PREDICTING CUSTOMER CHURN IN THE TELECOM INDUSTRY

TEAM 1A

TABLE OF CONTENTS

MEET OUR TEAM

FATIMA

TEAGAN WHITE

INEZ CHANG
(YING CHIH)

SHWETA

gender	object
SeniorCitizen	int64
Partner	int64
Dependents	int64
tenure	int64
PhoneService	int64
MultipleLines	object
InternetService	object
OnlineSecurity	object
OnlineBackup	object
DeviceProtection	object
TechSupport	object
StreamingTV	object
StreamingMovies	object
Contract	object
PaperlessBilling	int64
PaymentMethod	object
MonthlyCharges	float64
TotalCharges	float64
Churn	int64

OUR DATA

Kaggle Dataset
7,043 Rows or Customers
21 Columns or Features

DESCRIPTIVE COLUMNS

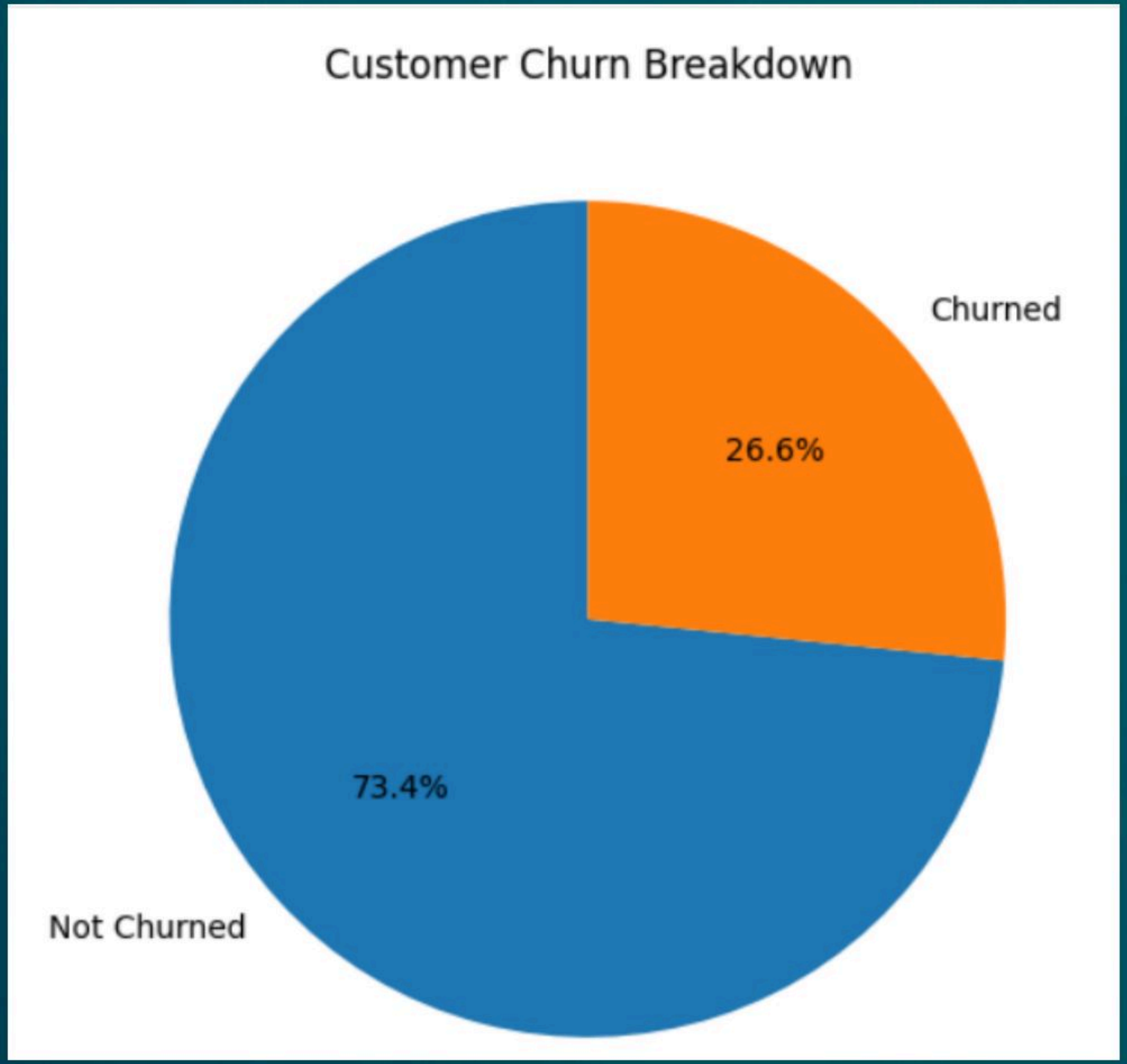
1. Services that each Customer signed up for
2. Demographics
3. Customer Account Information

TARGET VALUE

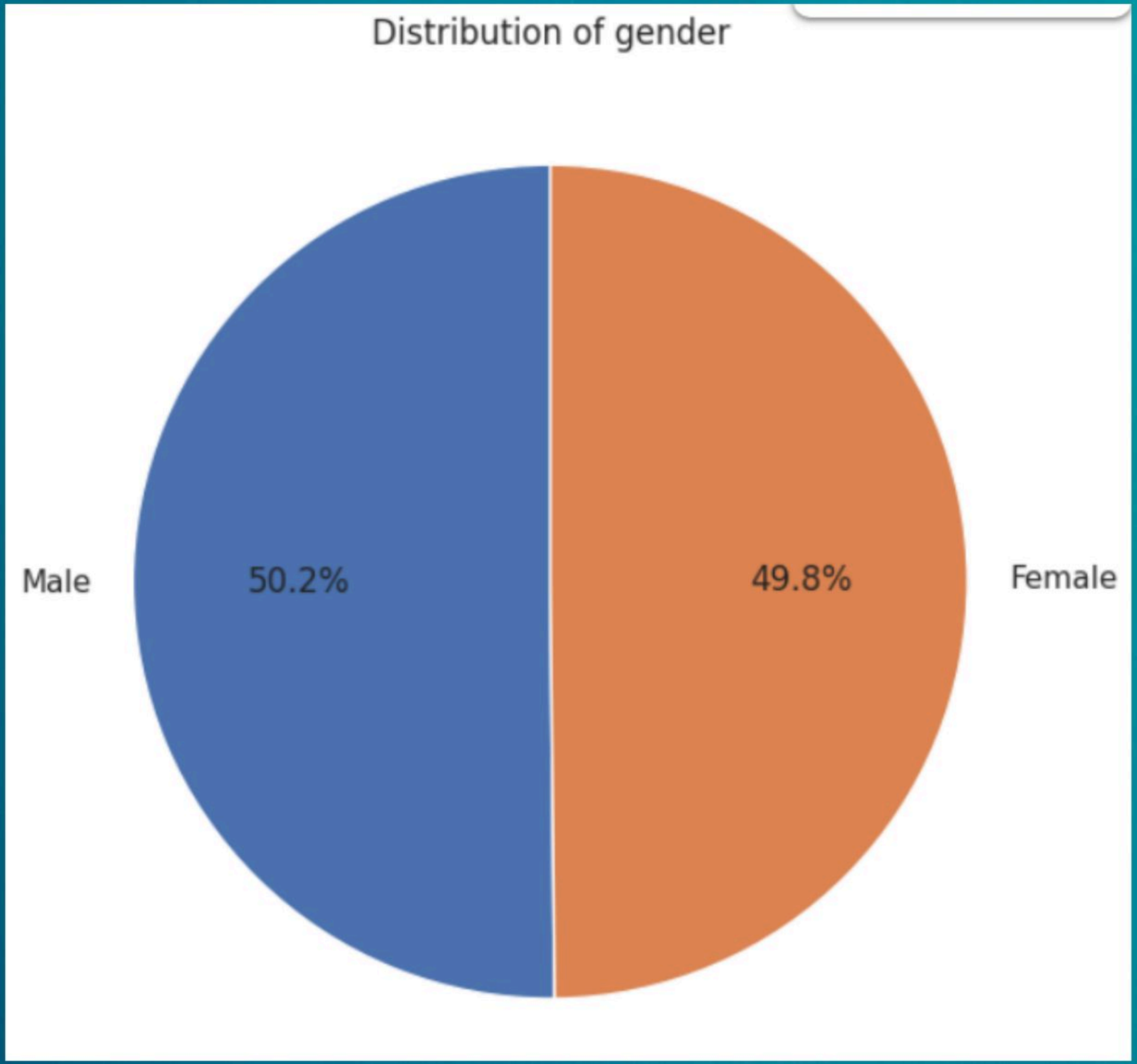
Customers who left within the last month – the column is called **Churn**

