- **Project Title: Customer Churn Prediction for a Telecom Company**
- **Name: Sobia Alamgir**
- **Course: Skilled Score by Zeeshan Usmani**
- **Date: 25-June-2025**

1. Title & Introduction

- **Title:** Customer Churn Prediction for a Telecom Company

- **Goal:** Predict whether a customer will churn based on synthetic data.

2. EDA (Exploratory Data Analysis)

★ Dataset shape, missing values
  - Number of rows: 10000
  - Number of columns: 9
  - No missing values in dataset

★ Key insights (e.g., churn is higher in month-to-month contracts)
  - **Tenure** ↔ **TotalCharges**: Strong positive correlation (**0.77**)
    → Longer-tenured customers spend more.

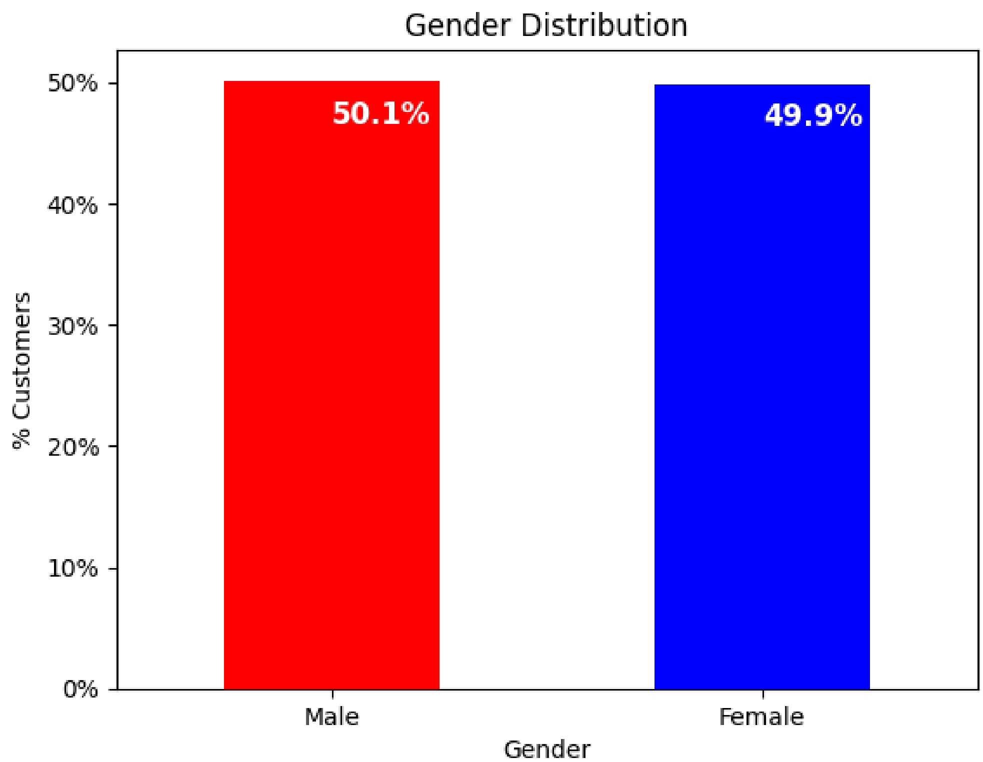
  - **Churn** ↔ **Contract_Two year**: Strong negative correlation (**-0.51**)
    → Customers on two-year contracts are less likely to churn.

  - **Churn** ↔ **Contract_Month-to-month**: Positive correlation (**0.51**)
    → Month-to-month customers are more likely to churn.

