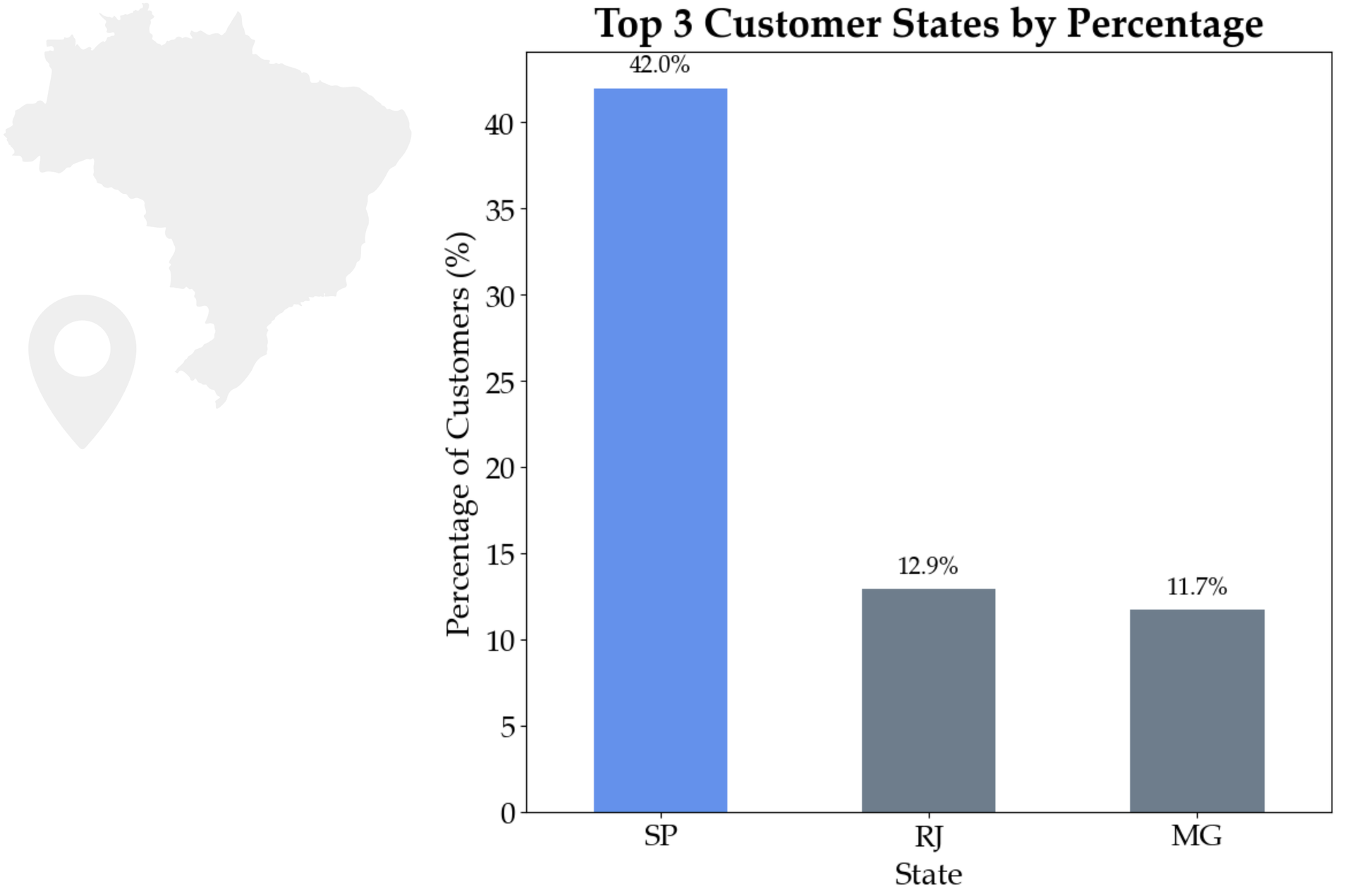
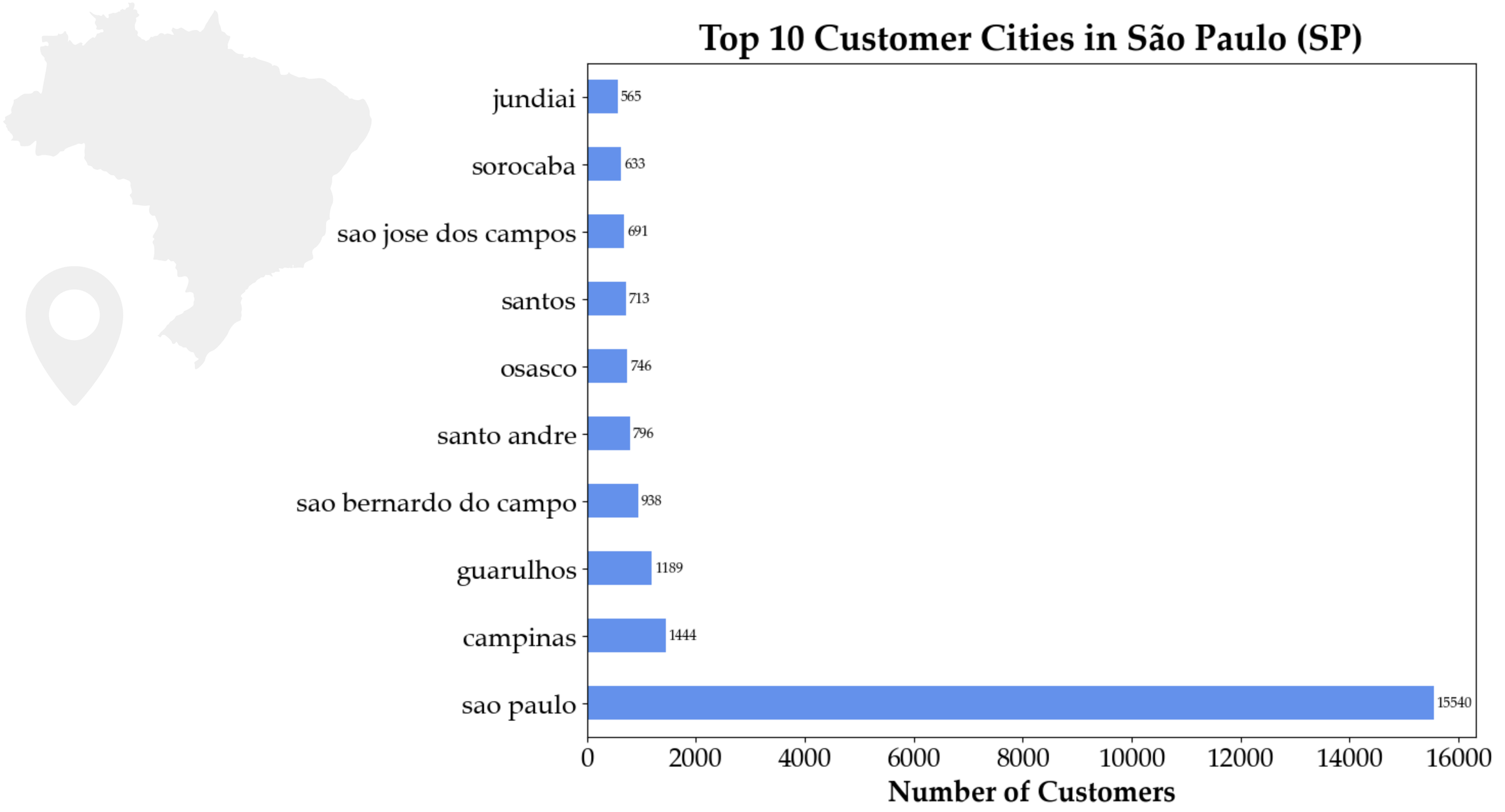
CUSTOMER SEGMENTATION

