E-COMMERCE RETAIL CASE STUDY

BY-
SONU VERMA
SIDDHARTH RAHATE

ABOUT US

Schuster, a multinational retail company, faces challenges with late vendor payments. Despite credit arrangements, some vendors consistently pay late, resulting in non-value-added activities, time loss, and financial impact. Schuster aims to predict the likelihood of late payments against open invoices to optimize its payment process.

CASE STUDY BY

Sonu Verma

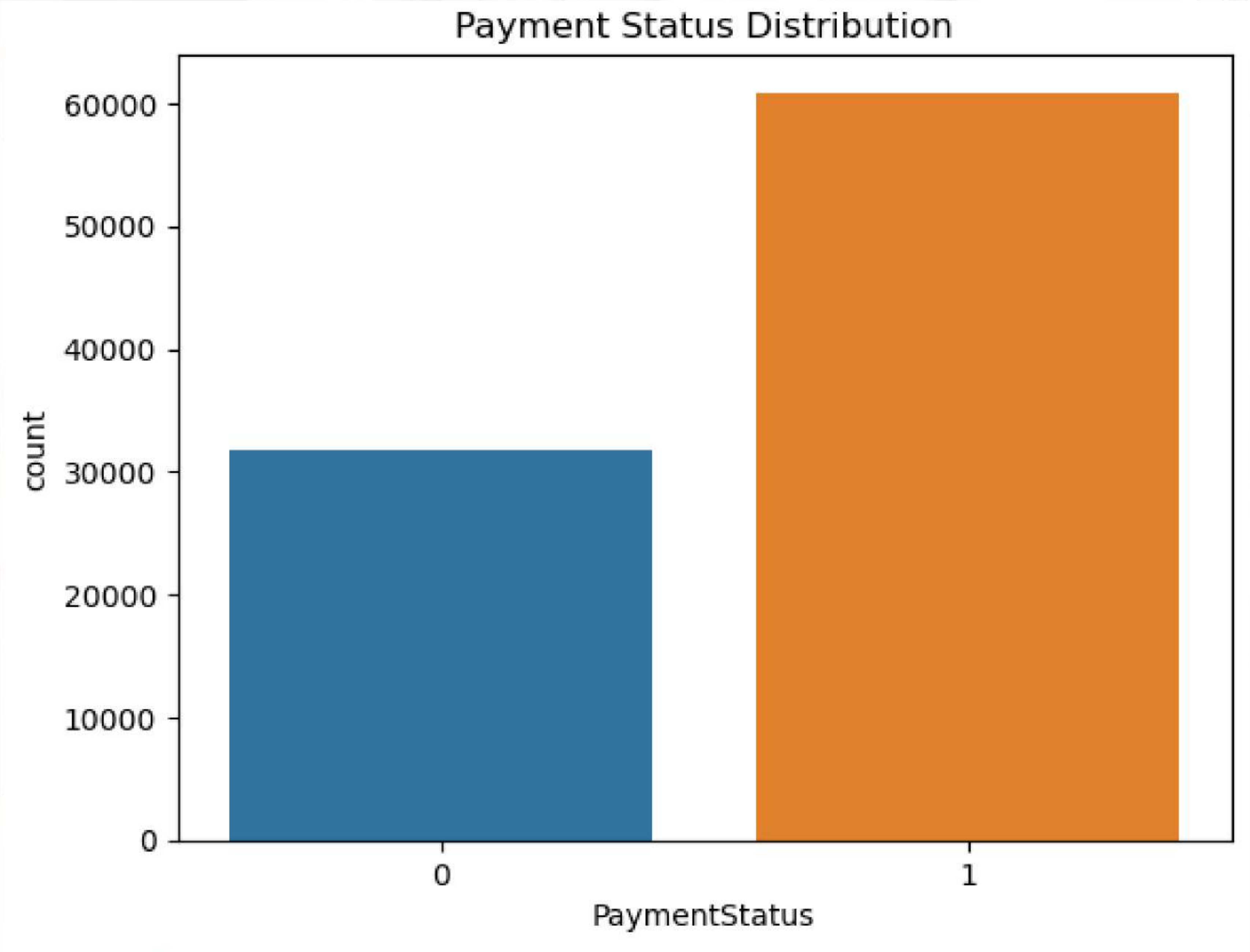
Siddharth Rahate

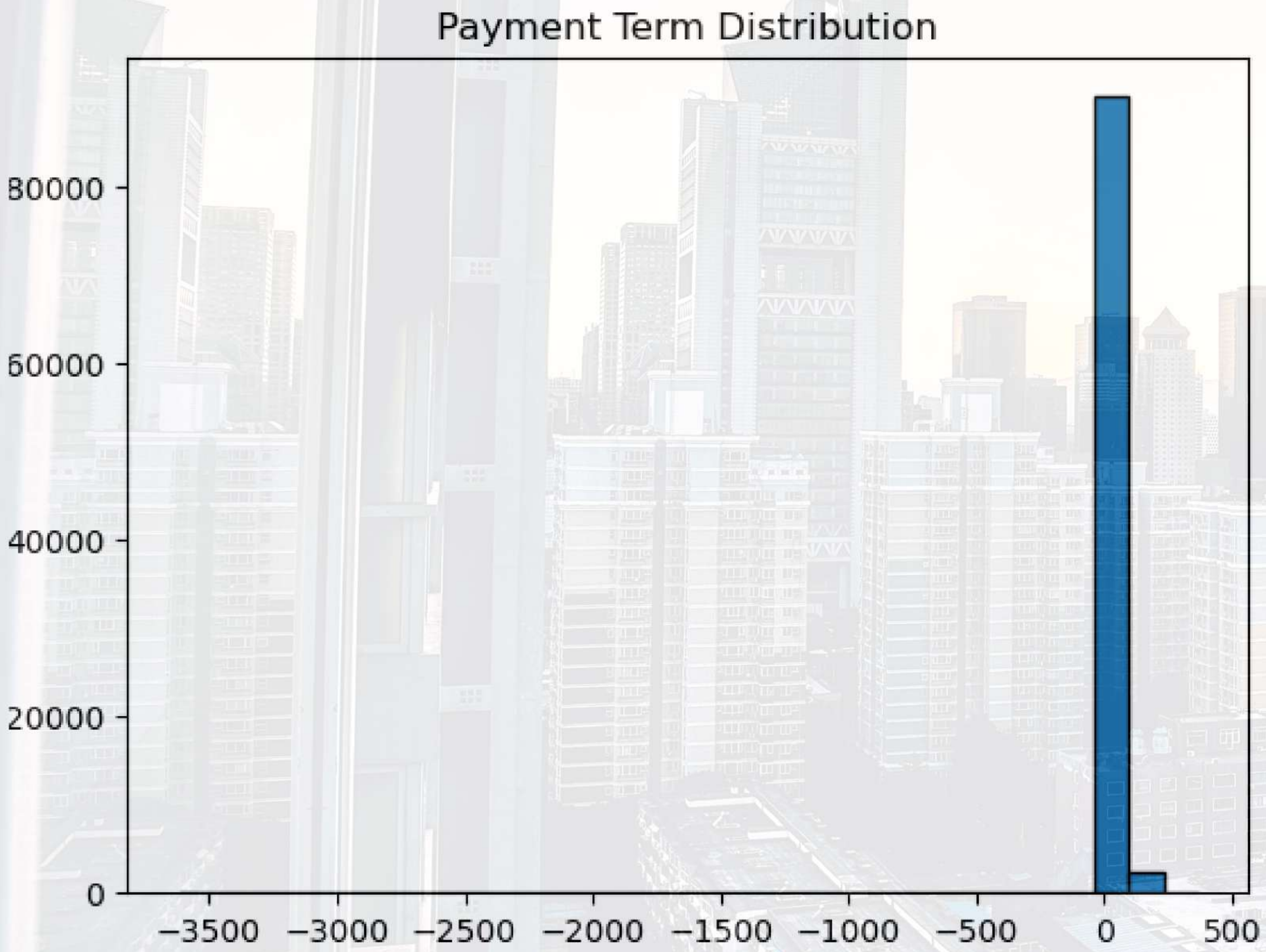
STEPS INVOLVED

- Data Understanding
- Data Cleaning
- Data Visualisation
- Customer Segmentation
- Model Building & Evaluation
- Prediction on Test Set
- Model Deployment

