TELCO-CUSTOMER-CHURN CHANGE

Li yueting 22227660
WONG WING YIN 22231137
Asma Mohammad 23203269

TABLE OF CONTENTS

purpose

- To predict whether a telecom contact will be cancelled or not
- Telecom can employ proactive tactics, including targeted marketing to lower churn rate (27%) and make better decisions with the use of predictive models.

Business Context

- **CUSTOMER RETENTION CHALLENGES:**
  - HIGH CHURN RATES IMPACTING REVENUE AND GROWTH.
  - IDENTIFYING AT-RISK CUSTOMERS IS CRUCIAL.

- **BUSINESS OBJECTIVES:**
  1. ENHANCE CUSTOMER RETENTION:
     - DEVELOP TARGETED RETENTION STRATEGIES.
     - REDUCE CHURN RATES THROUGH PROACTIVE ENGAGEMENT.
  2. IMPROVE CUSTOMER SATISFACTION:
     - UNDERSTAND CUSTOMER NEEDS AND PREFERENCES.
     - PERSONALIZE INTERACTIONS AND SERVICE OFFERINGS.

Contract	Churn
One year	No
One year	No
One year	No
Two year	No
One year	No
Two year	No
One year	No
Two year	No
One year	No
Two year	No
Two year	No
Two year	No
One year	No
Two year	No
One year	No
Two year	No
One year	No
Two year	No
Two year	No
Two year	No
One year	No
One year	No
One year	No
Two year	No
Two year	No
Two year	No
One year	No
One year	No
Two year	No
Two year	No
One year	No
Two year	No
Two year	No
One year	No

TotalCharges	Churn
8684.8	Yes
8672.45	No
8670.1	No
8594.4	No
8564.75	No
8547.15	No
8543.25	No
8529.5	No
8496.7	No
8477.7	No
8477.6	No
8476.5	No
8468.2	No
8456.75	No
8443.7	No
8436.25	No
8425.3	No
8425.15	No
8424.9	No
8405	No
8404.9	No
8399.15	No
8375.05	No
8349.7	No
8349.45	No
8337.45	No
8333.95	No
8332.15	No
8331.95	No
8317.95	No
8312.75	No
8312.4	No
8310.55	No
8309.55	No

observation

- contract: customer with permanent contract are less likely to leave
- total chart: customer who pay a higher amount are less likely to leave

Data Pre-processing

- Dataset contains 7043 customer records and 21 columns
- Data Cleaning Data is cleaned for analysis by handling missing values. converting data types
- Converted TotalCharges to numeric and handled missing values by removing rows with NaN(listwise Deletion)
- Dropped the non-numeric customerID column

NON-NUMERIC COLUMNS:

```
Before removing 'customerID':
   customerID  gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  7590-VHVEG  Female              0     Yes         No       1           No
1  5575-GNVDE    Male              0      No         No      34          Yes
2  3668-QPYBK    Male              0      No         No       2          Yes
3  7795-CFOCW    Male              0      No         No      45           No
4  9237-HQITU  Female              0      No         No       2          Yes
```

```
After removing 'customerID':
   gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  Female              0     Yes         No       1           No
1    Male              0      No         No      34          Yes
2    Male              0      No         No       2          Yes
3    Male              0      No         No      45           No
4  Female              0      No         No       2          Yes
```

HANDLING MISSING VALUES:

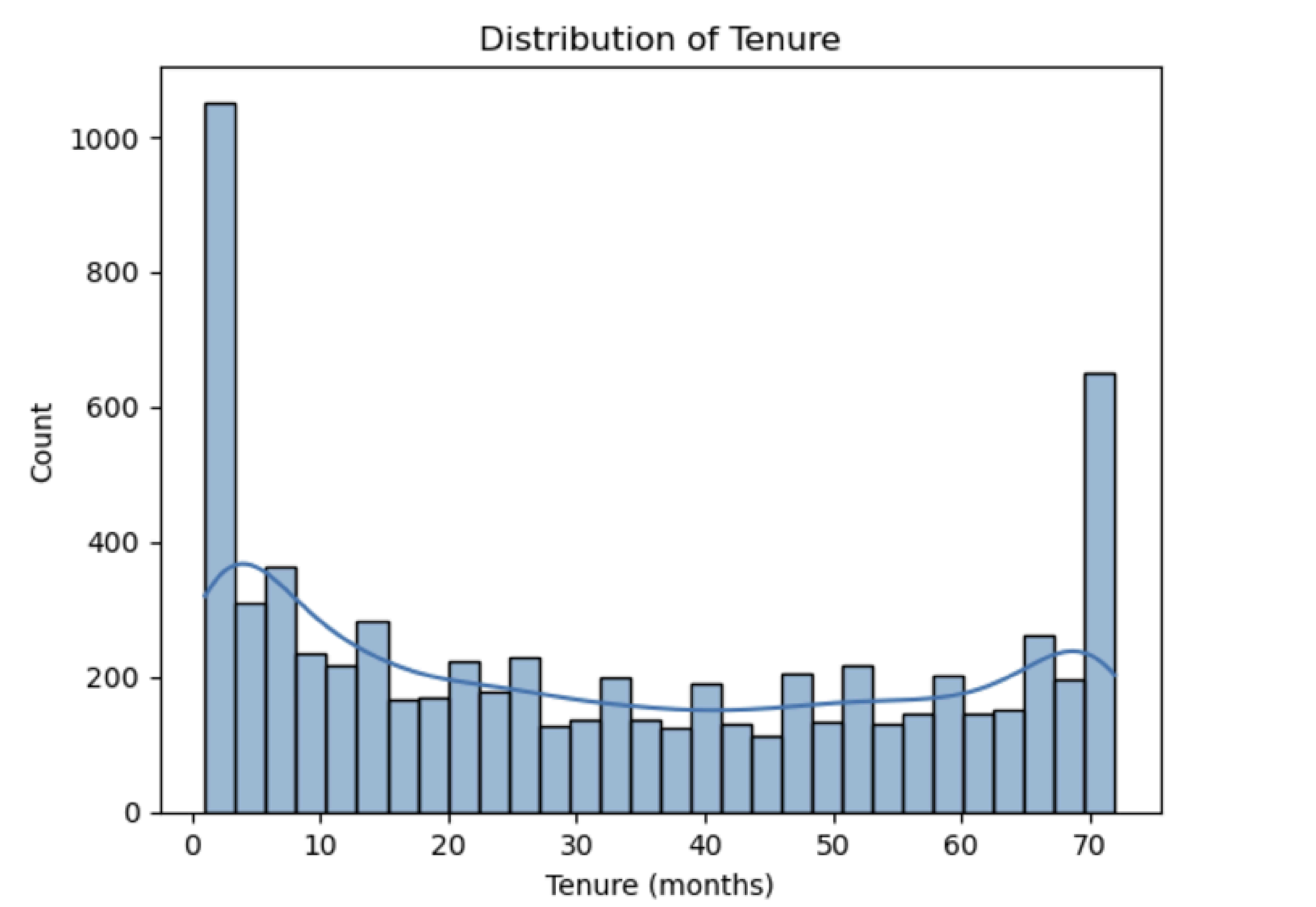
```
Total rows after cleaning: 7032
```

Data exploration analysis

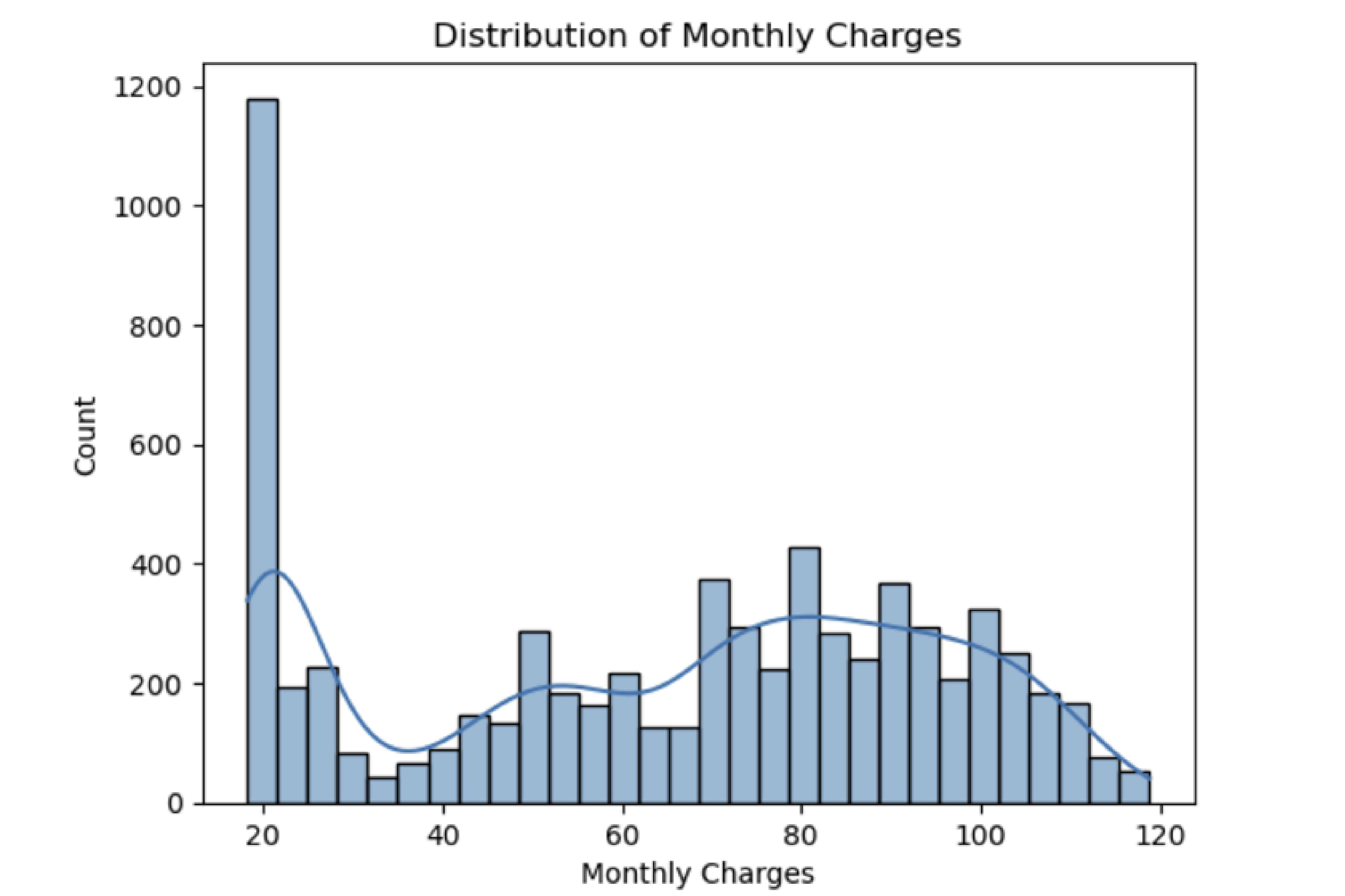
- Explore and understand the dataset, uncover patterns and identify key factors influencing customer churn
- Summary of the Statistics to understand distribution of numerical features