Observing Olist Customer's behavior

Data Analysis Team:

Ali Osama **Raghad Rashed** **Shuaib Baksh**


Top 3 Customer States by Percentage
42.0%
12.9%
11.7%
Percentage of Customers (%)
0
5
10
15
20
25
30
35
40
SP
RJ
MG
State


Top 10 Customer Cities in São Paulo (SP)
jundiai
sorocaba
sao jose dos campos
santos
osasco
santo andre
sao bernardo do campo
guarulhos
campinas
sao paulo
565
633
691
713
746
796
938
1189
1444
15540
0
2000
4000
6000
8000
10000
12000
14000
16000
Number of Customers

1

500 BD

9 month ago

1 Purchase

2

100 BD

2 month ago

3 Purchase

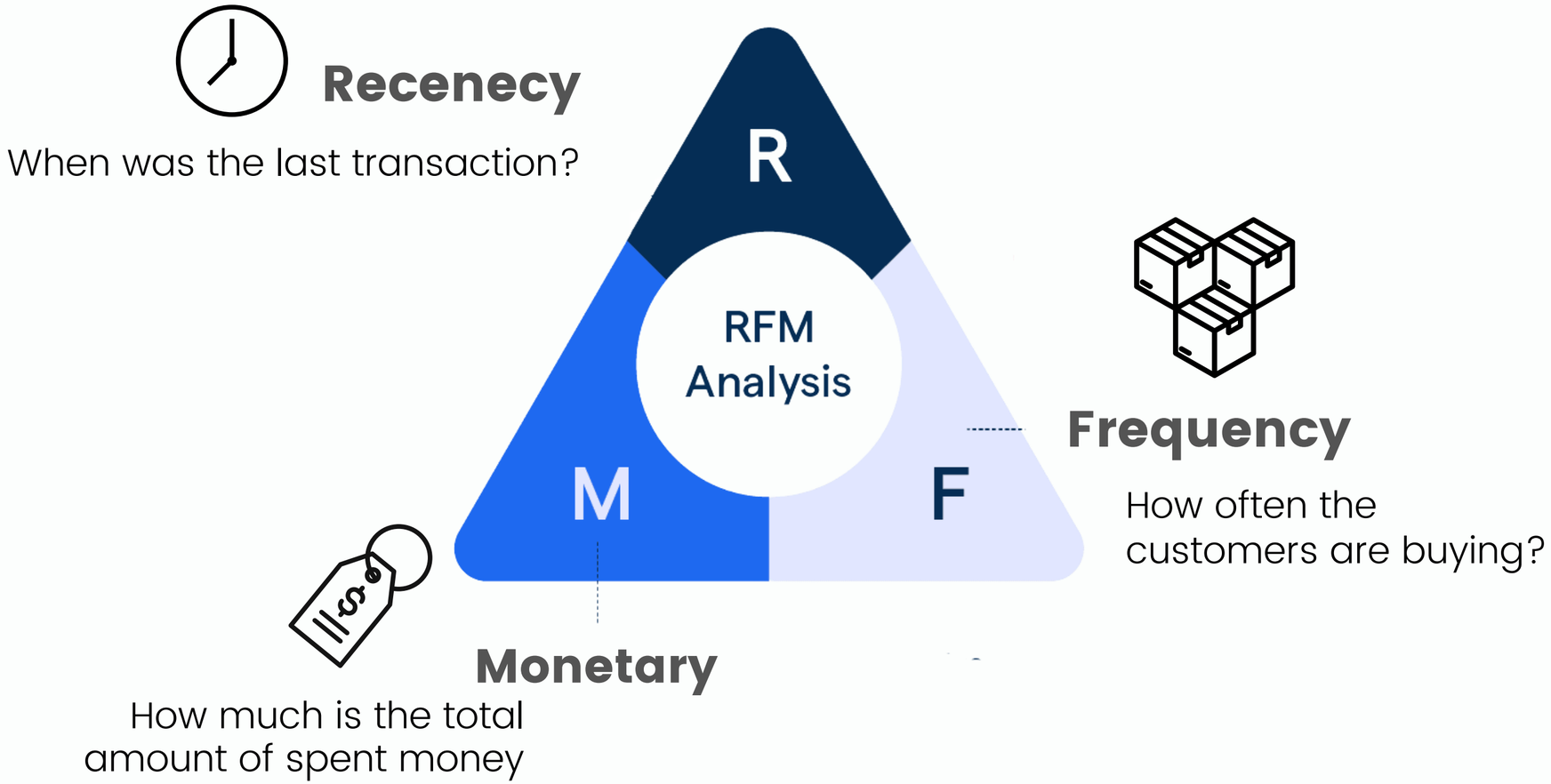
3

200 BD

3 days ago

1 Purchase

THE RFM MODEL



Key findings

01

In average, recency is 244 days (8 months)

02

75% of all customers made just 1 purchase

03

Most Customers Spend Small Amounts (Median ₹113)

04

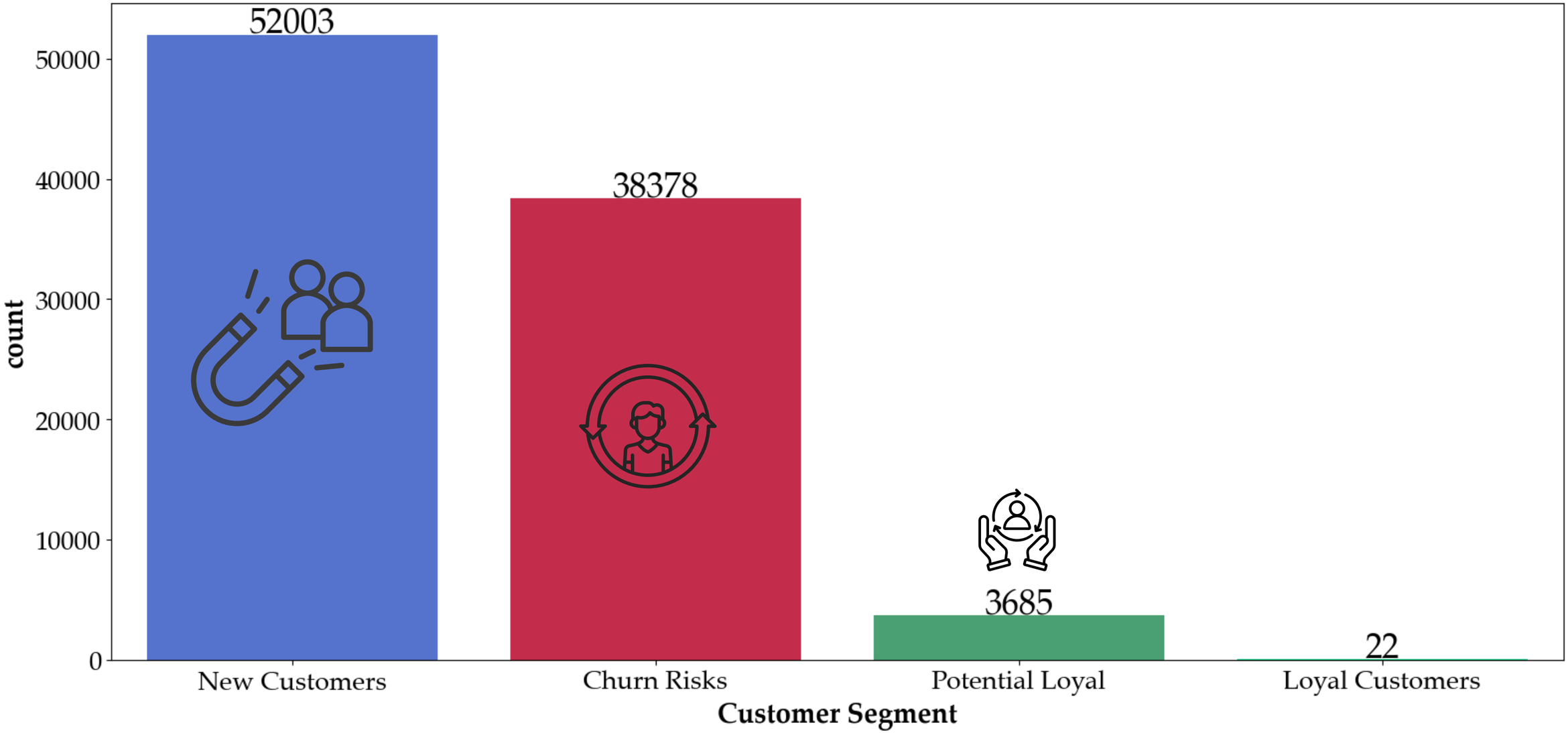
Few Big Orders Drive Up the Average (₹214)

