CUSTOMER SEGMENTATION

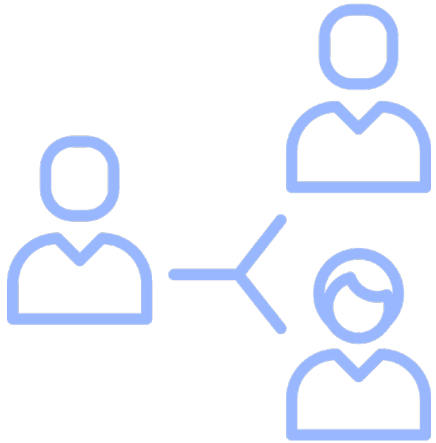
By : Ahmad Zaki Irfan

Contents

1. BACKGROUND & PROBLEM STATEMENT

Customer Segmentation is the process of dividing customers into different groups based on the same characteristics. Customer segmentation is an important analysis so that companies can adjust marketing strategies for each segment in order to increase company profits.

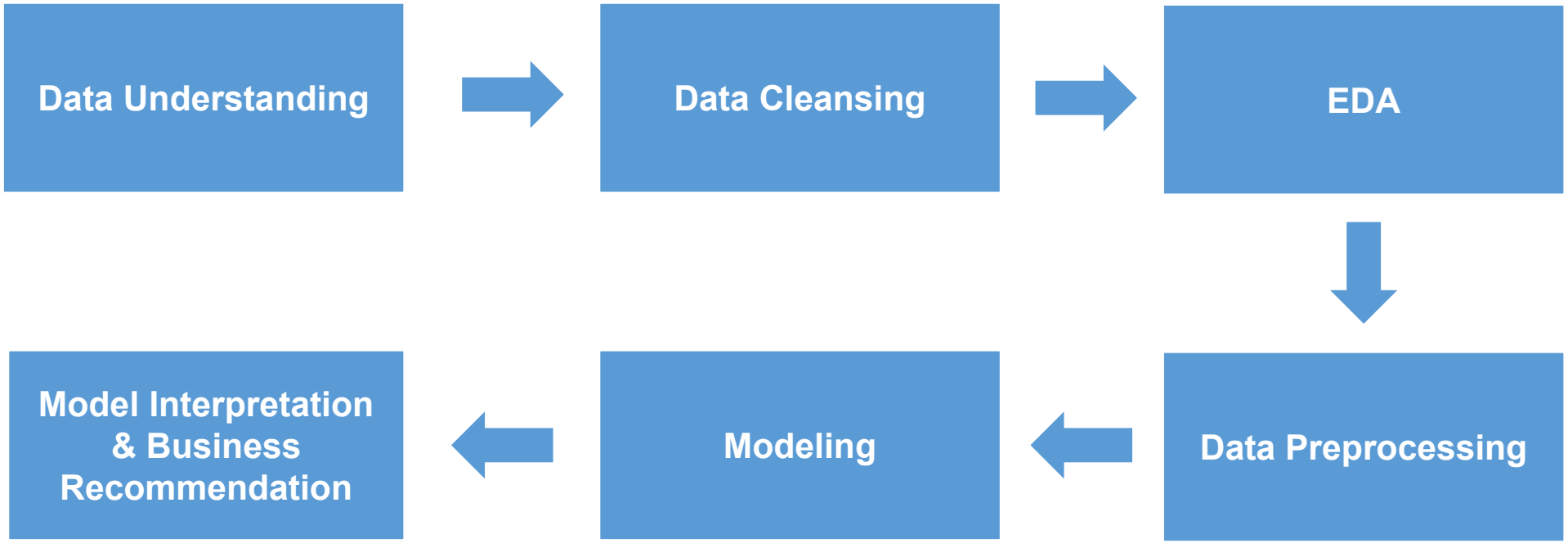
Customer Segmentation also helps to focus on the needs of each type of customer at a particular moment. Whether large or small, specific customer segments can be targeted based on a company's resources or needs.



1. BACKGROUND & PROBLEM STATEMENT

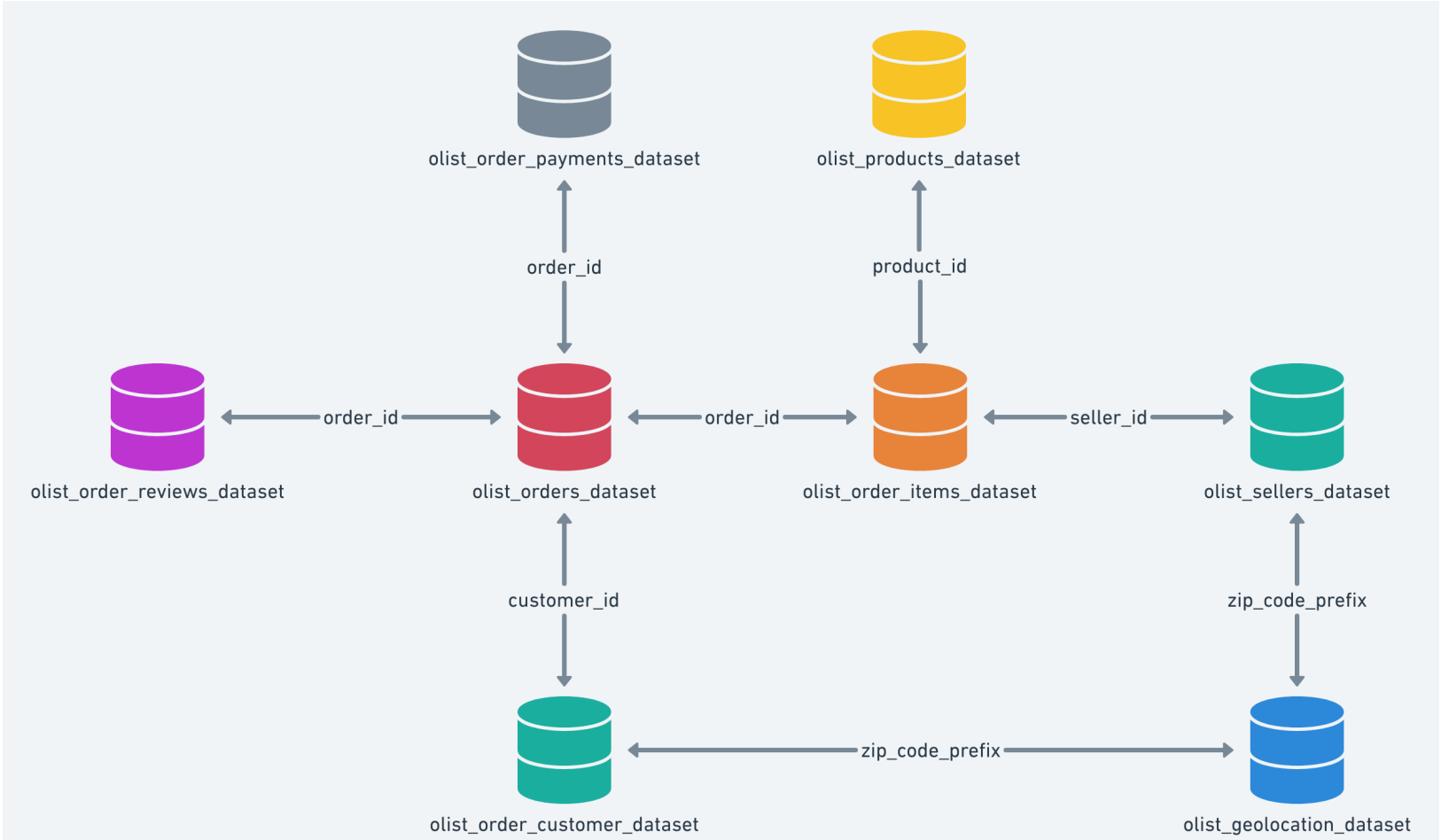
The purpose of this project is to segment customers based on Recency, Frequency and Monetary (RFM)