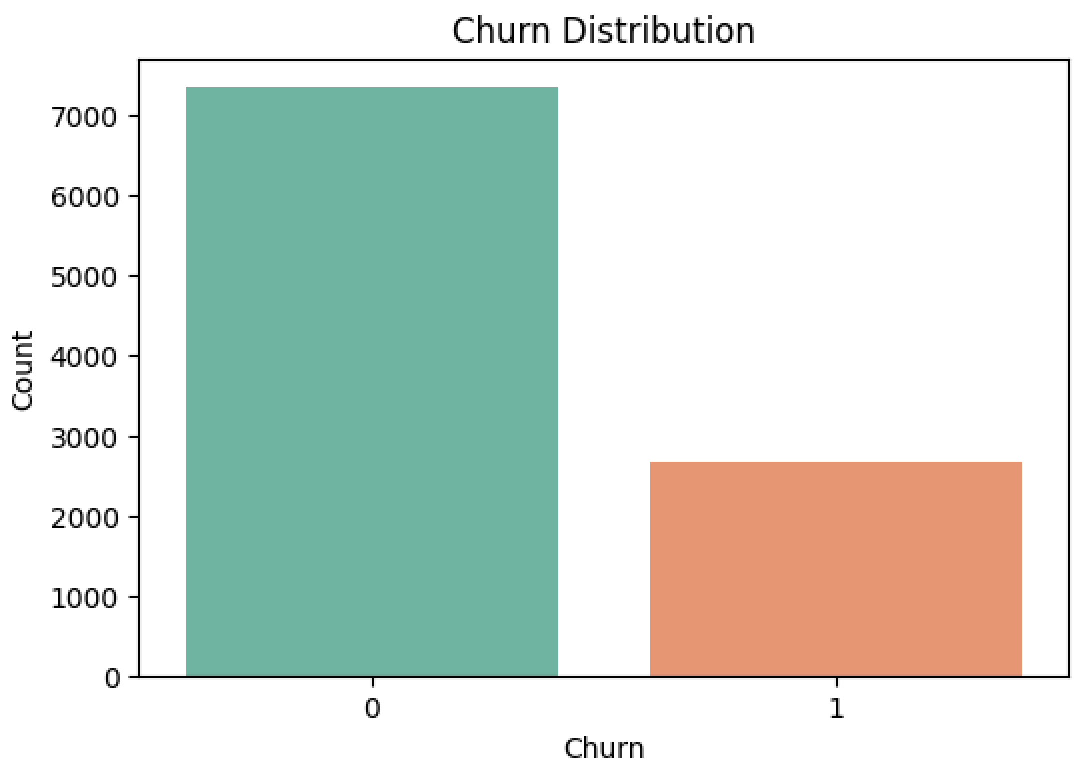
  - **Churn** ↔ **PaymentMethod_Electronic check**: Positive correlation (**~0.34**)
    → Customers using electronic checks tend to churn more.

