

UNVEILING CUSTOMER RETENTION DYNAMICS FOR BLU

PRESENTATION

MELLAS Anais
ROUINI Arij
SUBRAMANI Praveen



EXECUTIVE MANAGEMENT SUMMARY

Challenge:

- Predicting Factors Influencing Customer Review Scores on Orders.

Solution:

- Predict scores: Accurately forecast positive (4-5) or negative (1-3) reviews.
- Identify drivers: Pinpoint factors influencing the reviews.

Impact:

- Proactive resolution: Address potential issues before negative reviews.
- Data-driven decisions: Optimize business for improved customer experience.
- Enhanced retention: Drive growth through increased loyalty.



EXECUTIVE MANAGEMENT SUMMARY

Project set up:

- Using PySpark, DataBricks (handle large dataset)

Main conclusion:

- Negative reviews rate : 3.29 %
- Factors : Order approval time, Shipping cost, Photos quality, Payments types.

Proposed actions:

- Streamline Approval Processes and Enhance Availability of Products
- Improve Payment Systems
- Enhance visual appeal and engagement
- Optimize Shipping Costs and Product Pricing



BUSINESS SECTION



**Unlocking
Customer Insights**



**Actionable
Recommendations**

BUSINESS SECTION

Suggestion : Number of Positive & Neg reviews per Order status

Insights:

- Orders with the status “Unavailabe”, “Shipped”, and “Delivered” have a higher chance of receiving a negative review.

Business Actions:

- Provide tracking information and estimated delivery dates to manage expectations effectively.
- Reduce order processing time.
- Demonst commitment to customer satisfaction (offer apologies/incentives for issues).



BUSINESS SECTION

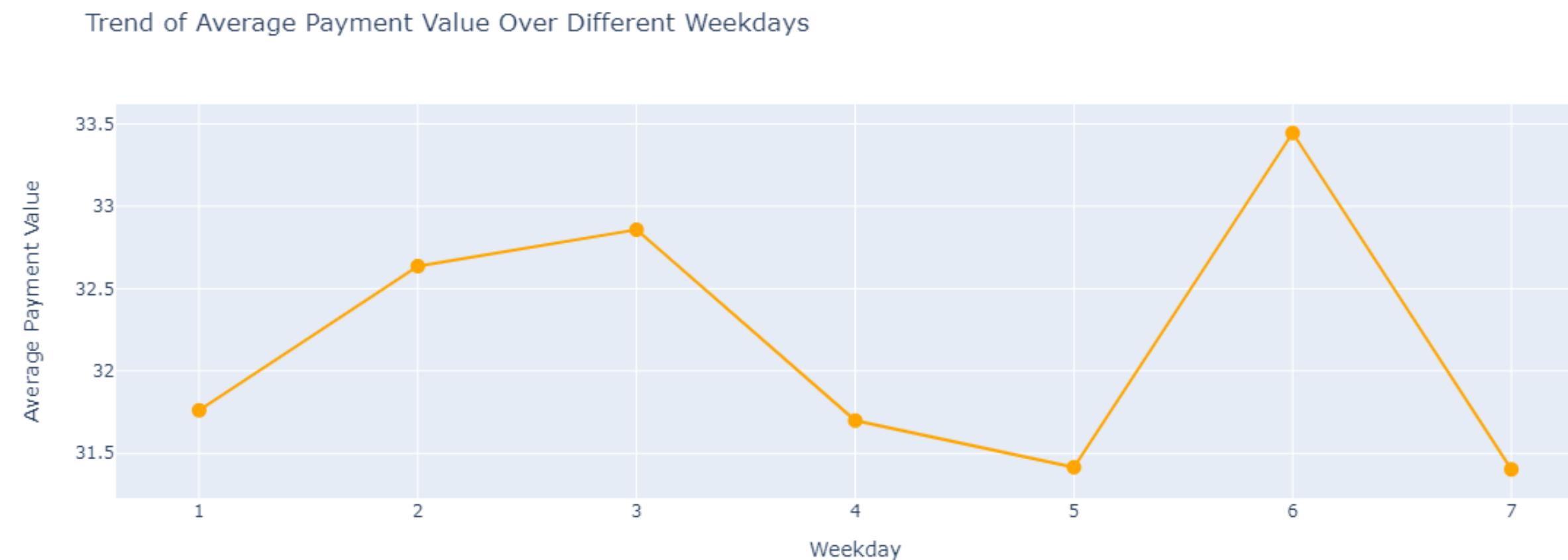
Revenue and Sales Analysis

Insights:

- Customers tend to place higher order values on Wednesday and Saturday.