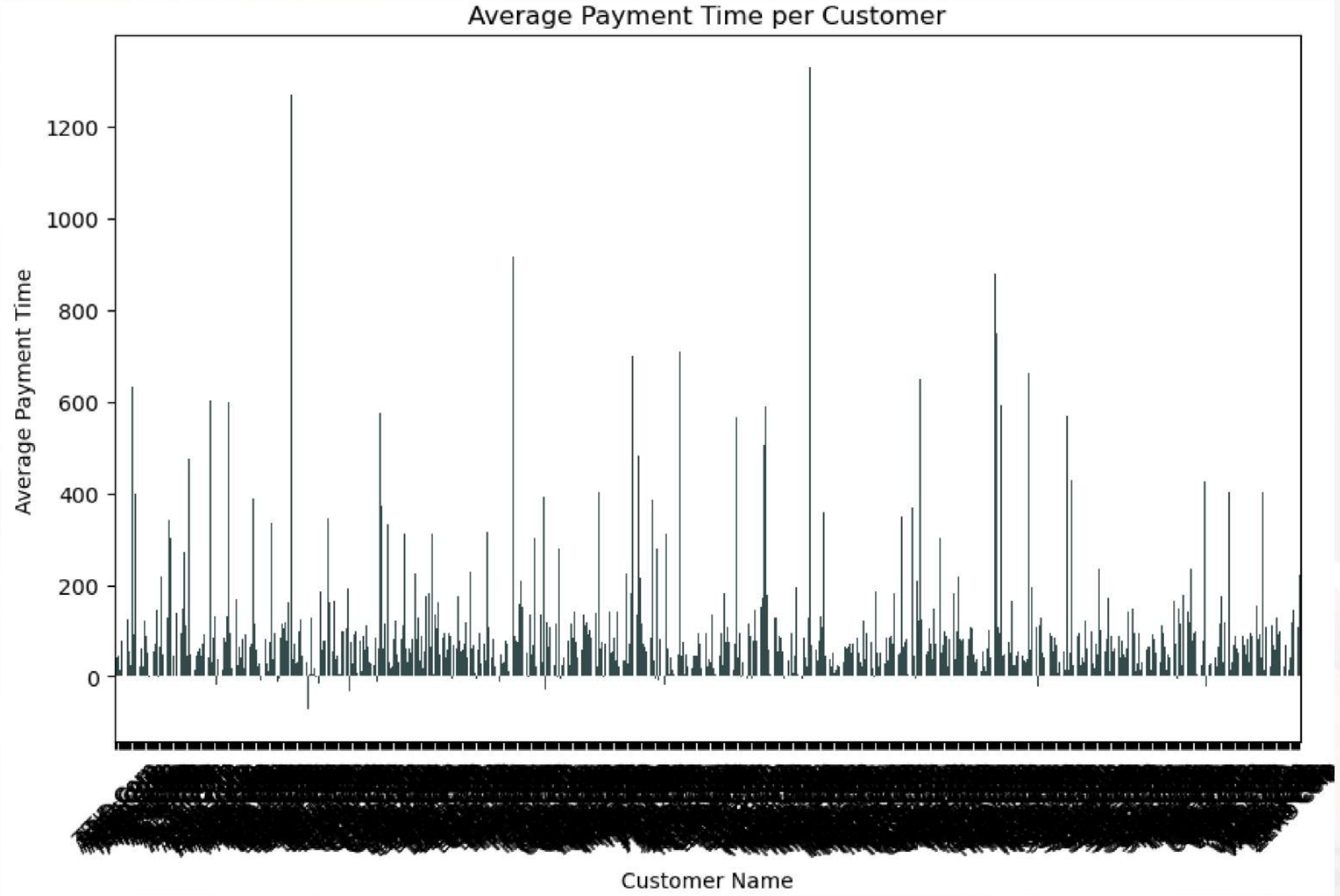
EDA INSIGHTS

The x-axis represents different payment statuses, where 0 corresponds to "not paid" and 1 corresponds to "paid." On the y-axis, we have the number of customers in each payment status category. The height of each bar reflects the count of customers with a specific payment status.