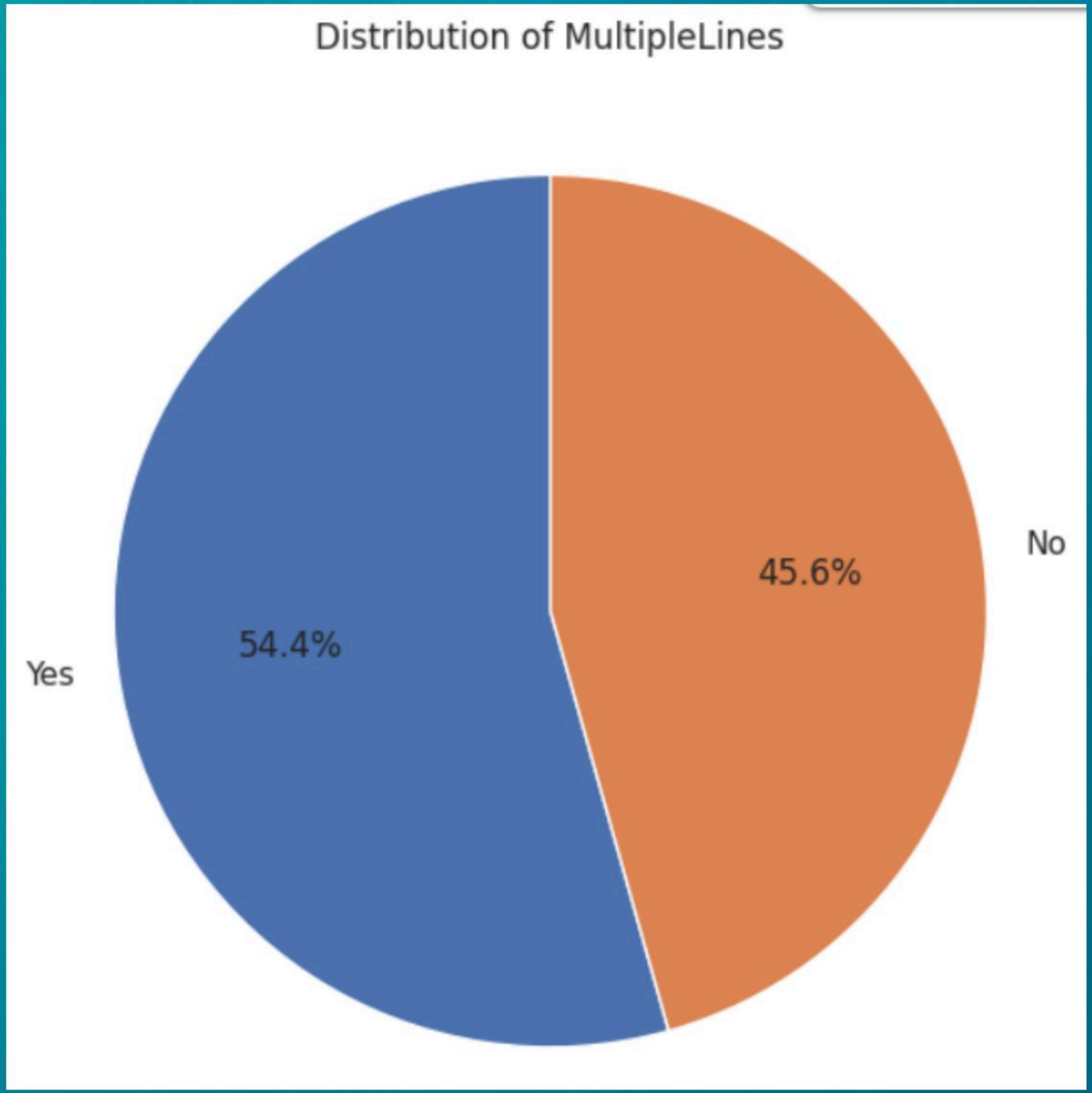
OUR PROBLEM

We are trying to predict Customer Churn using our dataset. By doing so, companies can act accordingly on high risk customers by offering promotions and incentives to stay.