2. METHODOLOGY



3. DATA UNDERSTANDING & DATA CLEANSING

Data source : *https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce*
(Merged 115,609 rows, 40 columns)



olist store

3. DATA UNDERSTANDING & DATA CLEANSING

Missing Value

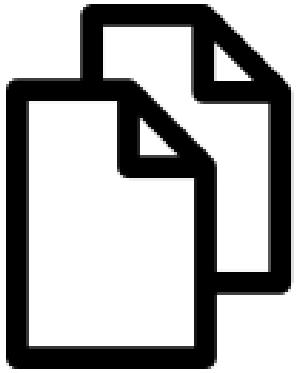
```
In [9]:
def missing_value_check (df) :
    null_number = (df.isnull().sum()/len(df))*100
    return null_number.sort_values(ascending = False)
missing_value_check(df_merged)
```

```
Out[9]:
review_comment_title            88.062348
review_comment_message          57.697065
order_delivered_customer_date    2.075963
order_delivered_carrier_date     1.033657
order_approved_at                0.012110
product_height_cm                0.000865
product_weight_g                 0.000865
product_length_cm                0.000865
product_width_cm                 0.000865
customer_id                      0.000000
product_name_lenght              0.000000
```

Drop Feature

Drop Missing Value

3. DATA UNDERSTANDING & DATA CLEANSING

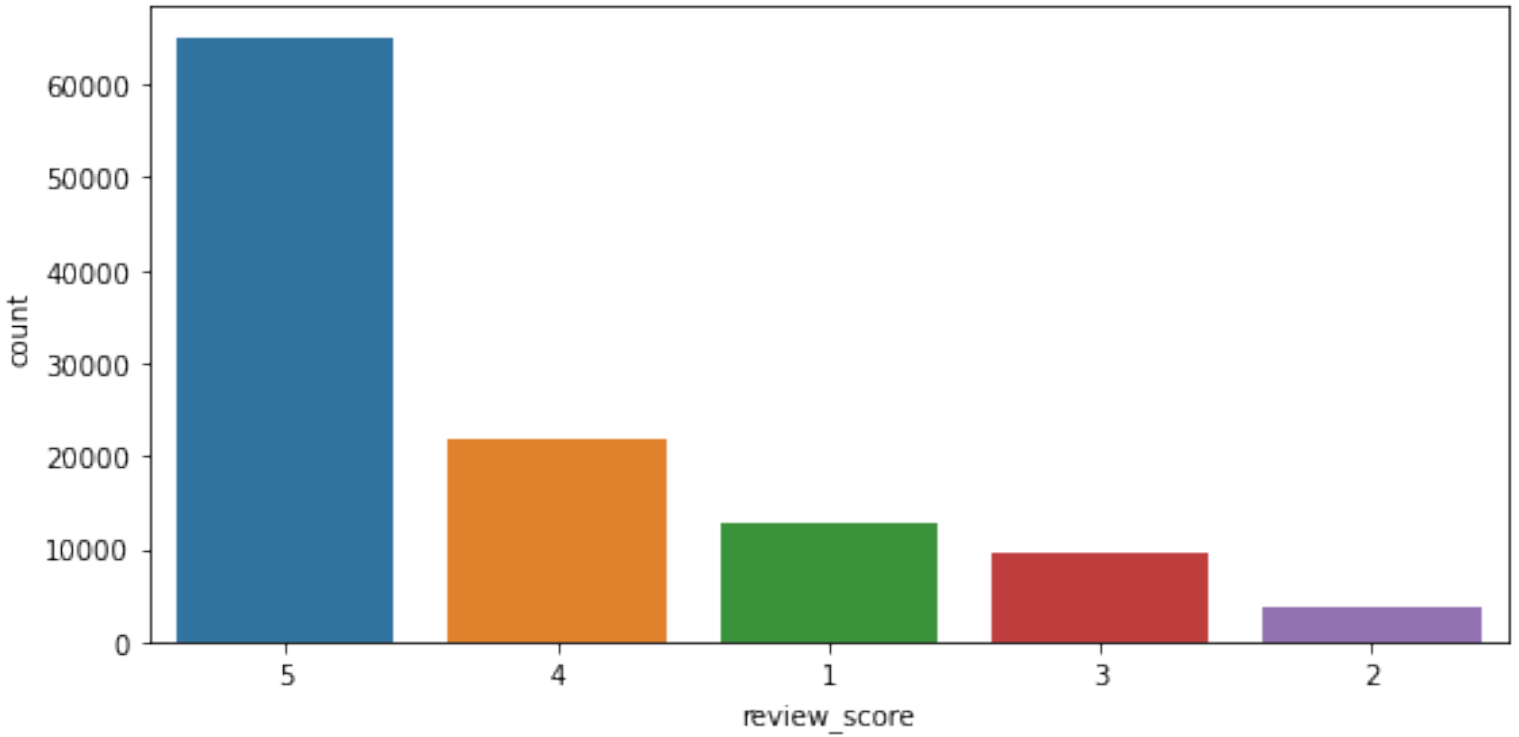


Duplicated Value

```
In [11]:  1  df_n.duplicated().sum()

Out[11]: 0
```

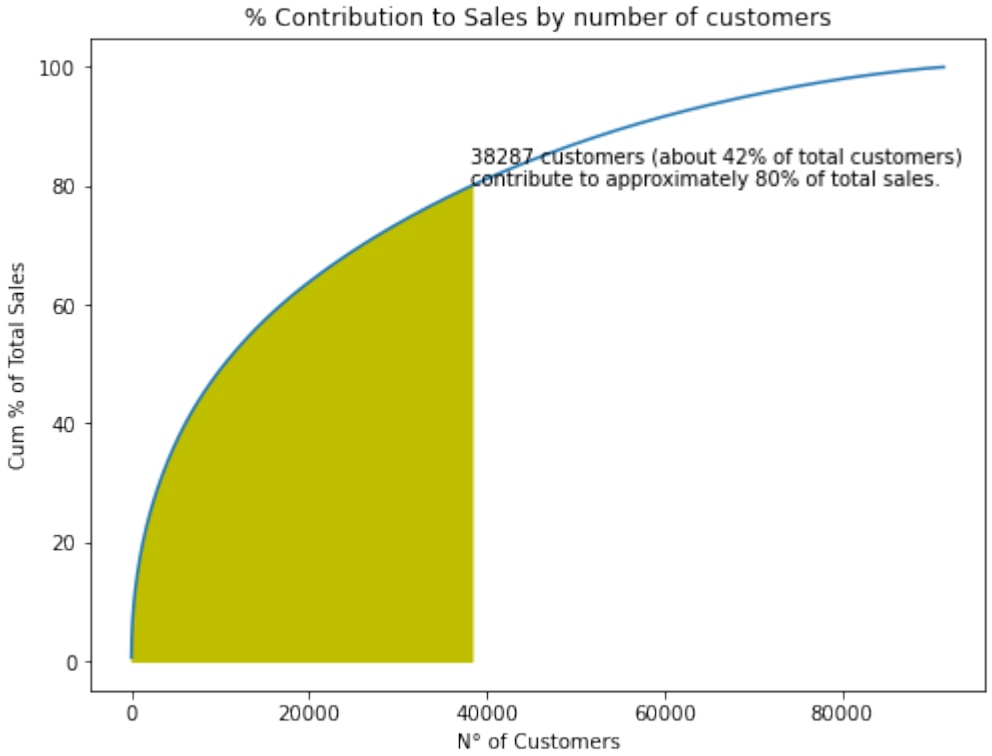
There are no duplicated value on dataset