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges
count	7032.000000	7032.000000	7032.000000	7032.000000
mean	0.162400	32.421786	64.798208	2283.300441
std	0.368844	24.545260	30.085974	2266.771362
min	0.000000	1.000000	18.250000	18.800000
25%	0.000000	9.000000	35.587500	401.450000
50%	0.000000	29.000000	70.350000	1397.475000
75%	0.000000	55.000000	89.862500	3794.737500
max	1.000000	72.000000	118.750000	8684.800000

EDA

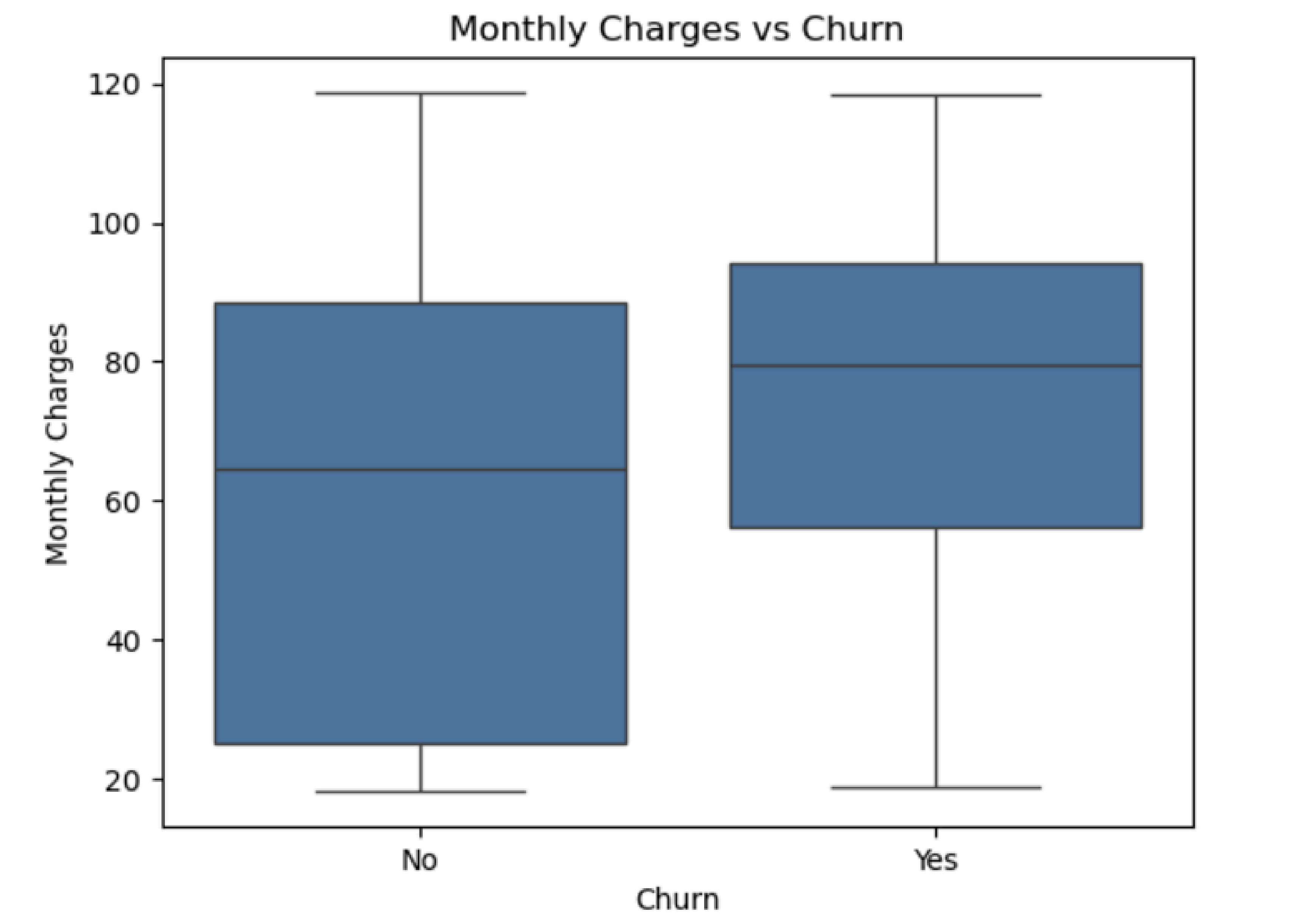


High count at 0 months: Indicate new customers
Peak around 70 months: shows good number of long-term customers



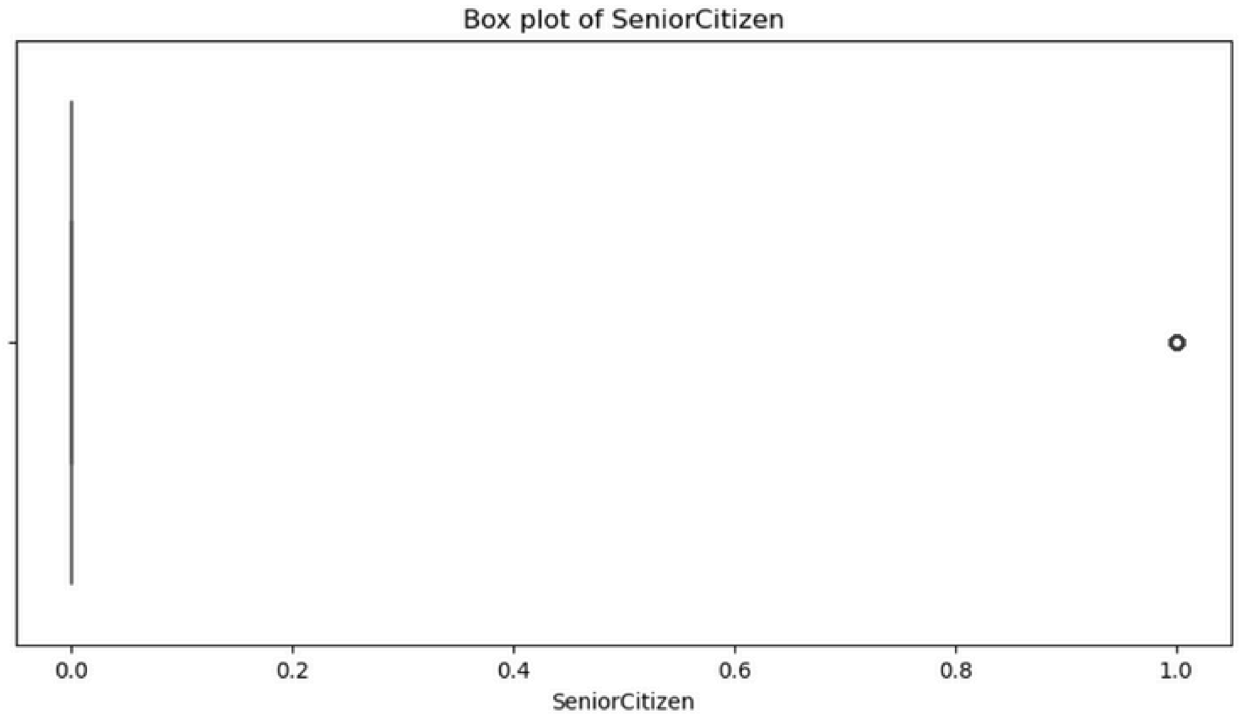
High count at low monthly charges $20
Spread at higher charges: When charged increases --> varied pricing plan and customer segments

