

PREDICTIVE ANALYSIS FOR CUSTOMER CHURN IN E-COMMERCE

Submitted by:
21 Bhakti Dwivedi
26 Simran Gupta
31 Tanvee Joshi
33 Archana Kalathiya

OVERVIEW

01

INTRODUCTION

02

PROBLEM STATEMENT

03

PROPOSED SYSTEM

04

WORKFLOW

05

H/W S/W REQUIREMENTS

06

IMPLEMENTATION

07

RESULTS

08

CONCLUSION

09

FUTURE SCOPE

INTRODUCTION

In the dynamic landscape of e-commerce, where competition is fierce and customer preferences are ever-evolving, businesses face the critical challenge of retaining their customer base. Customer churn, the phenomenon where customers cease their engagement with a brand, has a significant impact on a company's revenue and growth prospects. As such, the ability to predict and mitigate customer churn has become a strategic imperative for e-commerce enterprises. By analysing vast volumes of customer data encompassing behaviours, interactions, purchase histories, and more, predictive models can discern subtle patterns and indicators that signal potential churn. The introduction outlines the pivotal role that customer churn prediction plays in modern e-commerce operations. It highlights the economic implications of churn and underscores the potential of predictive analytics in minimising churn rates. This project encompasses various phases, including data collection, preprocessing, feature engineering, model selection, and evaluation. Each stage contributes to the overarching goal of creating accurate and reliable churn prediction models. By building and fine-tuning predictive models, we aim to predict which customers are likely to churn soon. Through this endeavour, we seek to demonstrate the potential of data-driven decision-making in mitigating churn rates and fostering customer loyalty.



PROBLEM STATEMENT

In the realm of e-commerce, the critical challenge of retaining customers has become increasingly complex due to shifting consumer preferences, intense competition, and the ease of switching between brands. Customer churn, the phenomenon where customers discontinue their engagement with a brand, poses a significant threat to the growth and profitability of e-commerce enterprises. The cost of acquiring new customers is substantially higher than retaining existing ones, making churn prevention a priority for businesses. The primary objective of this project is to explore the application of predictive analytics for customer churn in the e-commerce domain.



PROPOSED SYSTEM

Objective n° 1

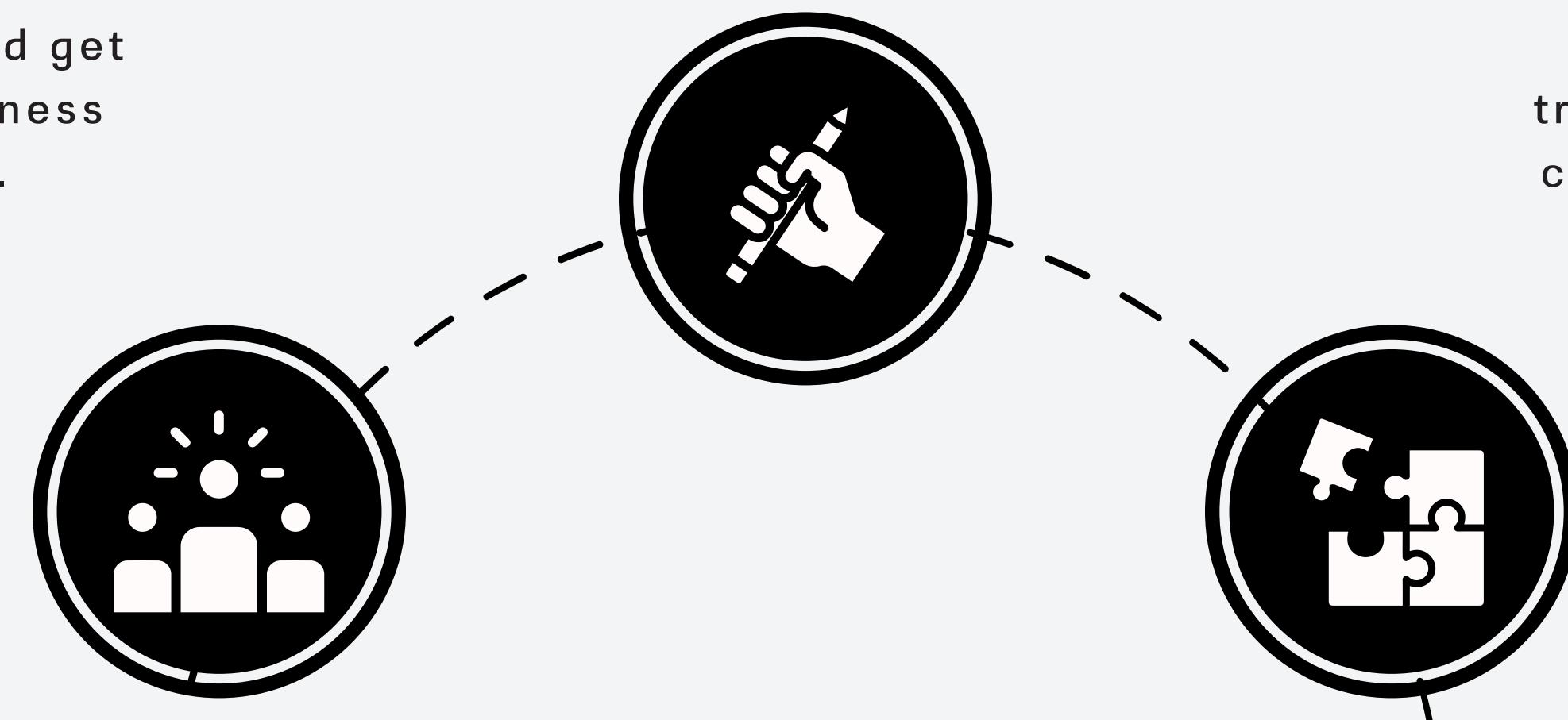
Predict the type of customer that has the potential to churn by identifying features to minimize customer churn rates and get the right business decisions.