SEGMENTATION

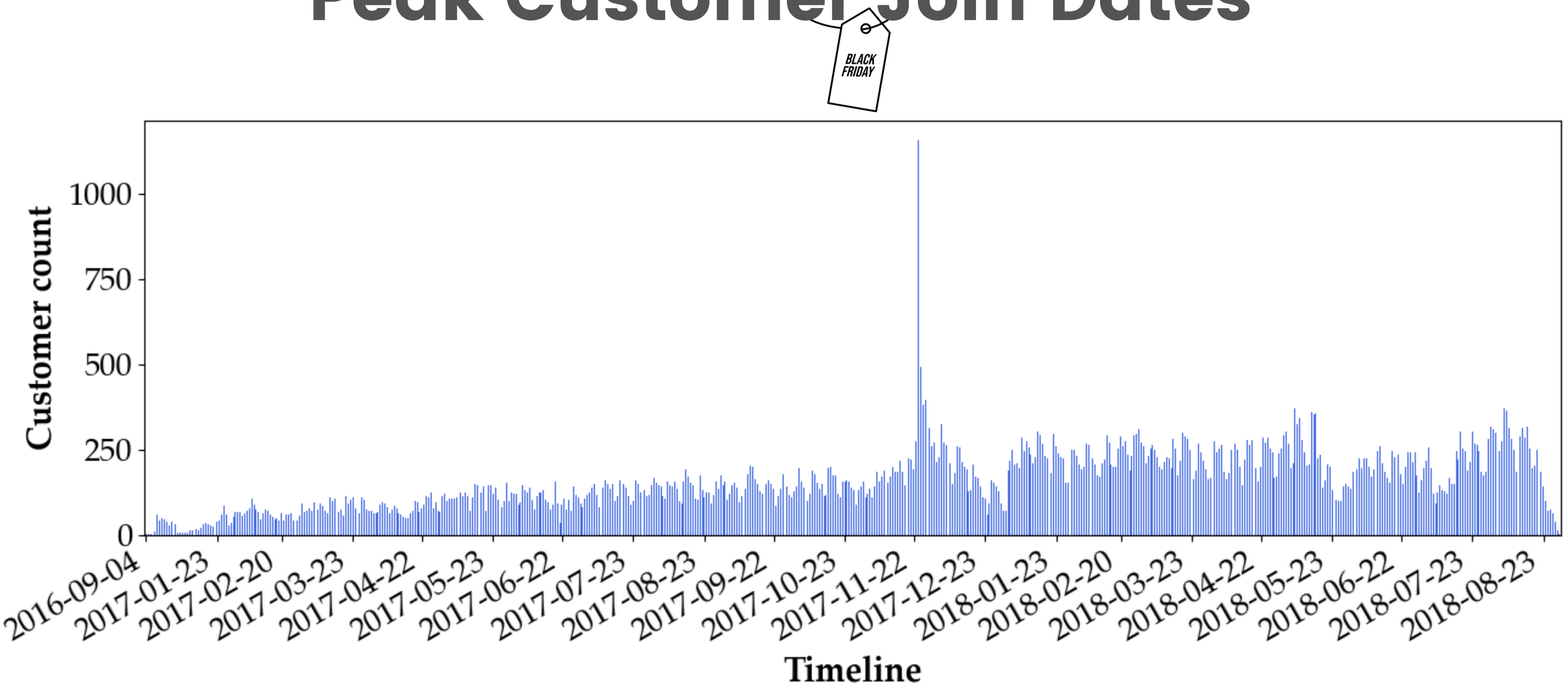
Tier	Behavior Summary
New Customers	Customers with low engagement who have transacted recently
Potential Loyal	Customers showing loyalty signals (Recent customers but with a good amount of spending/frequency)
Loyal Customers	Frequent buyers with high lifetime value (had the highest total transactions/monetary value)
Churn Risks	Inactive customers who may not return (Have made a transaction while ago with low frequency/monetary purchases)

Cluster	Recency	Frequency	Monetary	Tier
0	222.0	3.0	601.765	**Potential Loyal**
1	135.0	1.0	109.500	**New Customers**
2	227.5	10.0	8,716.88	**Loyal Customers**
3	382.0	1.0	106.970	**Churn Risks**