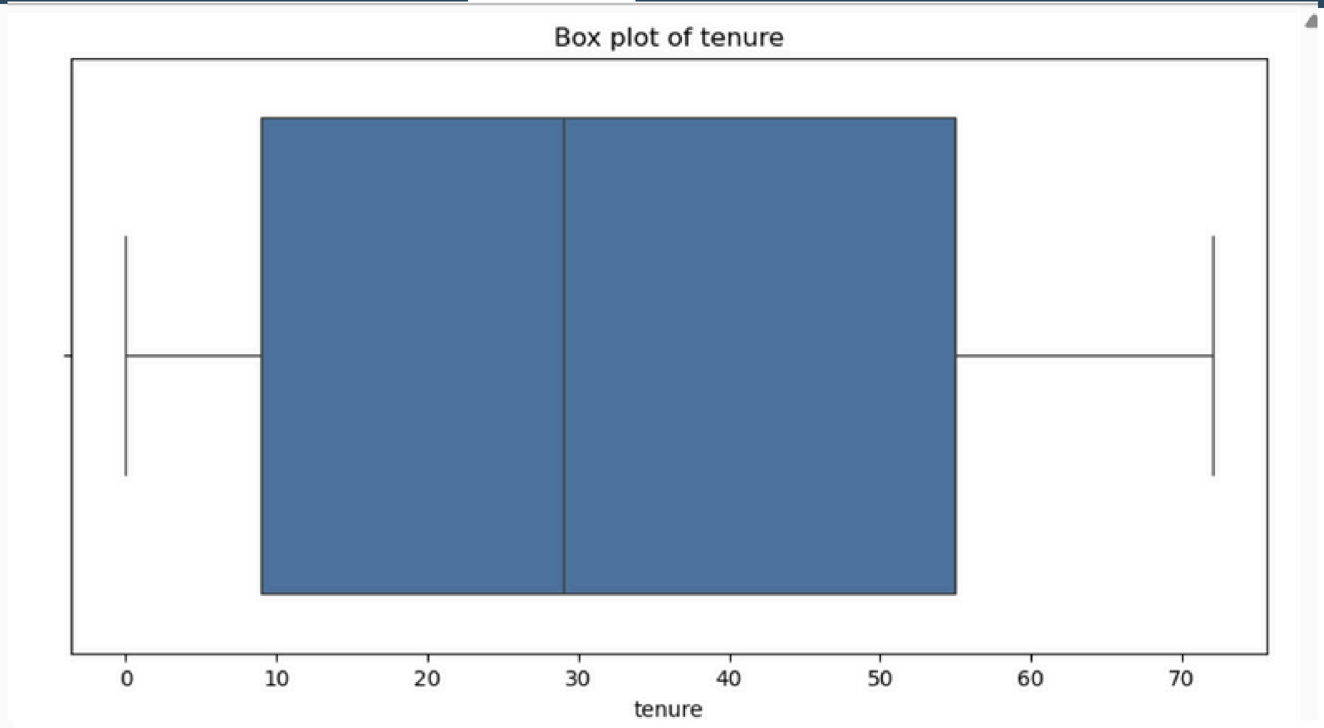
EDA

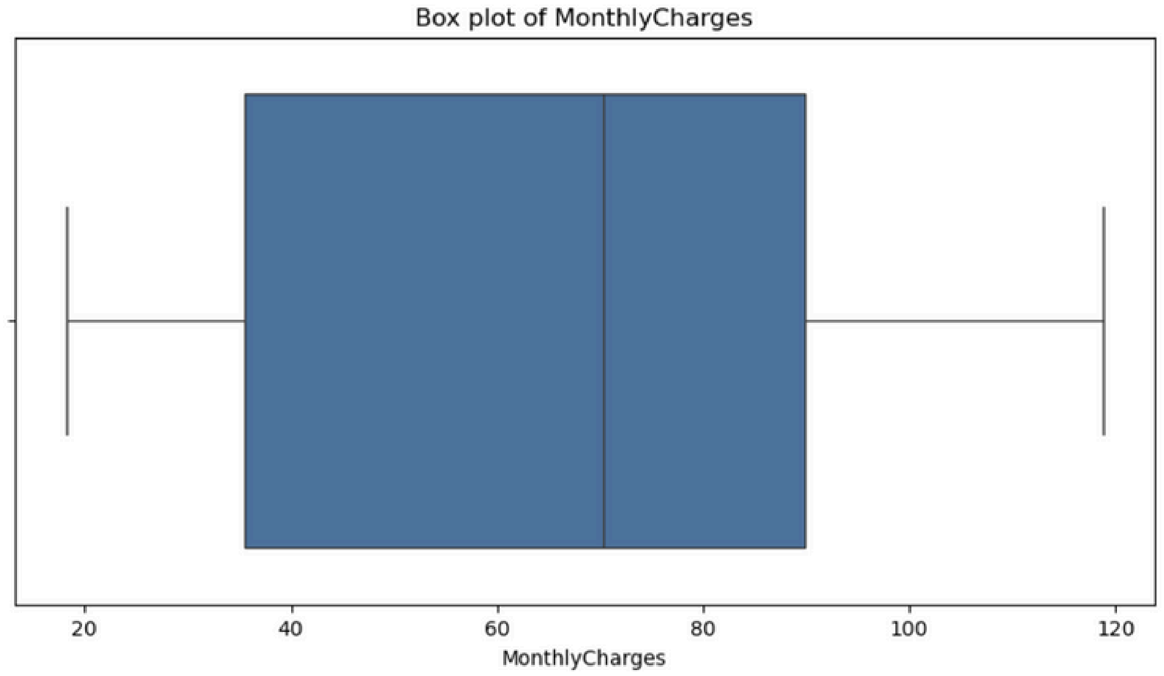


Monthly Charges vs Churn
Monthly Charges
120
100
80
60
40
20
No
Yes
Churn

Identifying the outliers







- MOST DATA POINTS ARE CONCENTRATED AT 0
- LARGE MAJORITY OF PEOPLPPE IN THIS DATASET ARE NOT SENIOR CITIZENS.