Objective n° 2

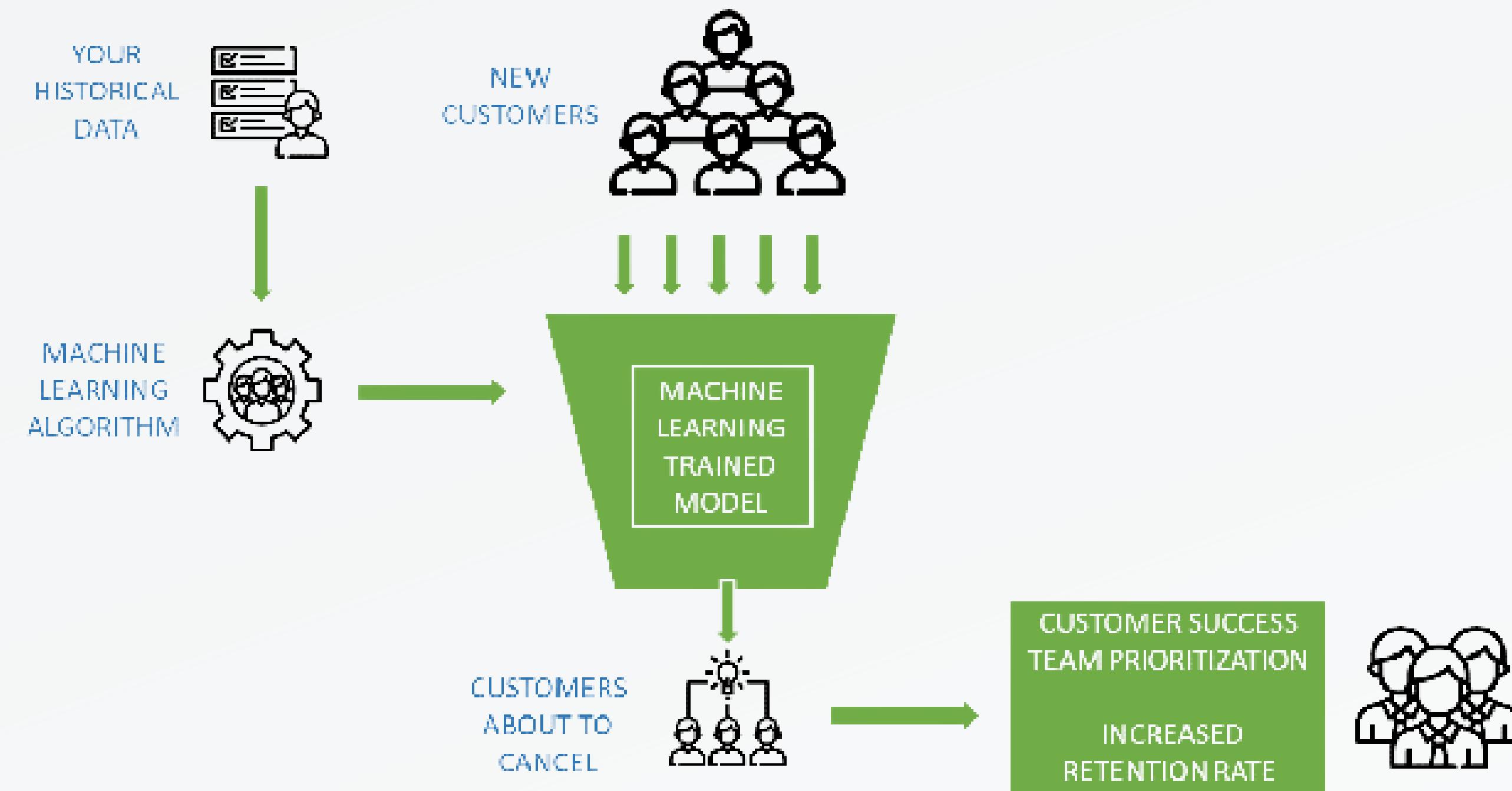
Create a model to predict users who have the potential to churn and understand what features cause users to churn and features that spur potential to minimize churn rates.

Objective n° 3

Optimizing the company's revenue by grouping users who have the potential to churn so that they can be given different treatments so that the churn rate decreases.



