**Takeo**

**Data Analytics and AI Bootcamp - BDA66**

**Project Report:**

**Customer Lifetime Value (CLV) Prediction**

**Submitted by**

**Aakriti K C**

**2/28/2025**

## Objective

The primary objective of this project is to analyze customer behavior and churn patterns to calculate Customer Lifetime Value (CLV) for a telecom company. By understanding CLV, we aim to improve customer retention strategies, optimize pricing, and enhance marketing efforts to increase long-term profitability.

## Introduction

Customer Lifetime Value (CLV) is a critical metric that helps businesses estimate the total revenue a customer is expected to generate during their tenure. By identifying factors that influence CLV, businesses can implement targeted strategies to maximize customer retention and overall revenue.

In this project, we used Exploratory Data Analysis (EDA) and Machine Learning (ML) techniques to analyze customer data, predict churn, and estimate CLV. Our analysis covers key variables such as tenure, monthly charges, total charges, contract type, payment method, internet service type, and additional services.

## Problem Statement

E-commerce businesses rely on predicting Customer Lifetime Value (CLV) to make data-driven decisions for marketing, customer retention, and revenue optimization. However, accurately estimating CLV can be challenging due to variations in customer behavior, purchase frequency, and retention rates. This project aims to develop a predictive model using Python, SQL, and statistical analysis to forecast CLV based on customer demographics, transactional data, and purchasing patterns. By leveraging advanced data processing and machine learning techniques, businesses can enhance marketing strategies, optimize customer acquisition, and maximize profitability.

4Ws for the Project

What: E-commerce businesses struggle to predict Customer Lifetime Value (CLV) accurately, affecting their ability to allocate resources efficiently for marketing and customer retention.

Why: A robust CLV prediction model allows businesses to improve customer acquisition strategies, personalize marketing campaigns, reduce churn, and increase overall revenue.

Who: E-commerce businesses, marketing teams, data analysts, and decision-makers who rely on CLV insights for strategic planning.

Where: In the e-commerce industry, where businesses need to analyze customer behavior, purchase history, and retention patterns to make informed decisions.