- SINGLE POINT AT 1, INDICATING OUTLIER. THIS POINT REPRESENTS SENIOR CITIZENS

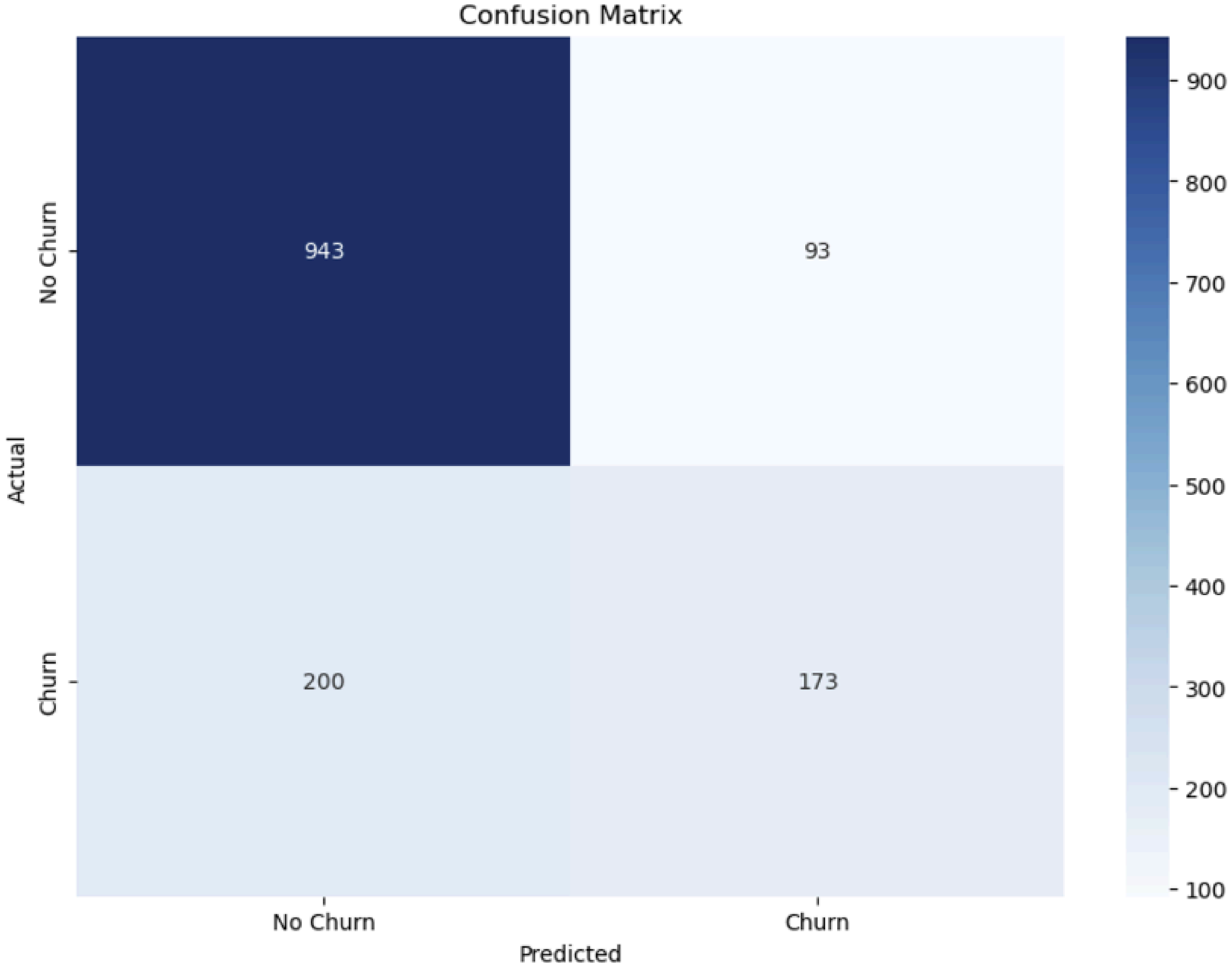
BUSINESS IMPLICATIONS:

UNDERSTANDING THE DISTRIBUTION OF TENURE HELPS IN ANALYZING CUSTOMER BEHAVIOR, IDENTIFYING SEGMENTS.

