

CUSTOMER SEGMENTATION

By : Ahmad Zaki Irfan



Contents

01.

**BACKGROUND &
PROBLEM
STATEMENT**

02.

METHODOLOGY

03.

**DATA
UNDERSTANDING &
DATA CLEANSING**

04.

**EXPLORATORY
DATA ANALYSIS**

05.

**DATA
PREPROCESSING &
MODELING**

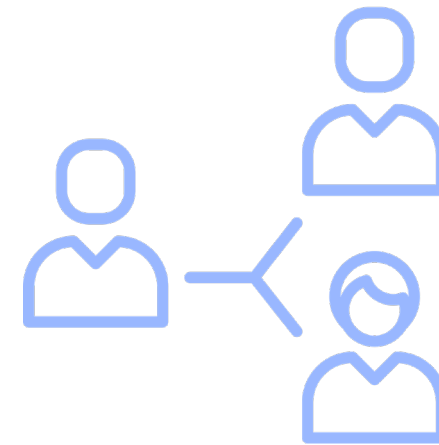
06.

**MODEL
INTERPRETATION &
BUSINESS
RECOMMENDATION**

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.

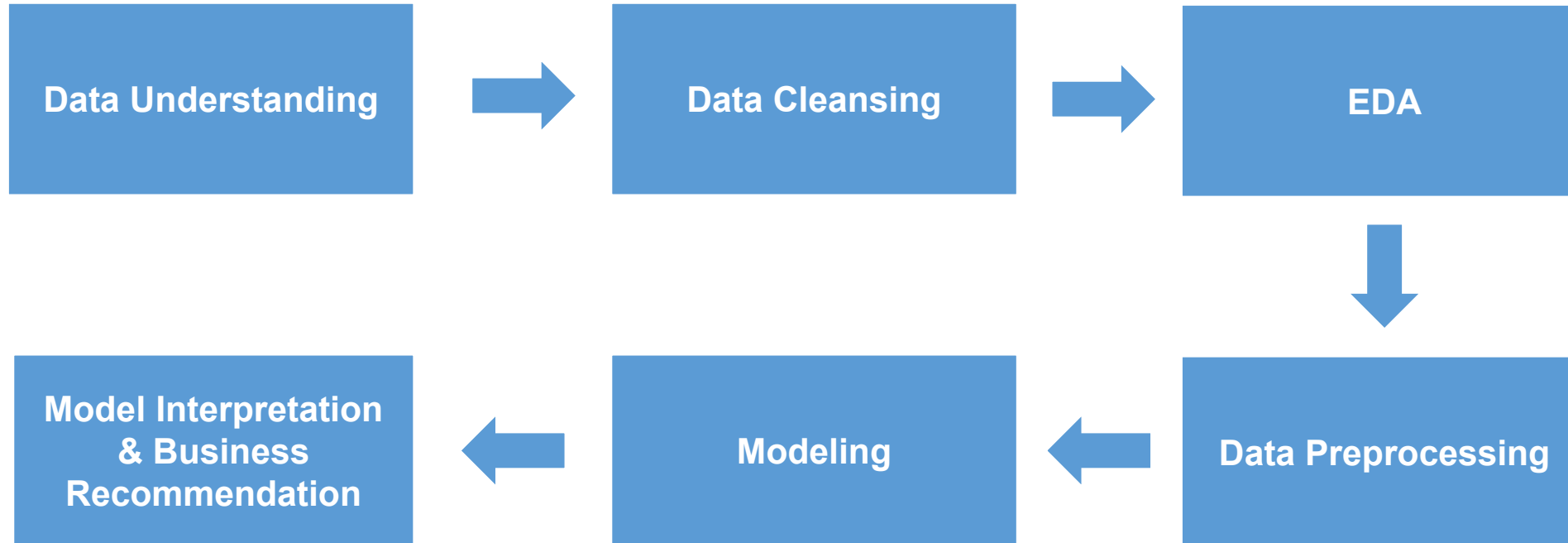


1. BACKGROUND & PROBLEM STATEMENT



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