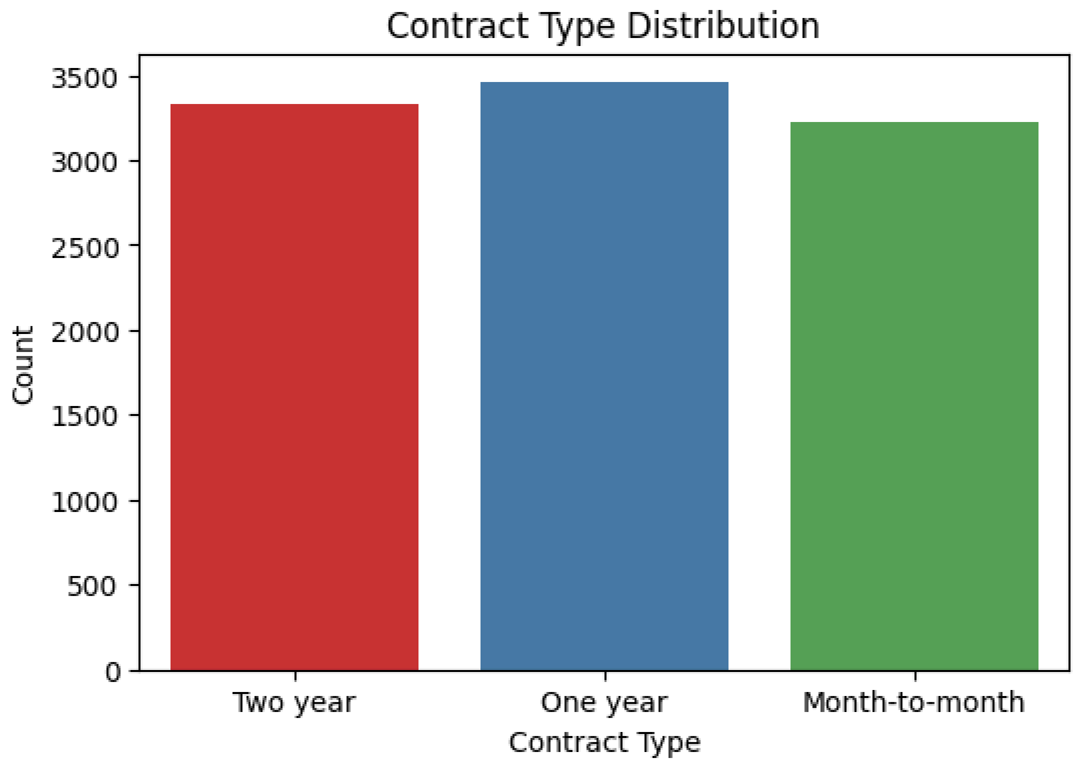
  - **Gender, SeniorCitizen**: Very weak or no correlation with churn
    → These features have little impact on churn prediction.

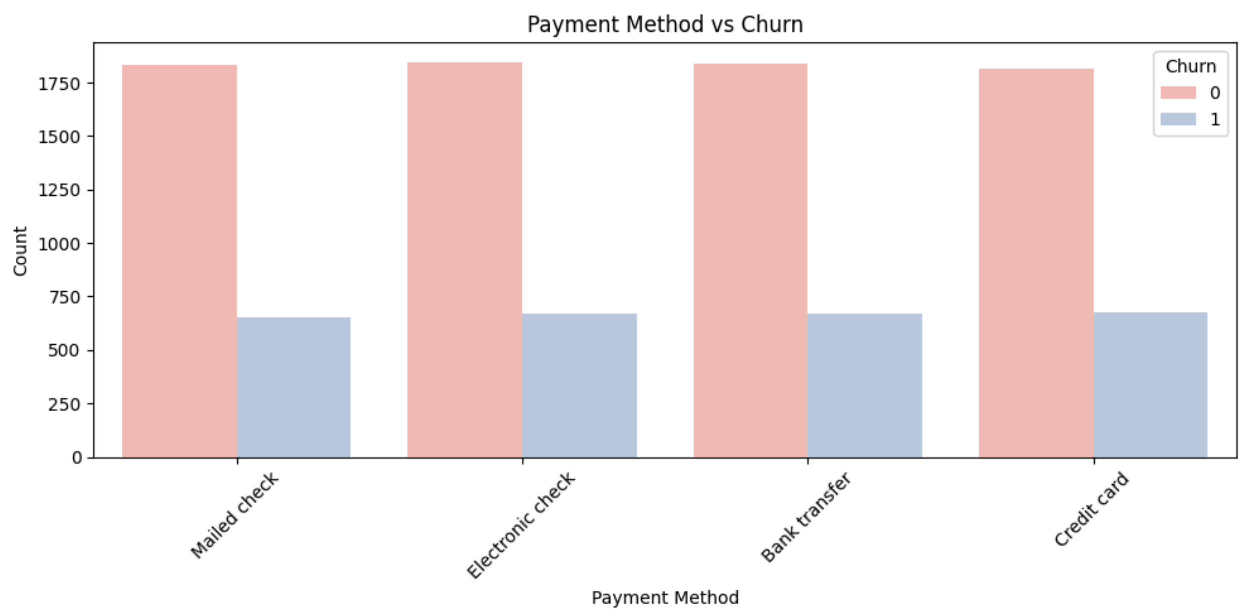
★ Visuals (e.g., churn distribution, contract types)
  1. Gender Distribution