Predictive Analytics

Confusion Matrix

BASLINE OF OUR PROJECT



Confusion Matrix

Accuracy: 0.7920511000709723

Classification Report:

	precision	recall	f1-score	support
0	0.83	0.91	0.87	1036
1	0.65	0.46	0.54	373
accuracy			0.79	1409
macro avg	0.74	0.69	0.70	1409
weighted avg	0.78	0.79	0.78	1409

Confusion Matrix

BASLINE OF OUR PROJECT

FOR CLASS 0 (NO CHURN)

- PRECISION =0.83
- RECALL = 0.91

FOR CLASS 1 (CHURN)

- **PRECISION =0.65**
- **RECALL = 0.46**

OVERALL MODEL ACCURACY= 0.79 WHICH IS RELATIVELY GOOD HOWEVER LOWER RECALL FOR CHURN SUGGESTS THAT THE MODEL COULD BE IMPROVED

BUSINESS IMPLICATION

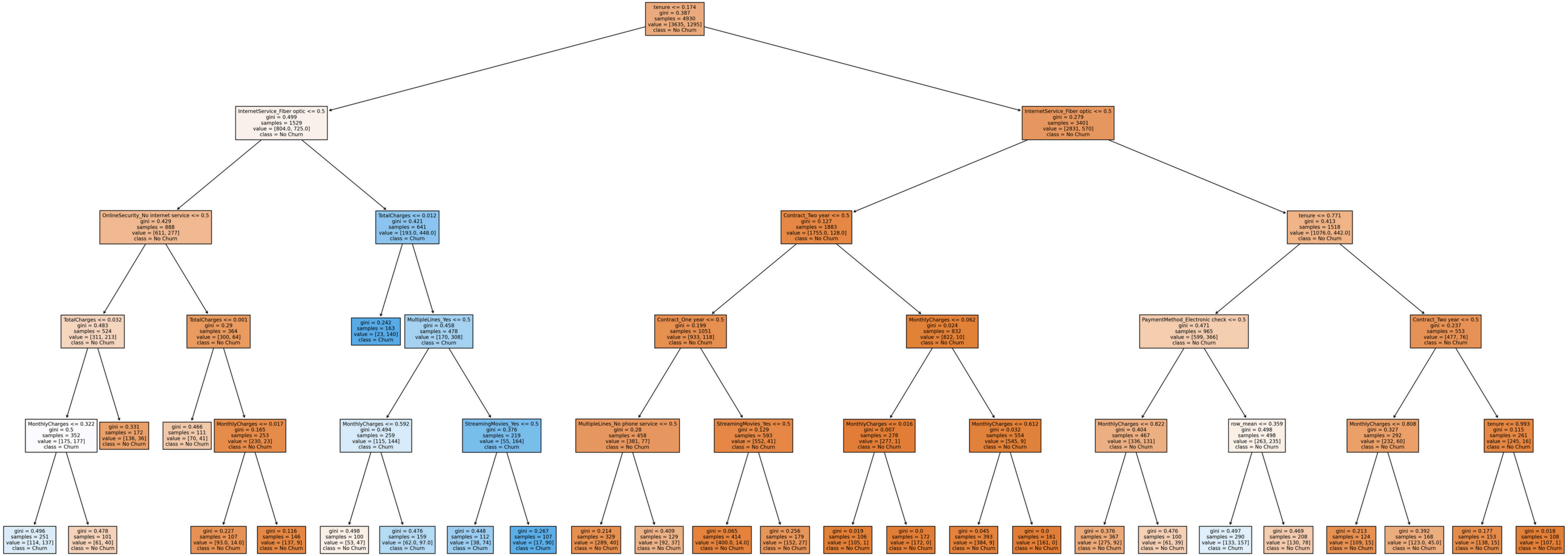
TARGETED INTERVENTIONS:

- FOCUS RETENTION EFFORTS ON CUSTOMER PREDICTED TO CHURN
- USE PRECISION AND RECALL TO PRIORTIZE INTERVENTIONS

IMPROVING MODEL PERFORMANCE:

- ADDITIONAL DATA COLLECTION OR FEATURE ENGINEERING
- OTHER MACHINE LEARNING ALGORITHMS

Decision Tree



Decision Tree

```
Accuracy: 0.7714691270404542
Classification Report:
```

	precision	recall	f1-score	support
0	0.82	0.88	0.85	1036
1	0.58	0.47	0.52	373
accuracy			0.77	1409
macro avg	0.70	0.68	0.69	1409
weighted avg	0.76	0.77	0.76	1409