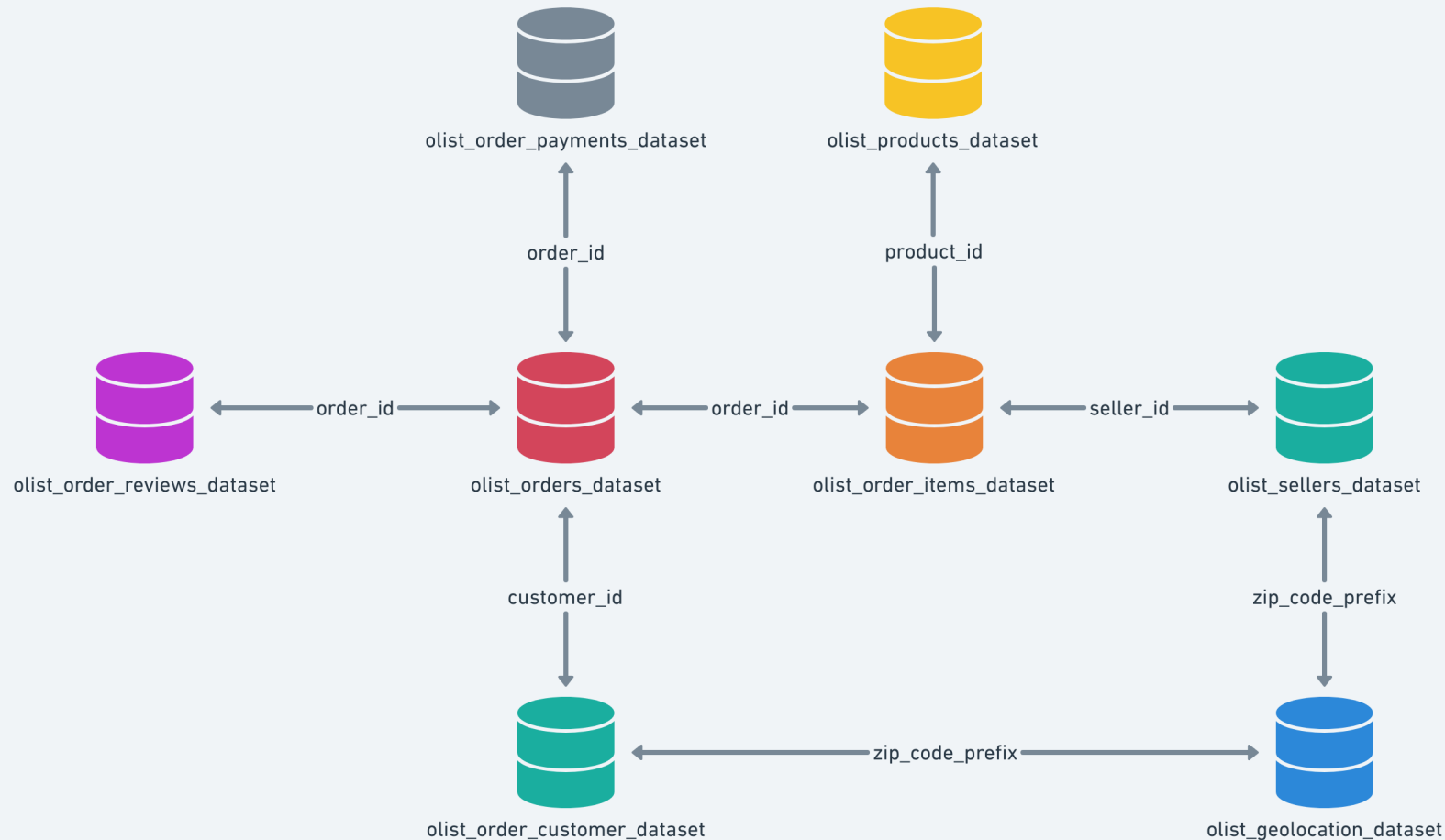
2. METHODOLOGY



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

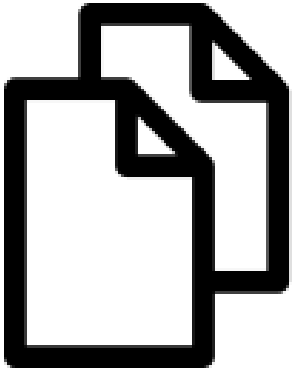
```
In [9]: 1 def missing_value_check (df) :  
2         null_number = (df.isnull().sum()/len(df))*100  
3         return null_number.sort_values(ascending = False)  
4 missing_value_check(df_merged)
```

```
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

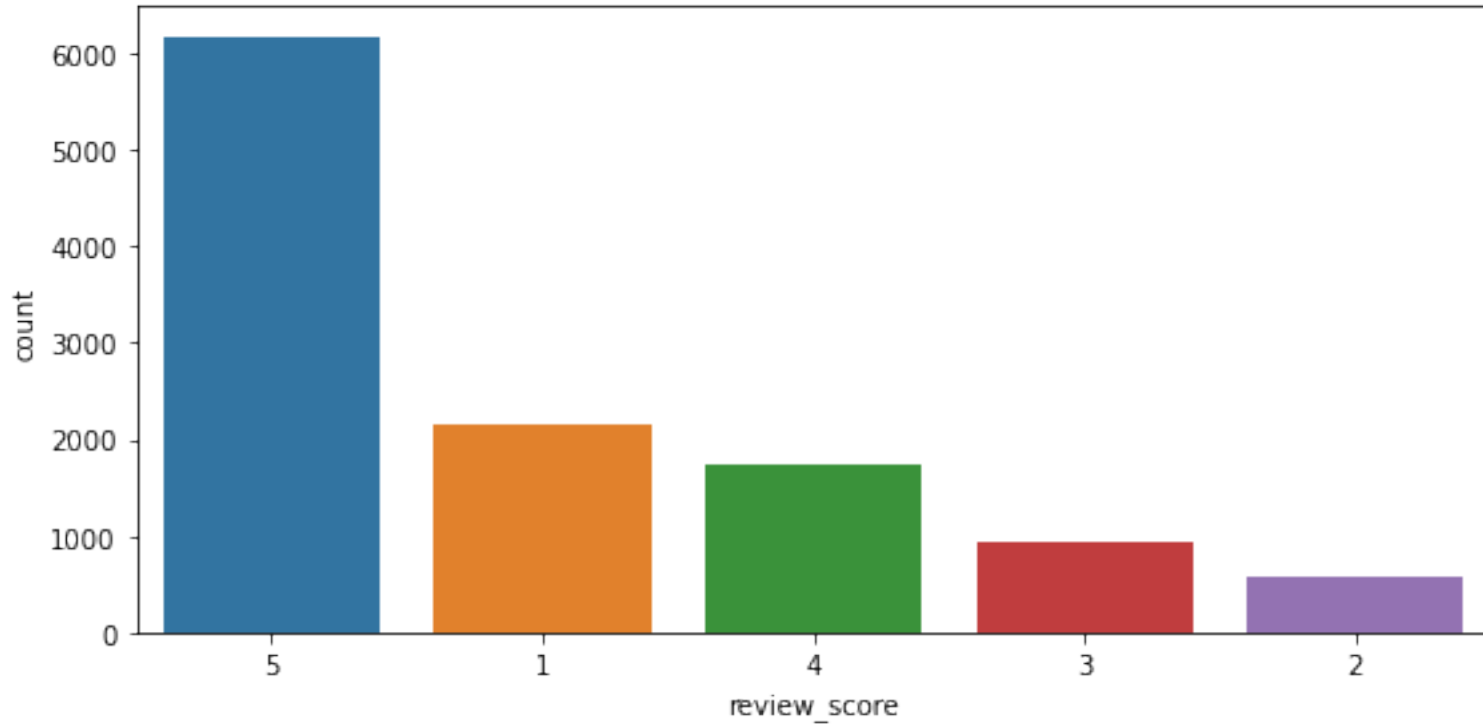


Duplicated Value