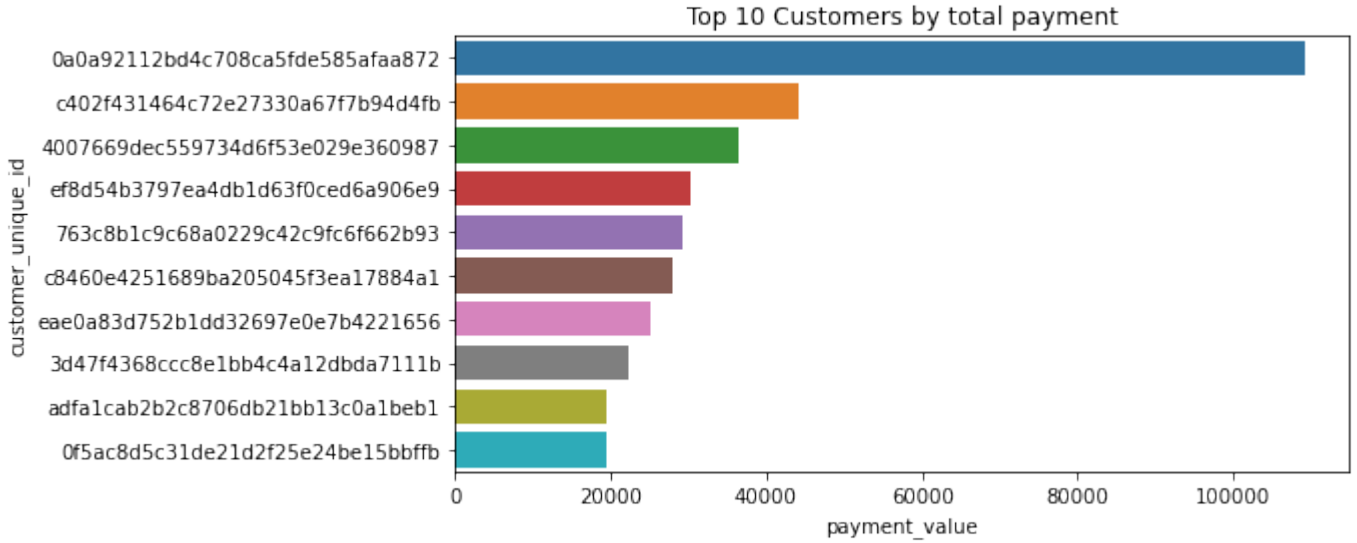
4. EXPLORATORY DATA ANALYSIS



76.7% of products sold have good reviews, but 23.3% of products get bad reviews. Need further investigation to find out the cause of the bad review, is it because of product quality? seller service? Or delivery?

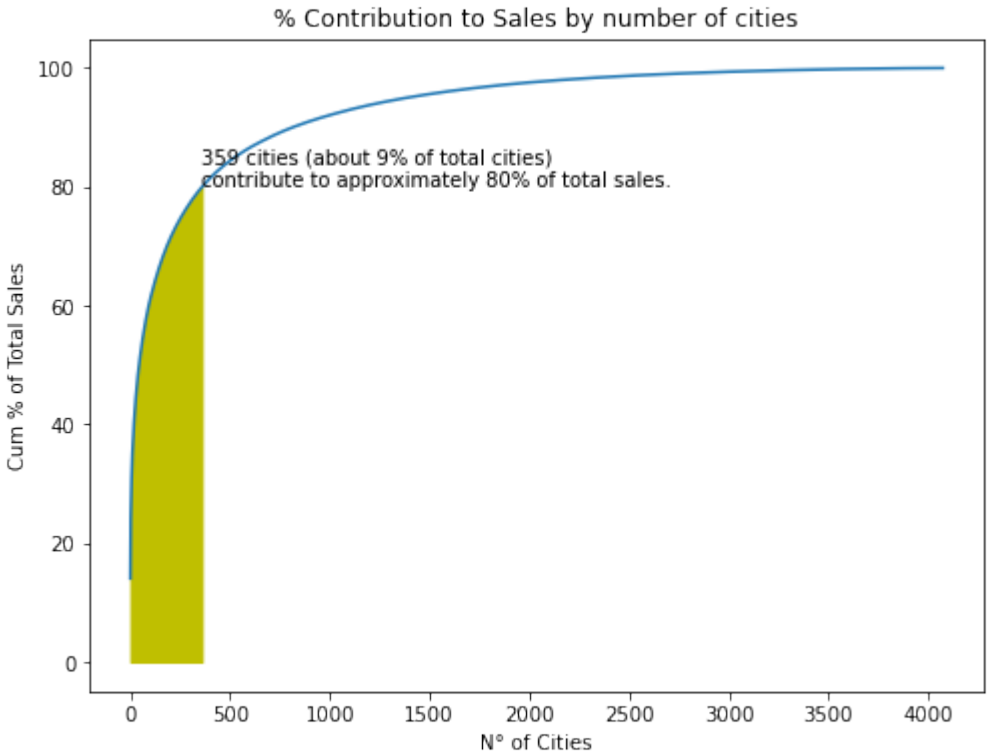
4. EXPLORATORY DATA ANALYSIS

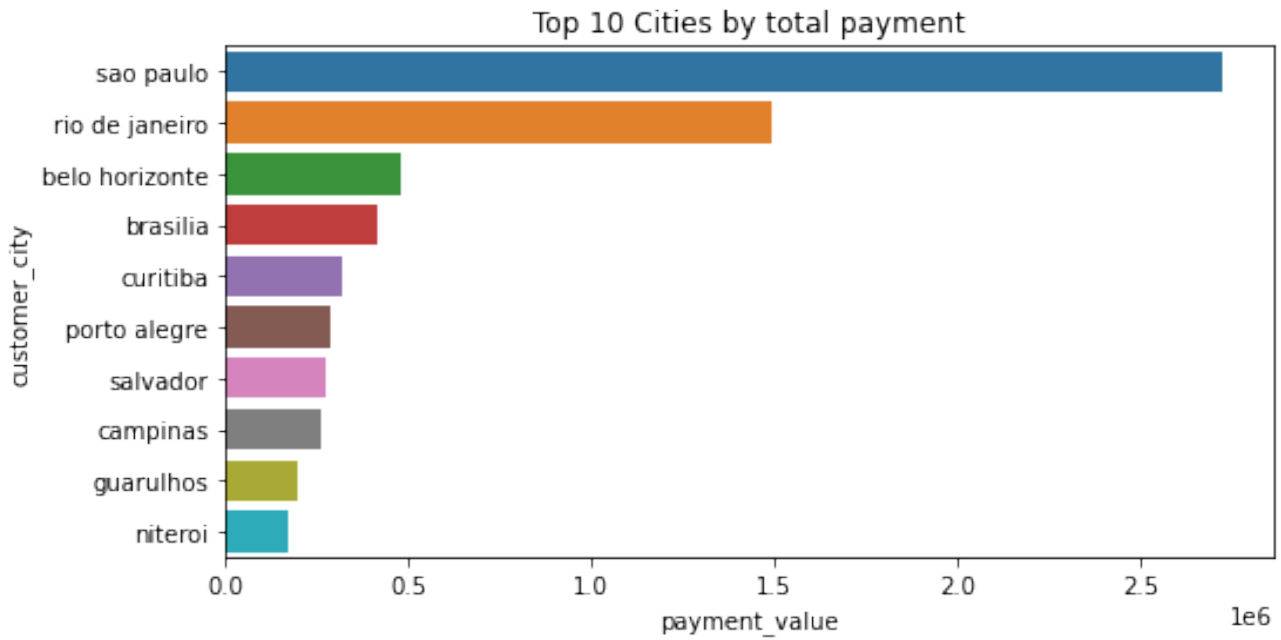




42% of customers affect 80% of sales, where the Top 10 customers have at least made sales of 19000 Brazilian Real.

