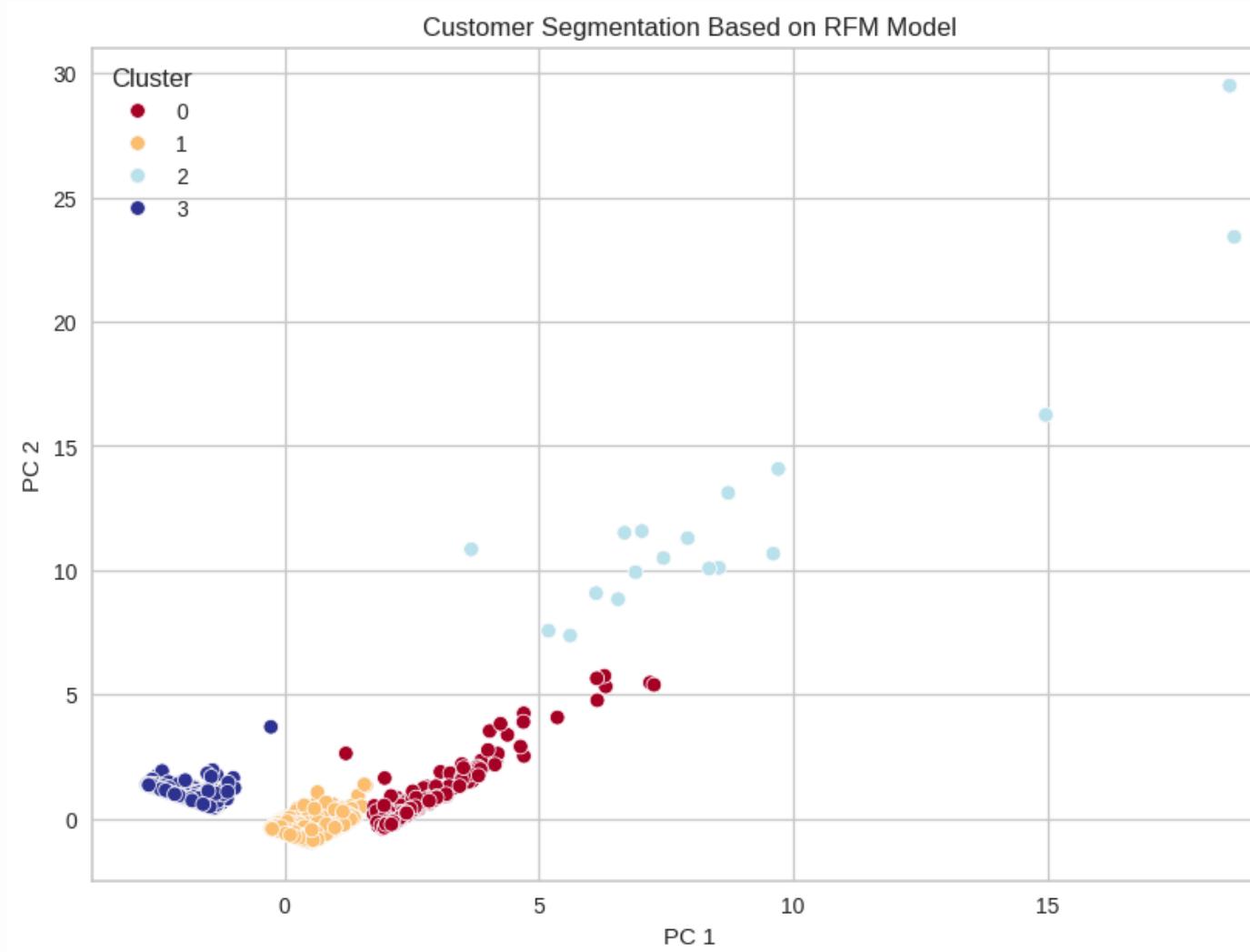
# K-MEAN CLUSTERING

After identifying the optimal number of clusters, which is 4, modeling is performed to determine the clusters.