WORKFLOW



WORKFLOW

Collection of
Data & Data
Exploration

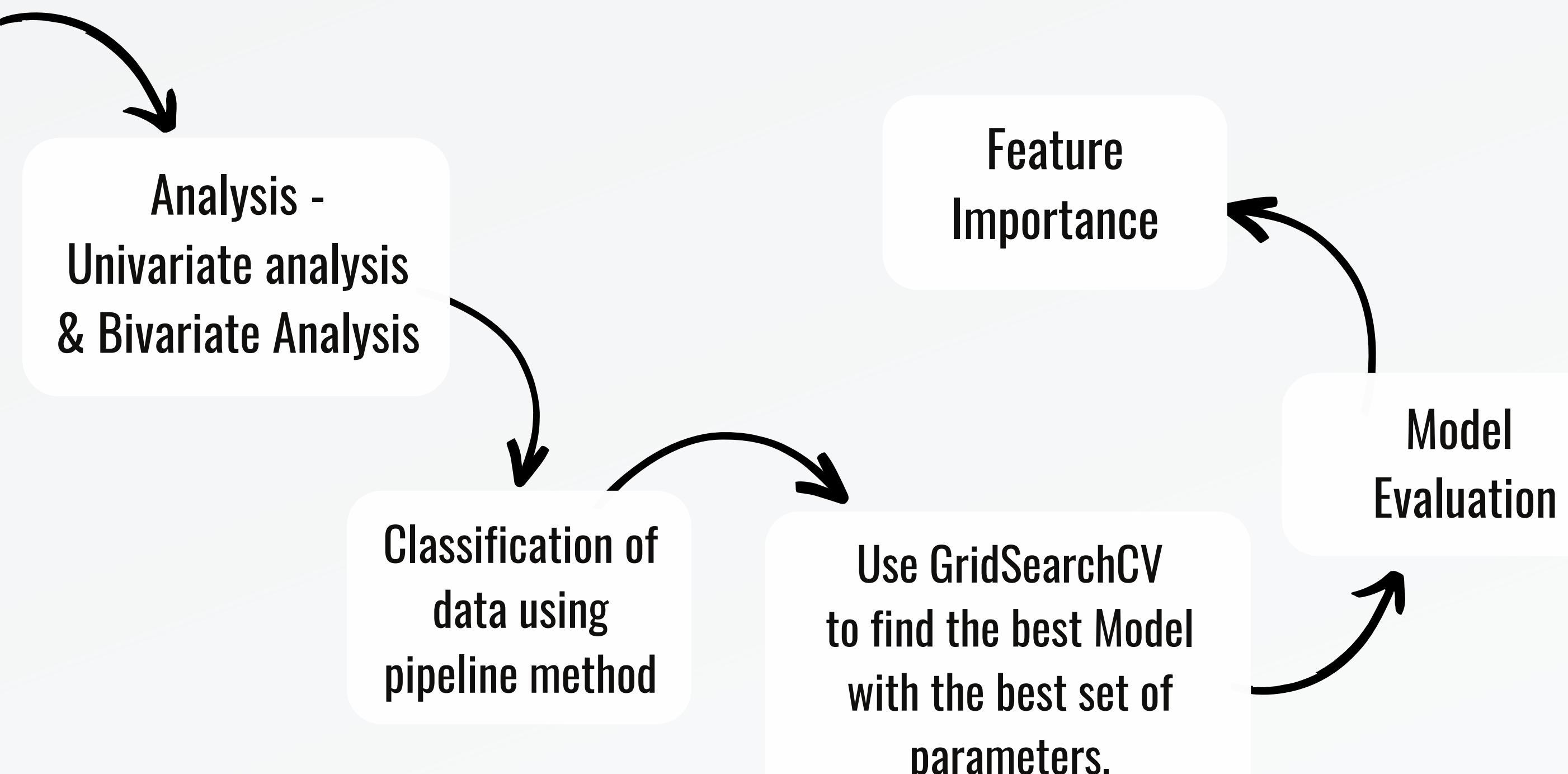
Analysis -
Univariate analysis
& Bivariate Analysis

Classification of
data using
pipeline method

Feature
Importance

Use GridSearchCV
to find the best Model
with the best set of
parameters.

Model
Evaluation



SOFTWARE AND HARDWARE REQUIREMENTS

Hardware Specification

Processor: Quad-core or higher

RAM: 8GB or higher

Storage: At least 256GB SSD

Internet Connection: Required for
downloading libraries and datasets.



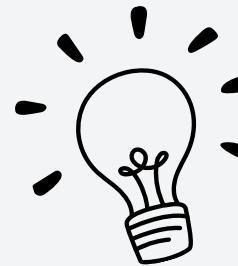
Software Specification

Operating System: Windows

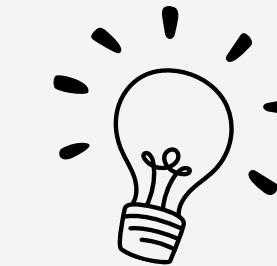
Programming Language: Python

Python Interpreter: Jupyter Notebook

Libraries: Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, XGBoost. These libraries are used for data manipulation, visualization, machine learning, and analysis.