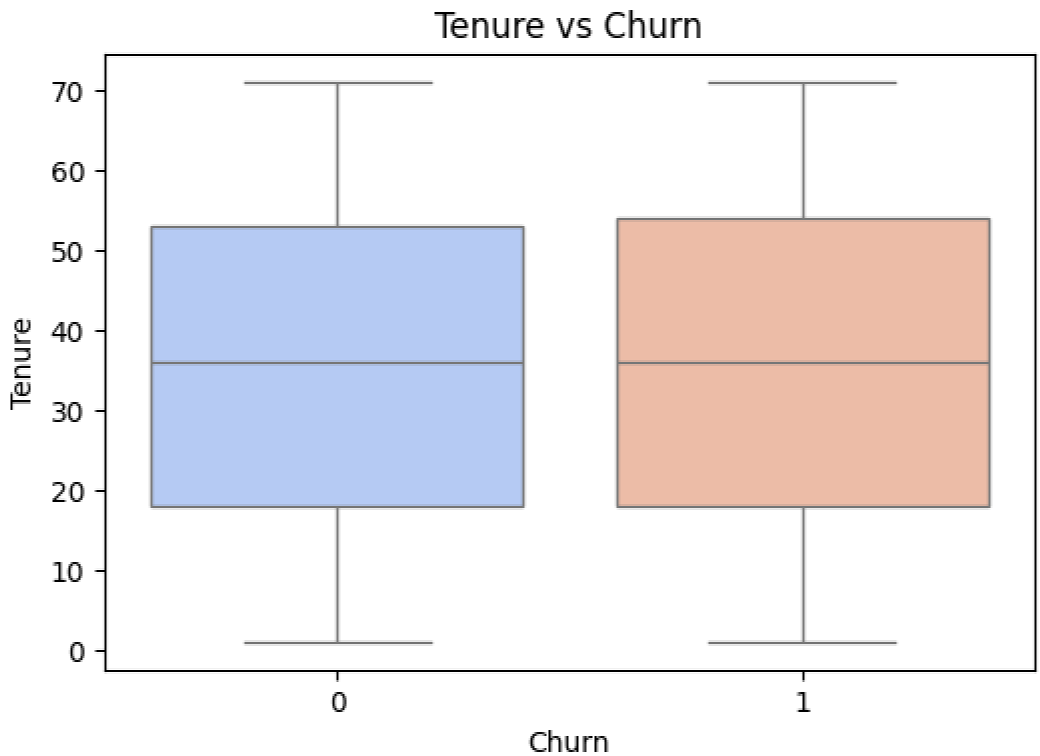

Gender Distribution
50.1%
49.9%
50%
40%
30%
20%
10%
0%
% Customers
Male
Female
Gender