Customers Distribution by Segments

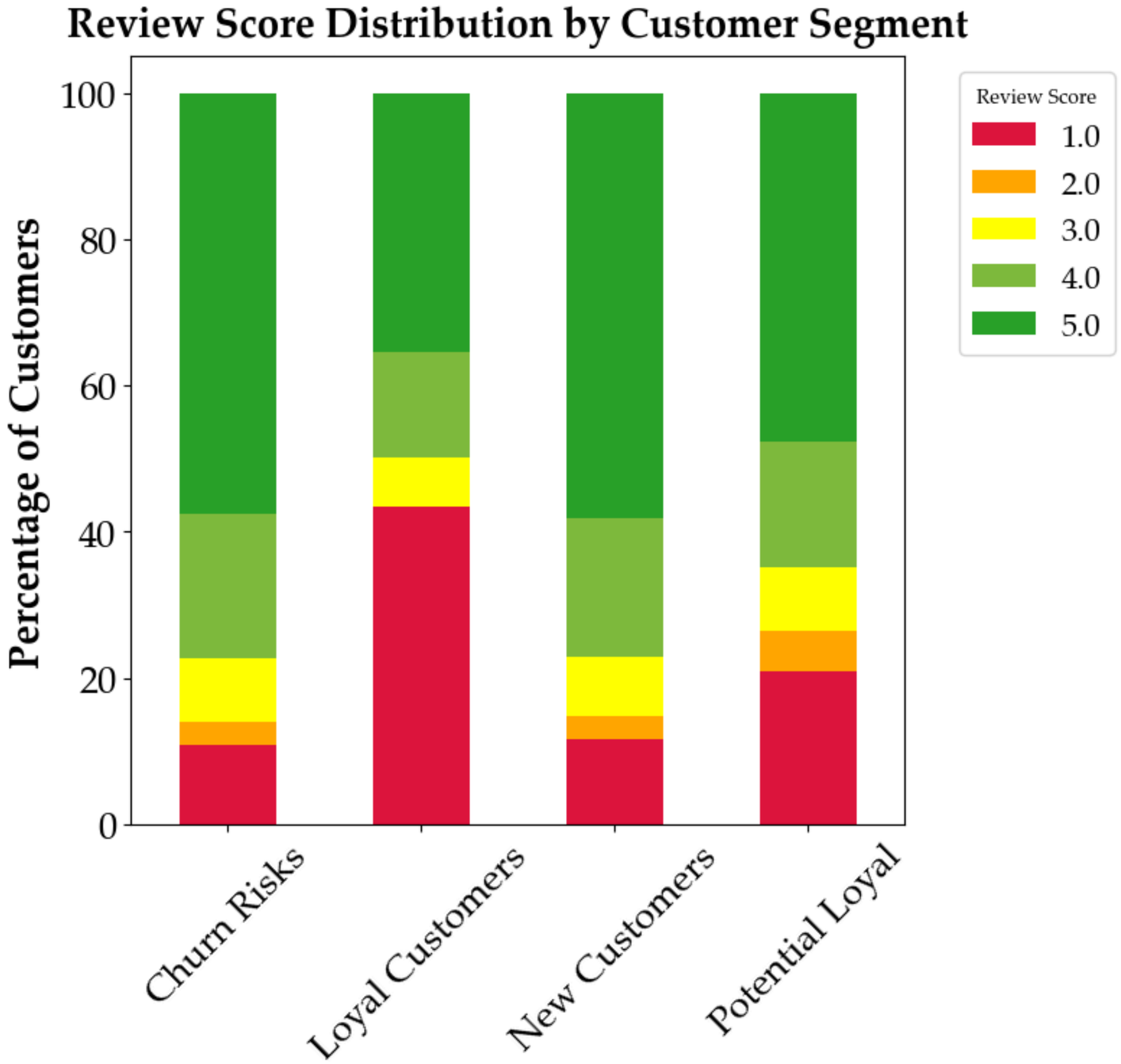


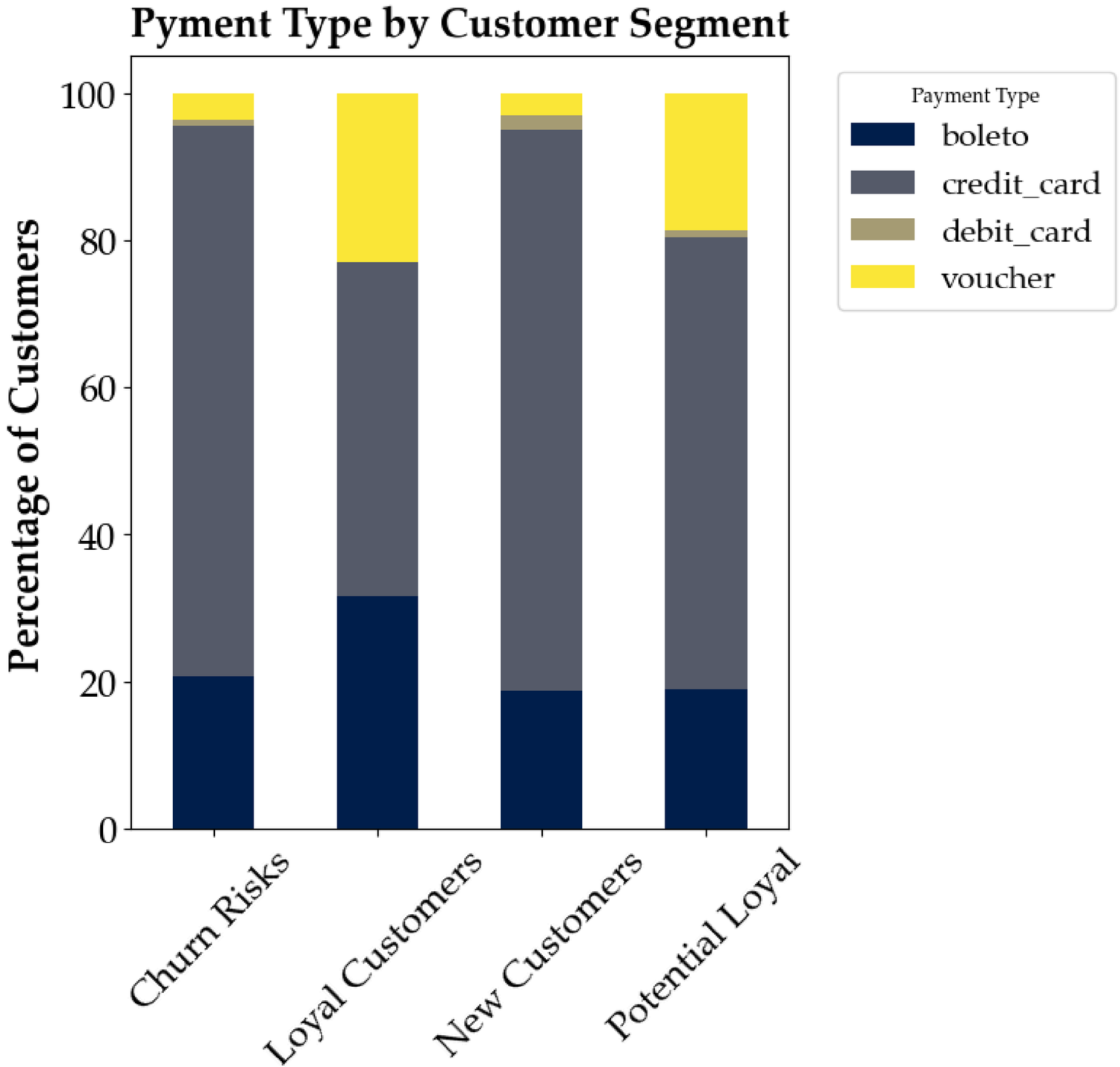
Peak Customer Join Dates

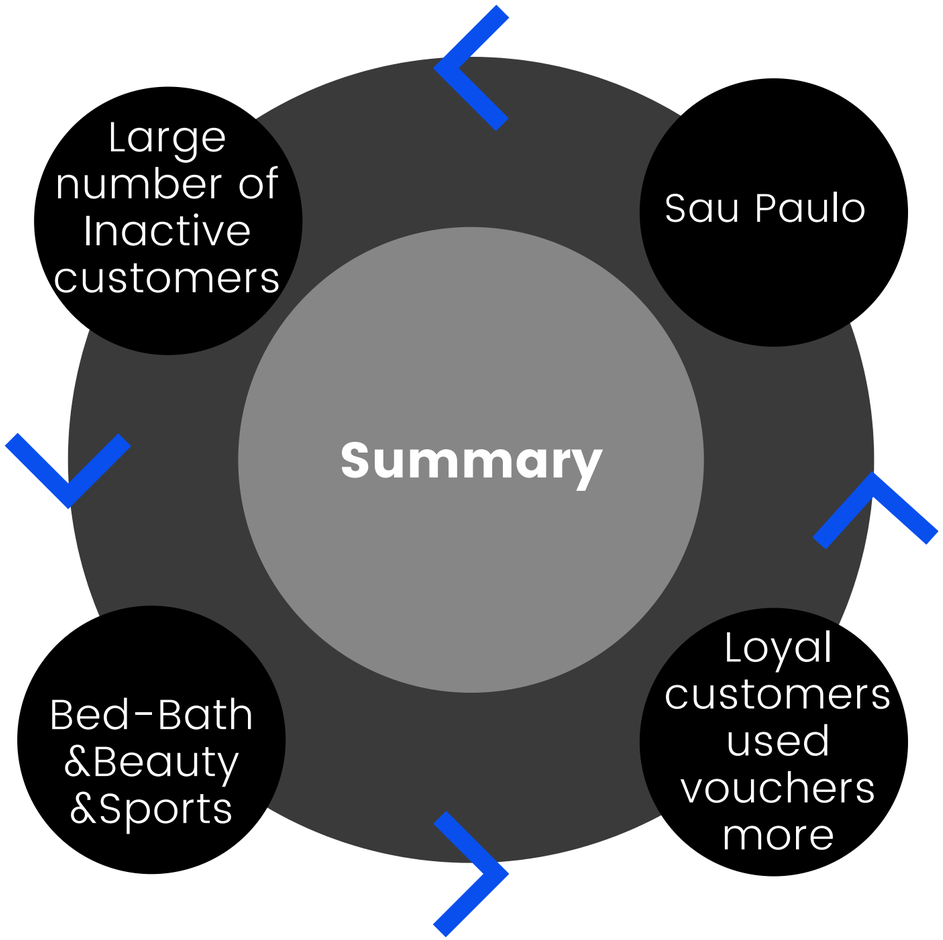


Customer Segment Price Preferences (BRL R$)

Top 3 Preferred Product Categories by Customer Segment (%)


Review Score Distribution by Customer Segment
Percentage of Customers
0
20
40
60
80
100
Churn Risks
Loyal Customers
New Customers
Potential Loyal
Review Score
1.0
2.0
3.0
4.0
5.0