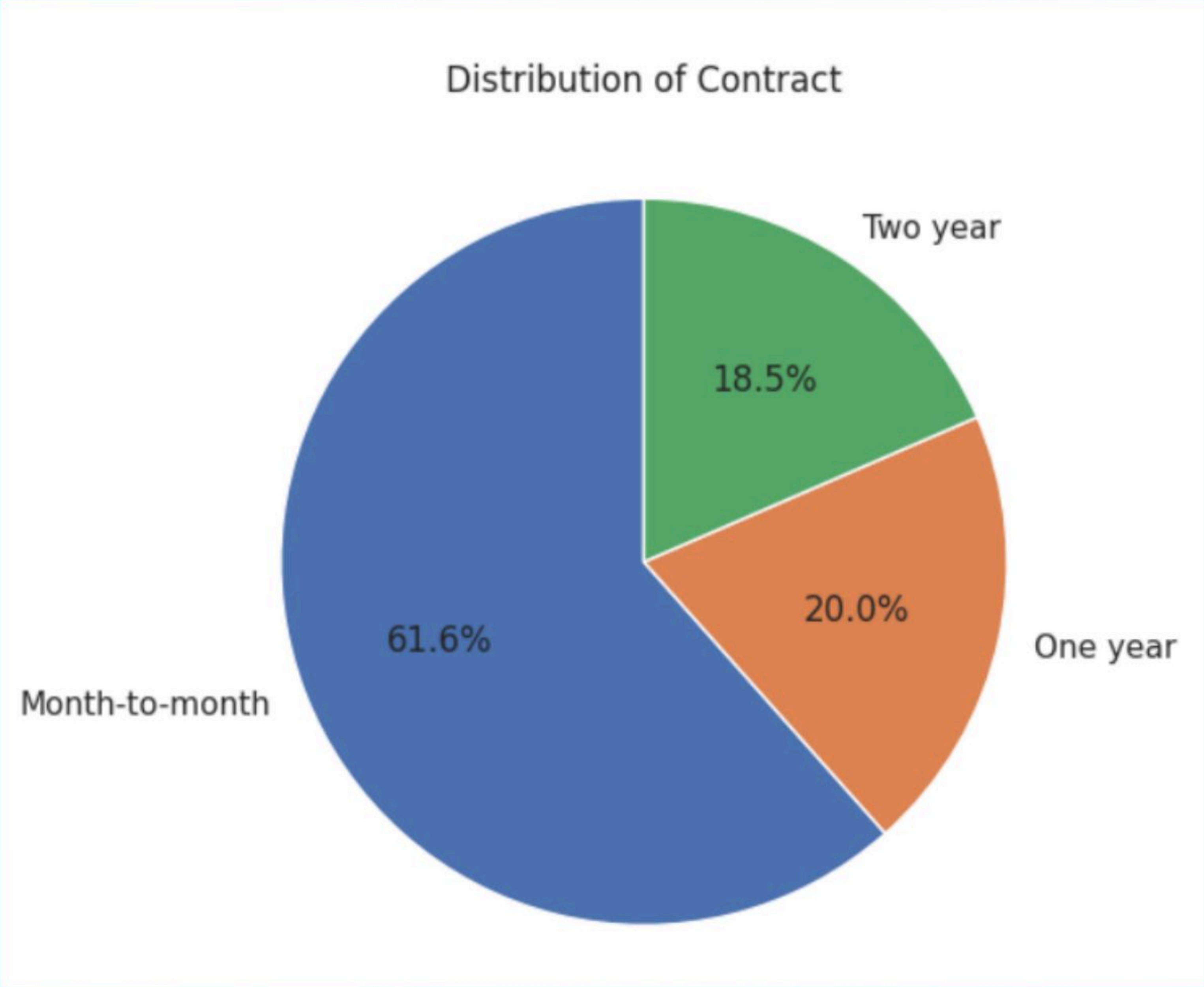


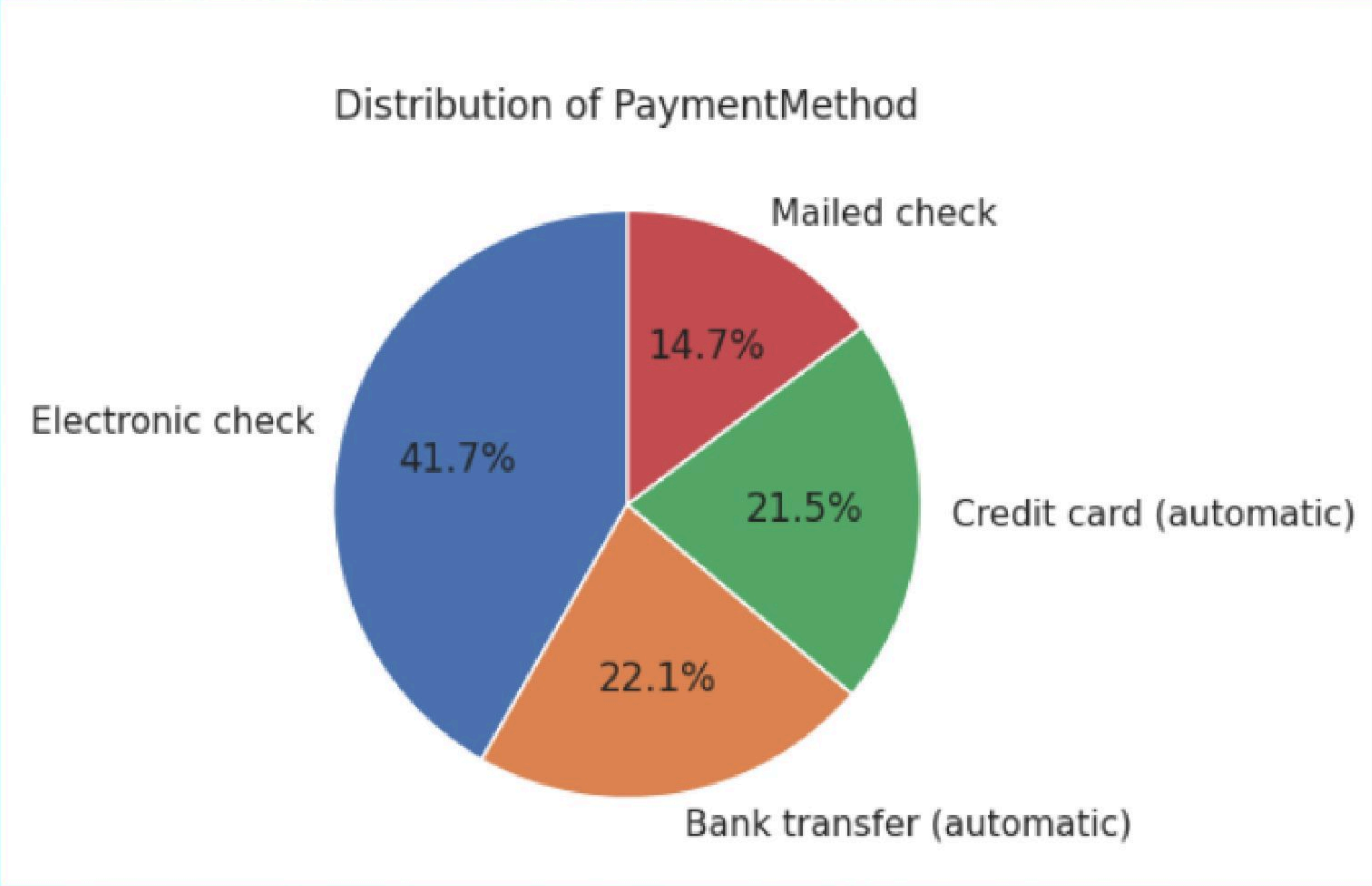
DATA VISUALIZATIONS



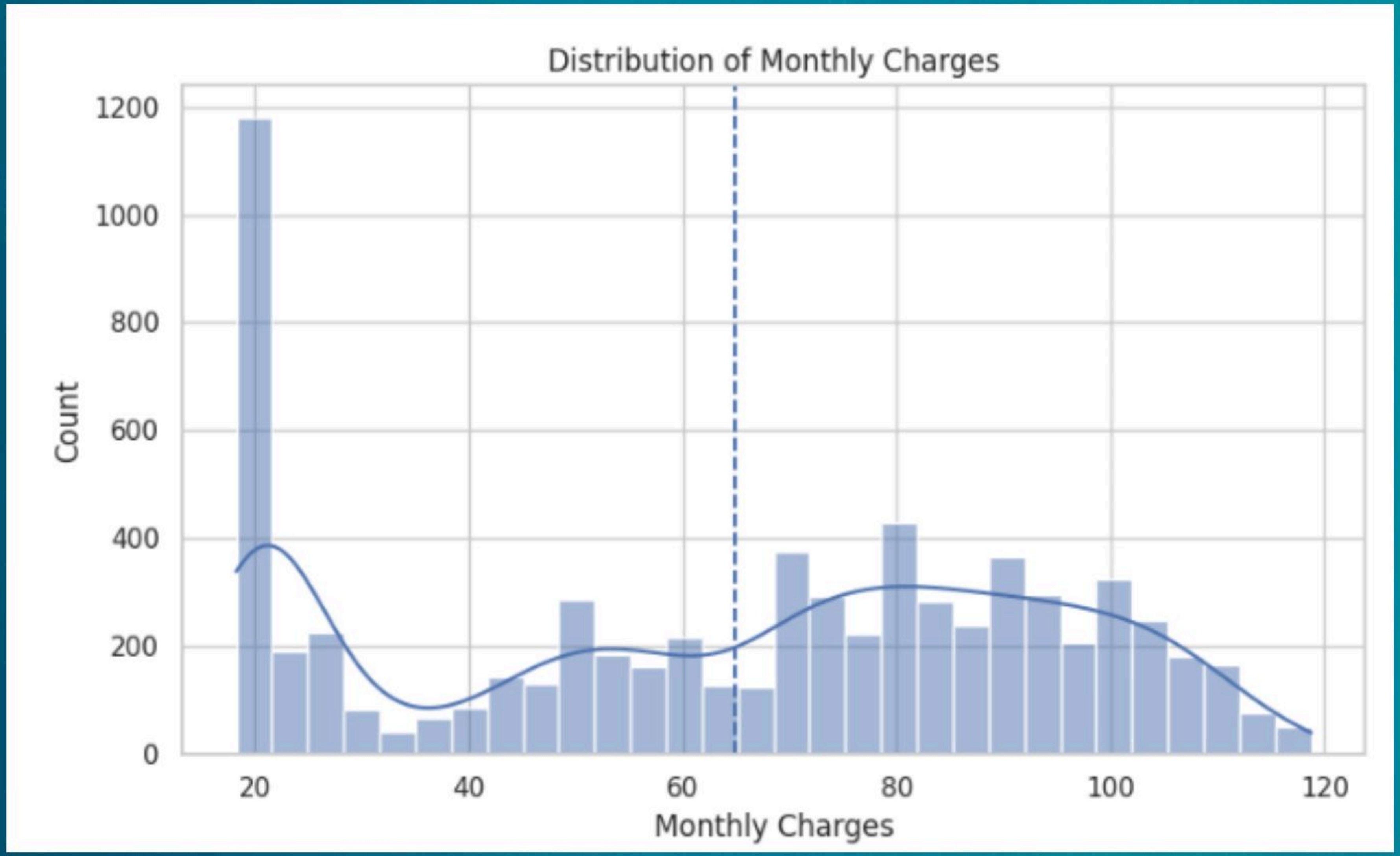


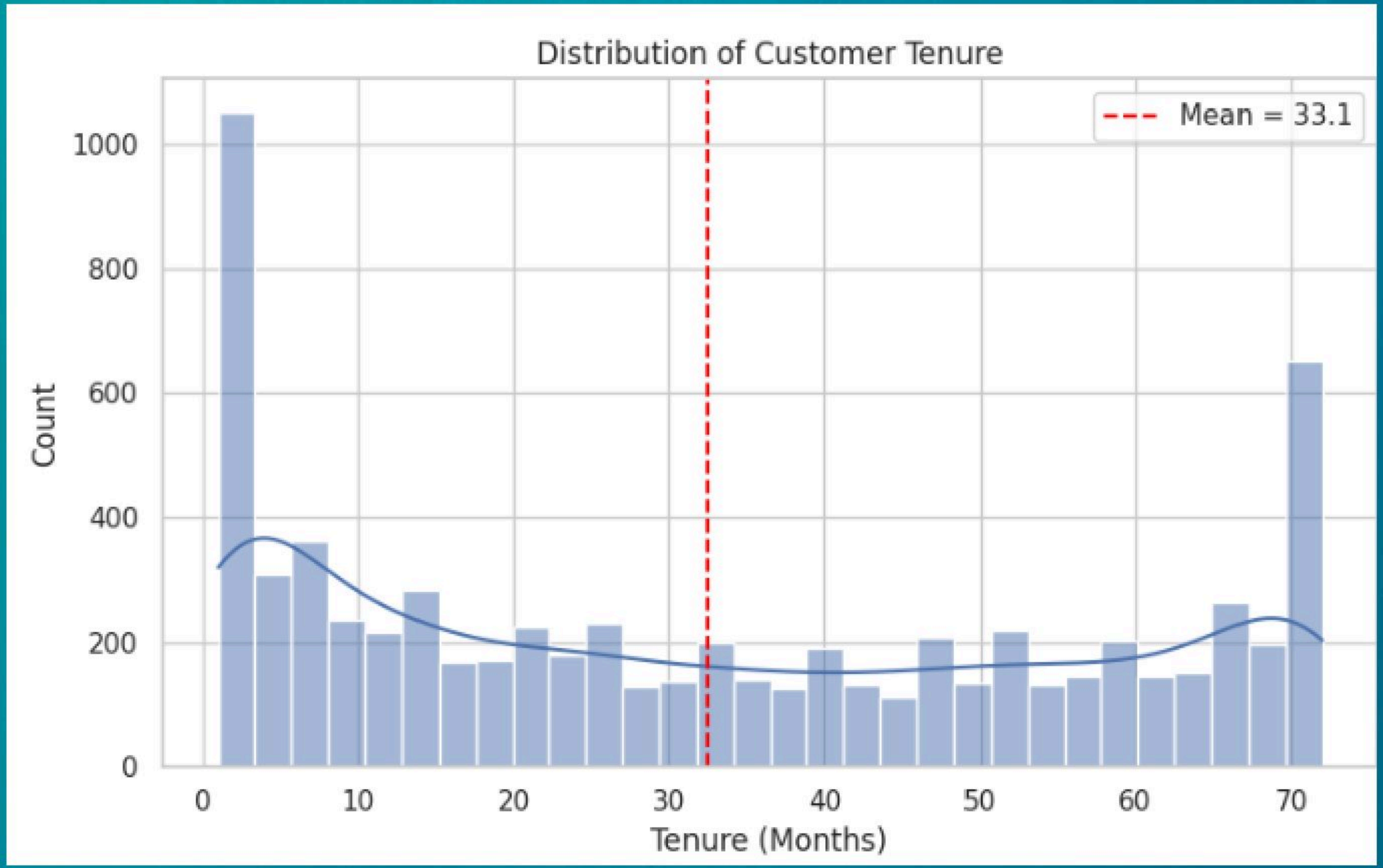
DATA VISUALIZATIONS





DATA VISUALIZATIONS



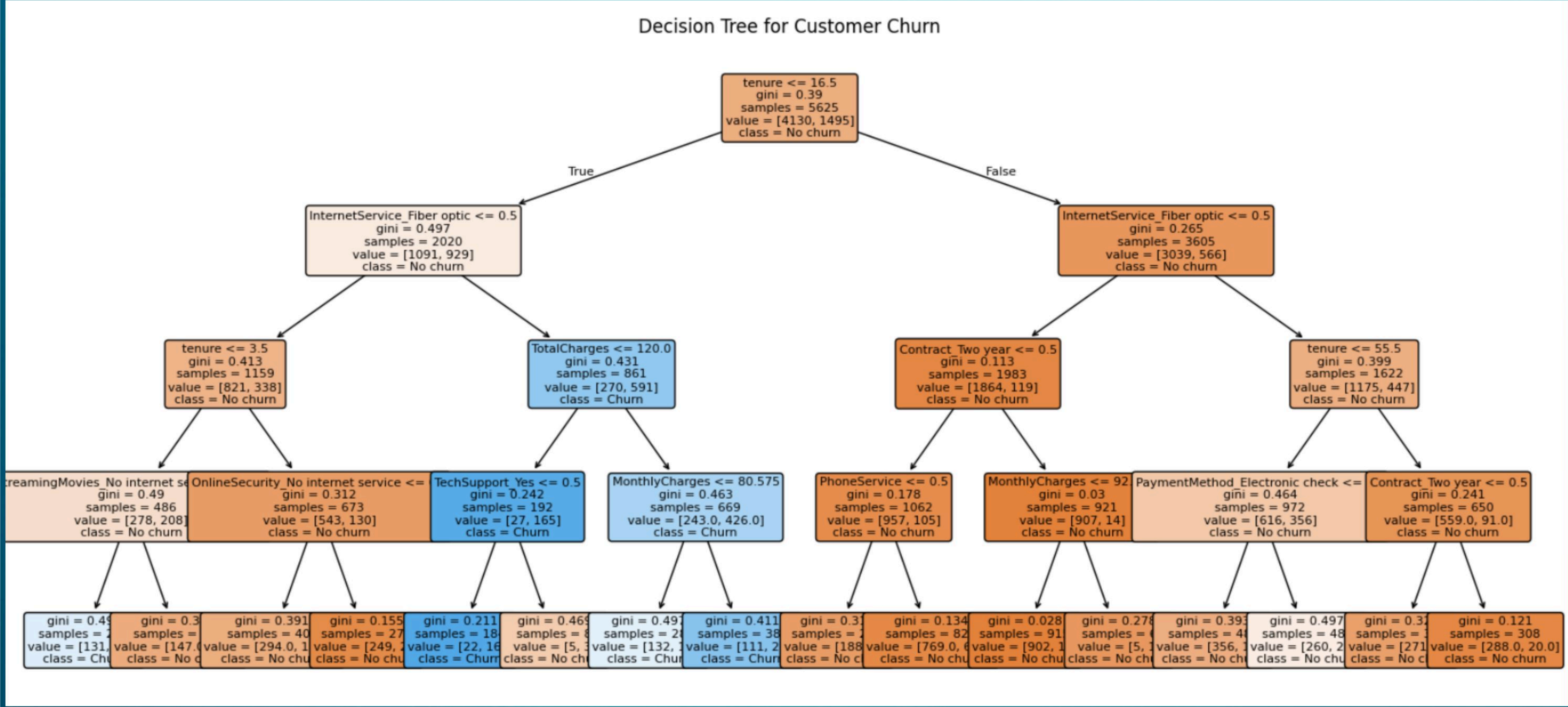


LOGISTIC REGRESSION

```
=== Logistic Regression Performance on Test Set ===
Accuracy:  0.8038379530916845
Precision: 0.6475903614457831
Recall:    0.5748663101604278
ROC-AUC:   0.8356184416915582

Confusion Matrix:
[[916 117]
 [159 215]]

Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.89      0.87      1033
           1       0.65      0.57      0.61       374

    accuracy                           0.80      1407
   macro avg       0.75      0.73      0.74      1407
weighted avg       0.80      0.80      0.80      1407
```