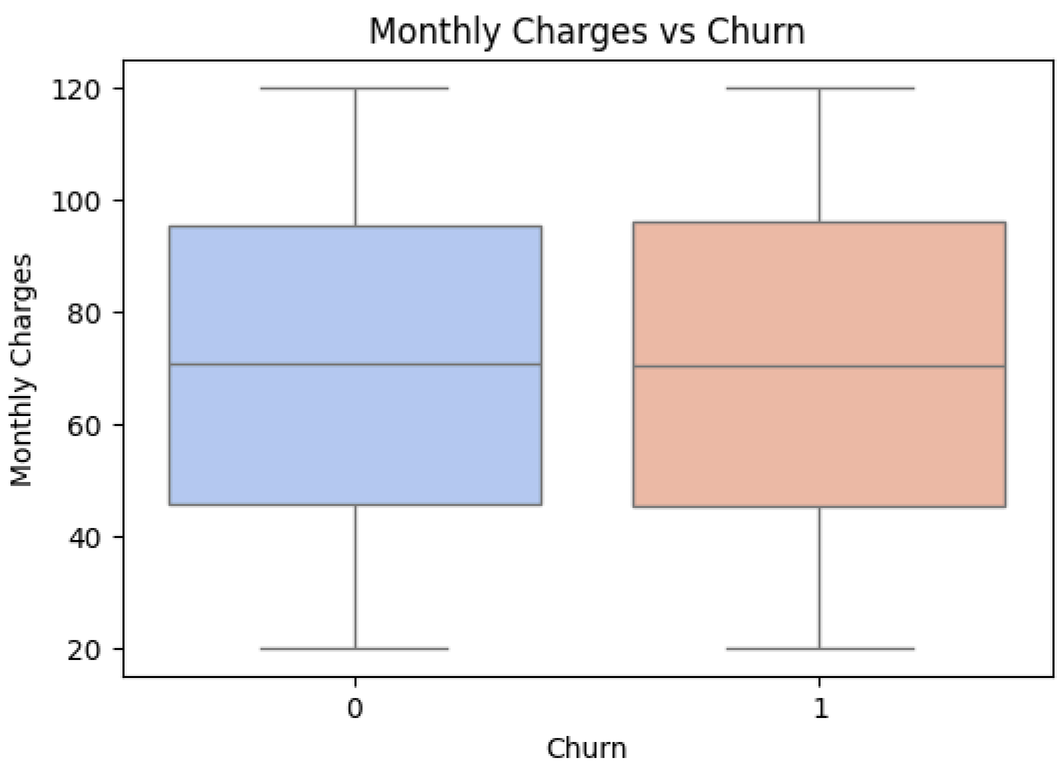
2.


Churn Distribution
7000
6000
5000
4000
3000
2000
1000
0
Count
0
1
Churn


Contract Type Distribution
3500
3000
2500
2000
1500
1000
500
0
Count
Two year
One year
Month-to-month
Contract Type


Payment Method vs Churn
1750
1500
1250
1000
750
500
250
0
Count
Churn
0
1
Mailed check
Electronic check
Bank transfer
Credit card
Payment Method


Tenure vs Churn
70
60
50
40
30
20
10
0
Tenure
0
1
Churn


Monthly Charges vs Churn
120
100
80
60
40
20
Monthly Charges
0
1
Churn

🛠️ 3. Feature Engineering

- Encoding (e.g., `get_dummies`, OneHotEncoder)
  - `get_dummies()` is used on the features 'Gender', 'Contract', and 'PaymentMethod' to convert categorical variables into multiple binary columns for model training.

```
df_encoded = pd.get_dummies(df, columns =
['Gender','Contract','PaymentMethod'])
```

🤖 4. Modeling

- Data split method (e.g., `train_test_split`, StratifiedKFold)

- Models used: Logistic Regression, XGBoost, XGBoost with Optuna, Random Forest, SVM

- **Optuna** is an automatic hyperparameter optimization library that efficiently searches for the best parameter combinations using intelligent sampling and pruning strategies.

📈 5. Evaluation Metrics

1. Logistic Regression
Accuracy: 0.74
AUC-ROC: 0.50
Confusion Matrix: [[1472, 0], [528, 0]]
Predicted only the majority class.

2. XGBoost Classifier
Accuracy: 0.71
AUC-ROC: 0.50
Confusion Matrix: [[1386, 86], [497, 31]]
Predicted both classes, but performance still low.

3. XGBoost with Optuna
Accuracy: 0.74
AUC-ROC: 0.50
Confusion Matrix: [[1472, 0], [528, 0]]
Same as Logistic Regression; no improvement with Optuna.

4. Random Forest Classifier
Accuracy: 0.68

AUC-ROC: 0.50
Confusion Matrix: [[1301, 171], [462, 66]]
Captured both classes slightly better.

5. Support Vector Classifier (SVC)
Accuracy: 0.74
AUC-ROC: 0.50
Confusion Matrix: [[1472, 0], [528, 0]]
Predicted only one class.