Business Actions:

- Ensure that logistics can handle an increased volume of sales.
- Staffing consideration (packing/shipping)



BUSINESS SECTION

Negative review - Shipping cost

Insights:

- Orders with low shipping cost are more likely to receive a negative review.

Business Actions:

- Evaluate Shipping Cost Structure.
- Improve Shipping Quality and Time.
- Enhance Customer Service.

Distribution of Average Shipping Cost for Negative Predictions



BUSINESS SECTION

Payments Types

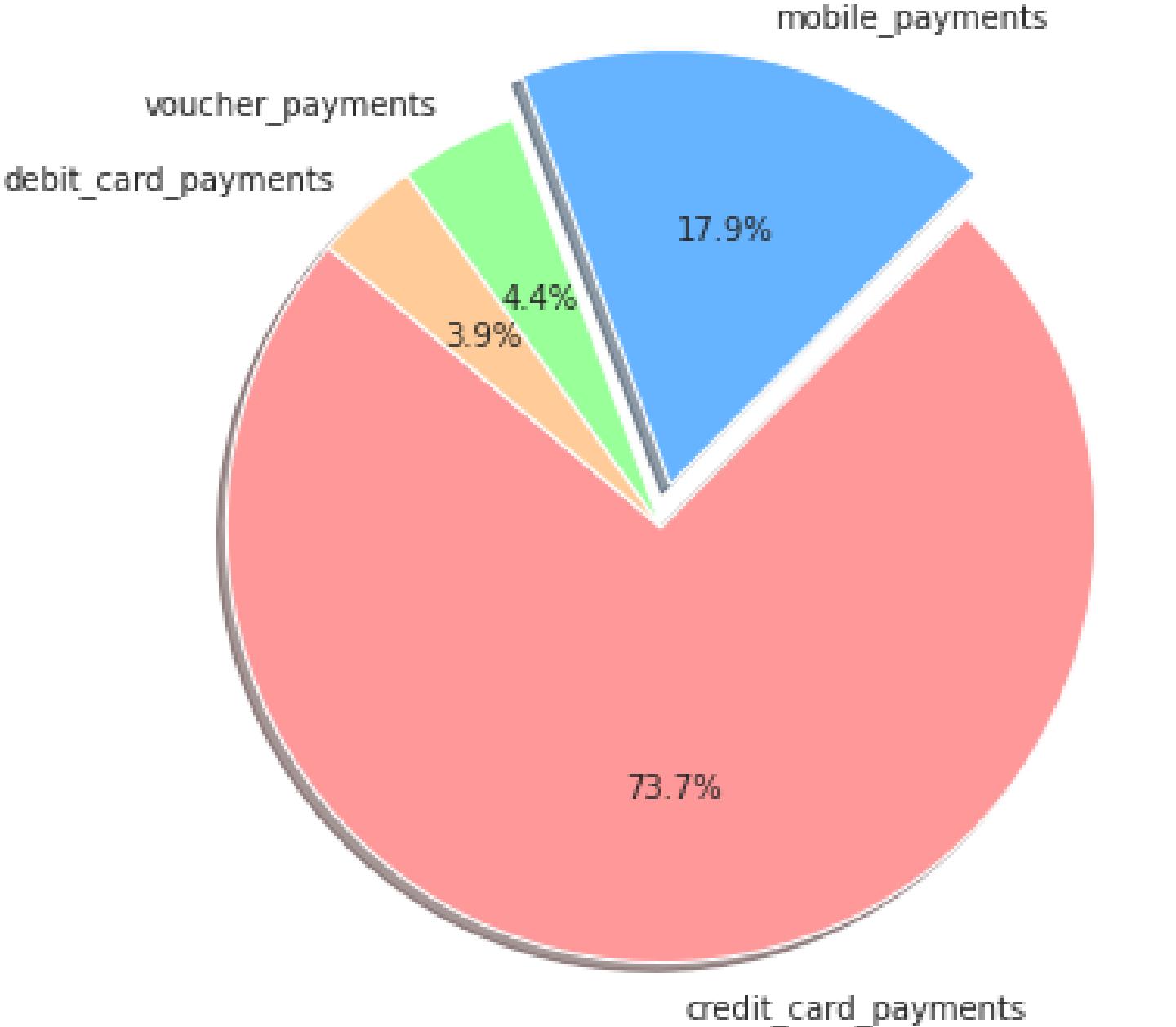
Insights:

- + 70% of order payments are done using credit card payments followed by mobile payments.

Business Actions:

- Expand the range of credit cards and digital wallets accepted to accommodate a wider customer base.
- Streamline the payment process for these two main methods.

Distribution of Payment Types Across Orders



BUSINESS SECTION

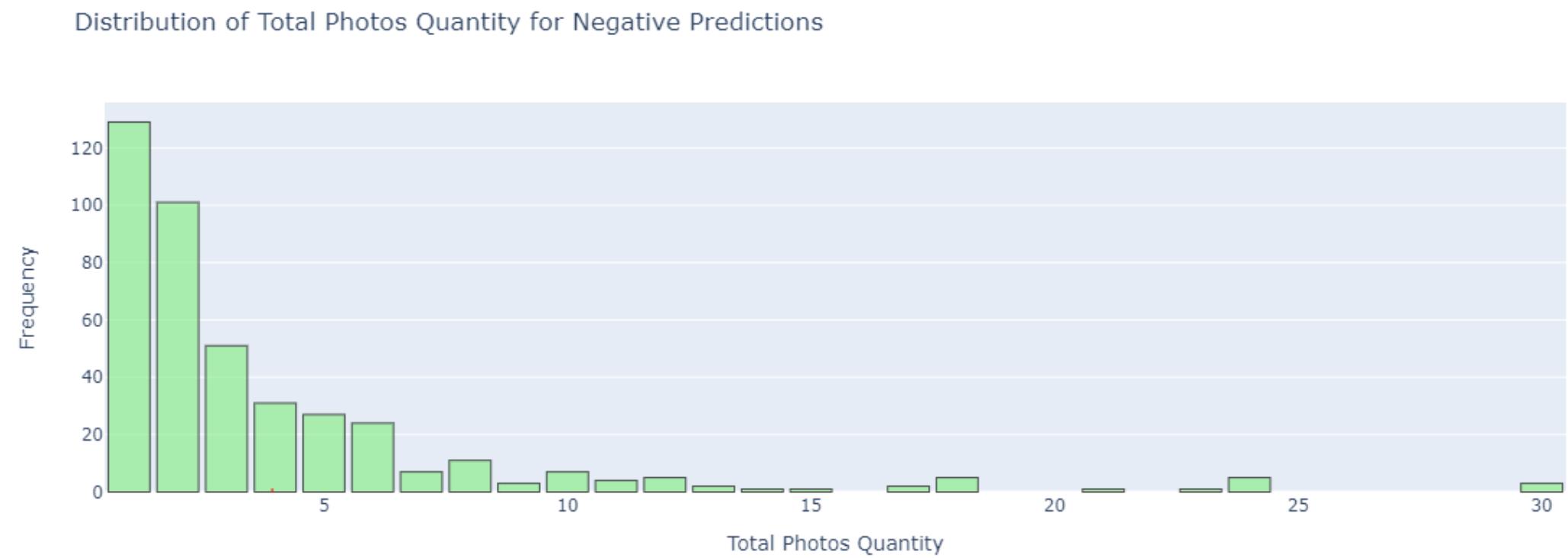
Negative review - Number of Photos

Insights:

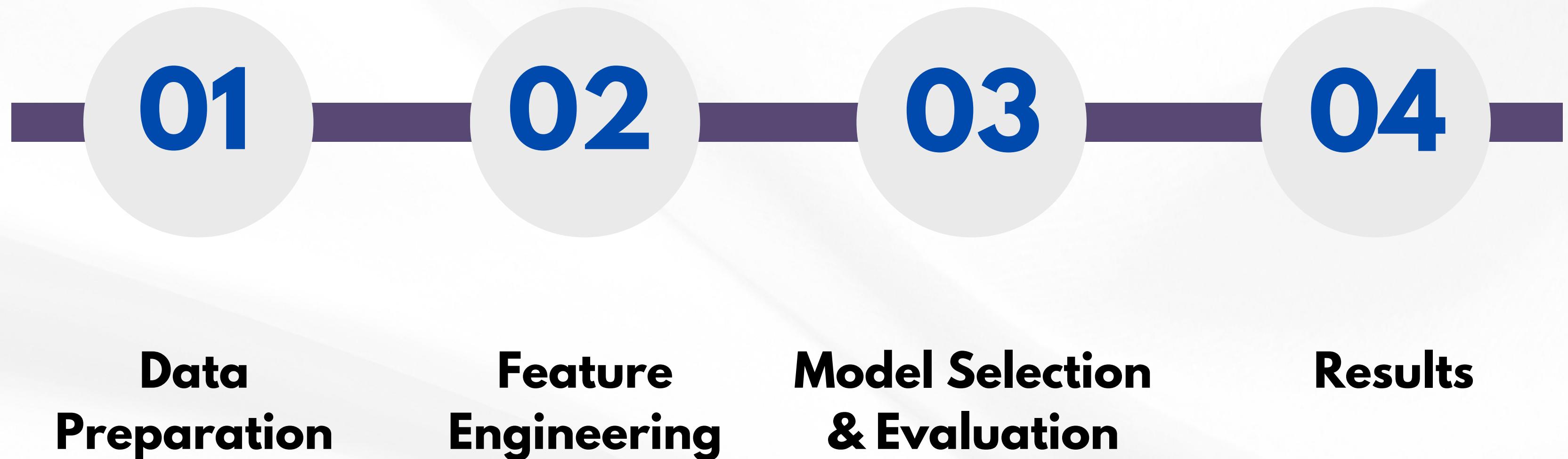
- Products with least Photos trend to have more negative reviews.

Business Actions:

- Standardize photography protocols.
- Enhance Photo Descriptions.
- Improve Product Page Layout.



TECHNICAL SECTION



TECHNICAL SECTION

Data Preparation

Data Cleaning:

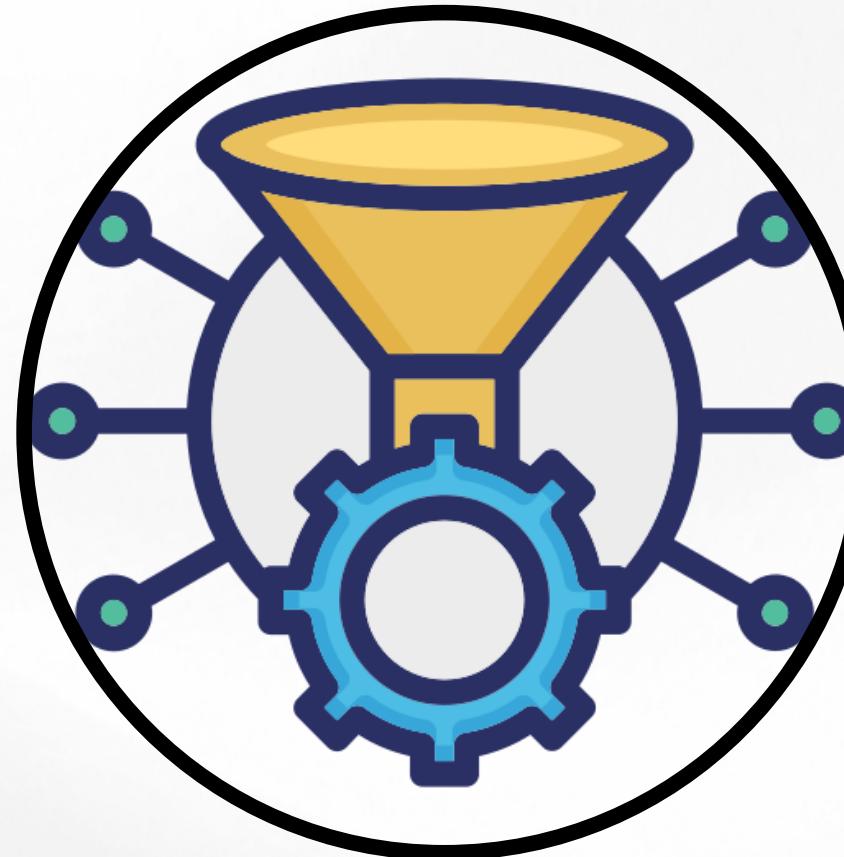
- Duplicates removed
- Handling missing values
- Consistency in data types