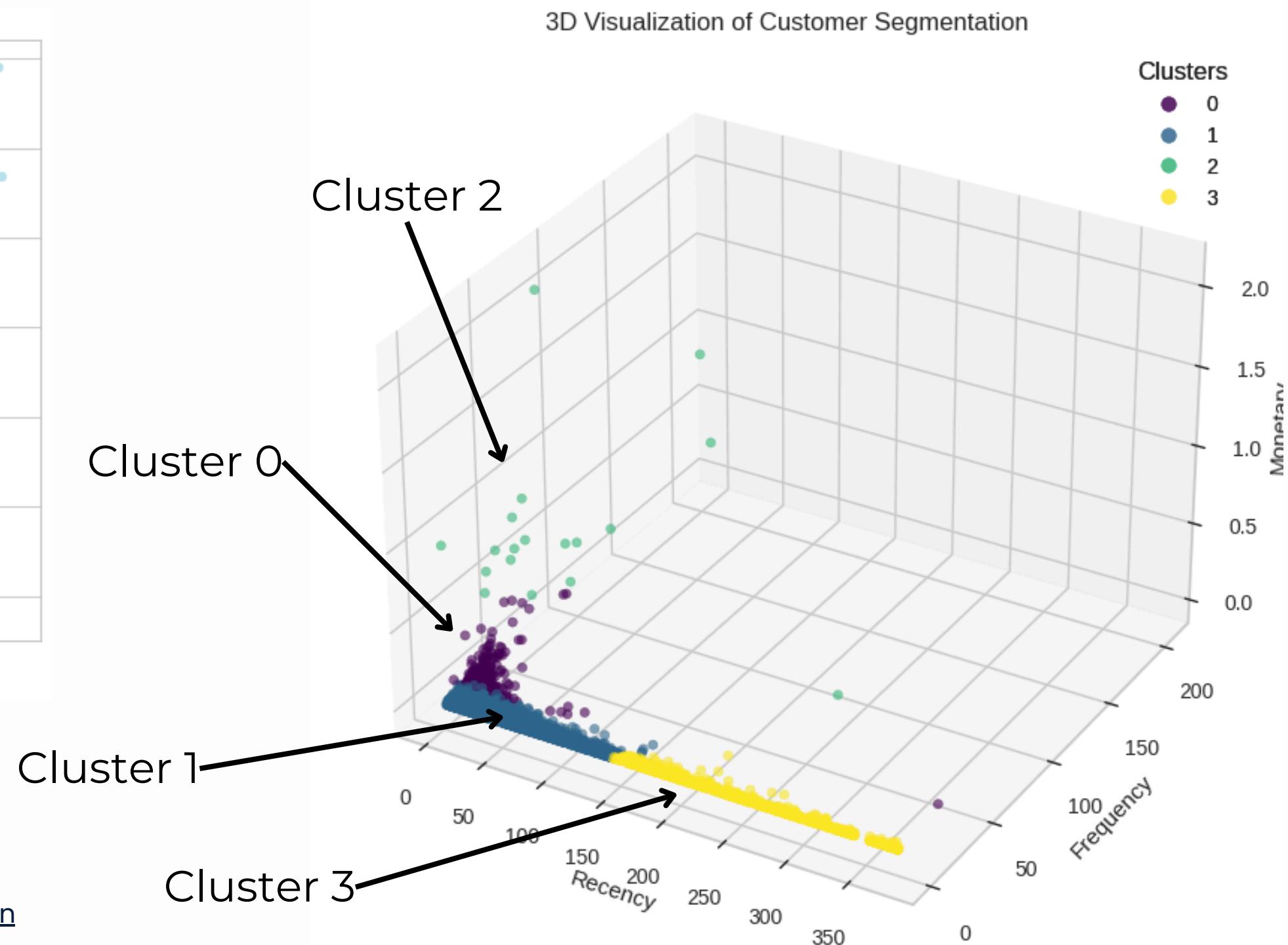
```
▶ from sklearn.cluster import KMeans  
  
# fit model  
kmeans = KMeans(n_clusters=4, random_state = 123)  
kmeans.fit(rfm_features_std.values)  
rfm_features_std['Cluster'] = kmeans.labels_  
  
▶ rfm_features_std.head()  
  
Recency Frequency Monetary Cluster  
0 1.287572 -0.431944 -0.217001 3  
1 1.199118 -0.431944 -0.244313 3  
2 1.769156 -0.431944 -0.140113 3  
3 2.584901 -0.431944 -0.243482 3  
4 0.786331 -0.431944 -0.242018 3
```

# VISUALIZE CUSTOMER SEGMENTATION BASED ON RFM

After fitting the model, visualize the results to evaluate the segmentation of each cluster.