```
In [11]: 1 df_n.duplicated().sum()  
Out[11]: 0
```

There are no duplicated value on dataset

4. EXPLORATORY DATA ANALYSIS

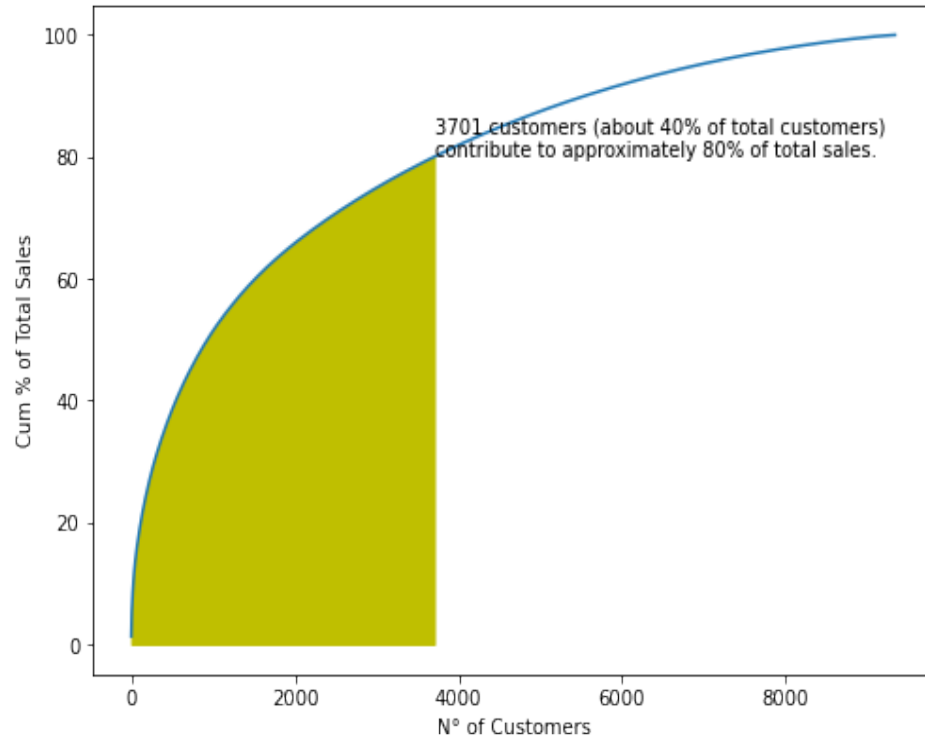


68% of products sold have good reviews, but 32% of products get bad reviews. Need further investigation to find out the cause of the bad review, is it because of product quality? seller service? Or delivery?

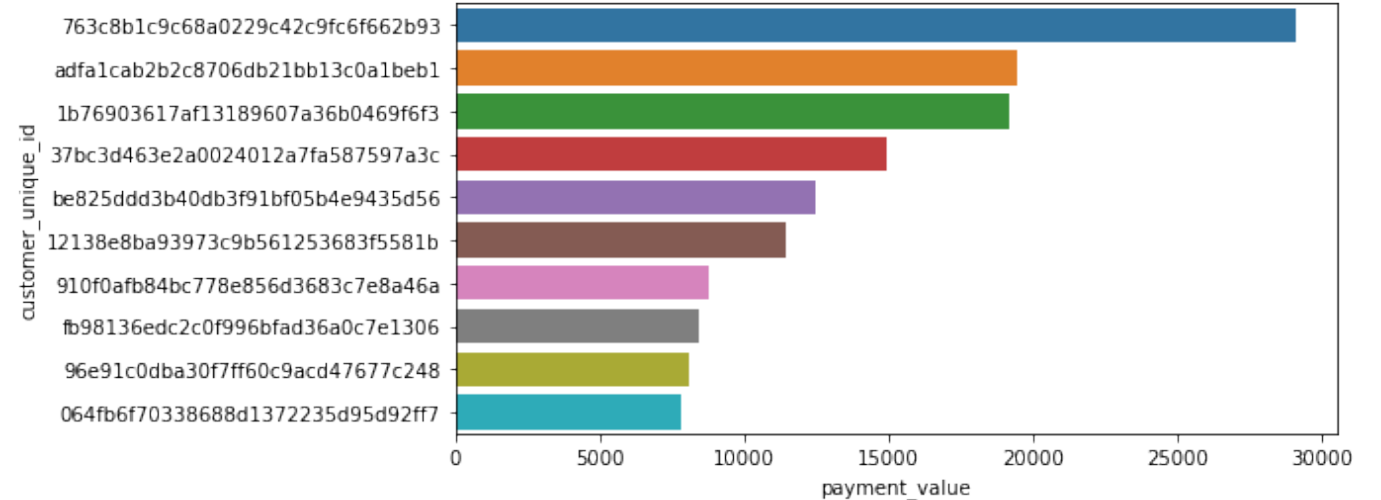


4. EXPLORATORY DATA ANALYSIS

% Contribution to Sales by number of customers

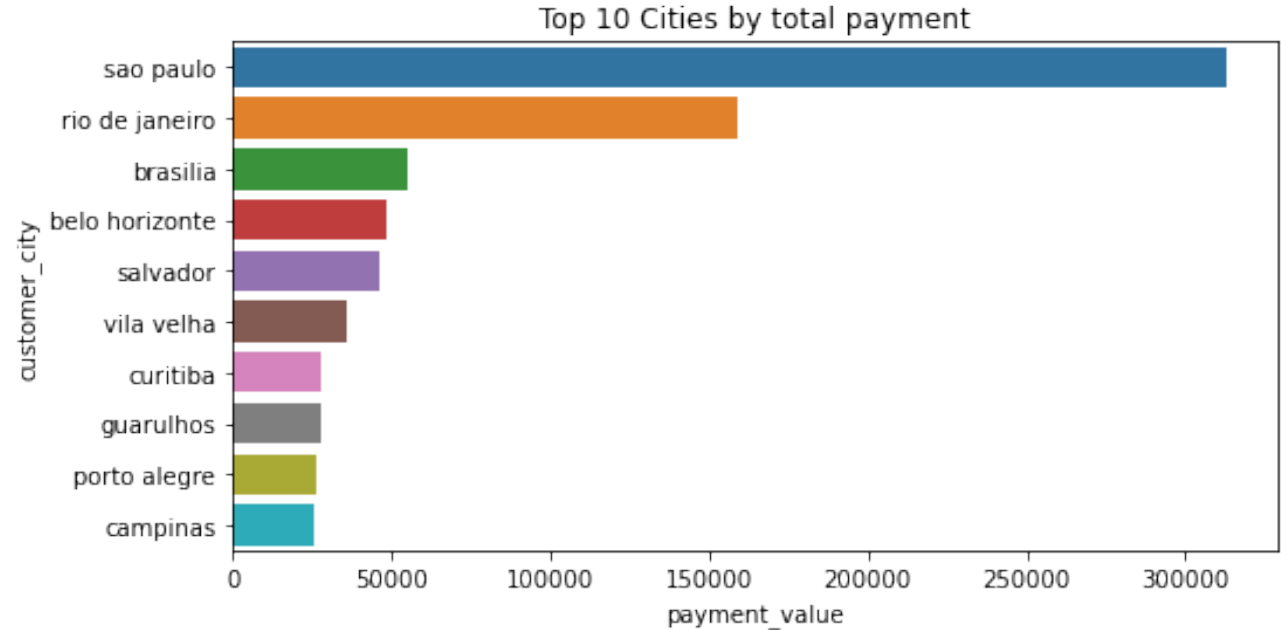
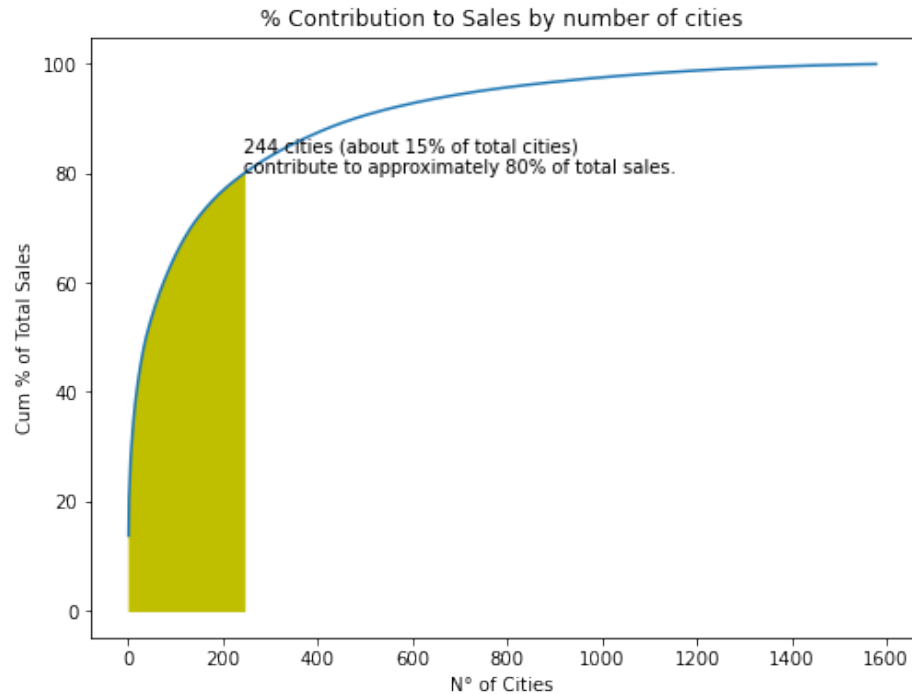


Top 10 Customers by total payment