📌 6. Final Results

- Best model and why

  Among all models tested, the **XGBoost Classifier (without Optuna)** performed slightly better. It was the only model that predicted both classes with some accuracy, as shown in its confusion matrix: `[[1386, 86], [497, 31]]`. While the AUC-ROC score remained low at 0.50, it still outperformed others by not defaulting to predicting only the majority class.

- Model performance summary

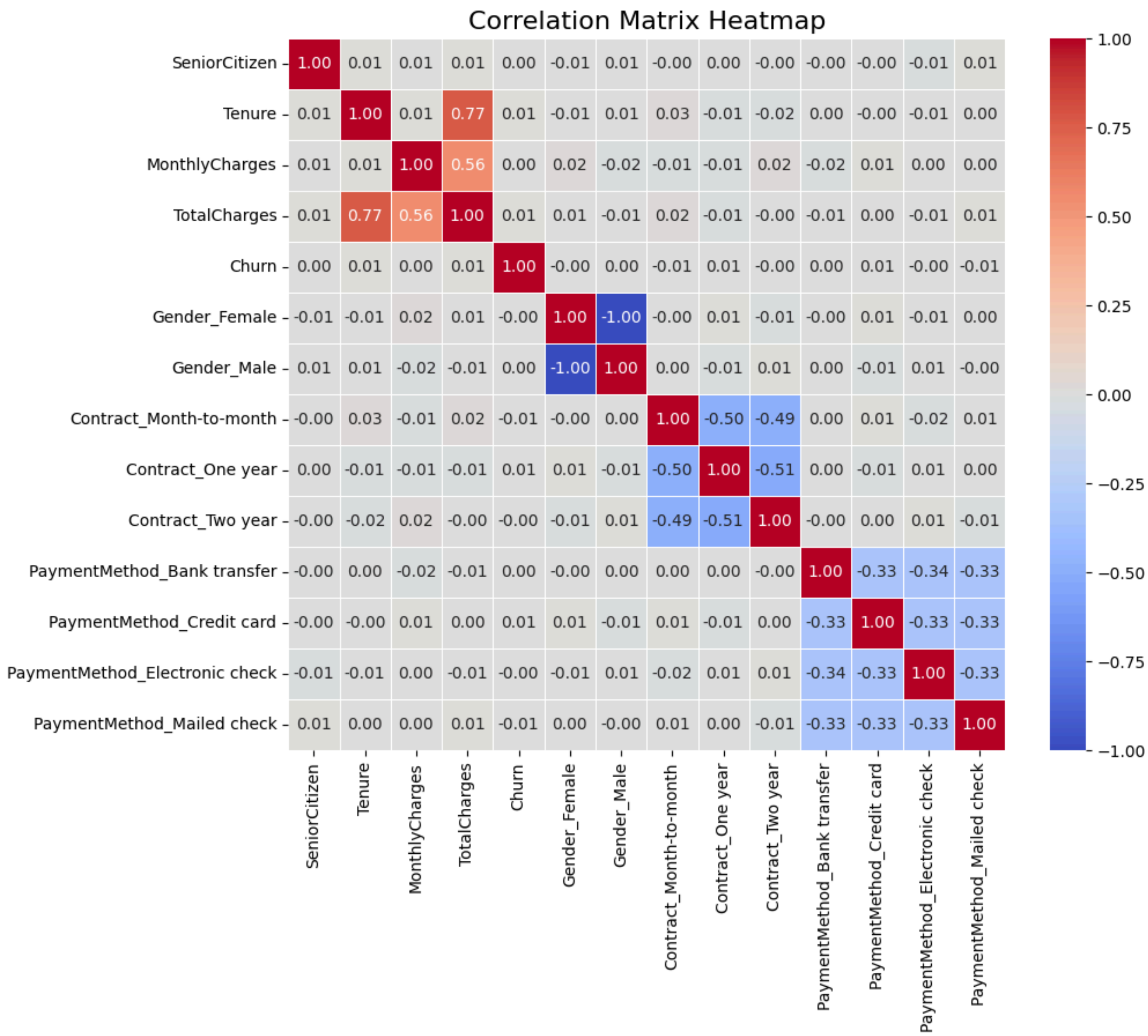
  All models had an AUC-ROC of 0.50. Logistic Regression, SVC, and Optuna-based models failed to classify churners. Random Forest and XGBoost showed slightly better class separation.