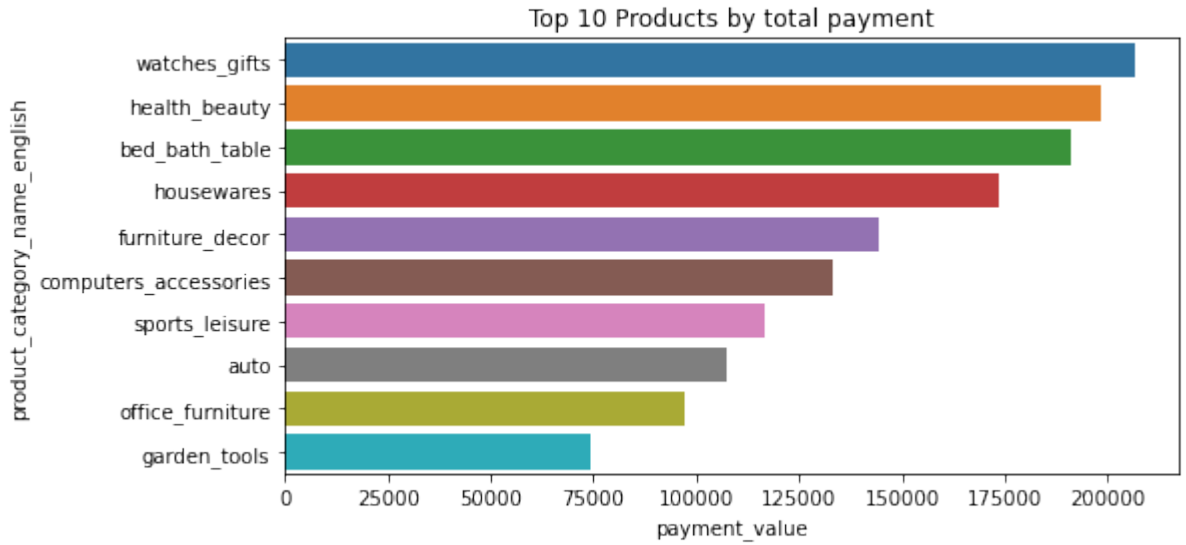
4. EXPLORATORY DATA ANALYSIS





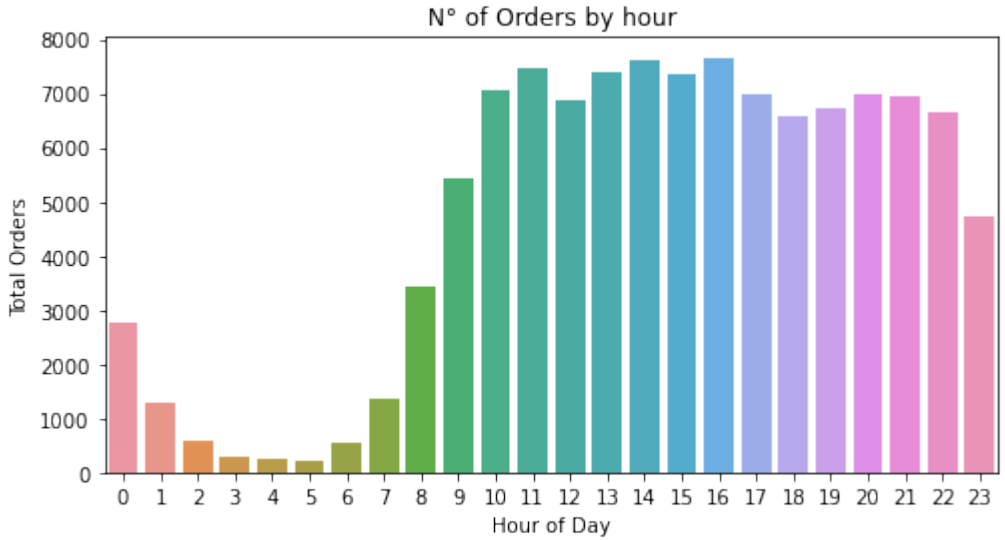
9% of cities affect 80% of sales, where the Top 10 cities have at least given sales of 200,000 Brazilian Real. Apart from the cities with the highest number of transactions, Sao Paolo and Rio de Janeiro are also the cities with the largest sale

4. EXPLORATORY DATA ANALYSIS

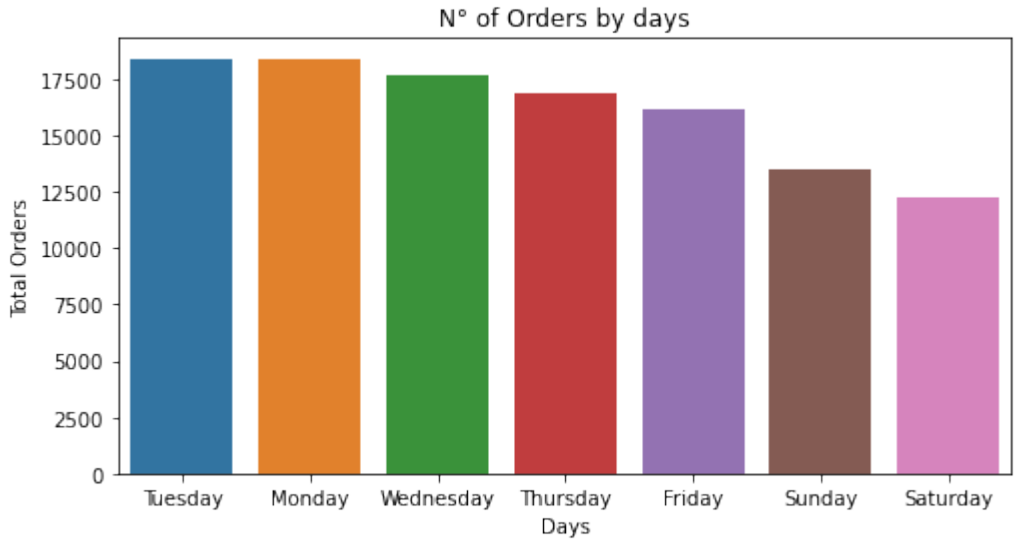


21% of product categories affect 80% of sales, where the Top 10 product categories have at least given sales of 700,000 Brazilian Real. Apart from product categoris with the highest number of transactions, bed_bath_table and healthy_beauty are also product categories with the largest sales

4. EXPLORATORY DATA ANALYSIS

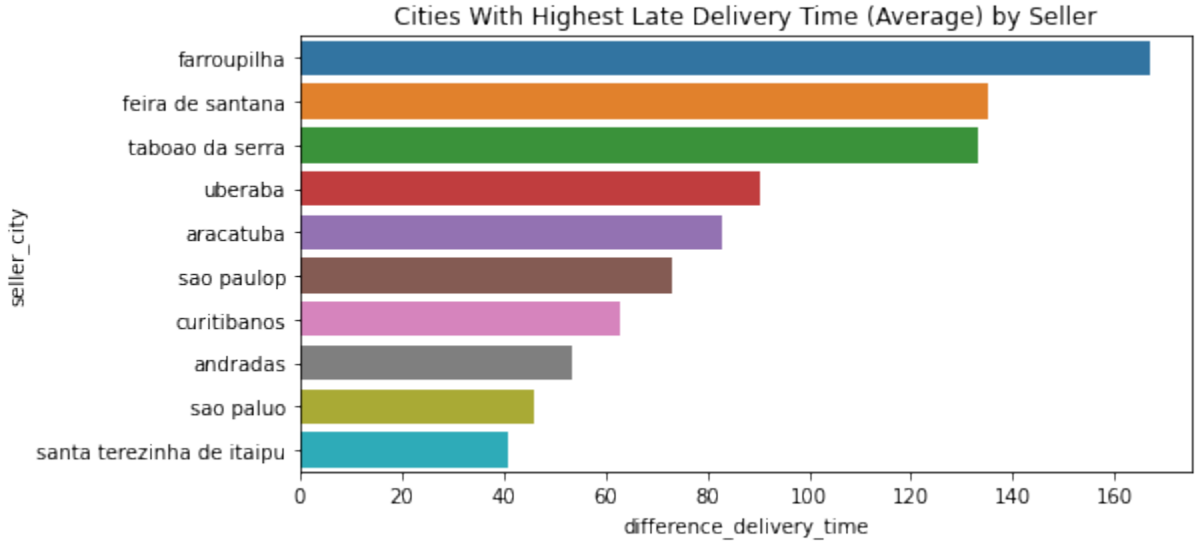


Starting at 7.00, transaction activity has started
and 16.00 is the time with the highest number of transactions.