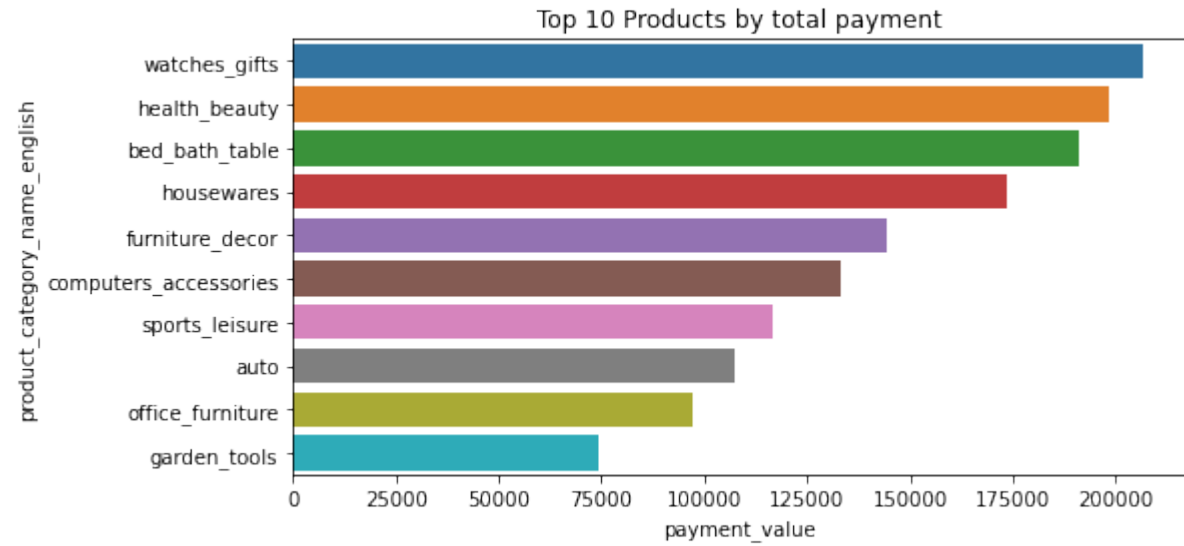
40% of customers affect 80% of sales, where the Top 10 customers have at least made sales of 6000 Brazilian Real.

4. EXPLORATORY DATA ANALYSIS



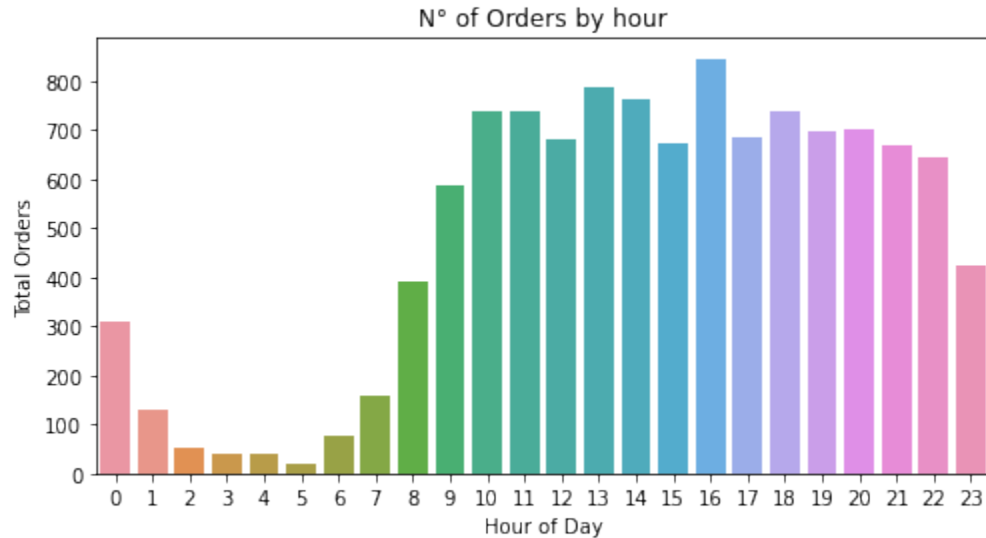
15% of cities affect 80% of sales, where the Top 10 cities have at least given sales of 25,000 Brazilian Real. Apart from the cities with the highest number of transactions, Sao Paolo and Rio de Janeiro are also the cities with the largest sale

4. EXPLORATORY DATA ANALYSIS

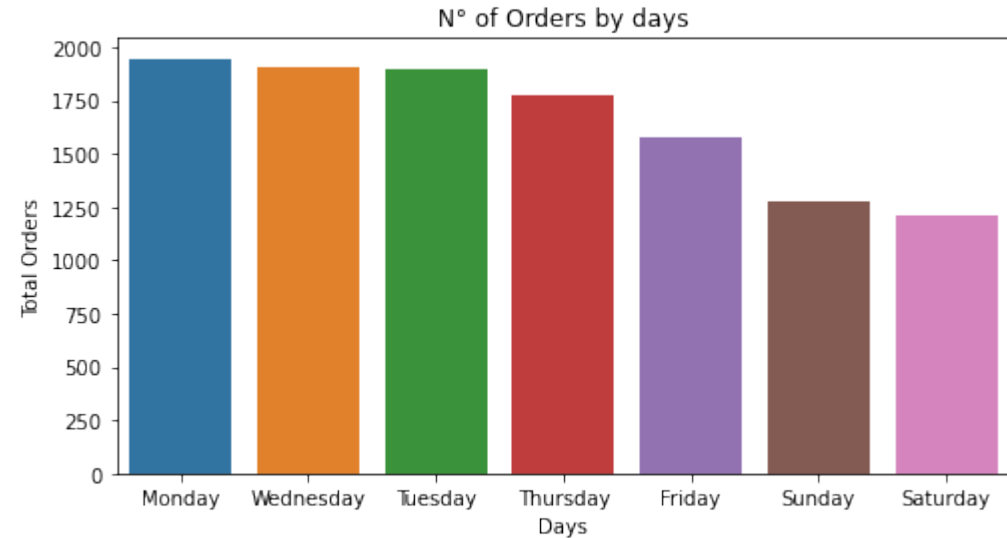


24% of product categories affect 80% of sales, where the Top 10 product categories have at least given sales of 