Visualizing with PCA --> 2D



For Better Visualization:  
[https://finalprojectcustsegment.streamlit.app/Customer\\_Segmentation](https://finalprojectcustsegment.streamlit.app/Customer_Segmentation)

# PERCENTAGE OF CUSTOMERS IN EACH CLUSTER

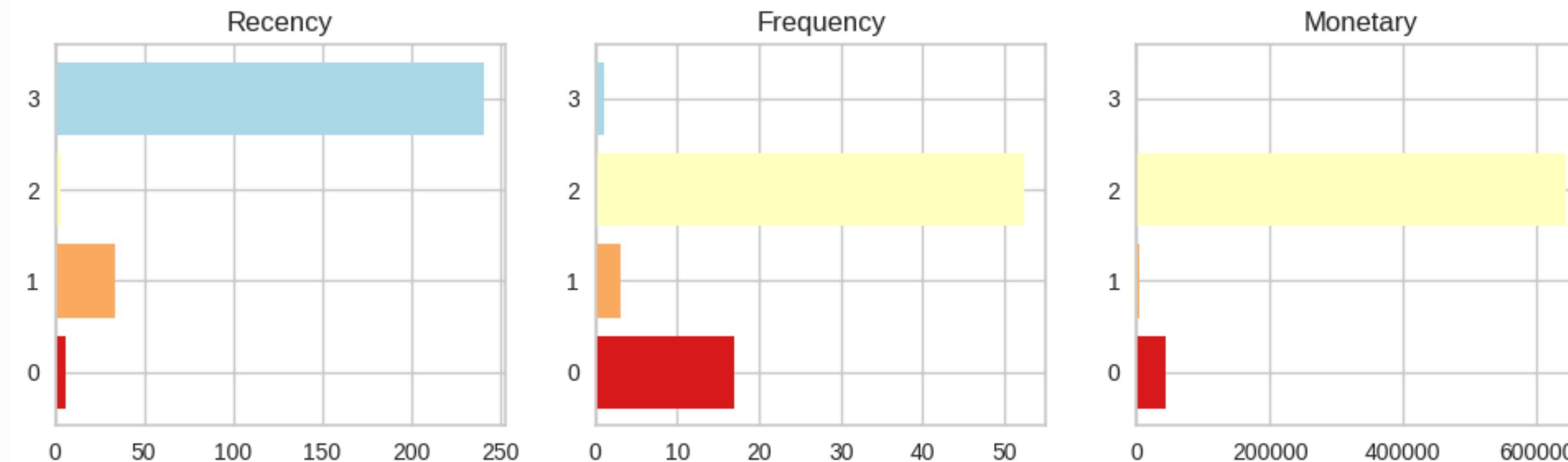
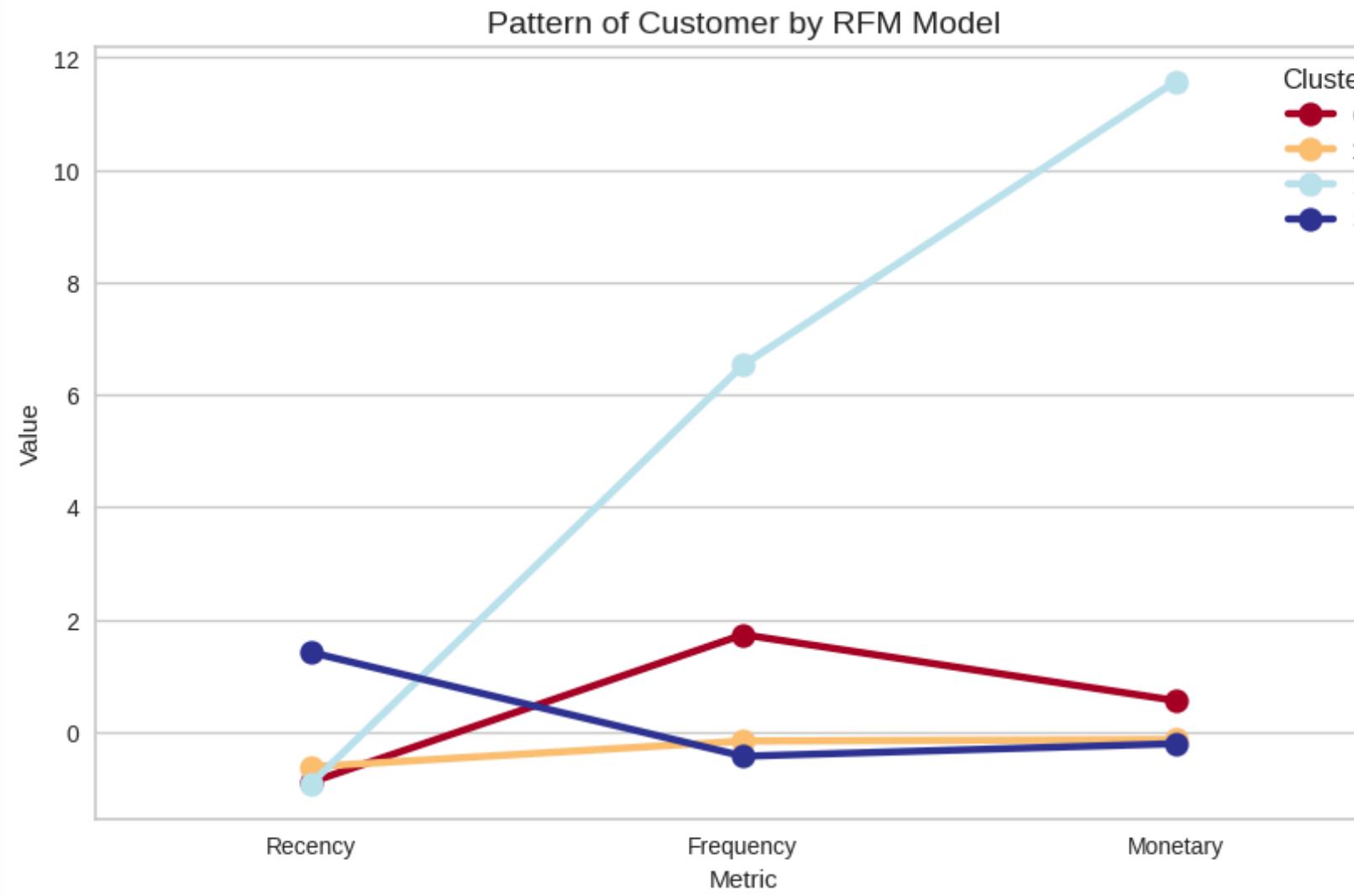