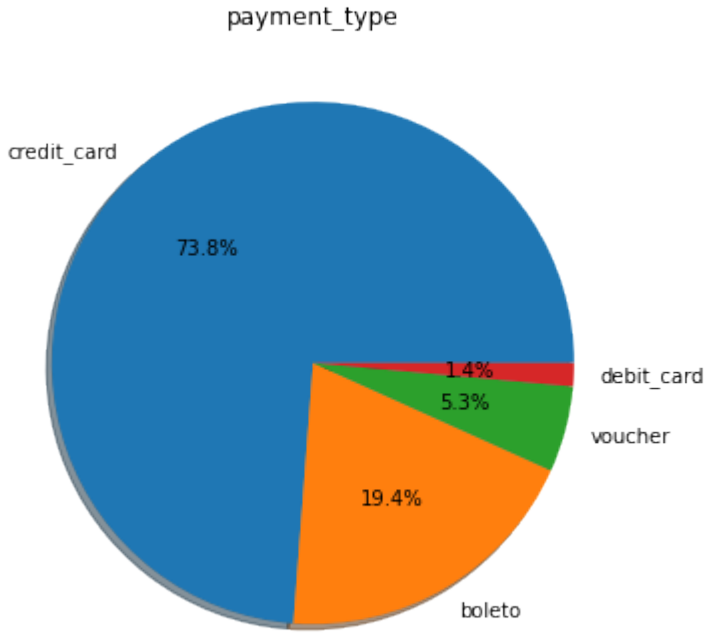


Tuesday is the day with the highest number of transactions and
tends to decrease towards the weekend.

4. EXPLORATORY DATA ANALYSIS



Farroupilha is a seller city with the highest number of late delivery of goods, companies need to evaluate goods delivery from the top 10 cities with the highest rate of late delivery of goods

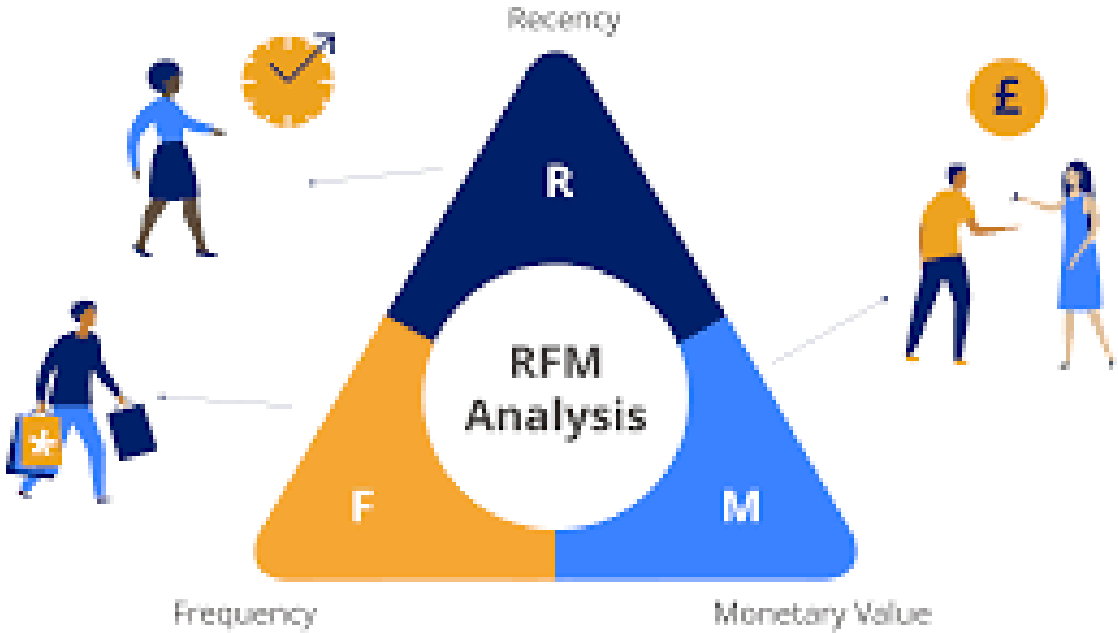


Payment for goods is dominated by using a credit card by 73.8%

5. DATA PREPROCESSING & MODELING

RFM ANALYSIS

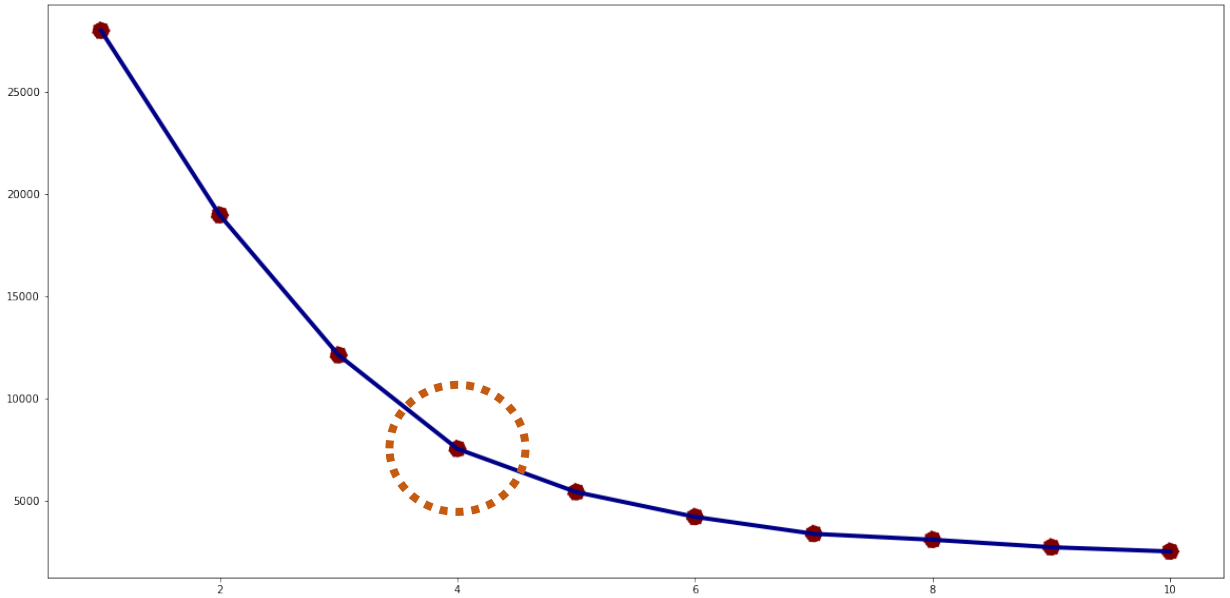
	count	mean	std	min	25%	50%	75%	max
Recency	91459.0	237.108311	152.573539	1.00	114.00	218.00	345.00	695.00
Frequency	91459.0	1.032867	0.206215	1.00	1.00	1.00	1.00	14.00
Monetary	91459.0	212.307969	631.327848	9.59	63.83	112.84	202.76	109312.64



Recency: How recently has the customer made a transaction with us
Frequency: How frequent is the customer in ordering/buying some product from us
Monetary: How much does the customer spend on purchasing products from us.