75,000 Brazilian Real. Apart from product categories with the highest number of transactions, bed_bath_table, health_beauty watches_gifts and houseware are also product categories with the largest sales

4. EXPLORATORY DATA ANALYSIS

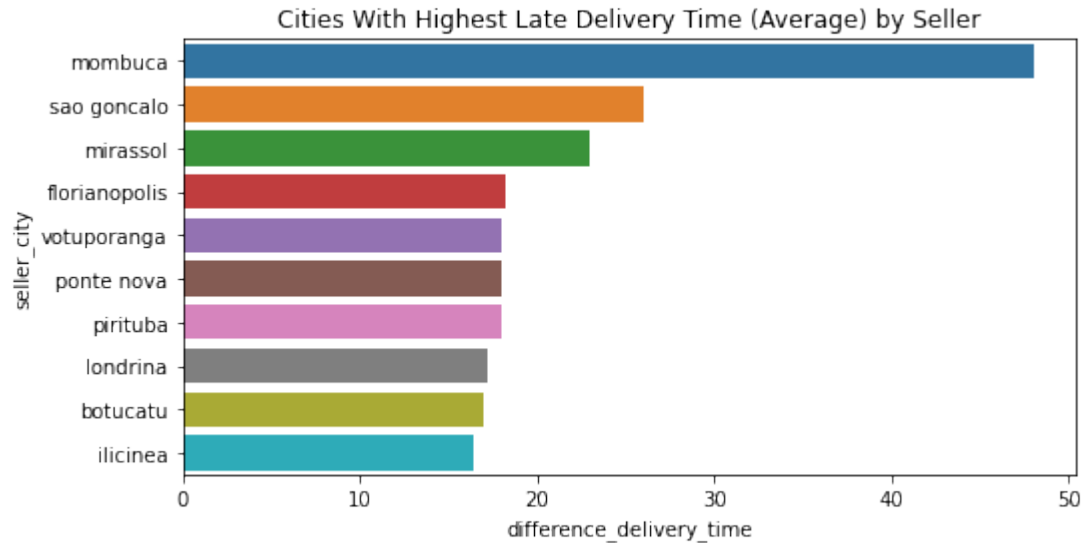


Starting at 7.00, transaction activity has started and 16.00 is the time with the highest number of transactions.

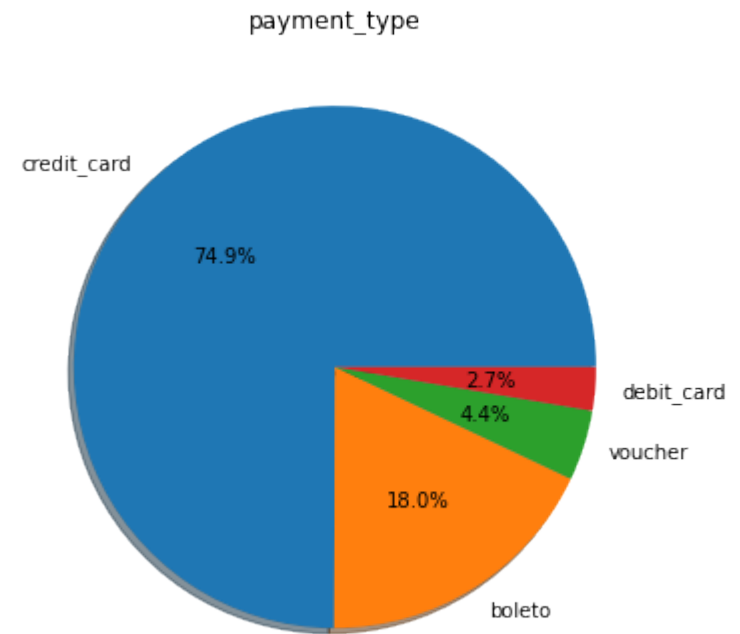


Monday is the day with the highest number of transactions and tends to decrease towards the weekend.

4. EXPLORATORY DATA ANALYSIS



Mombuca 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