🧾 7. Conclusion

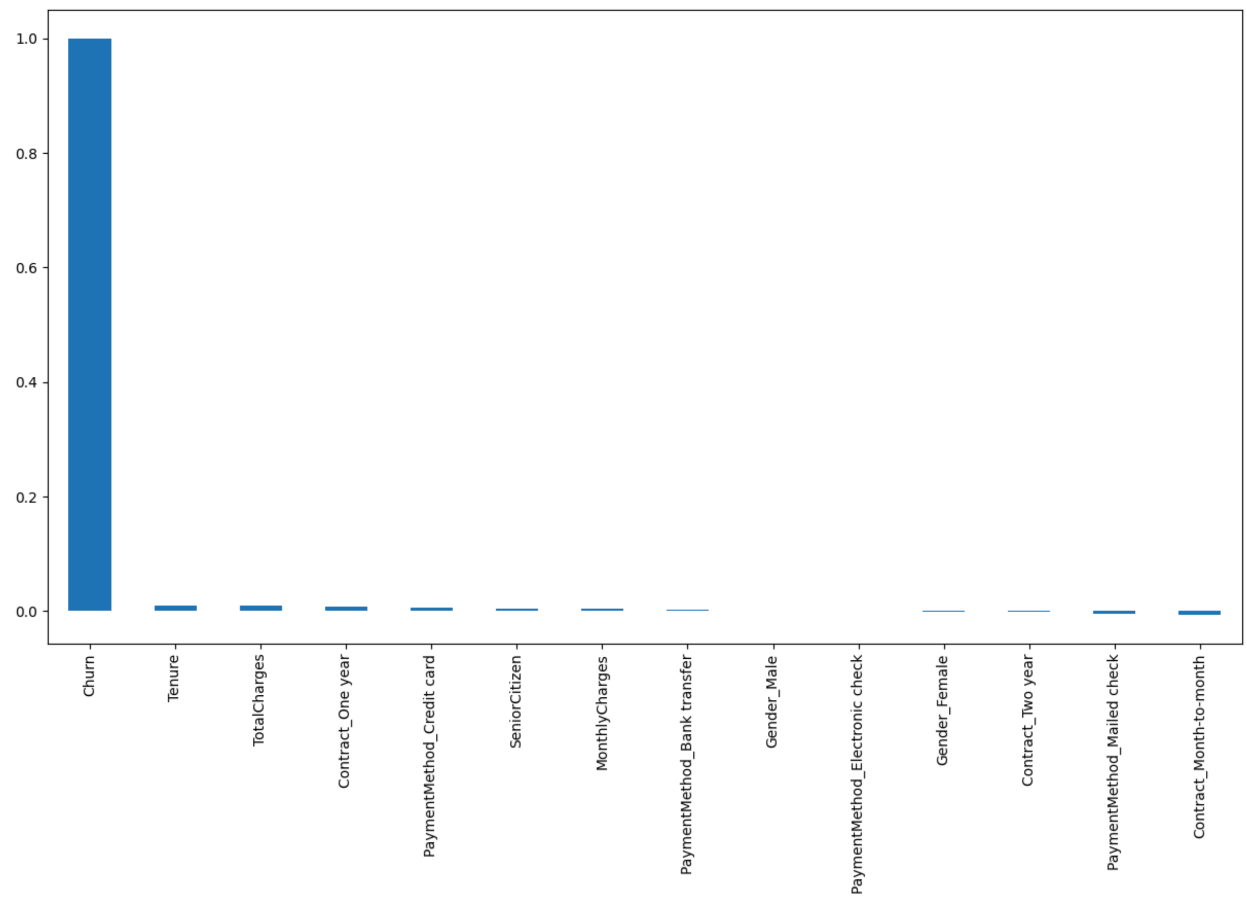
All models have AUC-ROC scores of 0.50, indicating they are not able to distinguish between churn and non-churn.