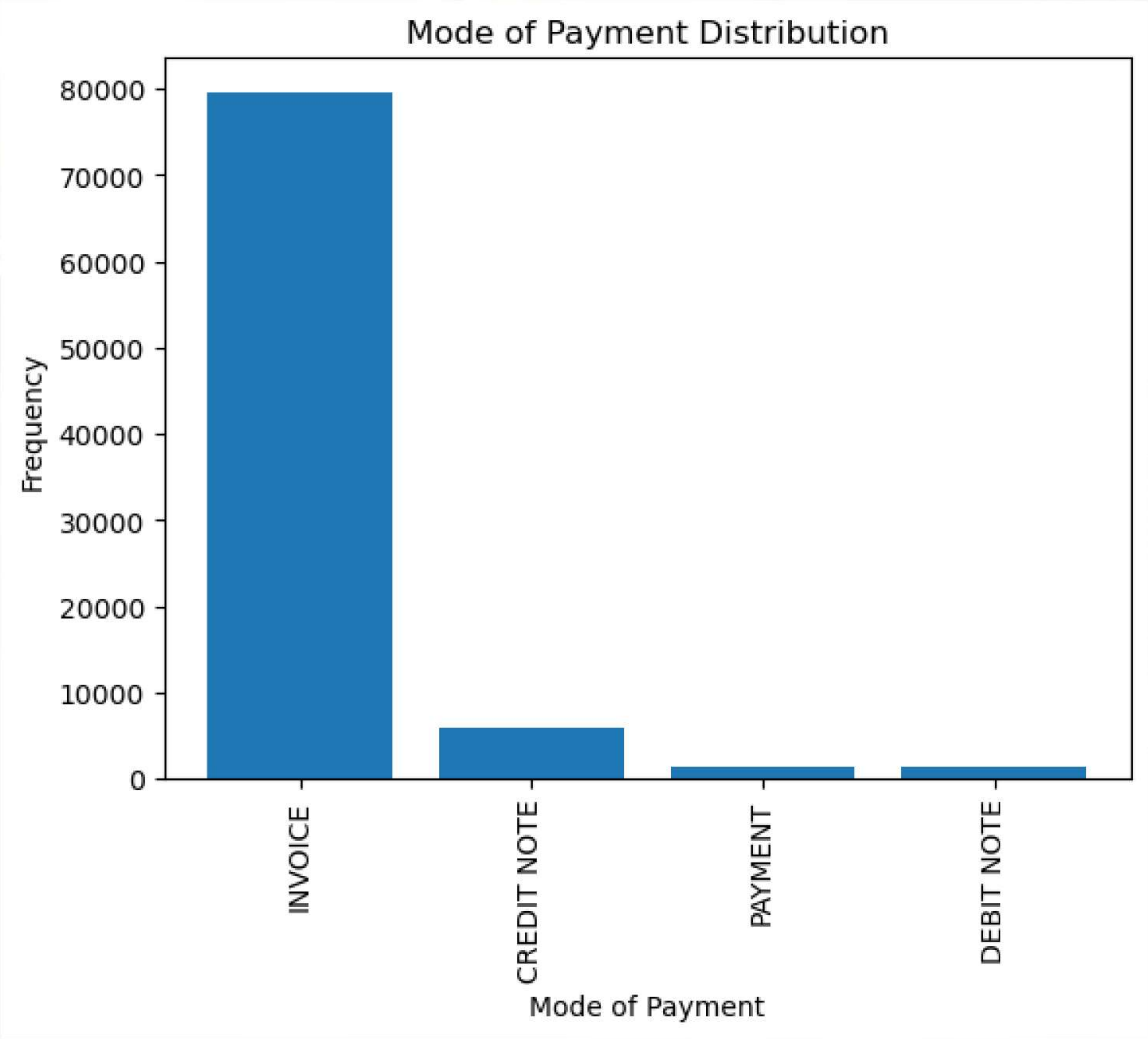


The bar chart displays the distribution of payment statuses. The x-axis represents two categories (typically 0 for "not paid" and 1 for "paid"), while the y-axis shows the count of occurrences for each payment status. Each bar corresponds to the number of customers with a specific payment status. The title of the plot is "Payment Status Distribution."



Payment Consistency:
A3 D Corp: High average payment time (45.4 days) with high variability.
ABC Corp: Consistent payment time (16 days, no variability).
General Payment Behavior:
3D D Corp and 6TH Corp: Moderate payment times (24-40 days).

Average Payment Time Differences:
ABDU Corp: Highest average payment time (˜77 days).
ABC Corp: Shortest average payment time (16 days).



The bar chart presents the frequency of different payment methods in the dataset. It helps identify popular and outlier payment modes, providing insights into customer preferences. This analysis can guide decisions related to payment processing and marketing strategies.