5. DATA PREPROCESSING & MODELING

K MEANS CLUSTERING

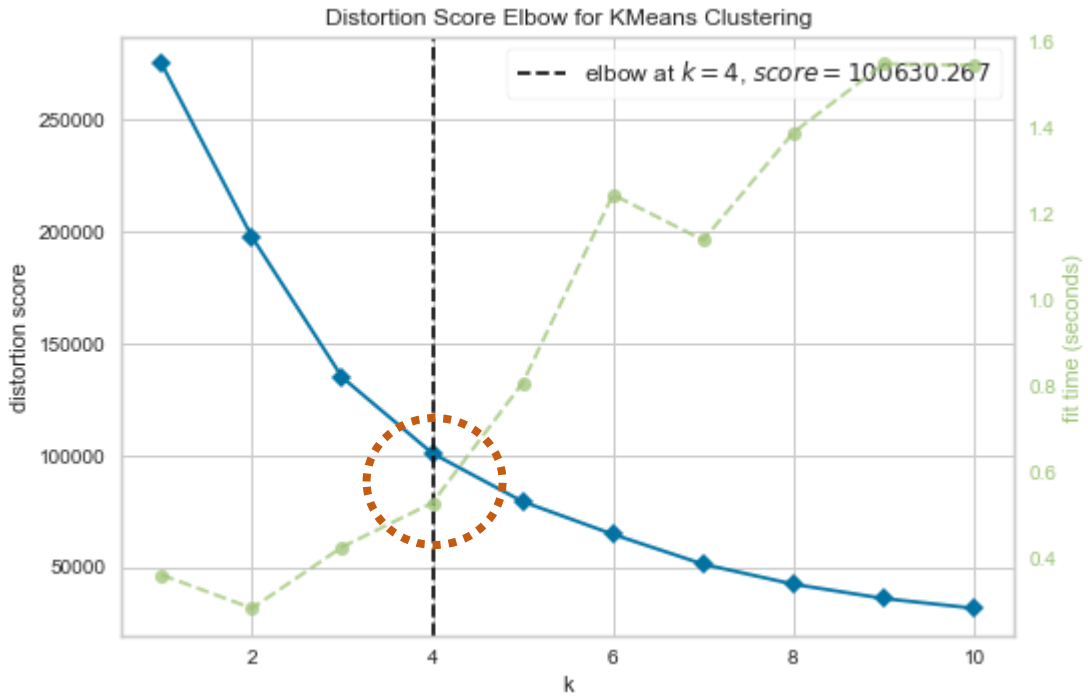


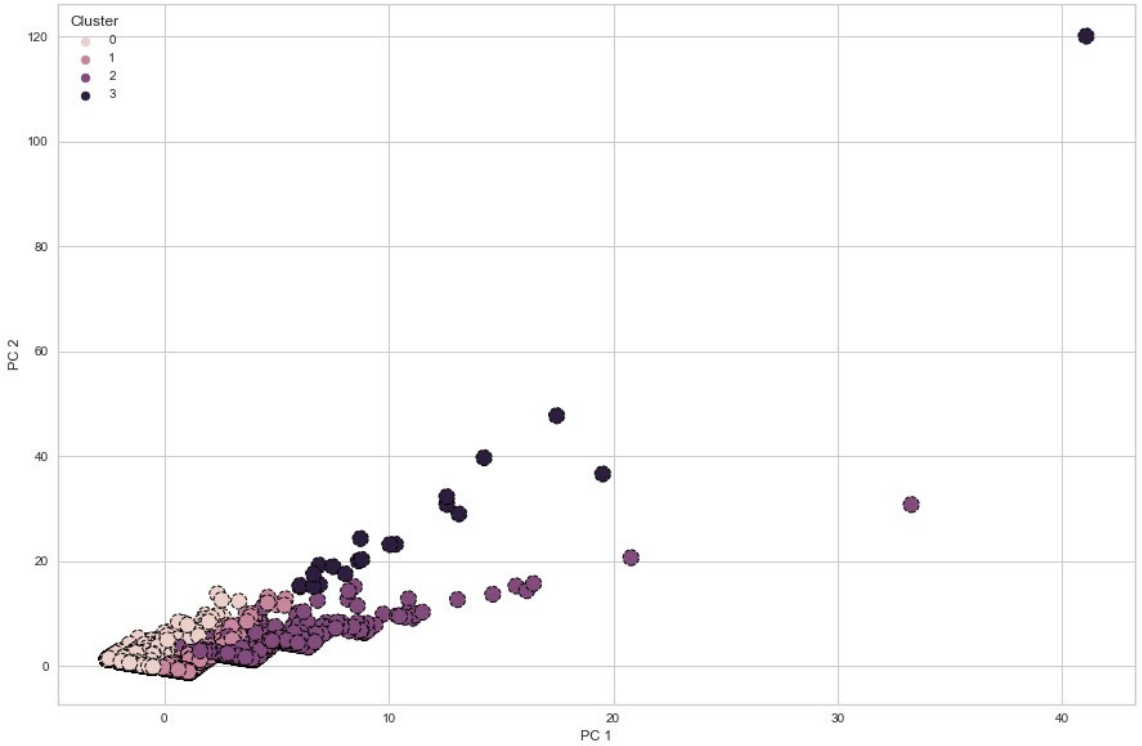

```python
inertia = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, random_state=42)
    kmeans.fit(data_std)
    inertia.append(kmeans.inertia_)


plt.figure(figsize=(20, 10))

# plt.plot(inertia)
sns.lineplot(x=range(1, 11), y=inertia, color='#000087', linewidth = 4)
sns.scatterplot(x=range(1, 11), y=inertia, s=300, color='#800000',  linestyle='--')
```

Based on the picture above, the selection of the initial number of clusters chosen is 4 clusters but it needs to be retested using KElbow