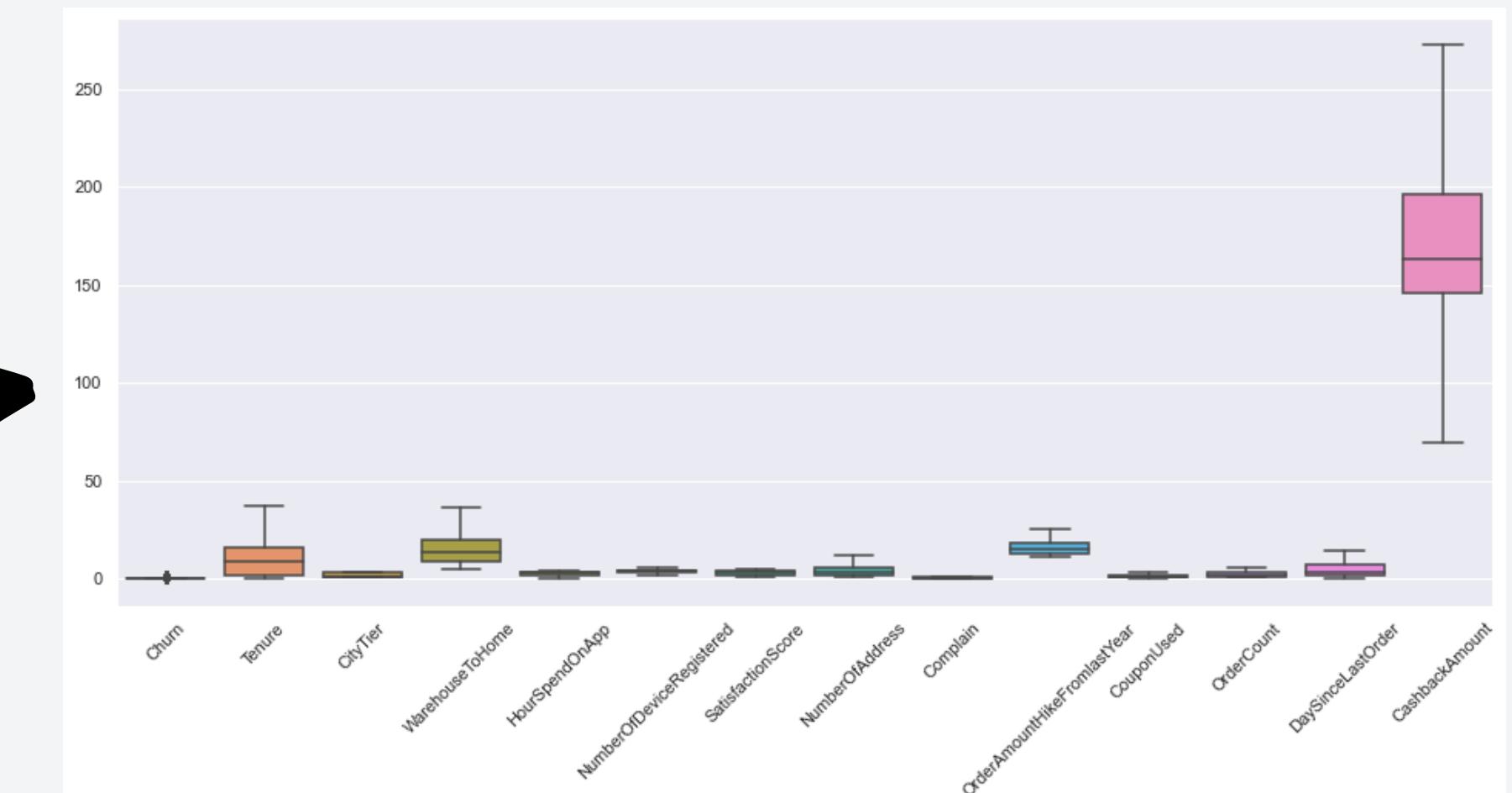
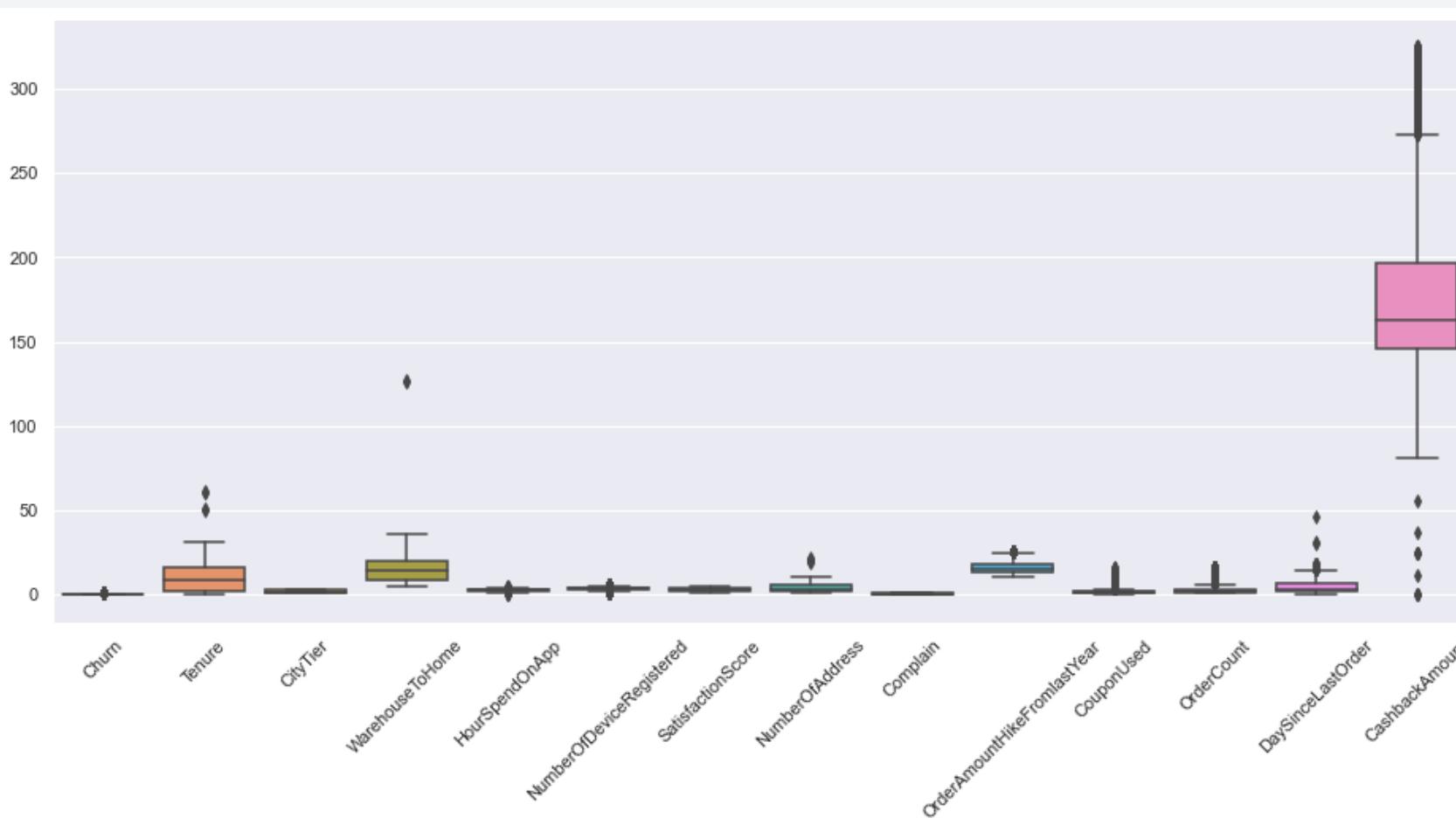