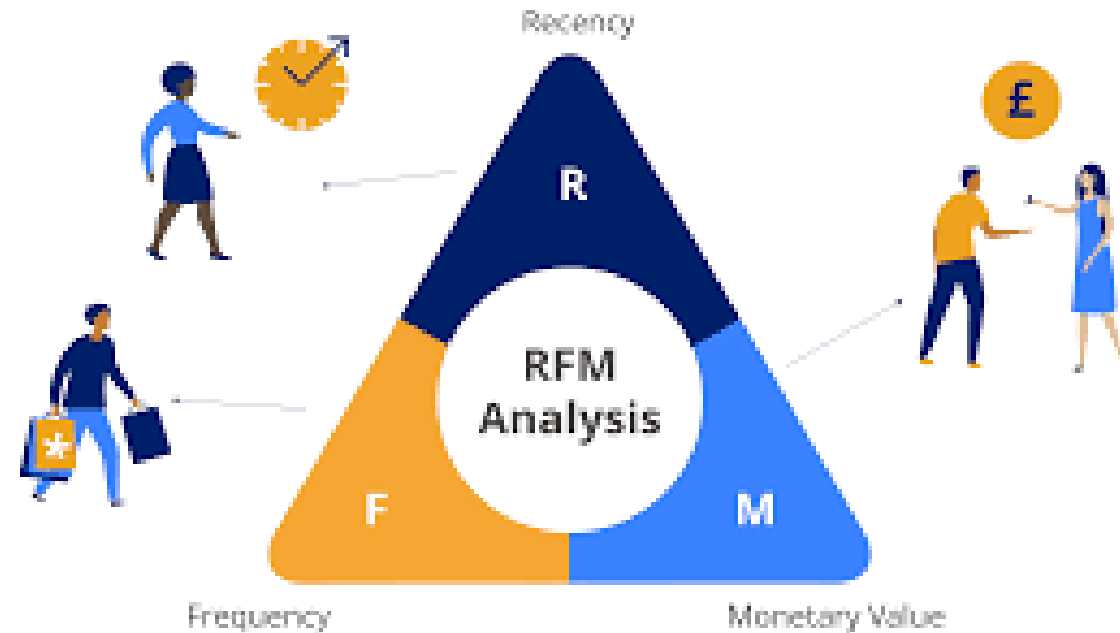


Payment for goods is dominated by using a credit card by 74.9%

5. DATA PREPROCESSING & MODELING

RFM ANALYSIS

	count	mean	std	min	25%	50%	75%	max
Recency	9330.0	73.867738	42.529298	1.00	36.00	74.000	111.0000	482.00
Frequency	9330.0	1.011040	0.108520	1.00	1.00	1.000	1.0000	3.00
Monetary	9330.0	245.632057	641.703657	13.89	68.18	122.415	217.5325	29099.52



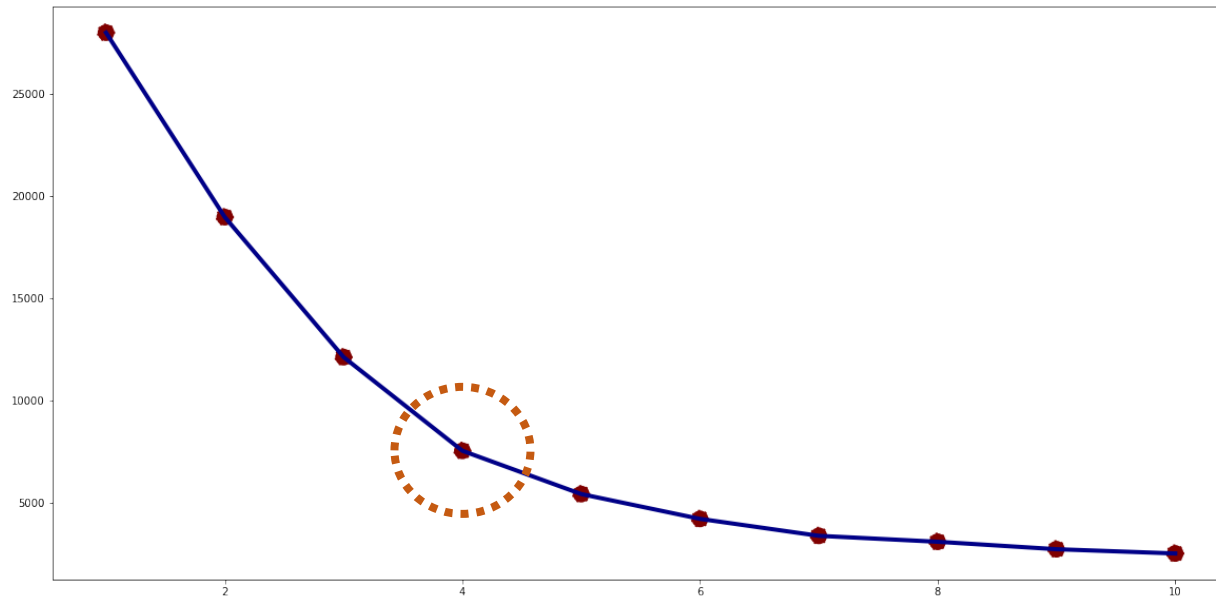
Recency: How **recently** has the customer made a **transaction** with us

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

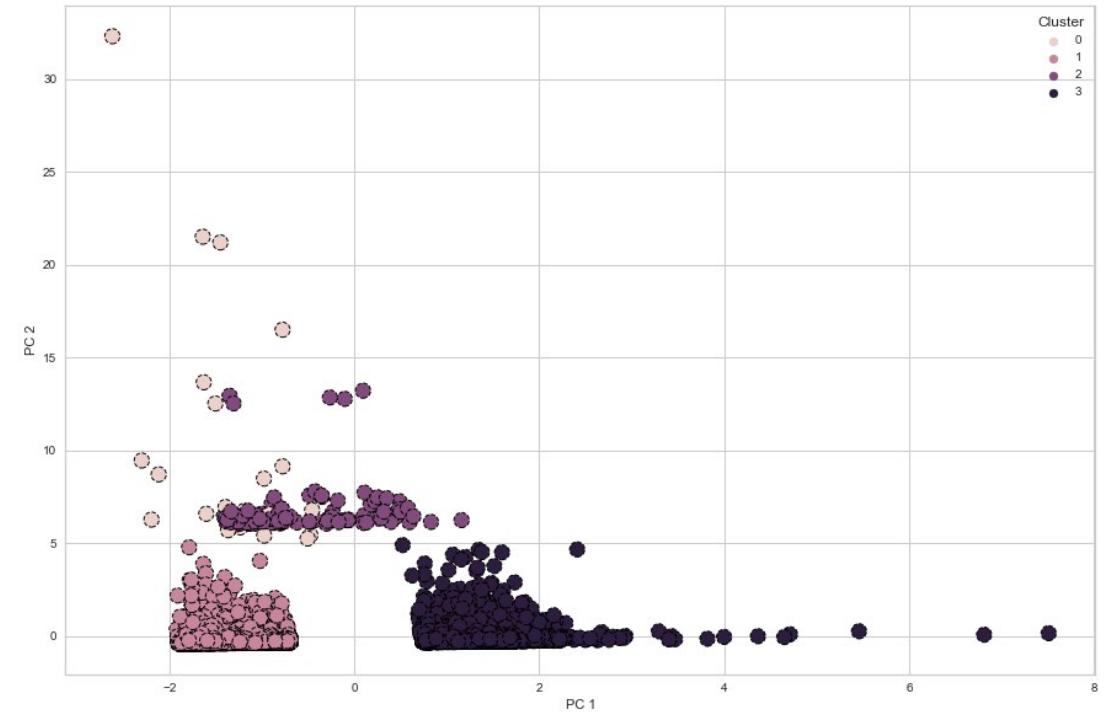
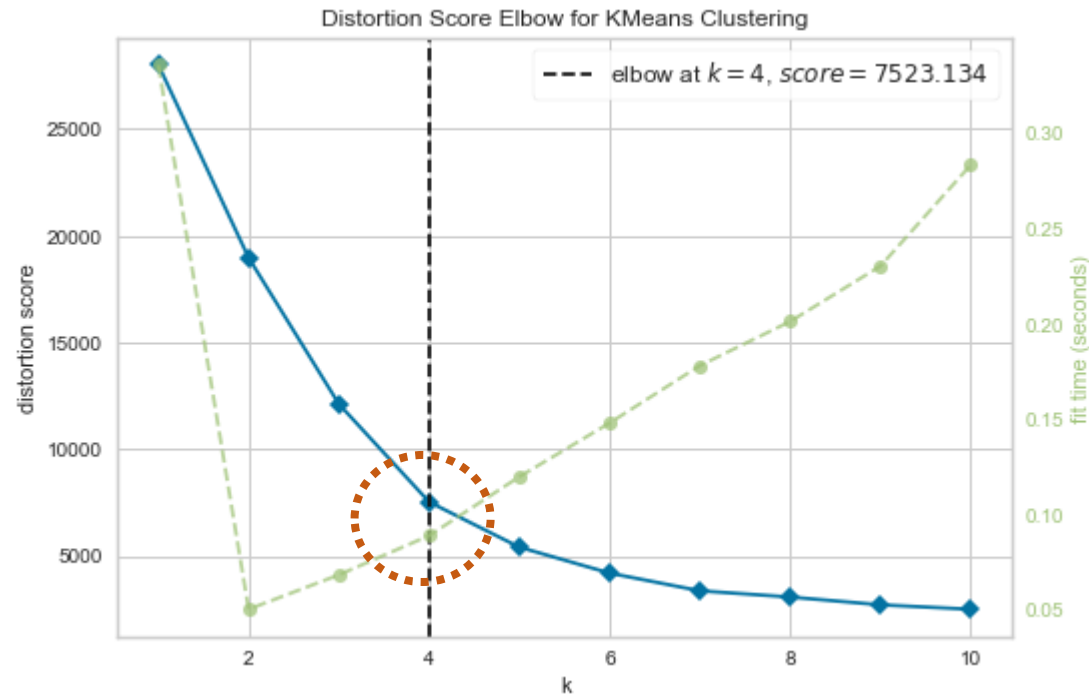