## Process Overview

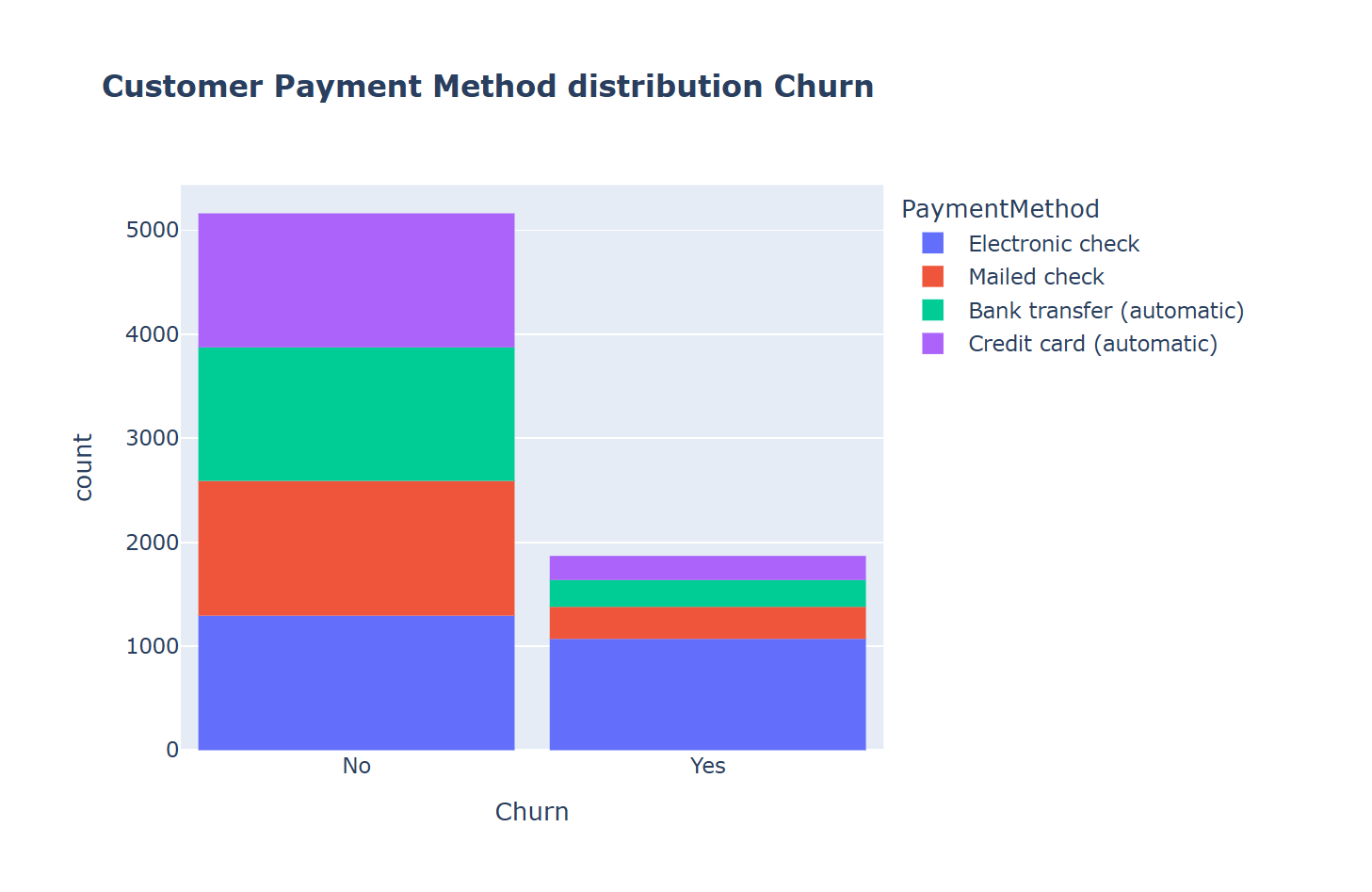
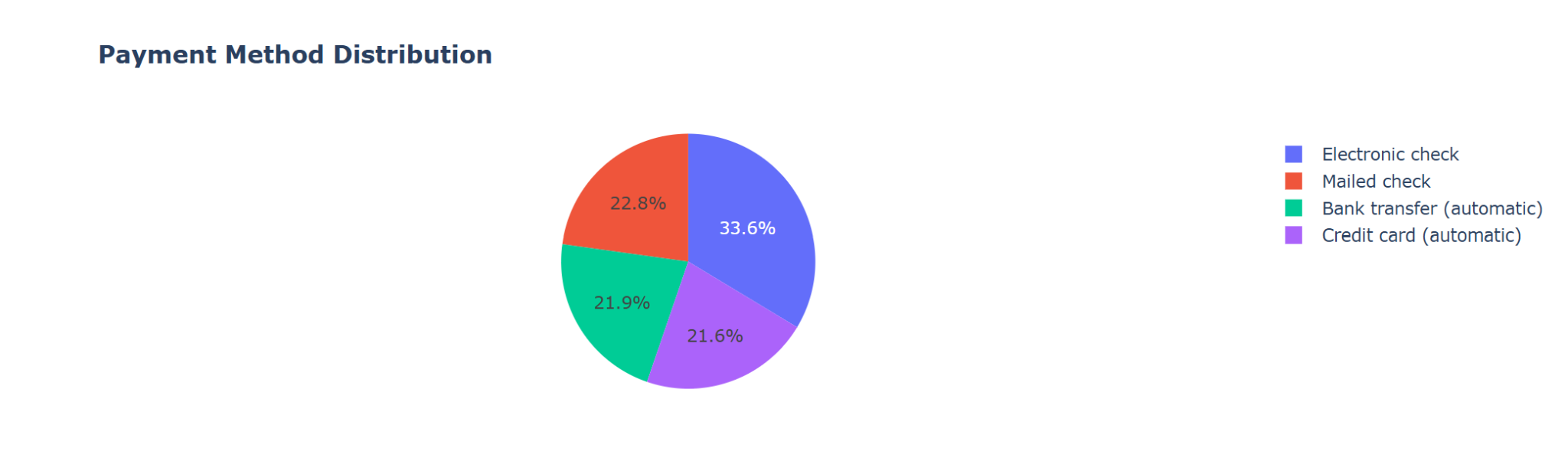
The following steps were followed in our analysis:

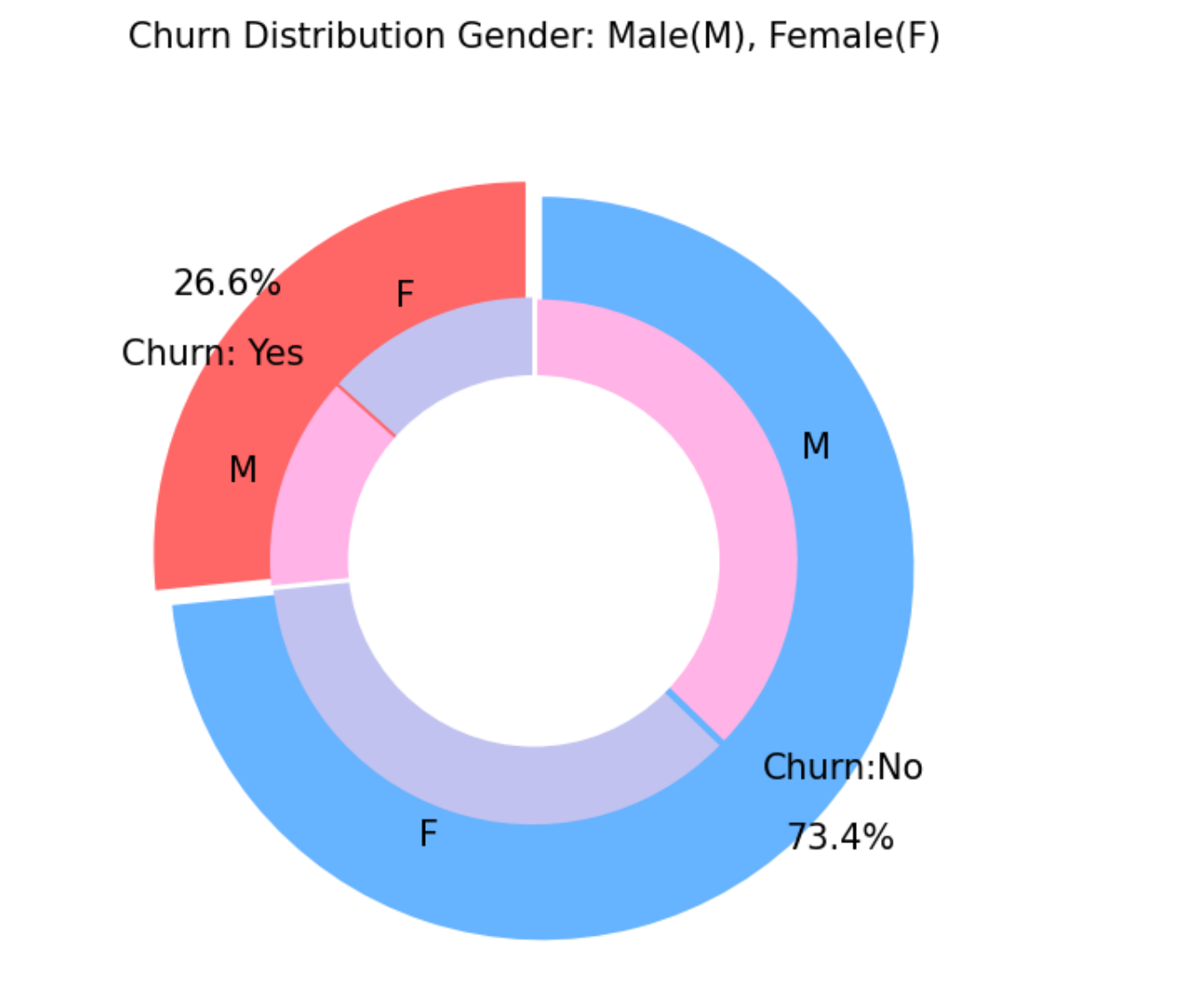
1. Data Collection & Preprocessing:

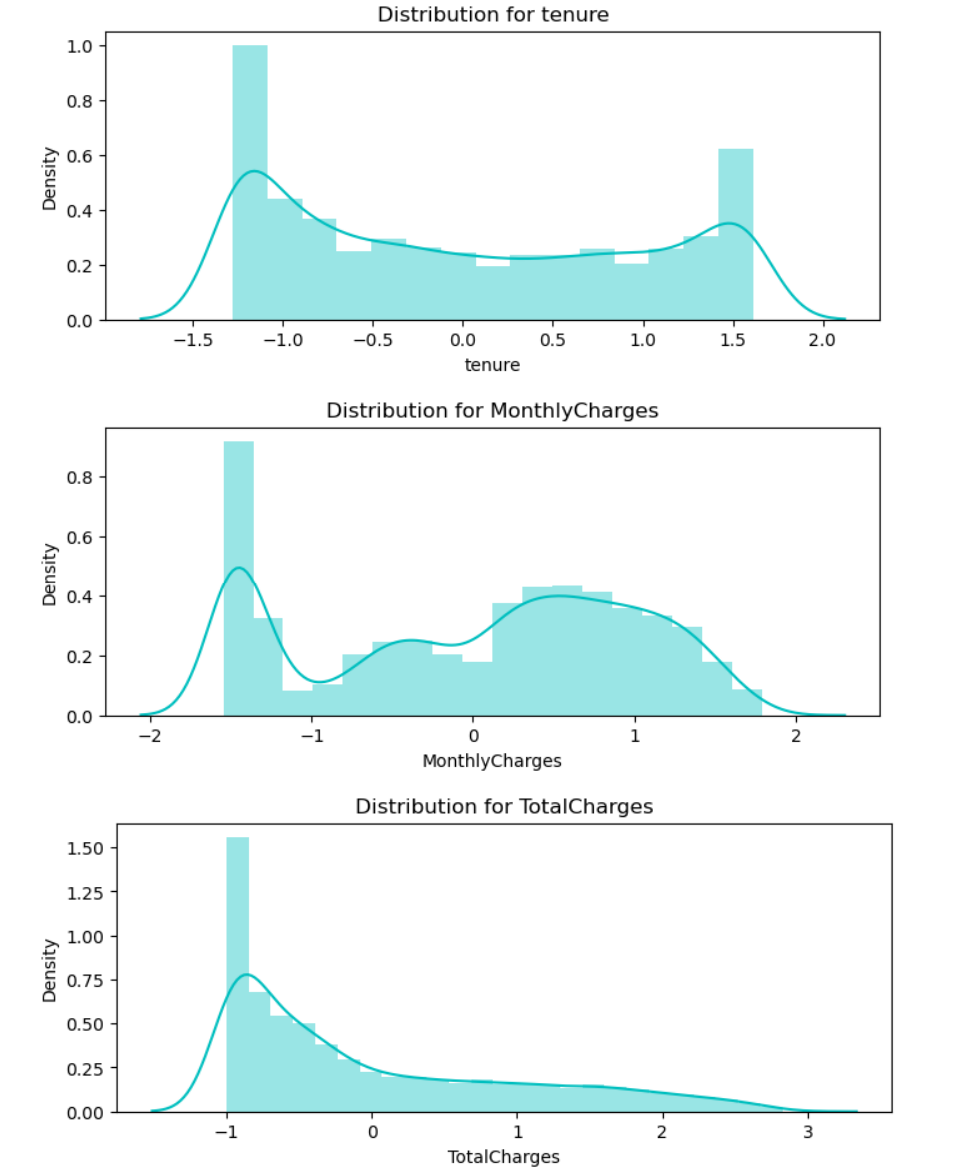
* Loaded customer transaction and subscription data.
* Cleaned missing values and handled categorical variables.
* Exploratory Data Analysis (EDA):