5. DATA PREPROCESSING & MODELING

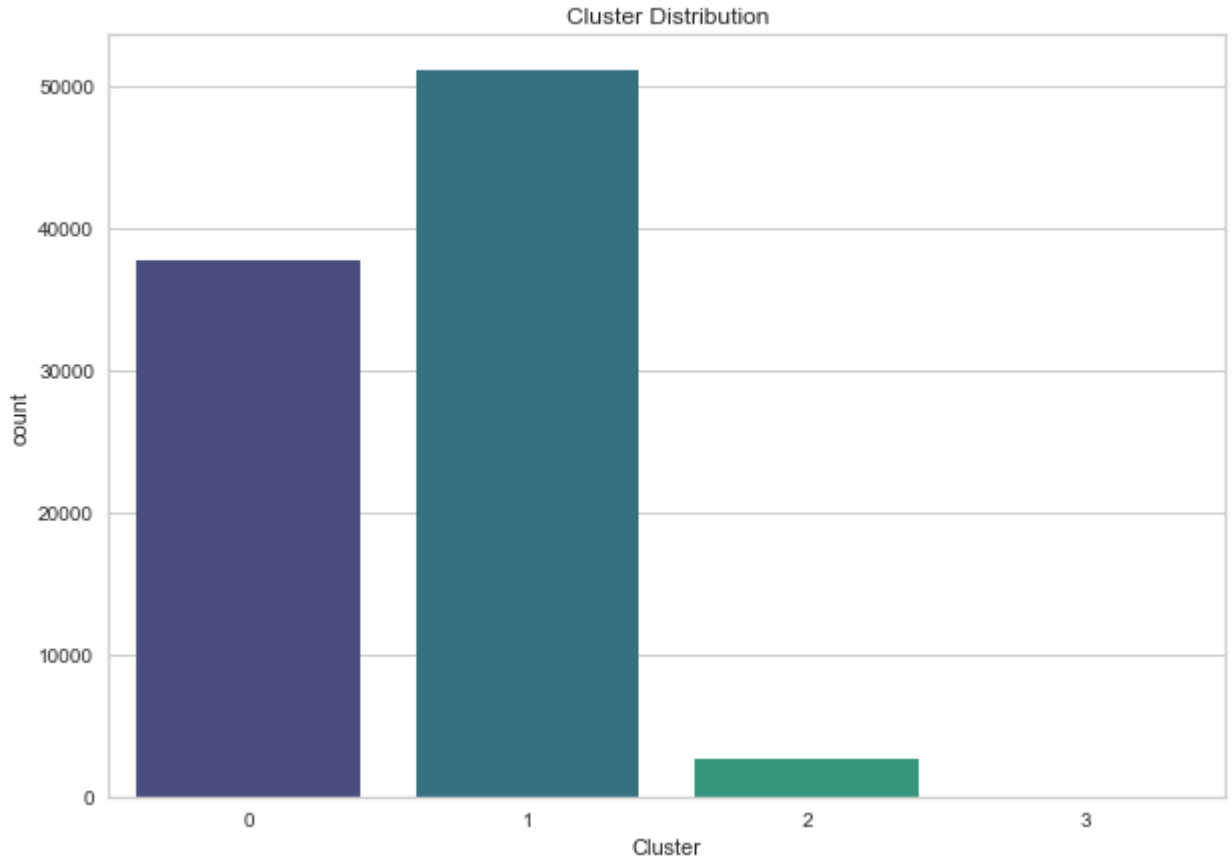
K MEANS CLUSTERING





Based on the KElbow image above, selecting the right number of clusters is 4 clusters according to the initial selection.

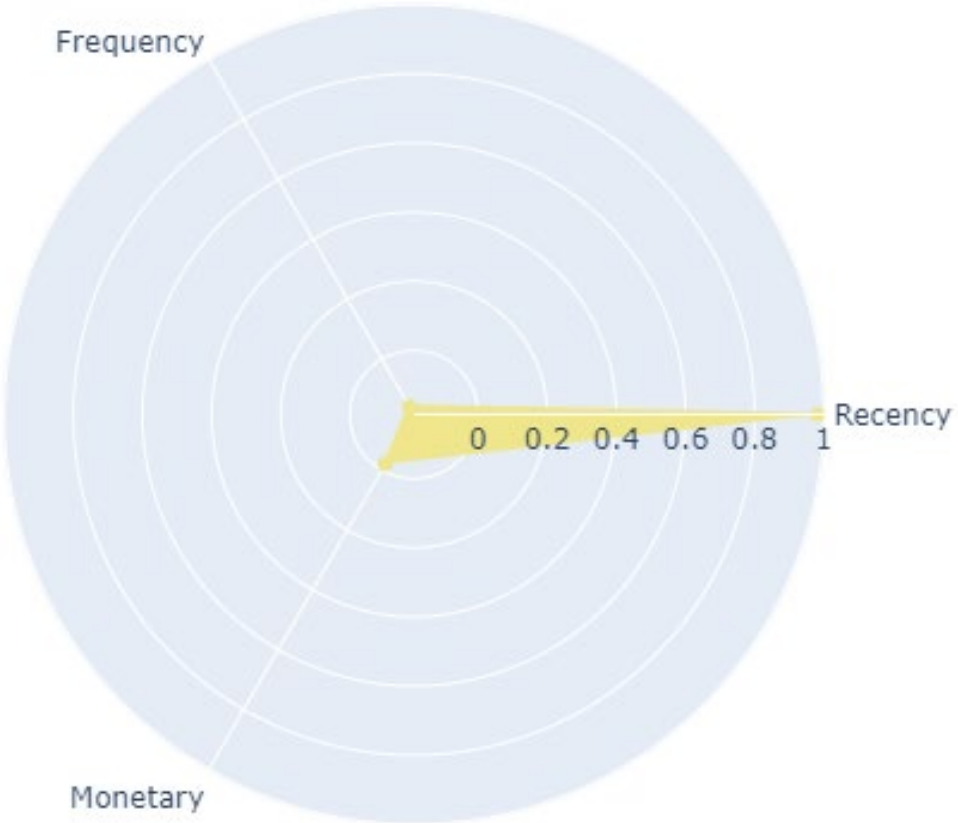
6. MODEL INTERPRETATION & RECOMMENDATION



The composition of customers is dominated by clusters 1 and 3, while clusters 0 and 2 are few.

6. MODEL INTERPRETATION & RECOMMENDATION

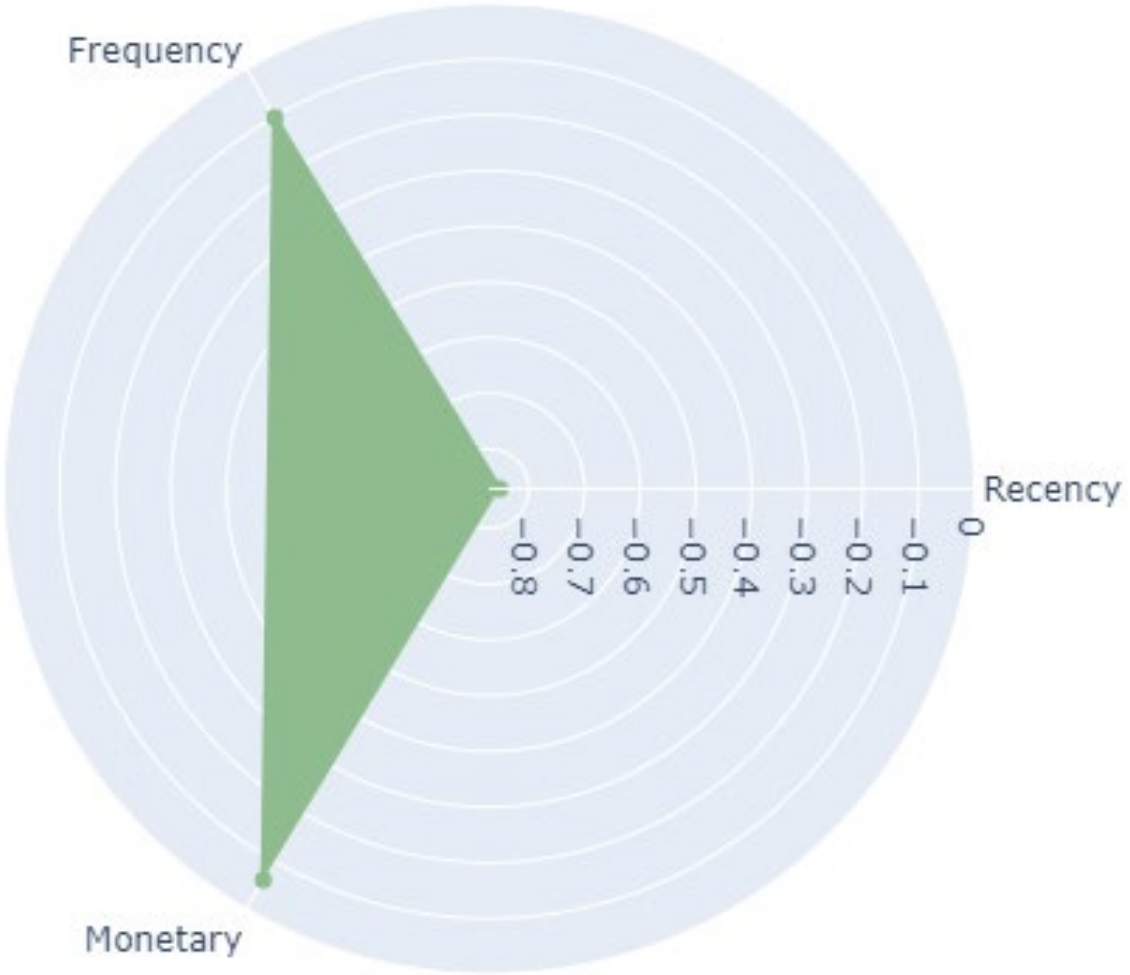
CLUSTER 0



This cluster is a lost customer with a high recency value and a small amount spent. Company can send promos via email for specific products with purpose they will visit and spend more money