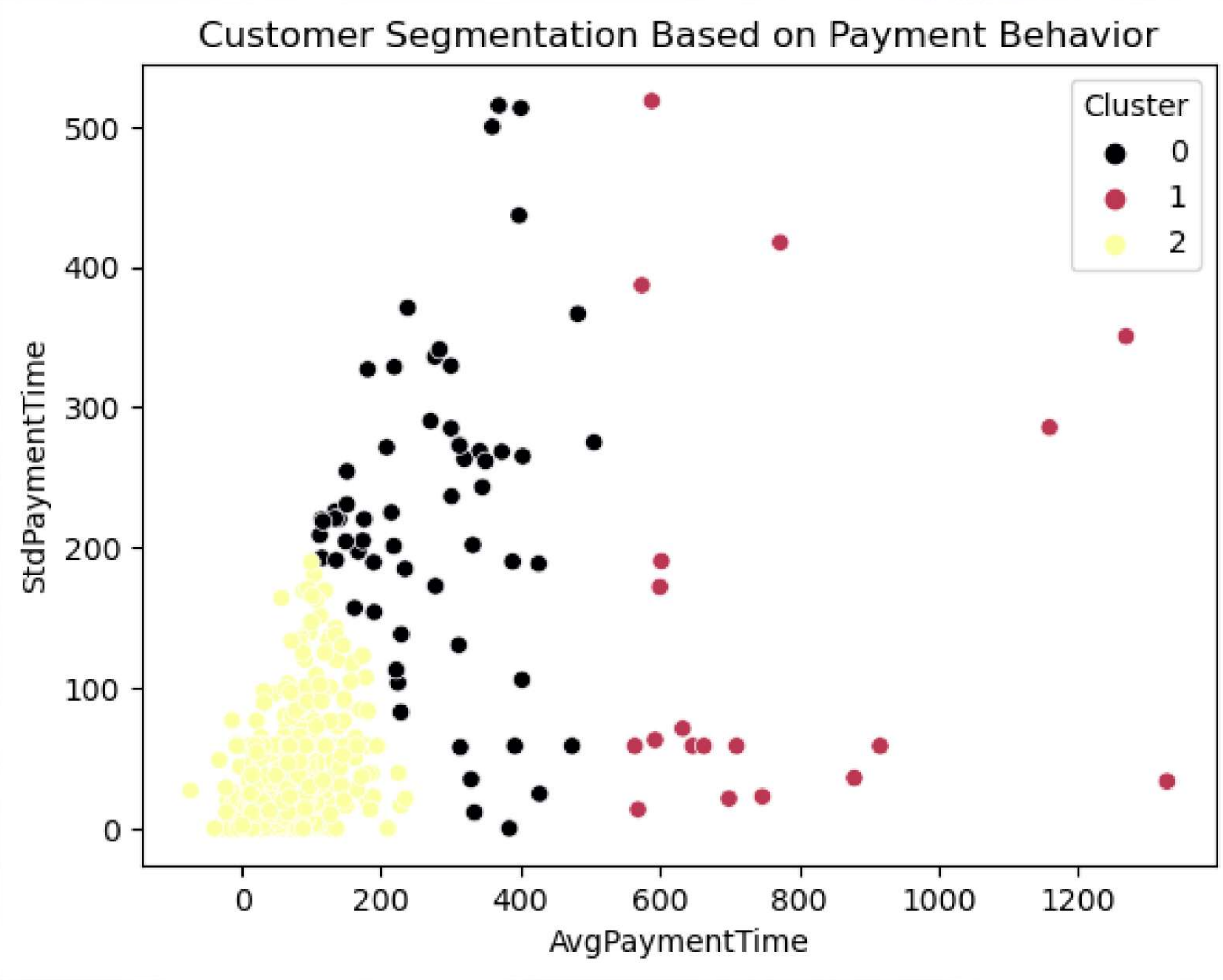
In this invoice has a major frequency mode of payment

CUSTOMER SEGMENTATION

Cluster 0: This group is characterized by significant payment delays, accompanied by a moderate standard deviation (SD). This suggests a degree of inconsistency in the payment timings within this cluster.

Cluster 1: This cluster predominantly consists of early payments, with a small SD indicating a high level of consistency in early payment behavior.

Cluster 2: This cluster exhibits minor payment delays. However, the slightly elevated SD indicates a wider spread in the payment timings, suggesting some variability within this group.