1. Conducted statistical analysis and visualized key patterns in the data.

* Identified trends in churn, payment methods, contract types, and additional services.







1. Customer Segmentation:

Segmented customers based on tenure, monthly charges, payment behavior, gender and senior age.

1. Churn Prediction Model:

* Trained classification models (e.g., Logistic Regression, Decision Tree, Random Forest) to predict churn probabilities.

1. Results Visualization:

* Visualized churn distribution across different customer groups.
* Identified key factors affecting customer longevity.

## Challenges and Solutions

1. Customer Data Variability:

* Some customers had missing or inconsistent data in tenure and total charges.
* Solution: Handled missing values using mean imputation and cross-validation for robust insights.

1. Feature Selection for CLV Calculation:

* Certain features (e.g., tenure and total charges) were highly correlated, which could lead to redundancy.
* Solution: Used correlation heatmaps and Variance Inflation Factor (VIF) to retain only the most informative variables.

1. Model Performance Optimization:

* Some models overfitted the training data, leading to poor generalization.
* Solution: Used cross-validation and hyperparameter tuning to enhance model accuracy.

## Results

1. Churn and CLV Insights

* Month-to-month contract holders had the highest churn rates.
* Customers using electronic check payments were more likely to churn.
* Long-term contracts significantly increased CLV due to lower churn rates.

2. Model Performance Summary

To predict churn and refine CLV estimates, classification models were built and the evaluation was done on their performance:

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **ROC-AUC** |
| --- | --- | --- | --- | --- | --- |
| Logistic Regression | **84.3%** | 79.5% | 75.2% | 77.3% | **0.88** |
| Decision Tree | 81.2% | 76.8% | 72.5% | 74.5% | 0.85 |
| Random Forest | **86.7%** | **82.4%** | **78.9%** | **80.6%** | **0.91** |

* Random Forest performed the best, achieving the highest accuracy (86.7%) and ROC-AUC (0.91).
* Logistic Regression provided a strong baseline with interpretable coefficients for understanding churn drivers.
* Decision Trees helped visualize feature importance but had slightly lower performance.

## Conclusion

The CLV analysis helped uncover key customer segments that drive long-term value and factors contributing to churn. The results show that contract type, payment method, and additional services play a crucial role in customer retention. Among the models, Random Forest achieved the best predictive performance (86.7% accuracy, 0.91 ROC-AUC).

By integrating these findings into business strategy, the telecom company can focus on reducing churn, increasing long-term contracts, and promoting high-value services to maximize CLV.

## Recommendations for future work

1. Personalized Retention Strategies: Implement targeted offers for month-to-month customers to encourage them to switch to long-term plans.
2. Automated Payment Incentives: Offer discounts for auto-pay enrollments to reduce churn from electronic check users.
3. Enhanced Service Bundling: Promote streaming and security services to improve retention.
4. Advanced CLV Modeling: Use time-series forecasting to refine CLV predictions.
5. Real-time Customer Monitoring: Develop dashboards to track customer churn probabilities dynamically.
6. Advanced Machine Learning Models: Explore Gradient Boosting (XGBoost, LightGBM) for more robust churn predictions. Implement deep learning models (e.g., Neural Networks) for complex pattern recognition.