	feature	coef	abs_coef
3	tenure	-1.340230	1.340230
6	MonthlyCharges	-0.845896	0.845896
11	InternetService_Fiber optic	0.724660	0.724660
7	TotalCharges	0.632460	0.632460
26	Contract_Two year	-0.597358	0.597358
25	Contract_One year	-0.310502	0.310502
22	StreamingTV_Yes	0.248959	0.248959
24	StreamingMovies_Yes	0.235700	0.235700
10	MultipleLines_Yes	0.214014	0.214014
28	PaymentMethod_Electronic check	0.182509	0.182509
5	PaperlessBilling	0.142627	0.142627
14	OnlineSecurity_Yes	-0.137256	0.137256
20	TechSupport_Yes	-0.118646	0.118646
2	Dependents	-0.105459	0.105459
12	InternetService_No	-0.087816	0.087816

DECISION TREE



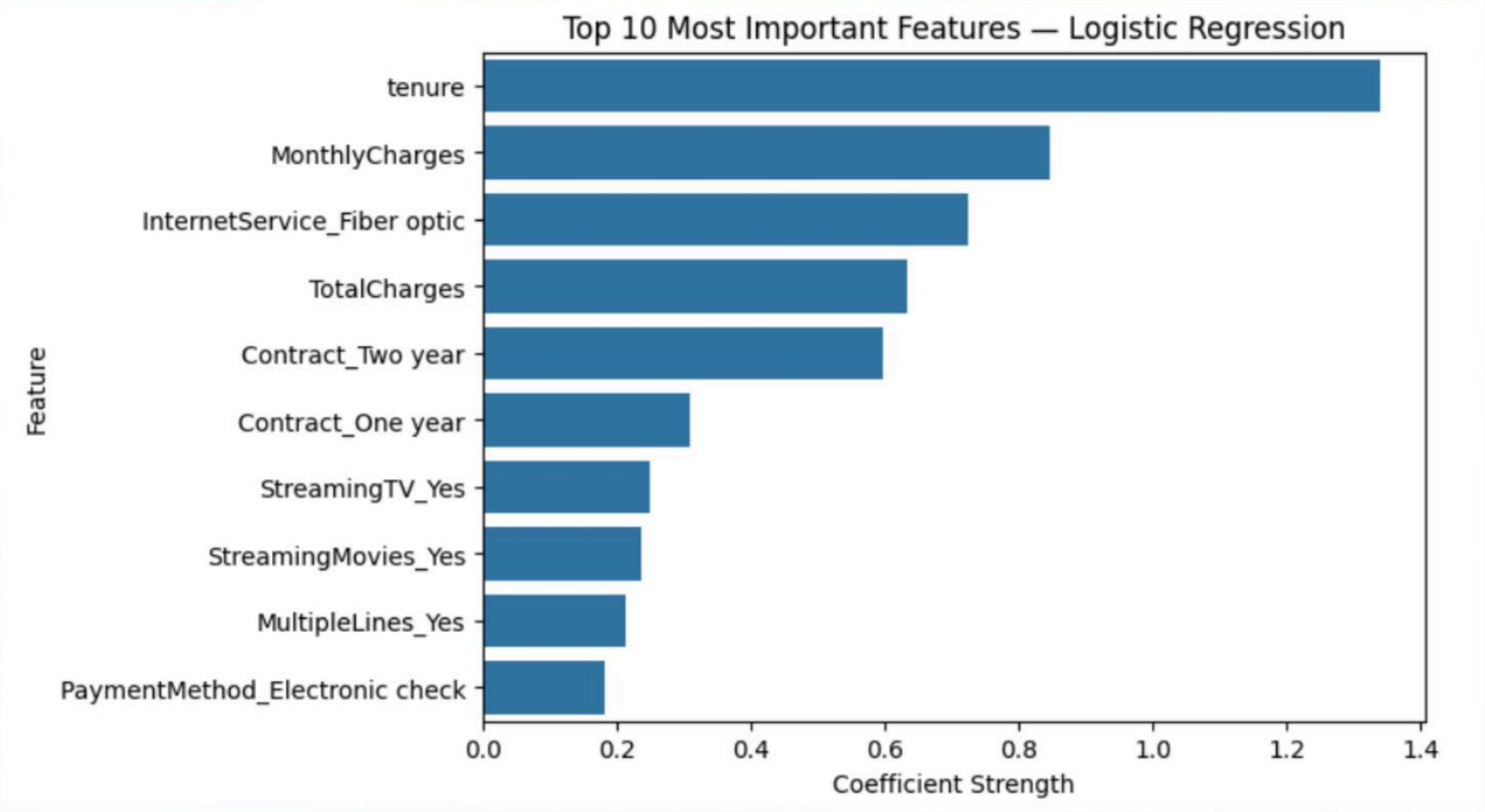
DECISION TREE RESULTS

```
=== Decision Tree Performance on Test Set ===
Accuracy:  0.783226723525231
Precision: 0.6131147540983607
Recall:    0.5
ROC-AUC:   0.8200350984361007

Confusion Matrix:
[[915 118]
 [187 187]]

Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.89      0.86      1033
           1       0.61      0.50      0.55       374

    accuracy                           0.78      1407
   macro avg       0.72      0.69      0.70      1407
weighted avg       0.77      0.78      0.78      1407
```

COMPARING THE TWO MODELS


```
=== Logistic Regression CV Accuracy ===
Fold scores: [0.80668088 0.80739161 0.79587482 0.80227596 0.81294452]
Mean accuracy: 0.8050335601003316
Std dev: 0.005698144632501366

=== Decision Tree CV Accuracy ===
Fold scores: [0.79175551 0.79175551 0.76671408 0.77596017 0.79800853]
Mean accuracy: 0.7848387608796094
Std dev: 0.011631500871054818
```