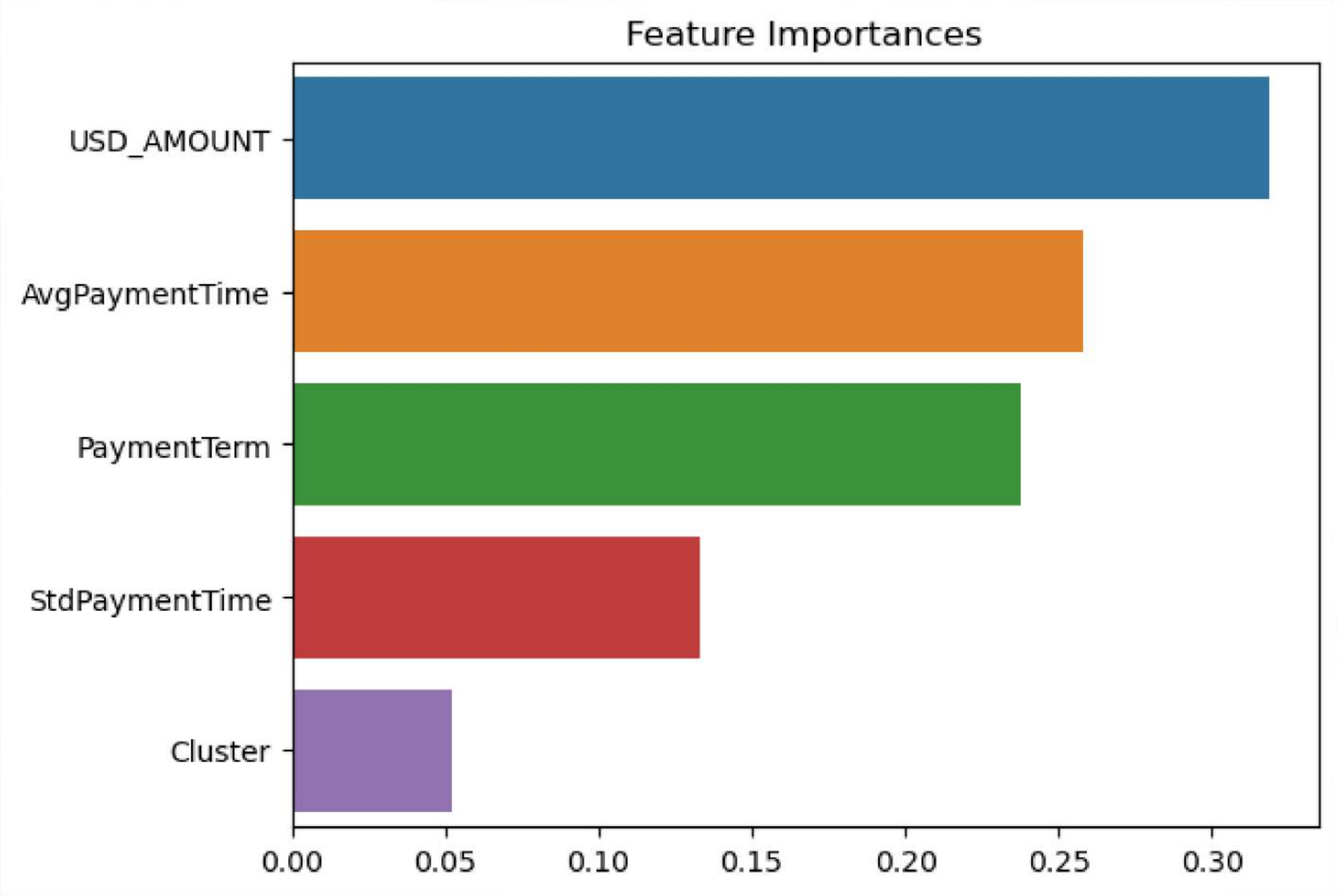


MODEL TRAINING

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00       1.0
           1       0.00      0.00      0.00       1.0

    accuracy                           0.00       2.0
   macro avg       0.00      0.00      0.00       2.0
weighted avg       0.00      0.00      0.00       2.0