6. MODEL INTERPRETATION & RECOMMENDATION

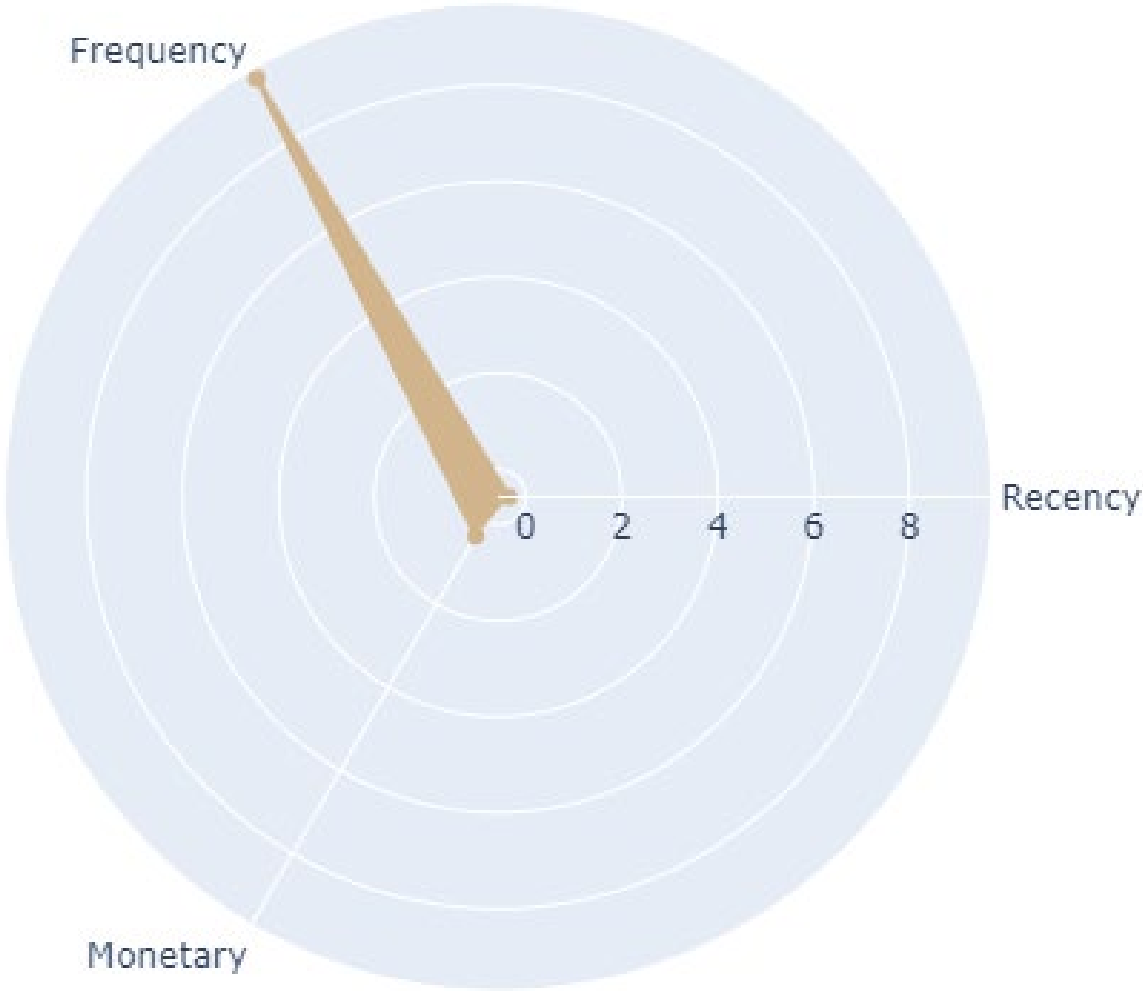
CLUSTER 1



This cluster is a new customer because the amount spent is low but has the smallest recency value, as the company has only been running for 2 years, this cluster dominates. company must start a good relation ship with a good corporate image.

6. MODEL INTERPRETATION & RECOMMENDATION

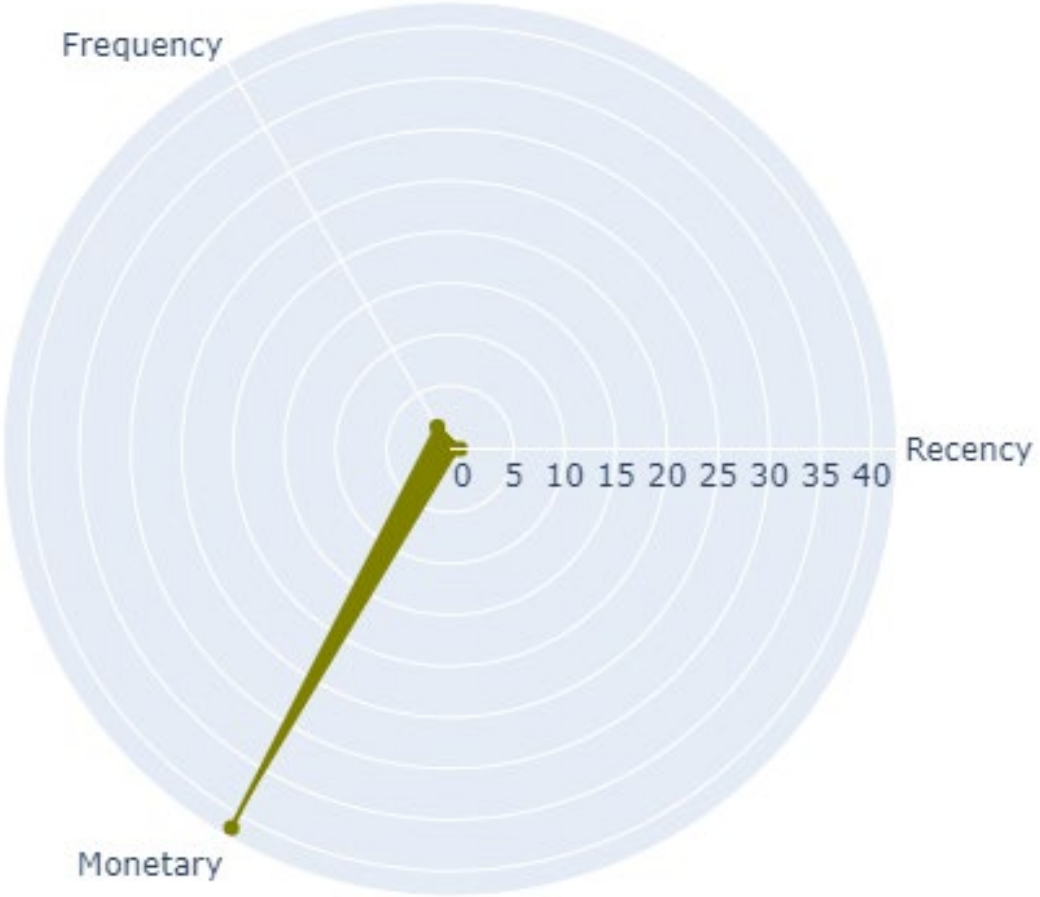
CLUSTER 2



This cluster is a loyal customer with the highest frequency compared to other clusters with the second largest spending and recency values that are still relatively low. company can offer products or services that have greater value and company need reviews from these customers so company can provide feedback.

6. MODEL INTERPRETATION & RECOMMENDATION

CLUSTER 3



This cluster is a promising customer with a large amount of spending compared to other clusters. Company can provide treatment in the form of discount offers and increase engagement by offering personal recommendations.

Thanks !