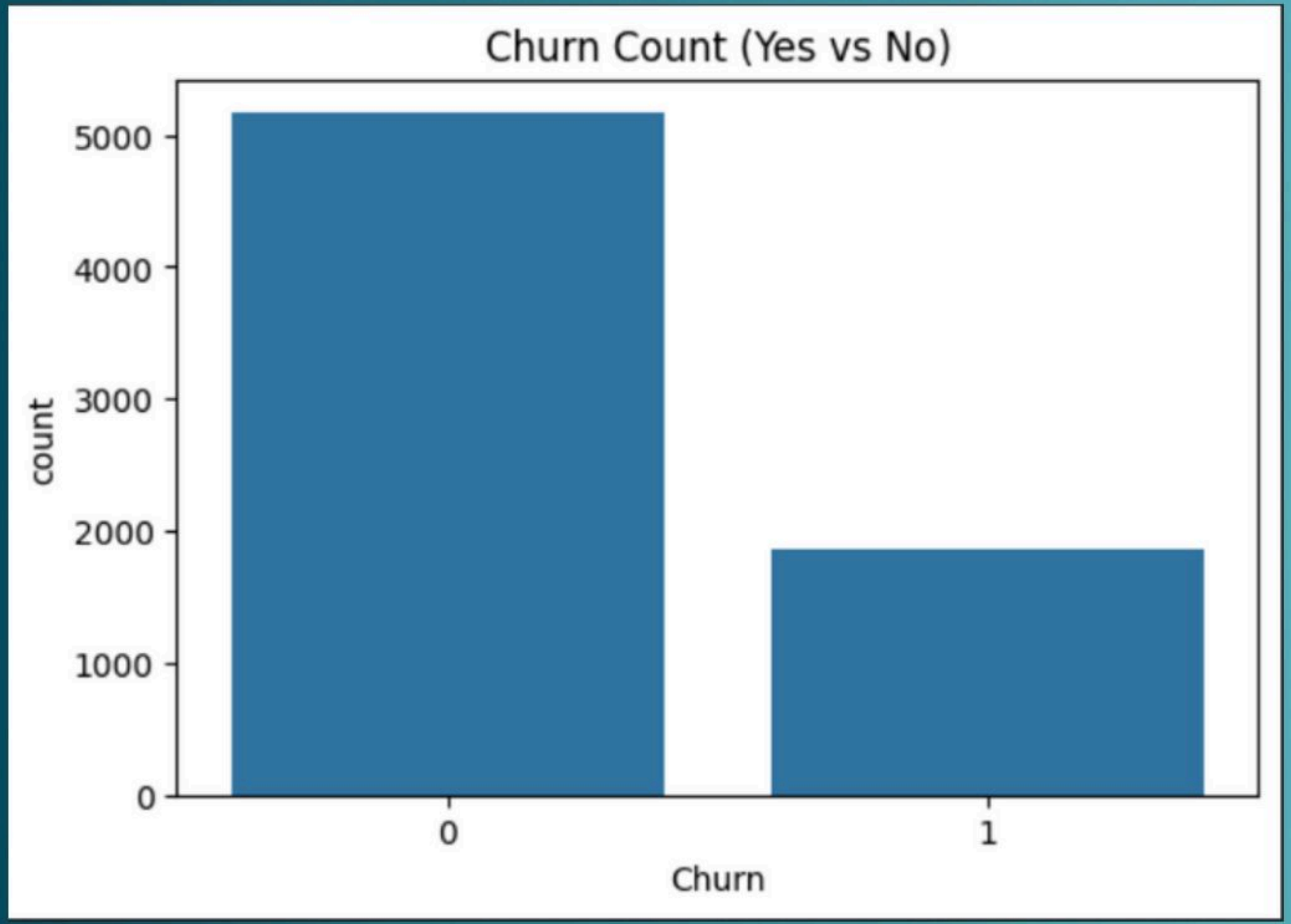
LOGISTIC REGRESSION FEATURE IMPORTANCE

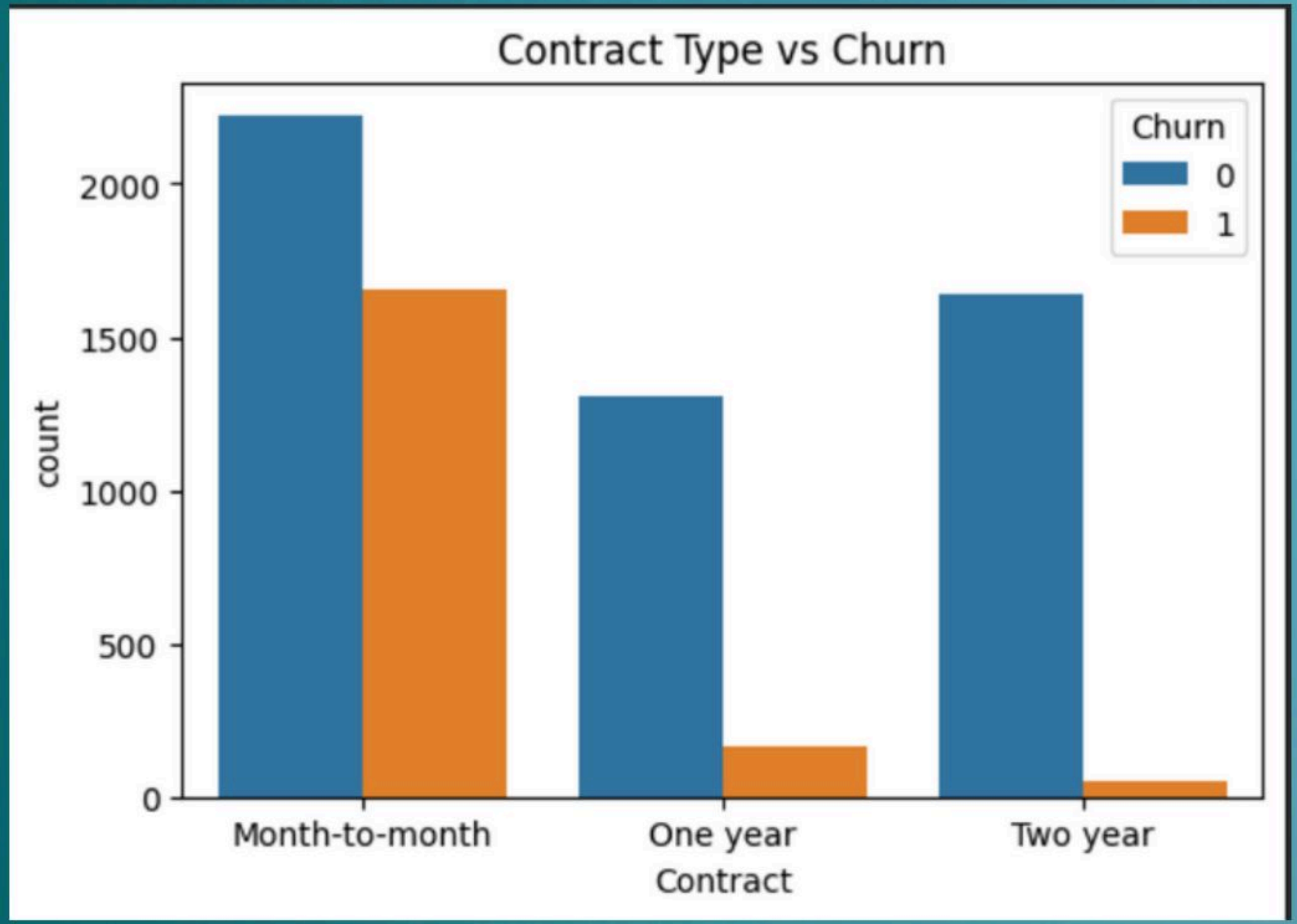


DECISION TREE FEATURE IMPORTANCE

	feature	importance
3	tenure	0.481266
11	InternetService_Fiber optic	0.369526
23	StreamingMovies_No internet service	0.029204
28	PaymentMethod_Electronic check	0.028802
7	TotalCharges	0.023061
26	Contract_Two year	0.021061
13	OnlineSecurity_No internet service	0.016797
6	MonthlyCharges	0.015899
4	PhoneService	0.008274
20	TechSupport_Yes	0.006110
5	PaperlessBilling	0.000000
2	Dependents	0.000000
0	SeniorCitizen	0.000000
1	Partner	0.000000
12	InternetService_No	0.000000

EXPLORATORY DATA





Churn Rate:

26.6% of customers churned

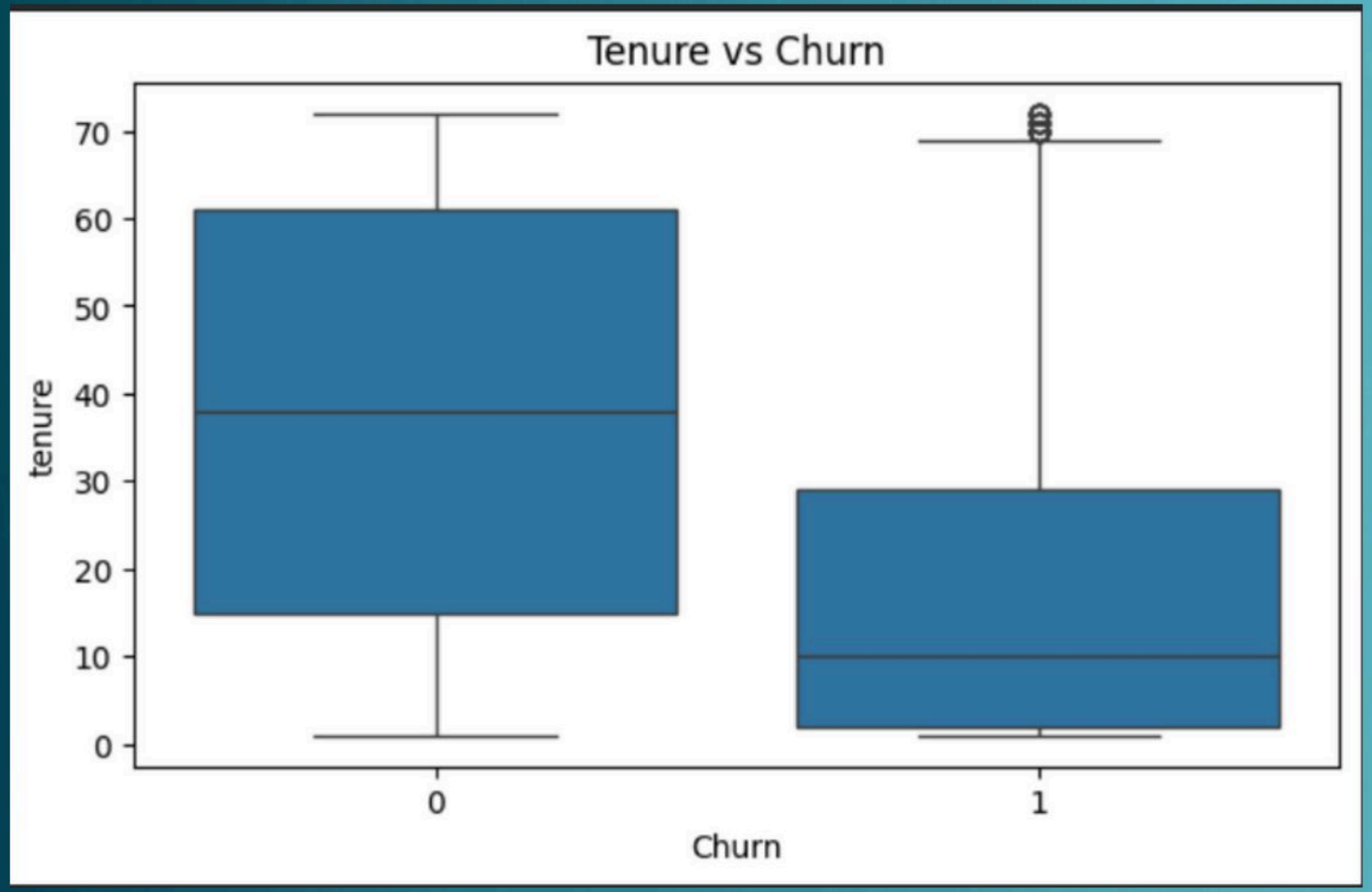
↓

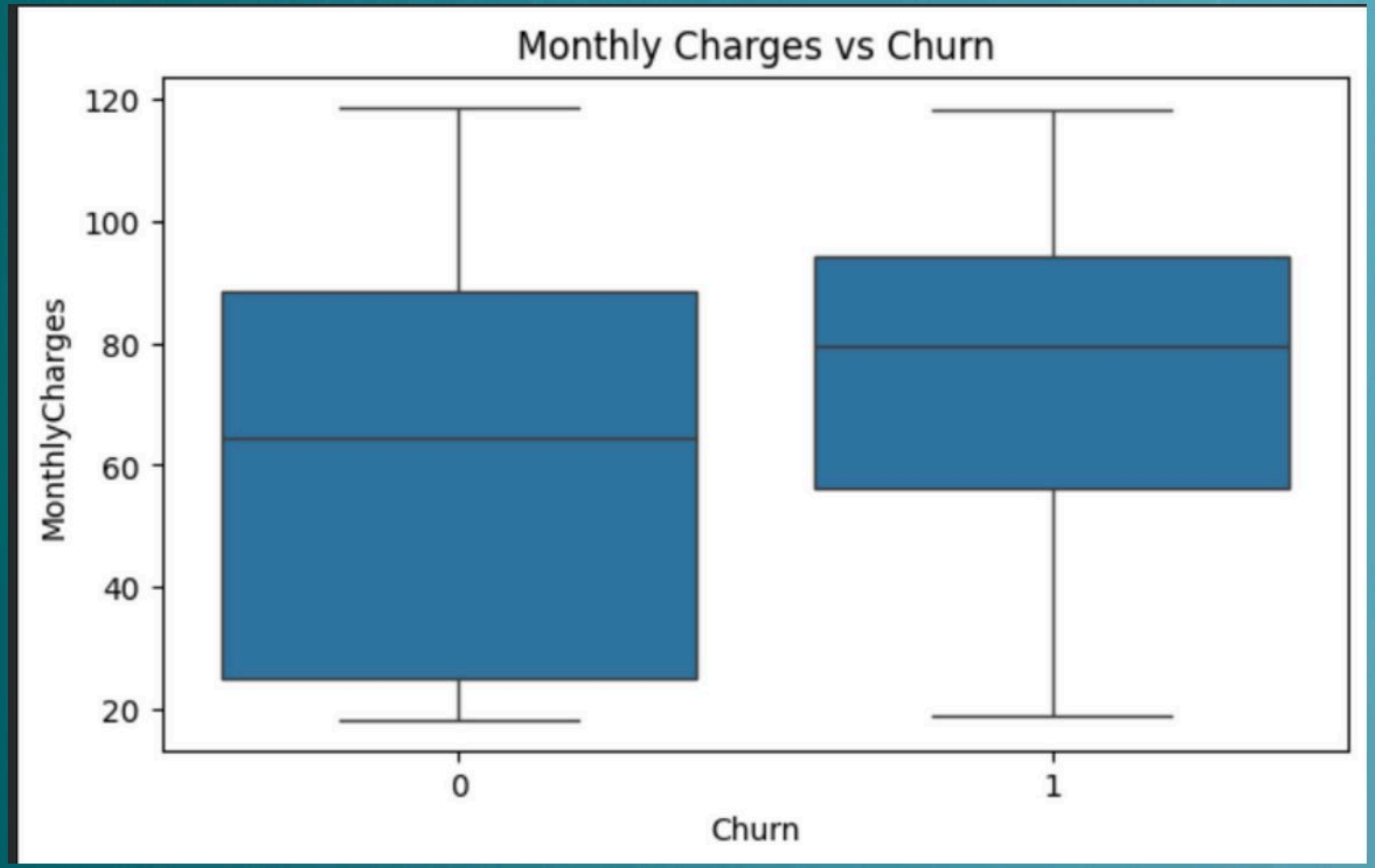
telecom churn is relatively high

Contract Type:

- Month-to-month customers churn the most.
- One-year and two-year contracts reduce churn significantly.

EXPLORATORY DATA





Tenure:

- Customers with short tenure churn much more.
- Long-tenure customers rarely churn