Cluster	count	percentage (%)
0	0	5.45
1	1	67.55
2	2	0.38
3	3	26.62

Cluster 1 has the highest percentage of customers.

Cluster 2 has a small percentage of customers.

# SEGMENTATION ANALYSIS



# SEGMENTATION ANALYSIS

## Cluster 0 : Frequent High Spender

- Customers in this cluster make purchases quite frequently and in relatively large amounts.
- They have recently made purchases (low Recency value), indicating that they are very active.

## Cluster 2 : Loyal High Spender

- Customers in this cluster are the most valuable with very high transaction values and very frequent purchases.
- They have recently made purchases, indicating a very high level of activity.

## Cluster 1: Regular Shopper

- Customers in this cluster have moderate transaction values and make purchases with moderate frequency.
- They have higher recency compared to some other clusters, indicating that they may not have made a purchase recently but are still fairly active.

## Cluster 3 : Inactive Low Spender

- Customers in this cluster have low transaction values and very low purchase frequency.
- They have not made purchases recently, indicating that they may no longer be active.

# REKOMENDASI

## Cluster 0 : Frequent High Spender

- Provide loyalty programs such as additional reward points or free shipping.
- Promote related products or more premium products that might interest them with special recommendations on the product page.
- Send personalized offers based on their purchase history via email.

## Cluster 1 : Regular Shopper

- Send special offers or discounts on special occasions or holidays to encourage them to shop more frequently.
- Provide special discounts for the next purchase, valid for a limited time.
- Offer incentives for customers to refer friends or family, such as discounts for each successful referral.

# REKOMENDASI

## Cluster 2 : Loyal High Spender

- Create a VIP program that provides exclusive benefits such as special gifts, exclusive customer service priority.
- Offer additional rewards for their loyalty, such as extra reward points or shopping vouchers.
- Send personal thank-you notes or small gifts to strengthen relationships and enhance loyalty.

## Cluster 3 : Inactive Low Spender

- Conduct a reactivation campaign with enticing offers, such as significant discounts or free shipping, to encourage them to shop again.
- Request feedback to understand why they've been inactive and what improvements can be made to regain their interest.
- Send personalized messages to remind them of products they've previously purchased or items they might be interested in.



# THANK YOU

*We look forward to working  
with you*

**CONTACT ME ON**



[LinkedIn](#)



[Github](#)



[m.alwannur@gmail.com](mailto:m.alwannur@gmail.com)