[[0 1]
 [1 0]]
Accuracy: 0.0
```

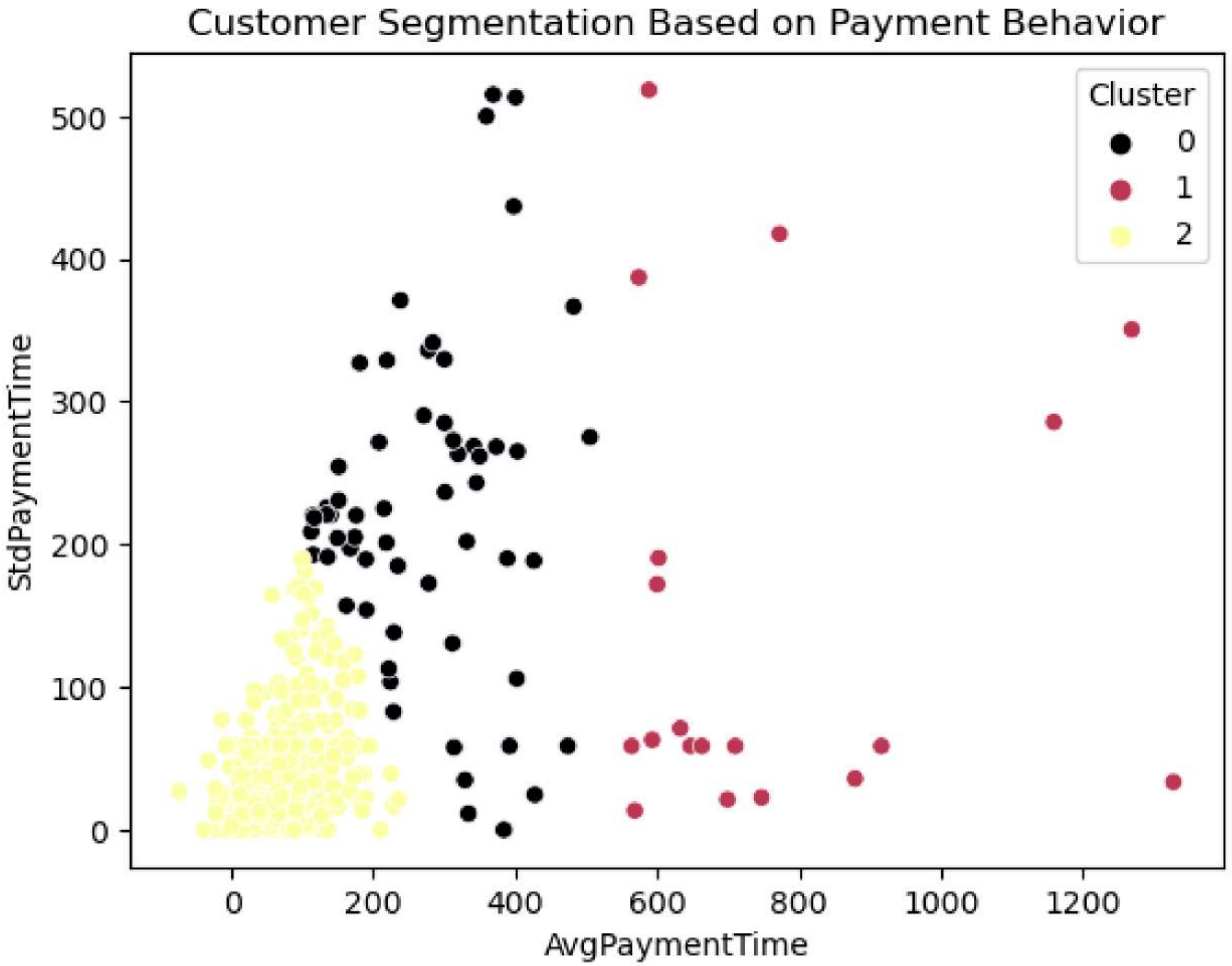
The Random Forest Classifier model, trained with a random state of 42, was tested on a dataset. Unfortunately, the model was unable to accurately predict instances for both class 0 and class 1. The confusion matrix reveals that an instance from class 0 was misclassified as class 1, and vice versa. Consequently, the model's accuracy score was 0.0, indicating that it failed to make any correct predictions on the test data. This suggests a need for model refinement or a review of the feature selection process.

MODEL EVALUATION & APPLICATION



These features are likely to be the most influential in determining the outcome of the model's predictions.

Model on Open Invoice Data



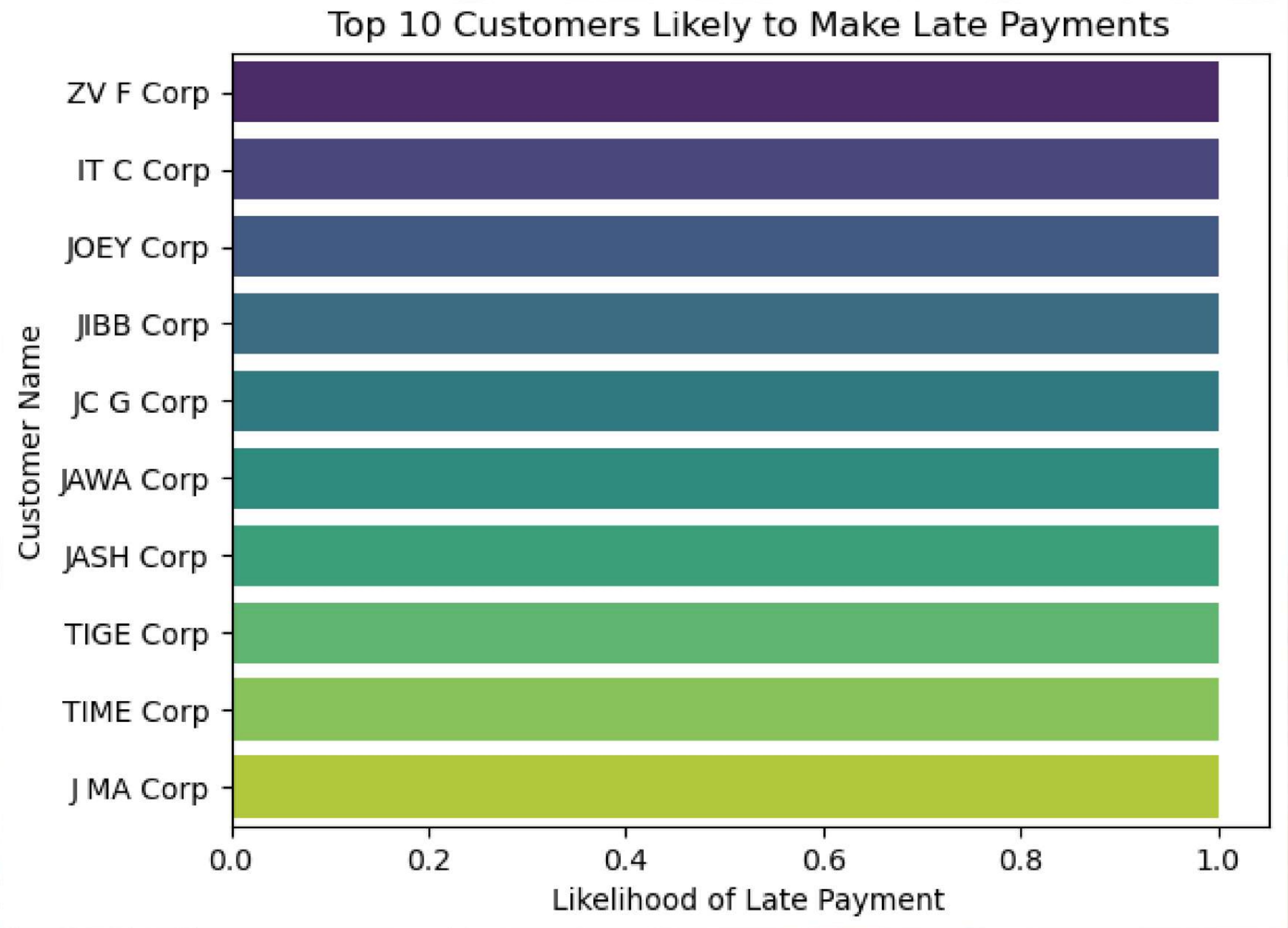