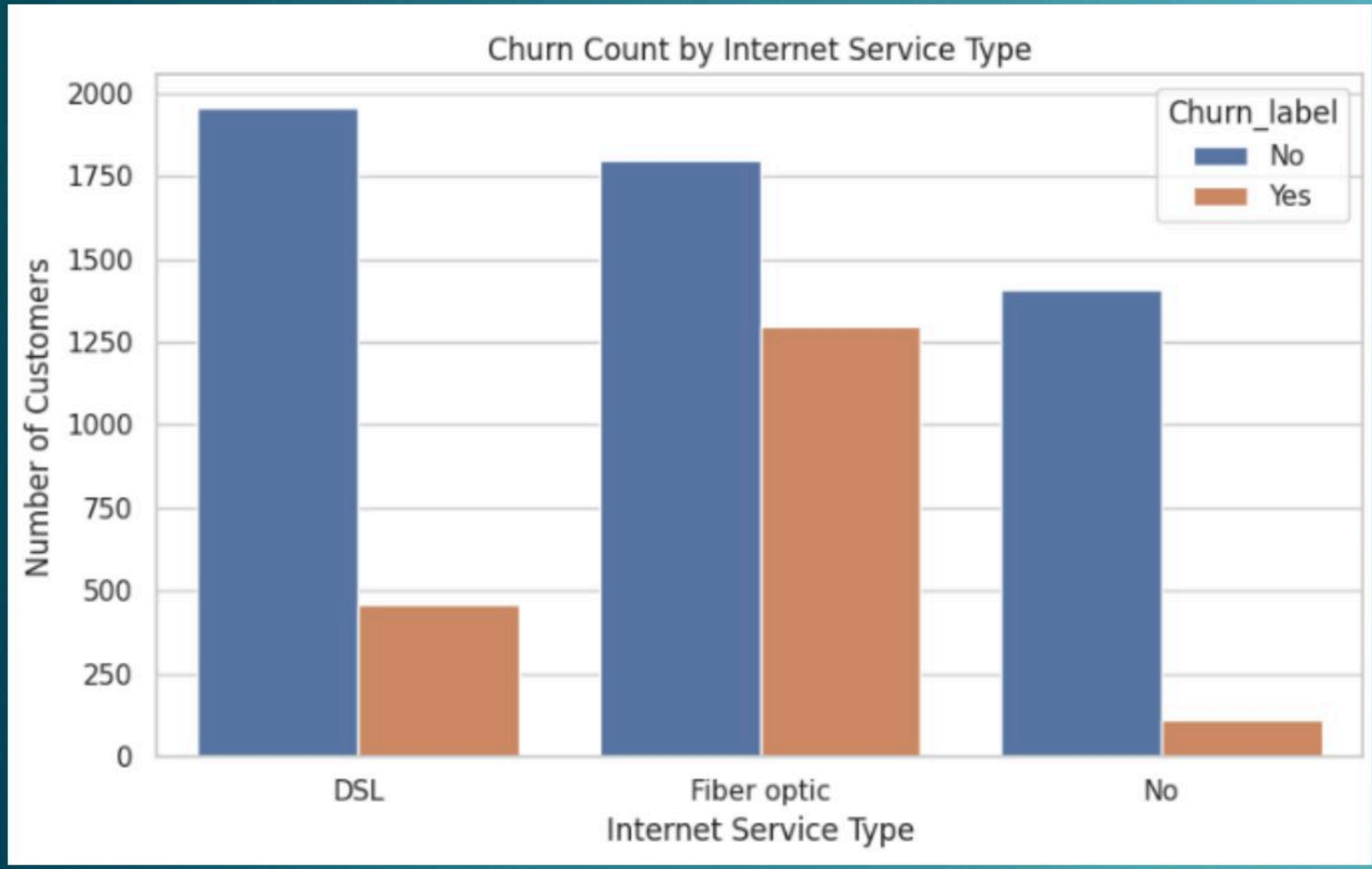
Monthly Charges:

Higher monthly charges

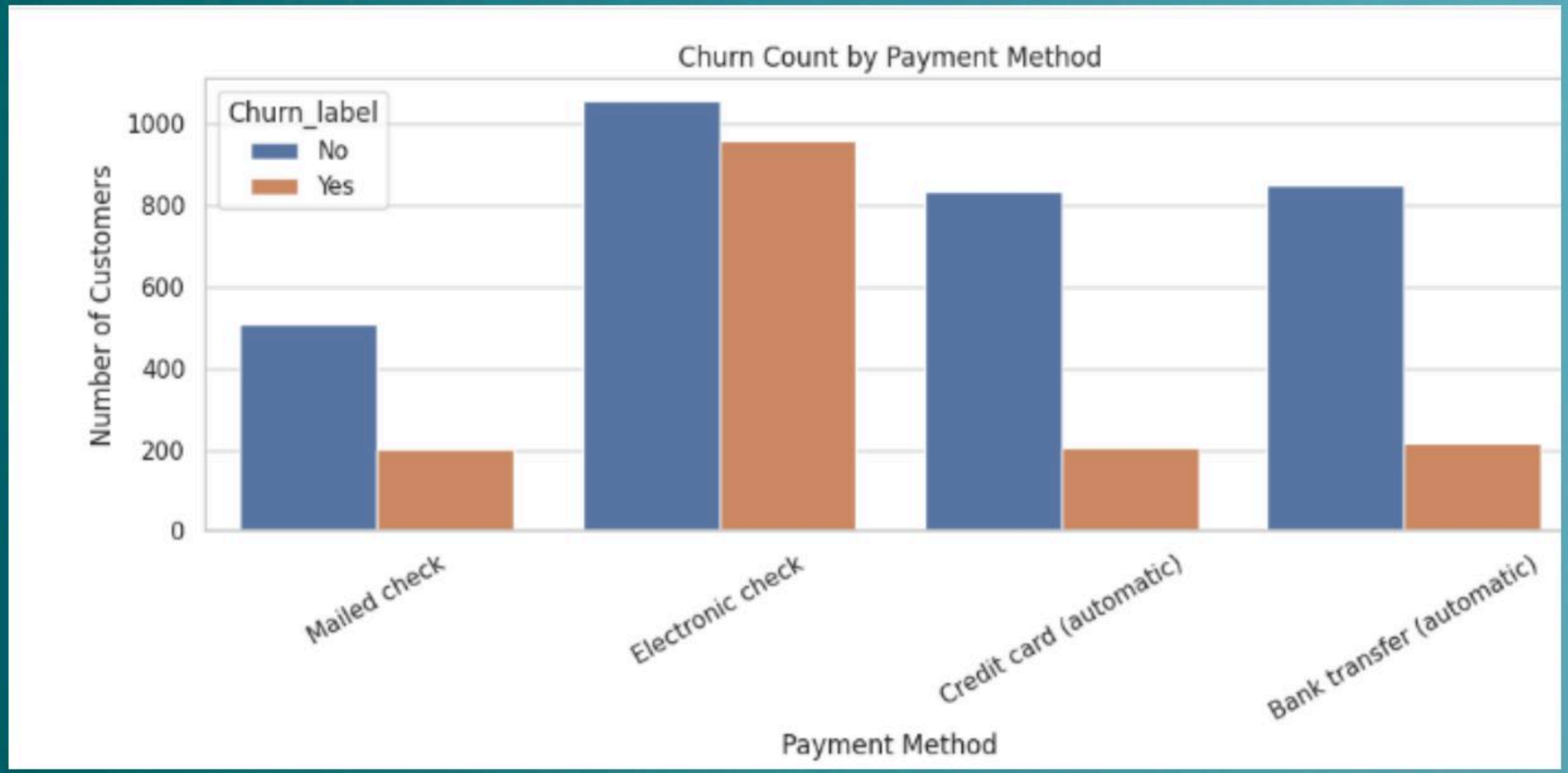
↓

more churn risk

EXPLORATORY DATA



Internet Service Type
Customers with Fiber Optic are much more likely to churn than other customers.



Payment Method
Electronic Check

↓

more churn risk

ACTIONABLE INSIGHTS

Target customers with **month-to-month contracts** for retention campaigns, they represent the highest churn risk group and can be converted to longer-term contracts with incentives.

Implement proactive customer service interventions for customers using **electronic check payments**, as this payment method shows higher churn correlation

Focus on customers with tenure **less than 12 months**—early engagement and satisfaction programs can significantly reduce first-year churn rates

LIMITATIONS OF OUR PROJECT

1. Limited Feature Scope: Dataset lacks behavioral data (call frequency, complaint history, customer service interactions) and external market factors
2. Class Imbalance Issue: Only 26.6% churn rate creates model bias toward predicting non-churn, potentially missing high-risk customers
3. Snapshot Data Only: No time-series analysis; cannot capture seasonal trends or changes in customer behavior over time
4. Model Accuracy Trade-offs: Logistic Regression (80.6%) vs Decision Tree (76.8%) show moderate performance; more complex models weren't explored
5. No Financial Impact Analysis: Models predict churn but don't include customer lifetime value (CLV) or cost-benefit analysis for retention strategies
6. Generalizability Concerns: Single-company dataset may not apply to other telecom providers; market changes require regular model retraining

THANK YOU

QUESTIONS?