🖼️ 8. Visualizations

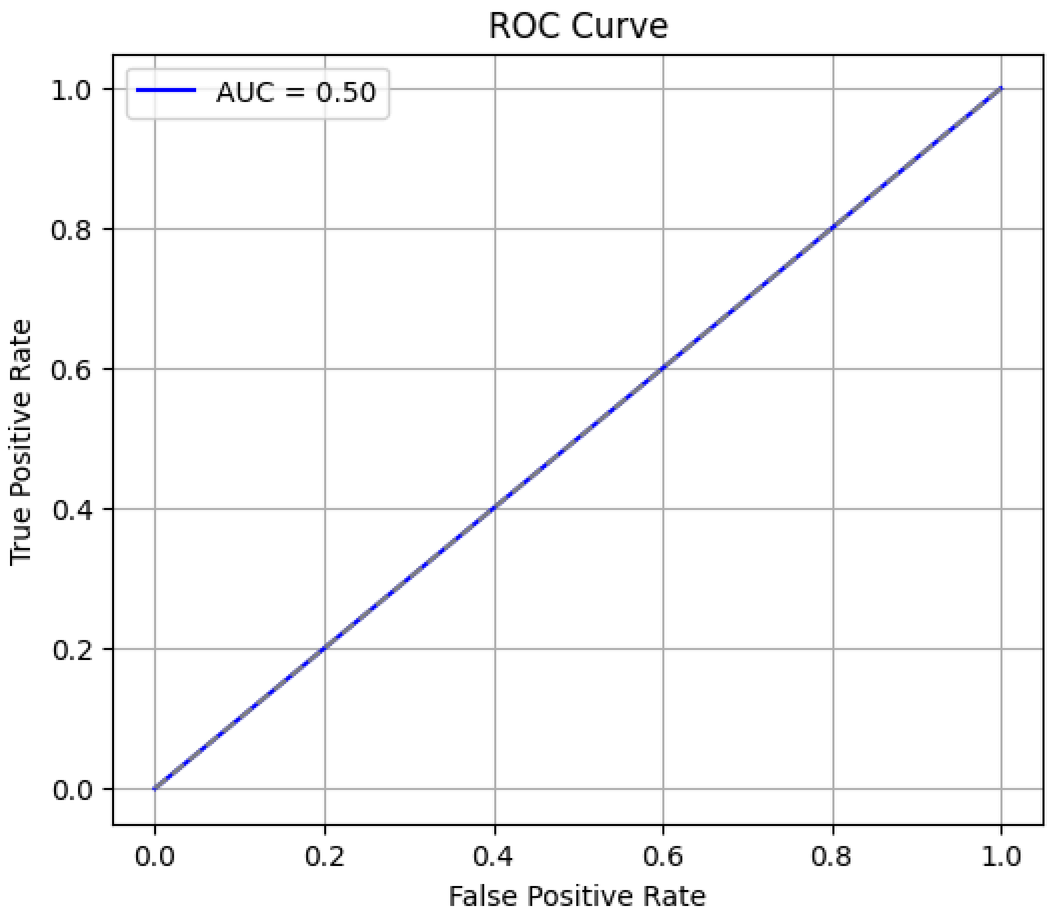
- Correlation matrix



- Feature importance


1.0
0.8
0.6
0.4
0.2
0.0
Churn
Tenure
TotalCharges
Contract_One year
PaymentMethod_Credit card
SeniorCitizen
MonthlyCharges
PaymentMethod_Bank transfer
Gender_Male
PaymentMethod_Electronic check
Gender_Female
Contract_Two year
PaymentMethod_Mailed check
Contract_Month-to-month

- ROC Curve



ROC Curve
AUC = 0.50
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0