IMPLEMENTATION



1. Data Import

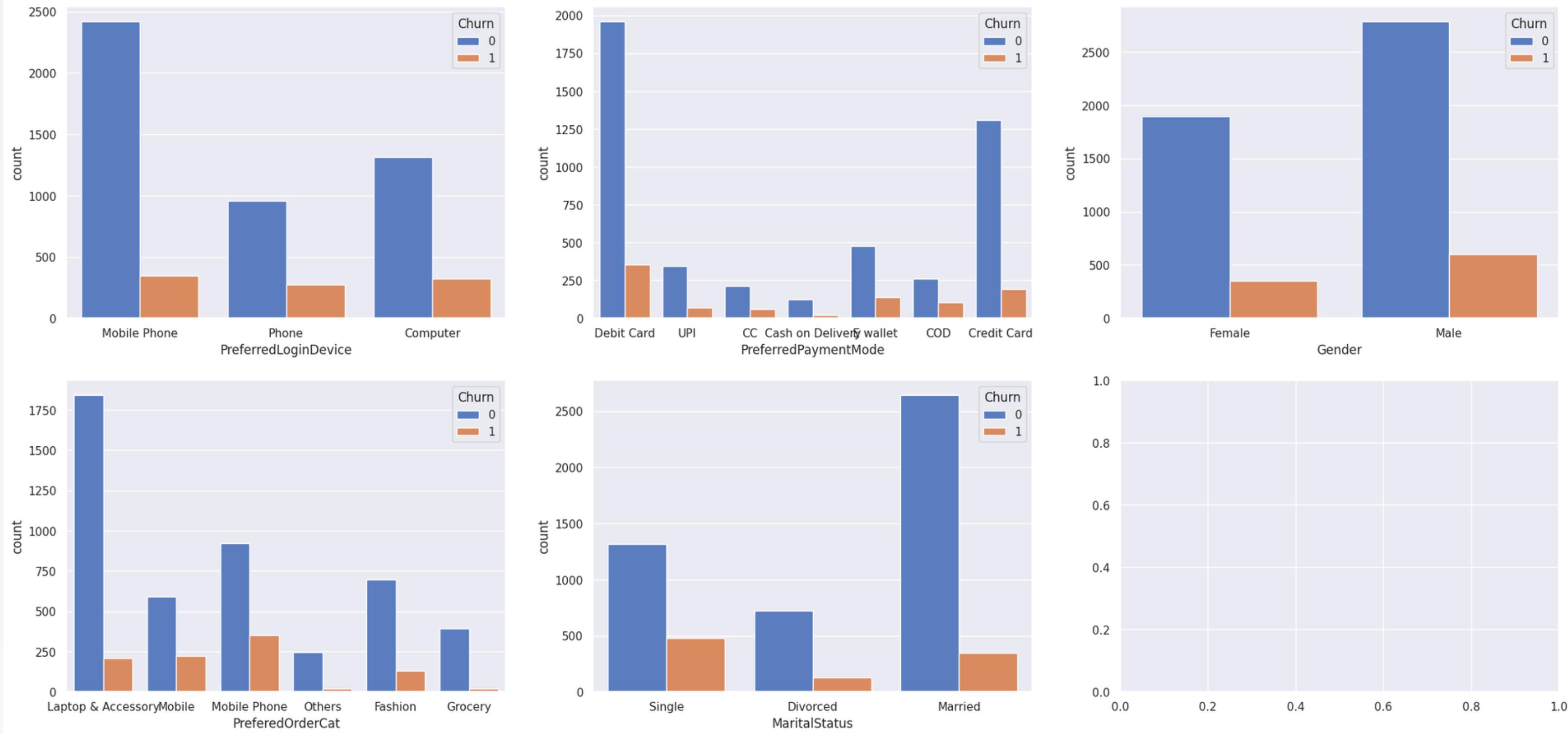
2. Data Cleaning



3. Data Exploration

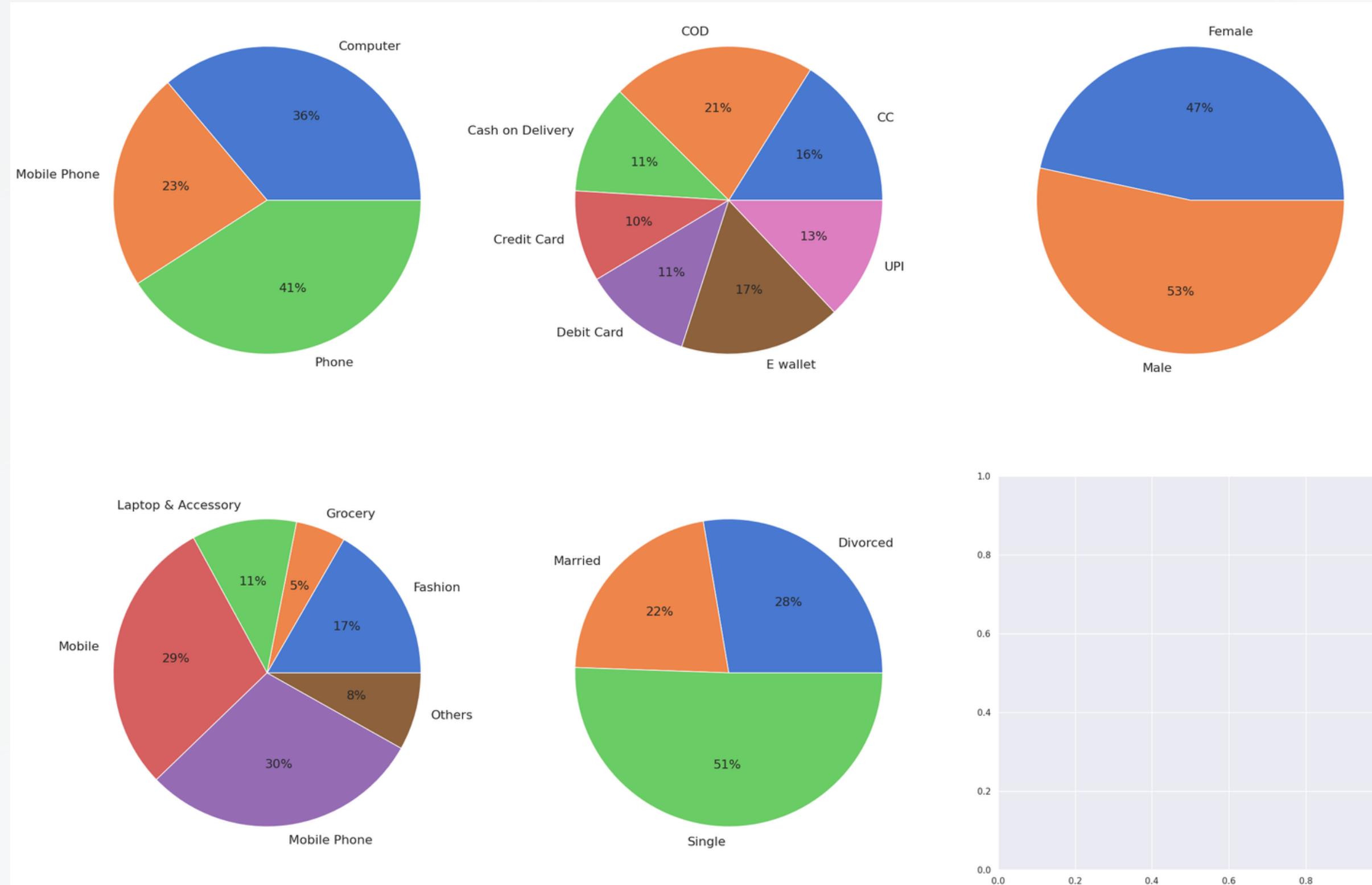
IMPLEMENTATION

4. Univariate Analysis



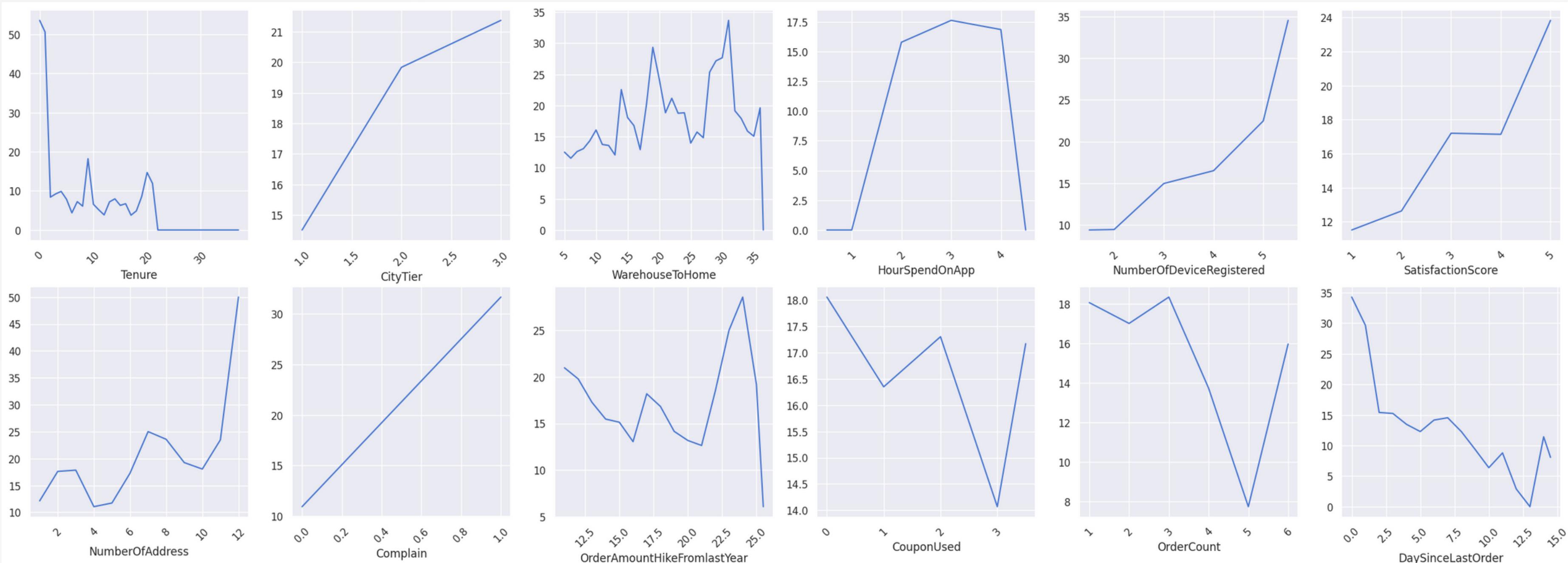
IMPLEMENTATION

4. Univariate Analysis



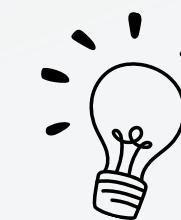
IMPLEMENTATION

4. Univariate Analysis

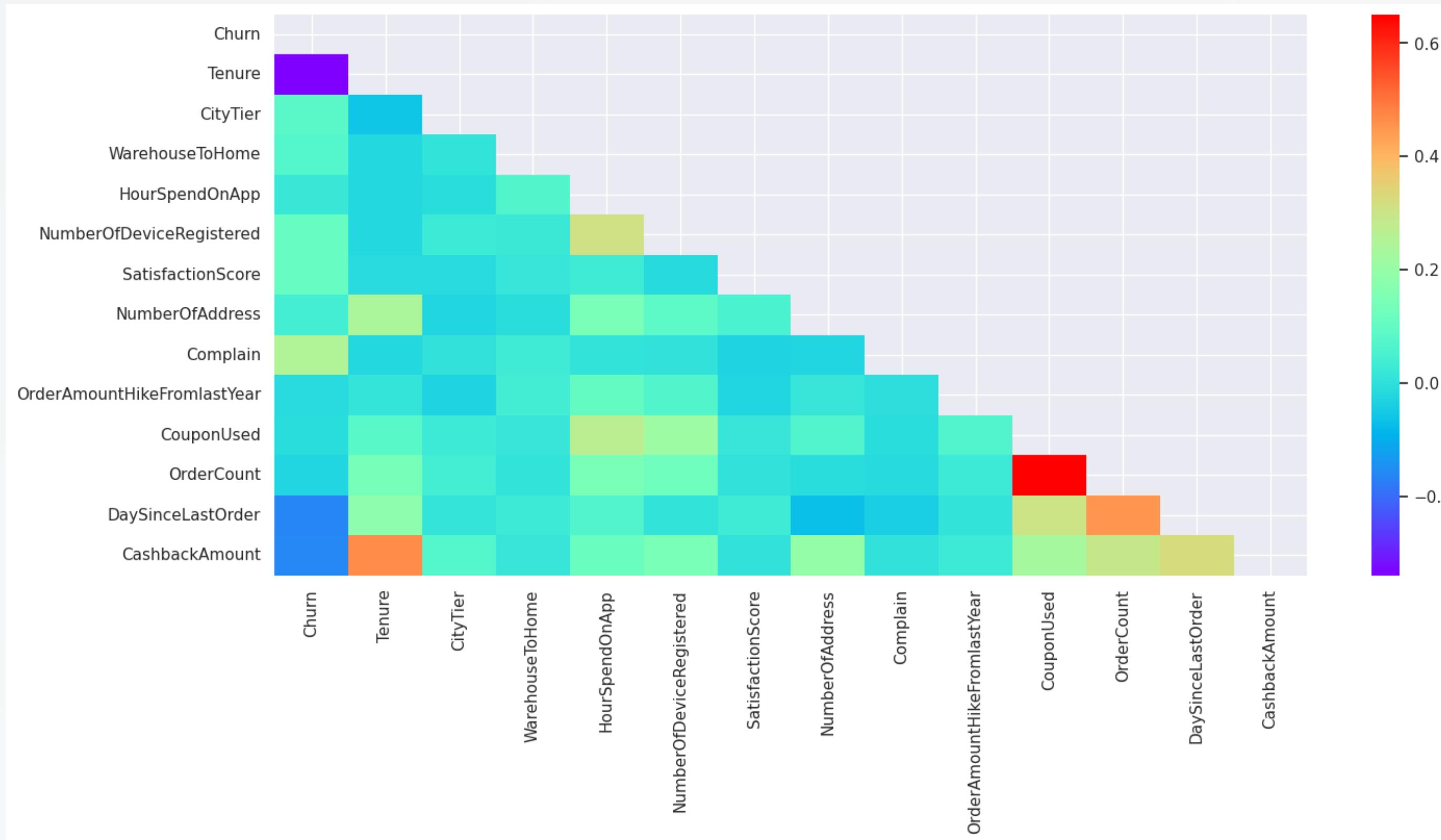


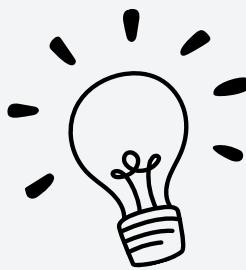


IMPLEMENTATION

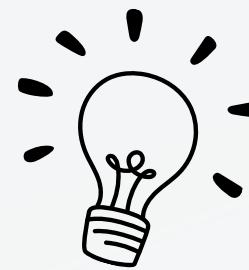


5. Bivariate Analysis





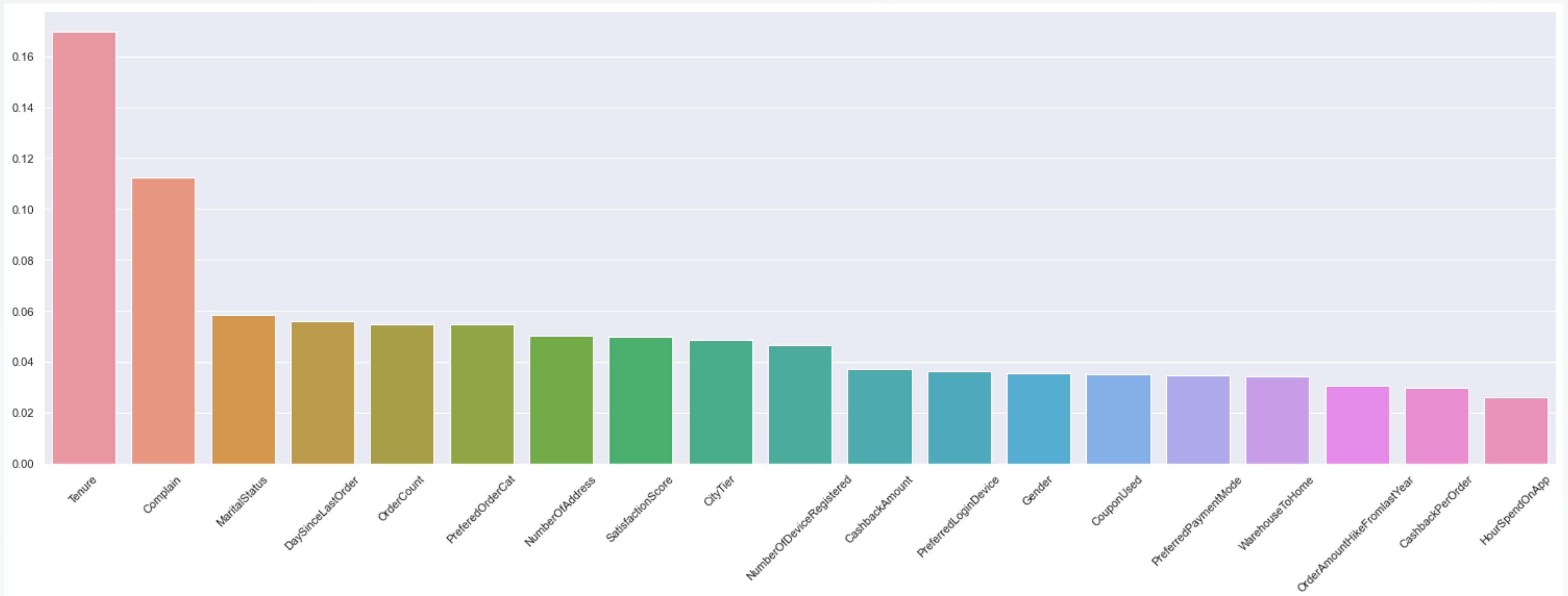
IMPLEMENTATION



6. Data Preprocessing
7. Customer Classifier and Pipeline
8. GridSearchCV for Model Selection
9. Model Evaluation

IMPLEMENTATION

10. Feature Importance



RESULTS