Timeline Preparation:

- Train dataset (Sep 2020-Jun 2022)
- Test dataset (Jul-Sep 2022)

Data Imputation:

- Numerical & categorical nulls filled with appropriate strategies.



TECHNICAL SECTION

Feature Engineering

Orders:

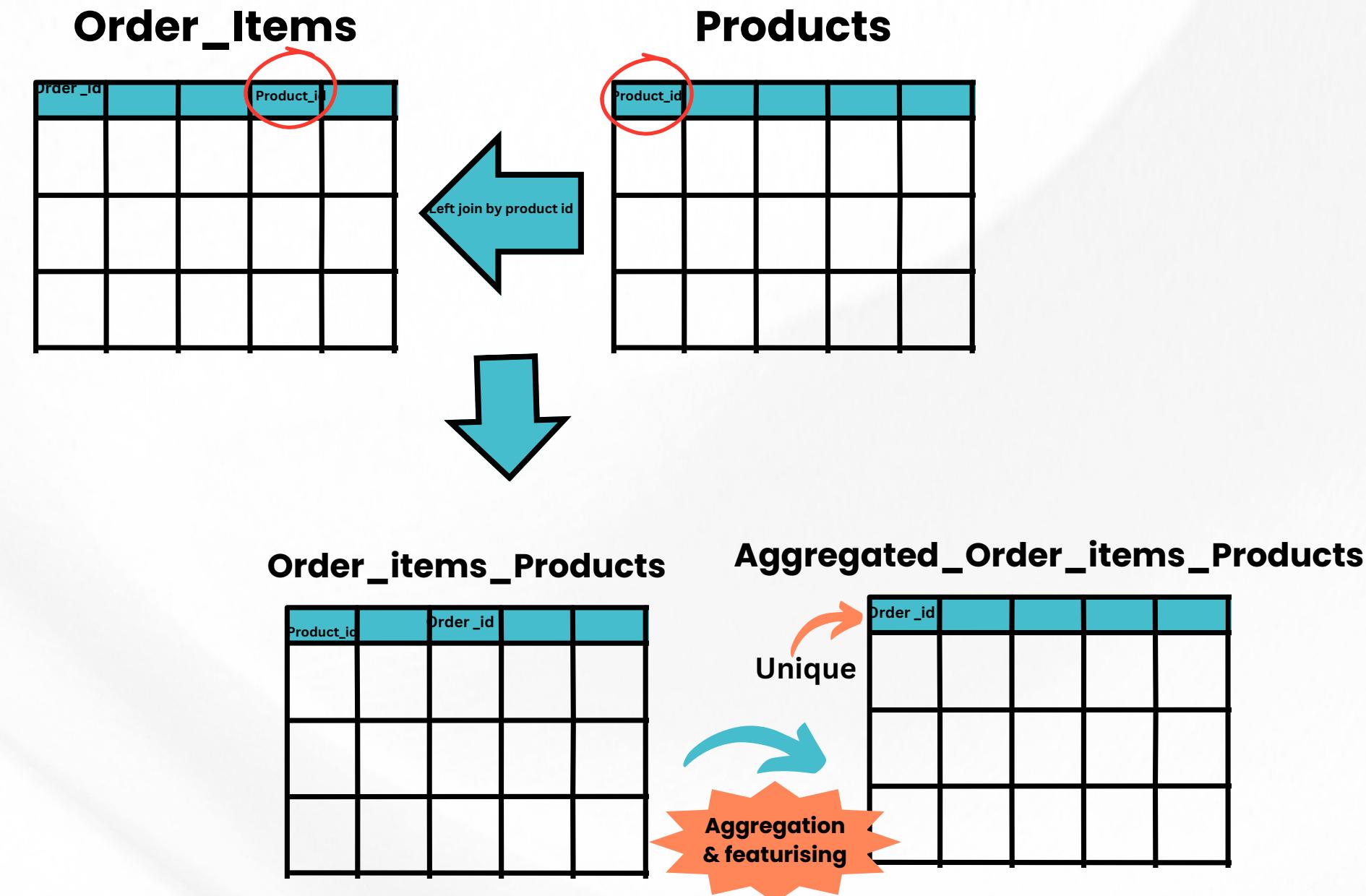
- New features: 'delivery_delay_days', 'purchase_weekday', 'total_shipping_cost' etc.