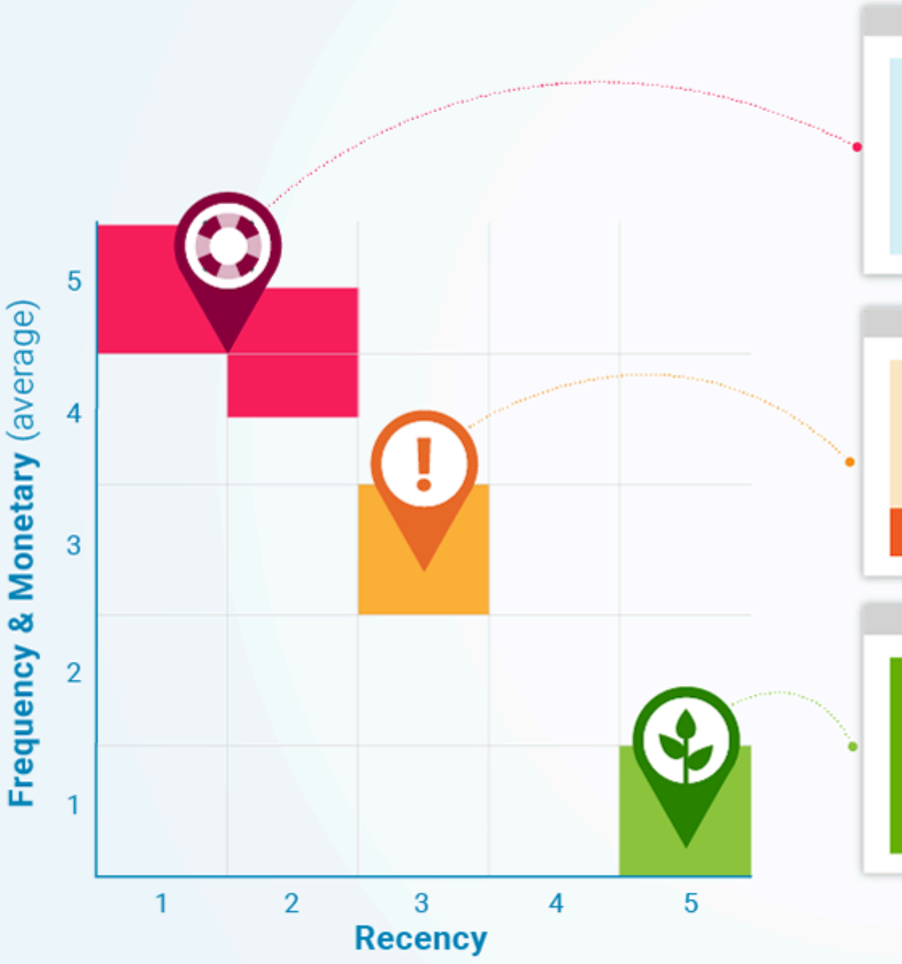

Pyment Type by Customer Segment
Percentage of Customers
0
20
40
60
80
100
Churn Risks
Loyal Customers
New Customers
Potential Loyal
Payment Type
boleto
credit_card
debit_card
voucher

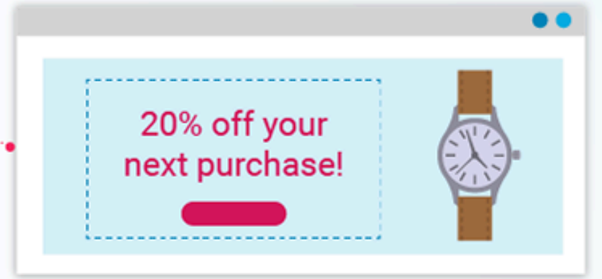

Large number of Inactive customers
Sau Paulo
Summary
Bed-Bath &Beauty &Sports
Loyal customers used vouchers more

Recommendation CRM

Tier	Recommended Strategy
New Customers	Welcome offers, second-purchase incentives, Retargeting (By ads)
Potential Loyal	Loyalty program enrollment, targeted deals,Urgency Tactics
Loyal Customers	VIP treatment, exclusive rewards, Asking for feedback
Churn Risks	Campaign via email, reactivation offers

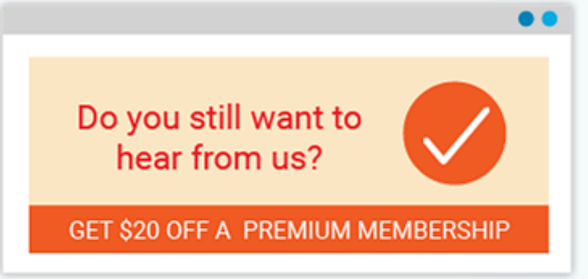
Examples of Campaigns based on Customer Segment





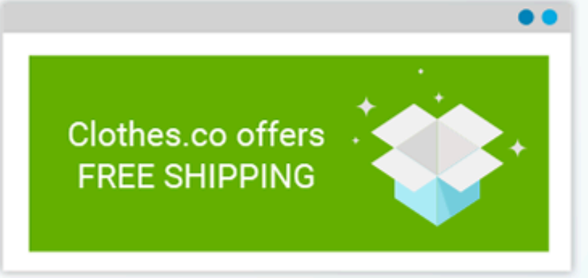
CANNOT LOSE SEGMENT

Win-back campaign



AT RISK SEGMENT

Re-engagement campaign



FRESH LEADS SEGMENT

Welcome series

Future Optimization

Olist Delivery

Can we predict the delivery time ?