1. **Confusion Matrix:** A confusion matrix is a useful tool for understanding the performance of a classification model. It provides a breakdown of predicted and actual class labels.

In our analysis:

True Positives (TP): $1.2e+03$

False Positives (FP): 11

False Negatives (FN): 31

True Negatives (TN): $2.1e+02$



RESULTS

2. **F1 Score:** The F1 score is a metric that combines precision and recall. It is especially useful when dealing with imbalanced datasets.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

In our analysis:

F1 Score: 0.915

3. **Training and Testing Scores:** Training and testing scores help assess the model's performance on the data it was trained on and unseen data, respectively.

Training Score (Accuracy): 0.9

RESULTS

Interpretation:

- The confusion matrix shows that our model correctly predicted 1.2e+03 instances as positive and 2.1e+02 instances as negative, but it made 11 false positive and 31 false negative predictions.
- The F1 score of 0.915 indicates a good balance between precision and recall.
- Our model achieved a high training accuracy of 0.96, indicating it fits the training data well

CONCLUSION

This comprehensive analysis serves as a guiding light for e-commerce businesses aiming to strengthen customer retention efforts. Leveraging advanced analytics, machine learning, and NLP, companies can not only predict customer churn but also devise personalised strategies to combat it. By proactively addressing churn drivers and enhancing the customer experience, e-commerce businesses can position themselves for sustained growth and success in a highly competitive market. As data continues to play a pivotal role in shaping business strategies, the insights presented in this analysis offer a valuable blueprint for e-commerce practitioners seeking to navigate the complex terrain of customer churn prediction and management. This proactive approach not only preserves revenue but also enhances customer satisfaction, fostering long-term success in the competitive e-commerce landscape.

FUTURE SCOPES

01

02

03

04

05

PERSONALIZED
CUSTOMER
ENGAGEMENT

REAL-TIME
CHURN
PREDICTION

ENHANCED
CUSTOMER
FEEDBACK
ANALYSIS

CROSS-
CHANNEL
INSIGHTS

ETHICAL
CONSIDERATI
ONS AND
DATA
PRIVACY

THANK YOU