```
1 inertia = []
2 for i in range(1, 11):
3     kmeans = KMeans(n_clusters=i, random_state=42)
4     kmeans.fit(data_std)
5     inertia.append(kmeans.inertia_)
6
7
8 plt.figure(figsize=(20, 10))
9
10 # plt.plot(inertia)
11 sns.lineplot(x=range(1, 11), y=inertia, color='#000087', linewidth = 4)
12 sns.scatterplot(x=range(1, 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

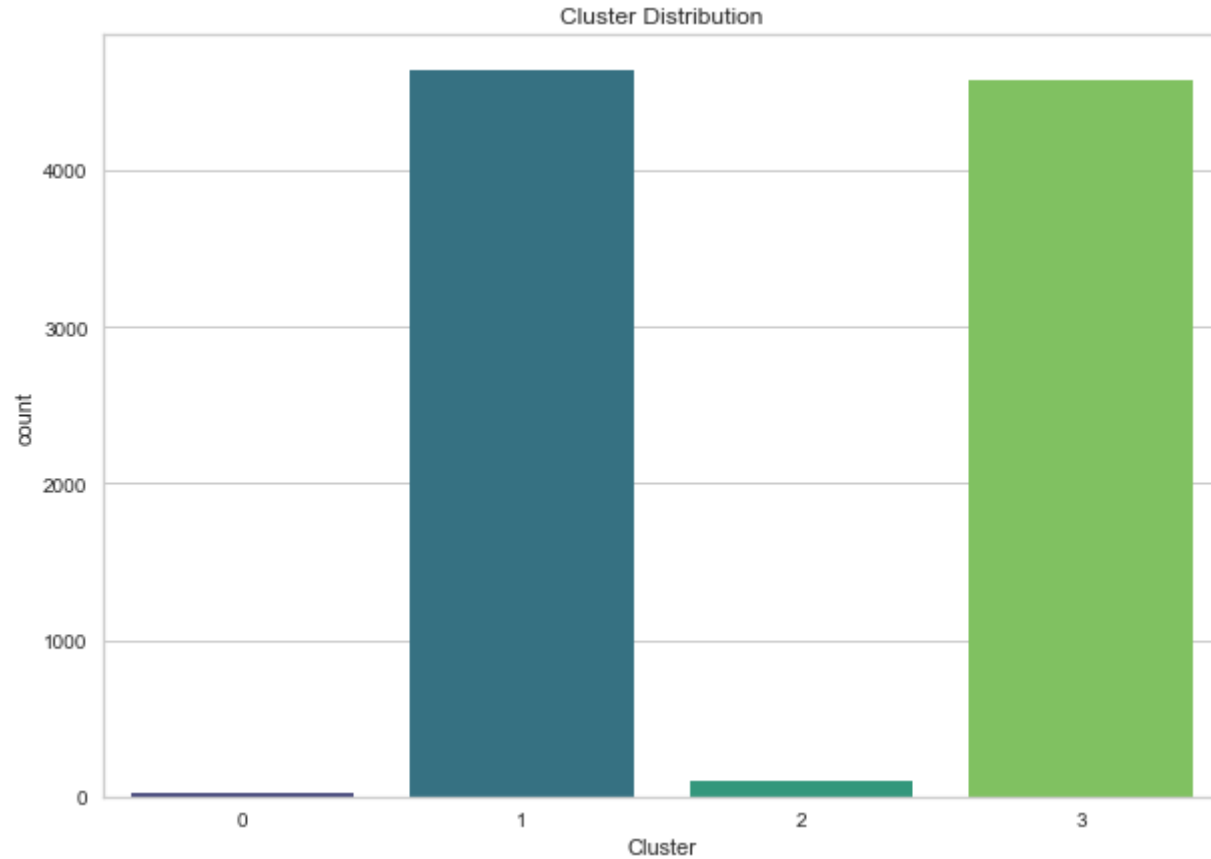
5. DATA PREPROCESSING & MODELING

K MEANS CLUSTERING



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

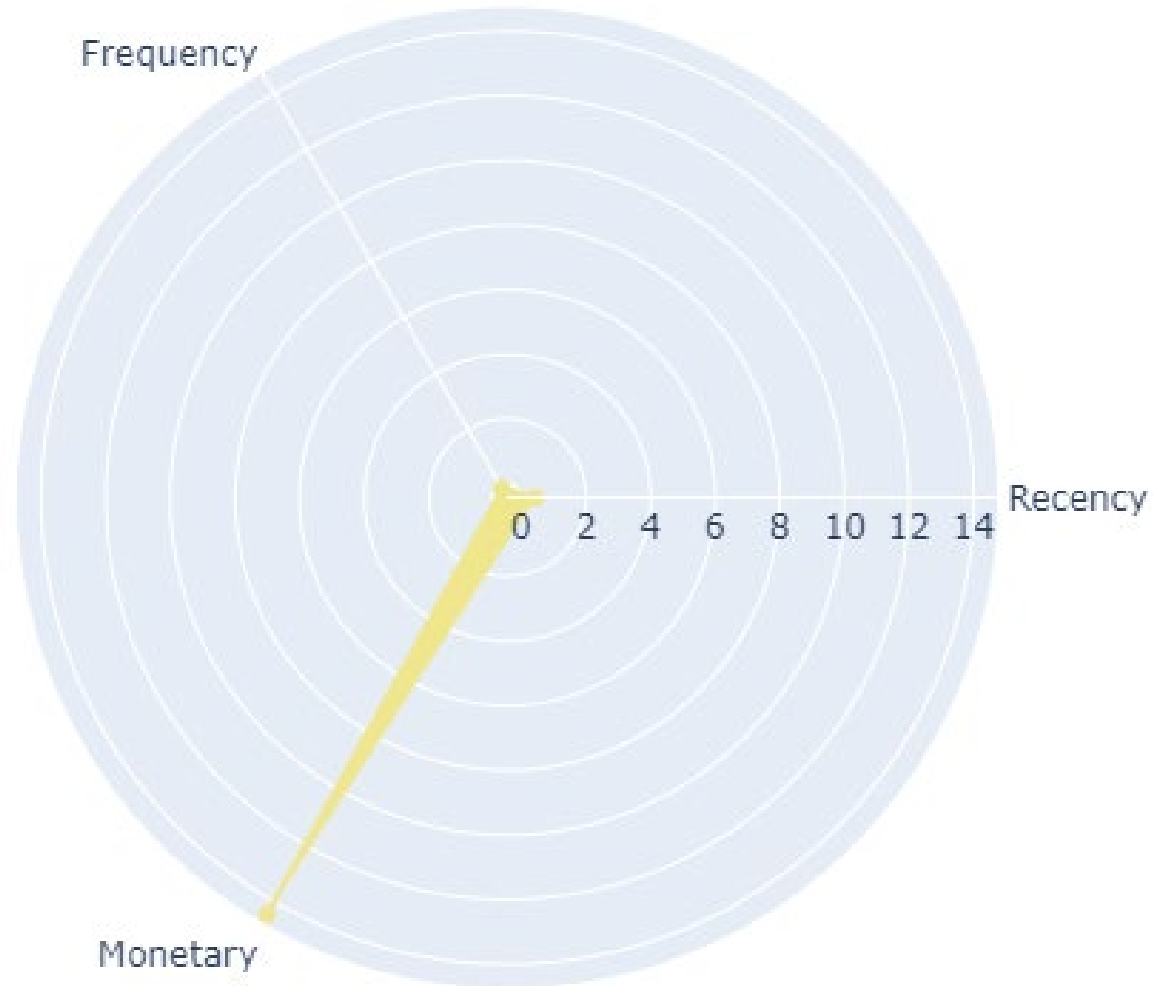
6. MODEL INTERPRETATION & RECOMMENDATION



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

6. MODEL INTERPRETATION & RECOMMENDATION

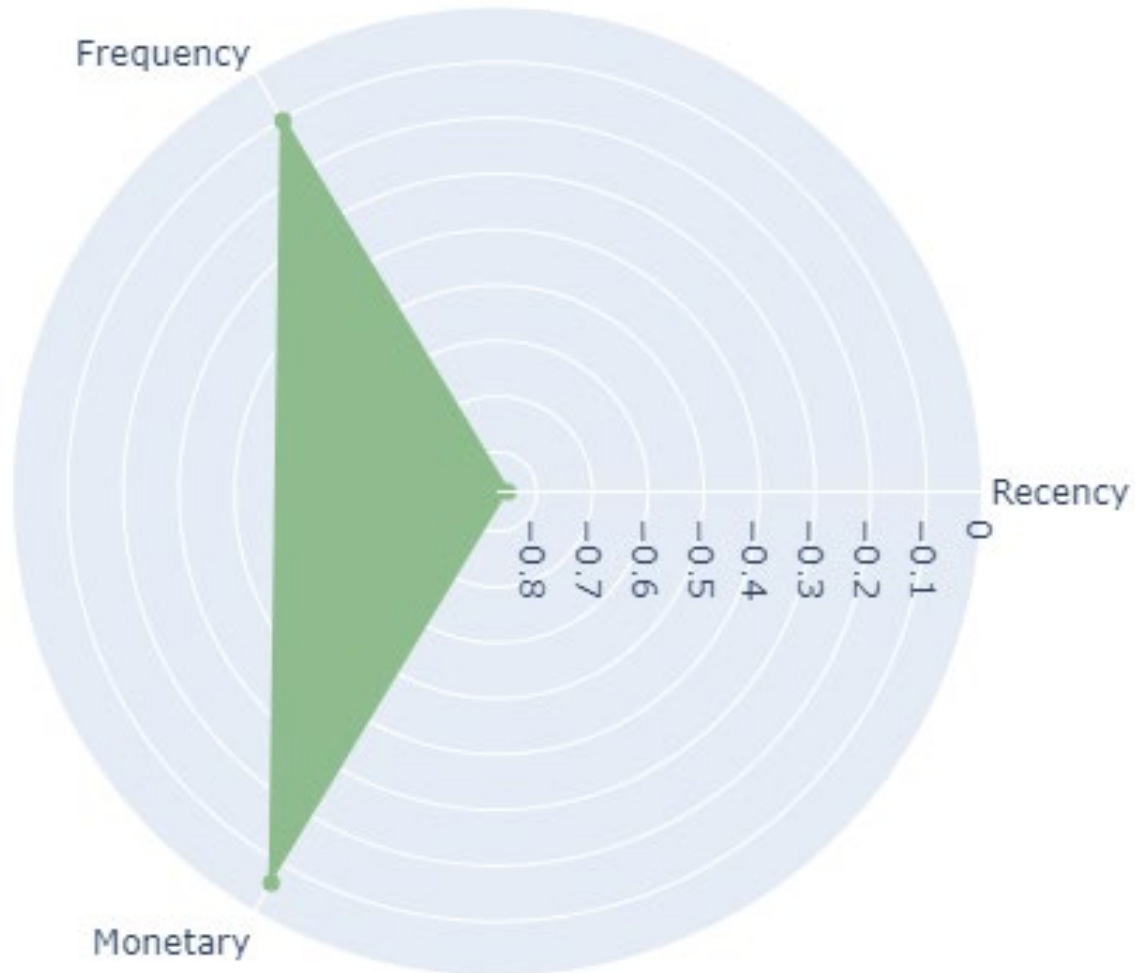
CLUSTER 0



This cluster is a 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.