Products & Order Items:

- Duplicates removed from products & order_items.
- Average-based features: 'average_price', etc.

Payments & Reviews:

- New payment features, 'review_encoded' handled.



Model Selection & Evaluation

Variables included in the predictive model:

- Target variable (e.g., prediction of positive or negative reviews)
- Predictor variables (e.g., purchase behavior, order status, payment methods)

LOGISTIC REGRESSION



RANDOM FOREST

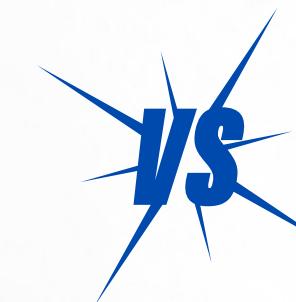


DECISION TREE

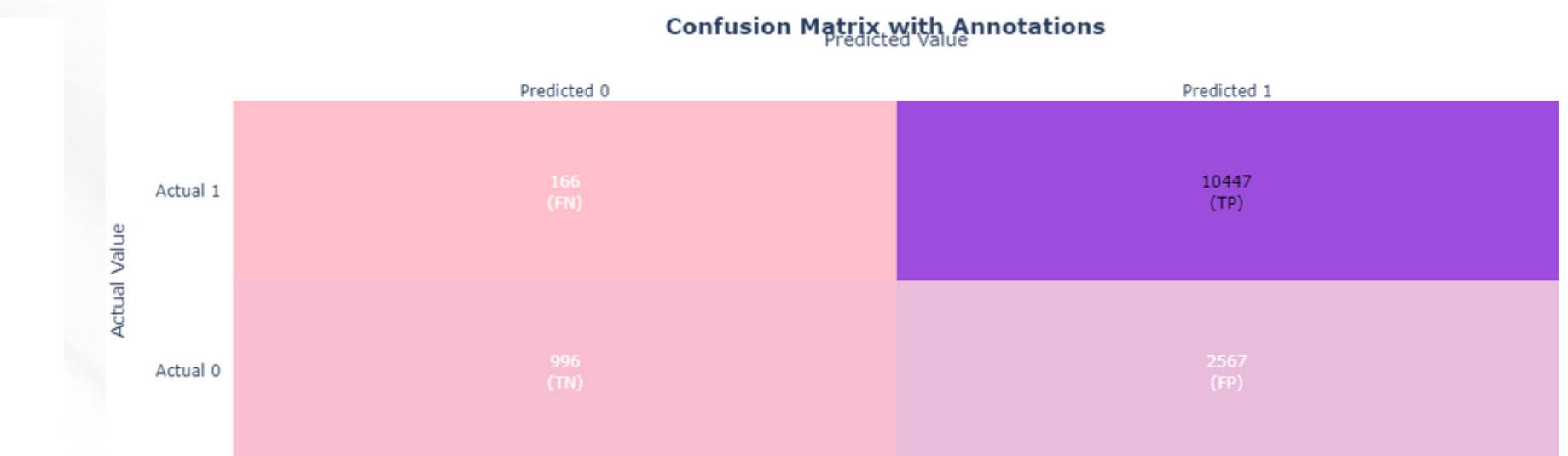
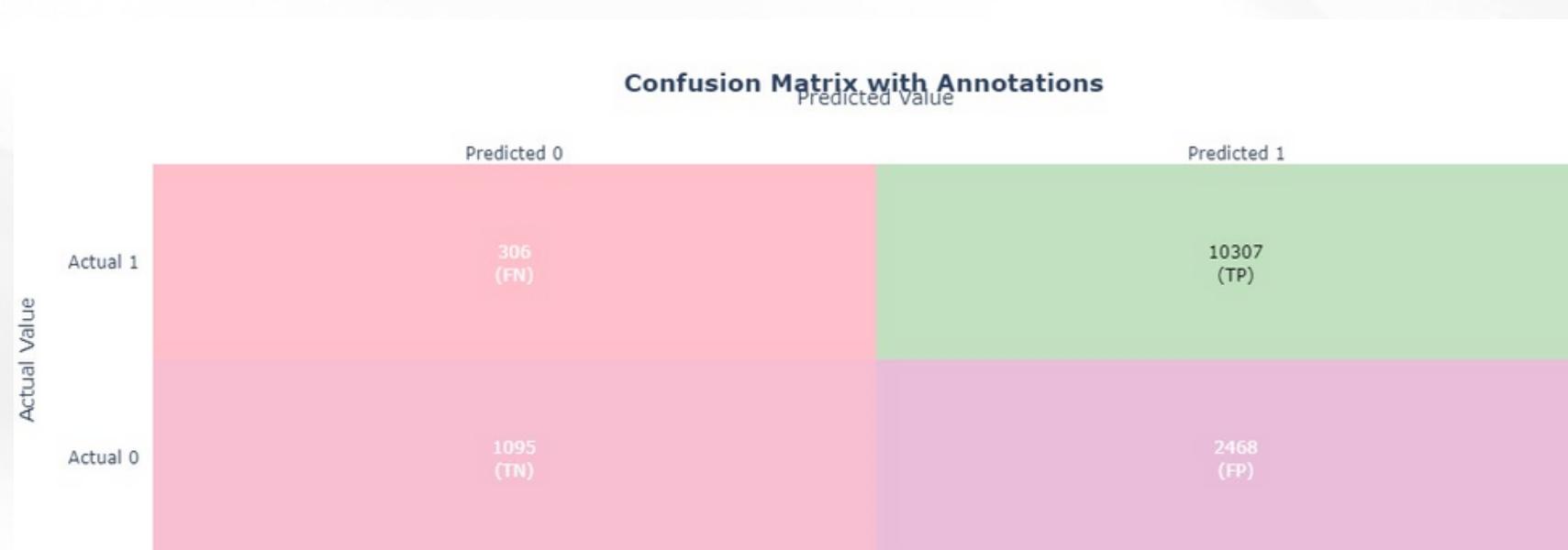


CONFUSION MATRIX

Logistic regression



Random Forest



Results

- Performance metrics of the predictive models: (e.g., accuracy, precision, recall, F1-score)

Logistics Regression



```
Accuracy: 0.8010016930022573
Precision (label 1.0): 0.801152300811
Recall (label 1.0): 0.981608768344789
F1 Score (label 1.0): 0.8822473598536
Weighted Precision: 0.800630907279834
Weighted Recall: 0.8010016930022573
Weighted F1 Score: 0.7561733102688737
Area under PR: 0.8007688463450099
```

Random Forest

```
Accuracy: 0.8151805869074492
Precision (label 1.0): 0.818651221
Recall (label 1.0): 0.97194872747
F1 Score (label 1.0): 0.8887378976
Weighted Precision: 0.81016135231
Weighted Recall: 0.8151805869074492
Weighted F1 Score: 0.7843122402659
```



Feature coefficients

Feature coefficients

	Feature	Coefficient	Rank
0	min_installments	0.316660	1
1	total_order_price	0.133645	2
2	total_photos_qty	0.023288	3
3	total_payment_value	0.008615	4
4	average_shipping_cost	0.006482	5
5	avg_payment_value	0.003822	6
6	average_description_length	0.003100	7

Discussion of limitations and areas for future improvement:

- Cross-Validation technique.
- Suggestions for refining the model through hyperparameters tuning
- Exploring additional variables for better prediction.

THANK YOU