Lets see the machine learning model!

lab? (5) - JupyterLab

Microsoft Edge - lab? (5) - Jupyte

localhost:8888/lab?

File Edit View Run Kernel Tabs Settings Help

Olist_Analysis (5).ipynb

Markdown

Share Notebook Python [conda env:base] *

Machine Learning part

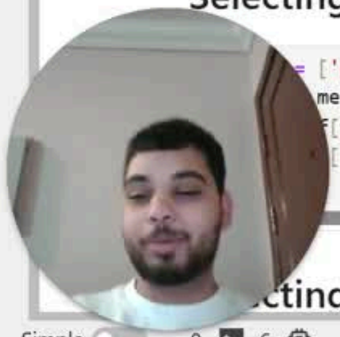
[31]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import matplotlib.pyplot as plt
import seaborn as sns
```

Converting data types and creating a new column

+ 1 cell hidden

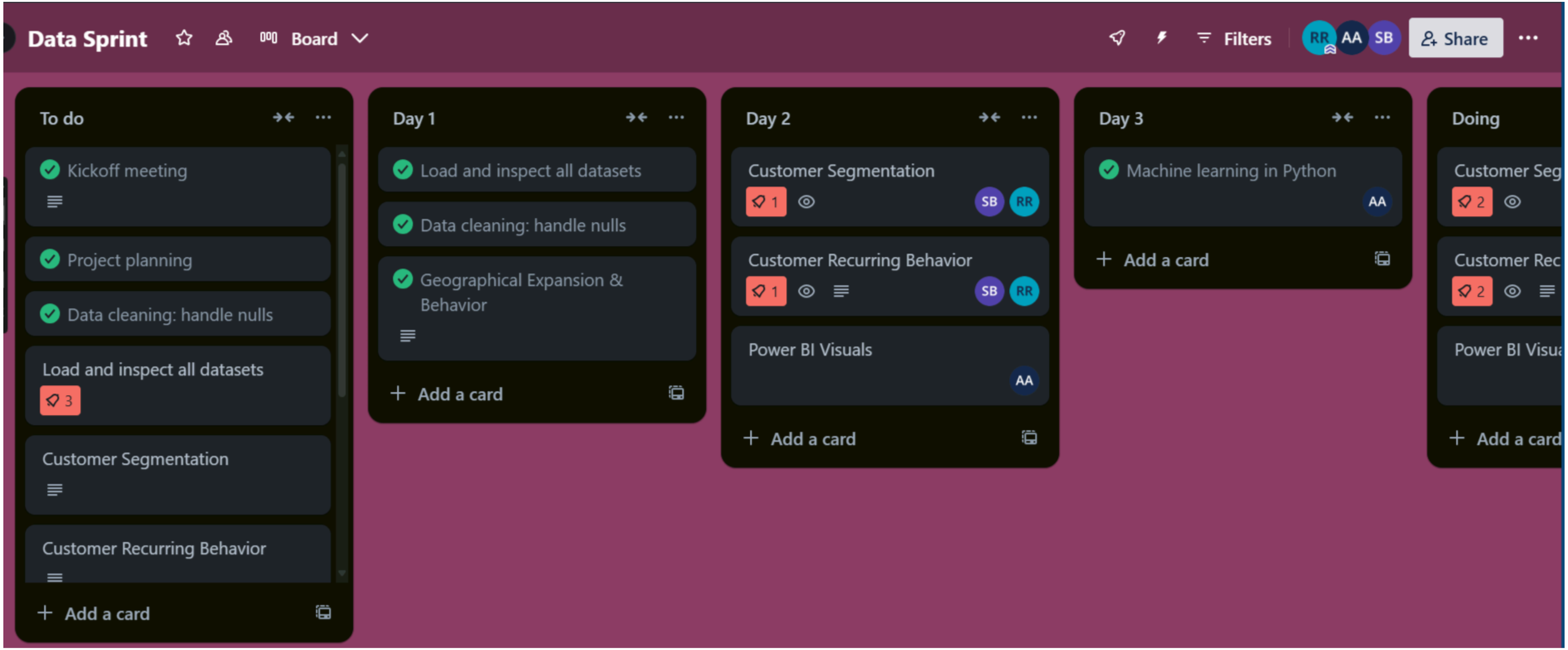
Selecting the features in the Model

```
= ['price', 'freight_value', 'payment_installments', 'payment_value', 'product_weight_g', 'product_length_cm', 'product_height_cm', 'pr
merged_df[features + ['delivery_time_days']].dropna()
[features]
['delivery_time_days']
```



cting the models and training them

Simple 0 $_ 6 Python [conda env:base] * | Idle Mode: Command Ln 1, Col 1 Olist_Analysis (5).ipynb 1


Data Sprint
Board
Filters
RR
AA
SB
Share
To do
Kickoff meeting
Project planning
Data cleaning: handle nulls
Load and inspect all datasets
3
Customer Segmentation
Customer Recurring Behavior
Add a card
Day 1
Load and inspect all datasets
Data cleaning: handle nulls
Geographical Expansion & Behavior
Add a card
Day 2
Customer Segmentation
1
SB
RR
Customer Recurring Behavior
1
SB
RR
Power BI Visuals
AA
Add a card
Day 3
Machine learning in Python
AA
Add a card
Doing
Customer Seg
2
Customer Rec
2
Power BI Visua
Add a card

Q&A