6. MODEL INTERPRETATION & RECOMMENDATION

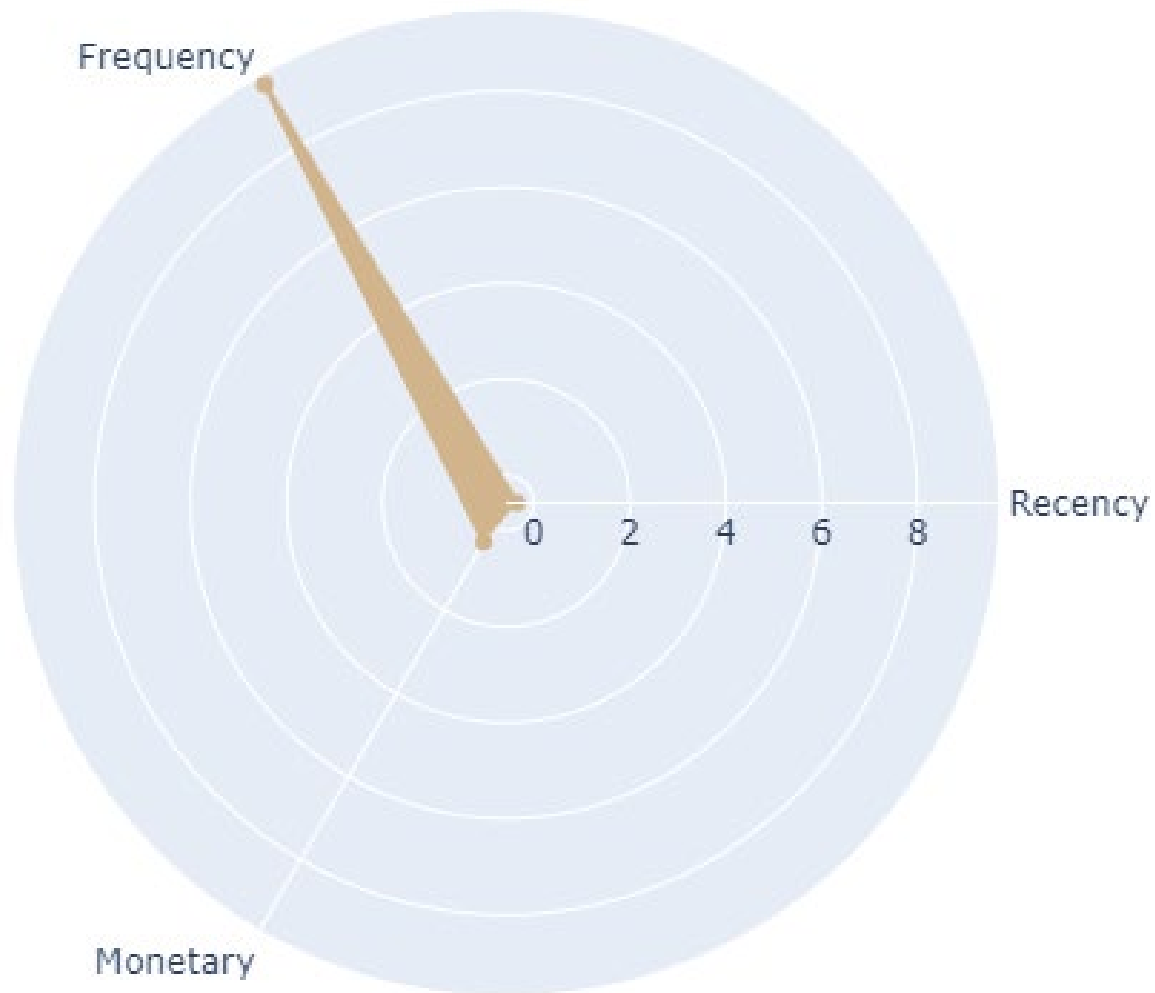
CLUSTER 1



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 relationship with a good corporate image.

6. MODEL INTERPRETATION & RECOMMENDATION

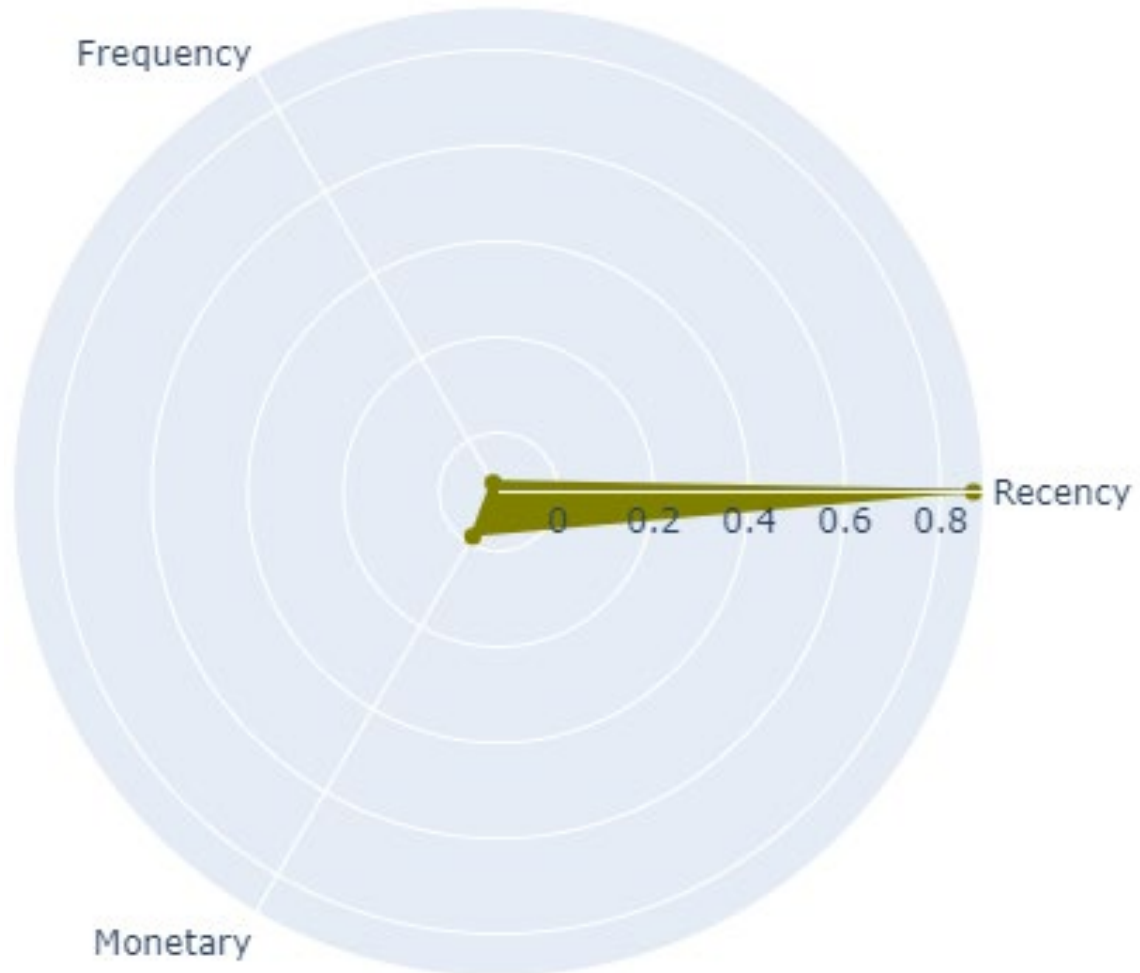
CLUSTER 2



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 lost customer with a high recency value and a small amount spent. Company can send promos via email for specific products but they are not the main focus.

Thanks !

Presentation template by

PPTMÖN.com

<http://pptmon.com/> Ctrl + Click to follow link