```
AvgPaymentTime     0
StdPaymentTime     0
dtype: int64
AvgPaymentTime     0
StdPaymentTime     0
dtype: int64
```

The output confirms that the 'AvgPaymentTime' and 'StdPaymentTime' columns in the dataset are complete with no missing values. Subsequently, KMeans clustering is applied to segment customers into three distinct groups based on their payment behavior.

The clusters are then visualized using a scatter plot. In this plot, each point corresponds to a customer. The x-axis represents the average payment time, while the y-axis denotes the standard deviation of payment time. Different colors are used to distinguish between the customer clusters.

This visualization provides a clear understanding of the different customer segments based on their payment patterns, which can be instrumental in shaping strategic decisions.

Results and Recommendations



Top Customers likely to make late payments

Key Predictors

1. Payment Term:

Description: The number of days given to a customer to settle an invoice.
Significance: Longer payment terms are associated with an increased risk of late payments. This is because customers have a longer window to potentially experience cash flow issues that could delay payment.

2. Average Payment Time (AvgPaymentTime):

Description: The average time it typically takes for customers to pay their invoices.
Significance: Customers with a history of slow payments are likely to continue this trend. This feature allows the model to identify customers who may require additional follow-up or stricter payment terms to ensure timely payment.

3. Standard Deviation of Payment Time (StdPaymentTime):

Description: The level of variation in how long it takes customers to pay their invoices.
Significance: Customers with inconsistent payment patterns pose a higher risk of late payments. This feature helps identify customers whose payment behavior is unpredictable, requiring closer monitoring or potential adjustments to credit limits.

Recommendations

1. Delinquency Prediction and Prioritization:
Develop a system to leverage model predictions to identify customers with a high risk of late payments.
Implement a prioritization mechanism to allocate resources for early intervention on high-risk customers.

2. Customer Segmentation for Collections:
Utilize existing clustering analysis results to segment customers based on payment behavior.
Define collection strategies tailored to the specific characteristics and needs of each customer segment.

3. Early Payment Incentives:
Evaluate the feasibility and impact of offering discounts or incentives for early payments.
Design a system to automate the application of these incentives during the payment process.

4. Improved Communication for Payment Reminders:
Implement a system to send automated reminders and notifications to customers approaching payment deadlines.
Ensure communication is clear, timely, and informs customers of potential consequences for late payments.

5. Performance Monitoring and Model Maintenance:
Establish processes to monitor key metrics related to payment collections, including delinquency rates and the effectiveness of collection strategies.
Regularly review and update the predictive model with new data to maintain its accuracy in identifying at-risk customers.

Conclusion

Deliverables:

- A comprehensive report outlining the data analysis, model development process, and evaluation results.
- A well-documented, functional late payment prediction model.
- Recommendations for collectors on using the model to identify and manage high-risk late payments.

Success Criteria:

- Increased percentage of on-time payments.
- Reduced average days to collect payments.
- Improved cash flow predictability.