- **Accuracy:**
  - **0.7714% of the samples are correctly classified**

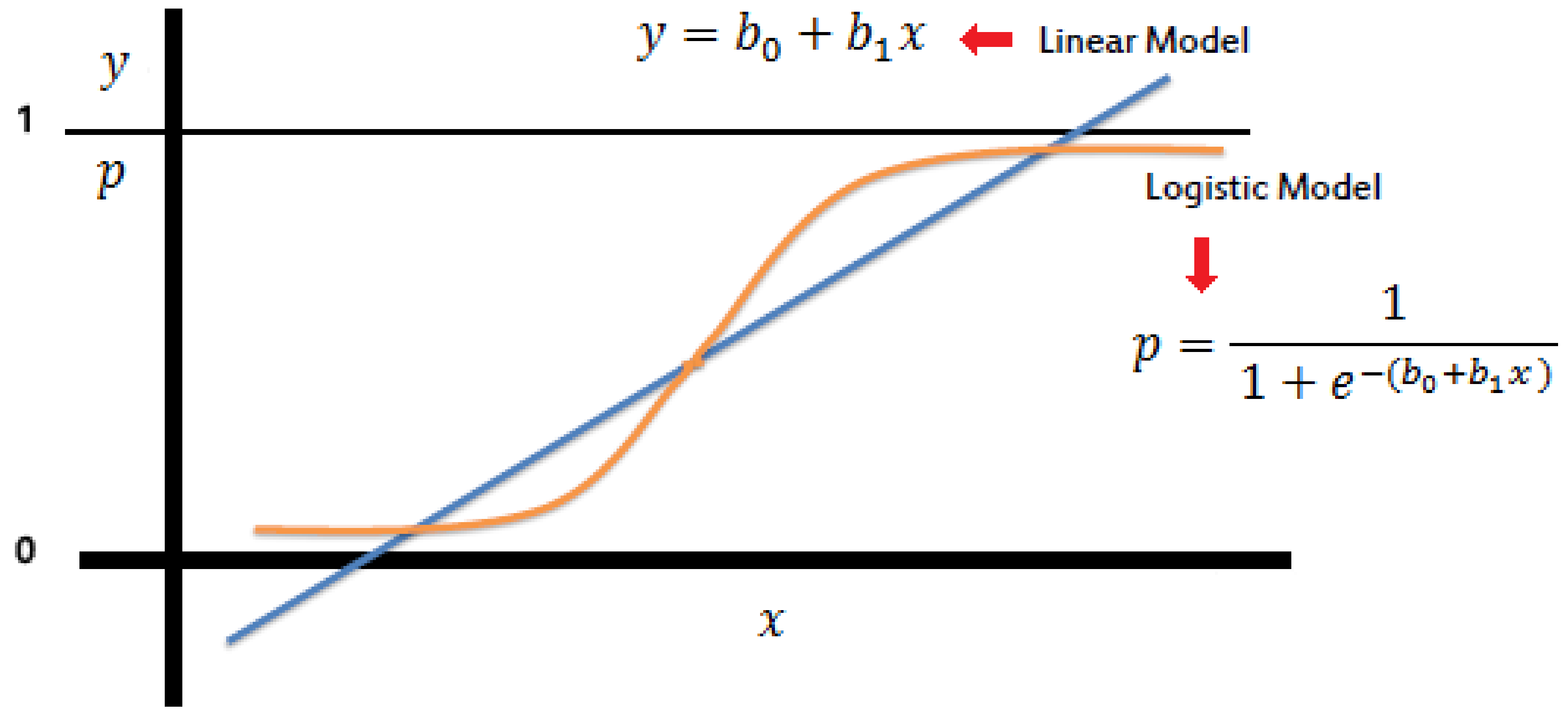
Classification Report:

- **Precision:**
  - **precision for class 0 is 0.82**
  - **for class 1 it is 0.58**
- **Recall:**
  - **The recall is 0.88 for category 0**
  - **0.47 for category 1**

Logistic Regression

What is Logistic Regression

- Data analysis technique:
  - uses mathematics > find the relationships between two data factors
- Uses this relationship to predict the value of one of those factors based on the other
- binary classification problems



Applying Logistic Regression to Telco-Customer

- Target variable: “Churn”
  - Yes
  - No

processing steps:

- Find categorical data
- Encoding categorical variables
- Fill in the null values
- Drop some less relative columns

Accuracy

```
Accuracy: 0.8204400283889283
Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.90      0.88      1036
           1       0.69      0.60      0.64       373

    accuracy                           0.82      1409
   macro avg       0.77      0.75      0.76      1409
weighted avg       0.81      0.82      0.82      1409
```

- **Accuracy:**
  - **82% of the samples are correctly classified**

Classification Report:

- **Precision:**
  - **precision for class 0 is 0.86**
  - **for class 1 it is 0.69**
- **Recall:**
  - **The recall is 0.90 for category 0**
  - **0.60 for category 1**

- **good performance**
- **need to improve for positive class**

Conclusion

01 Confusion Matrix

- precision(1):0.69
- recall(1):0.58
- accuracy:0.79

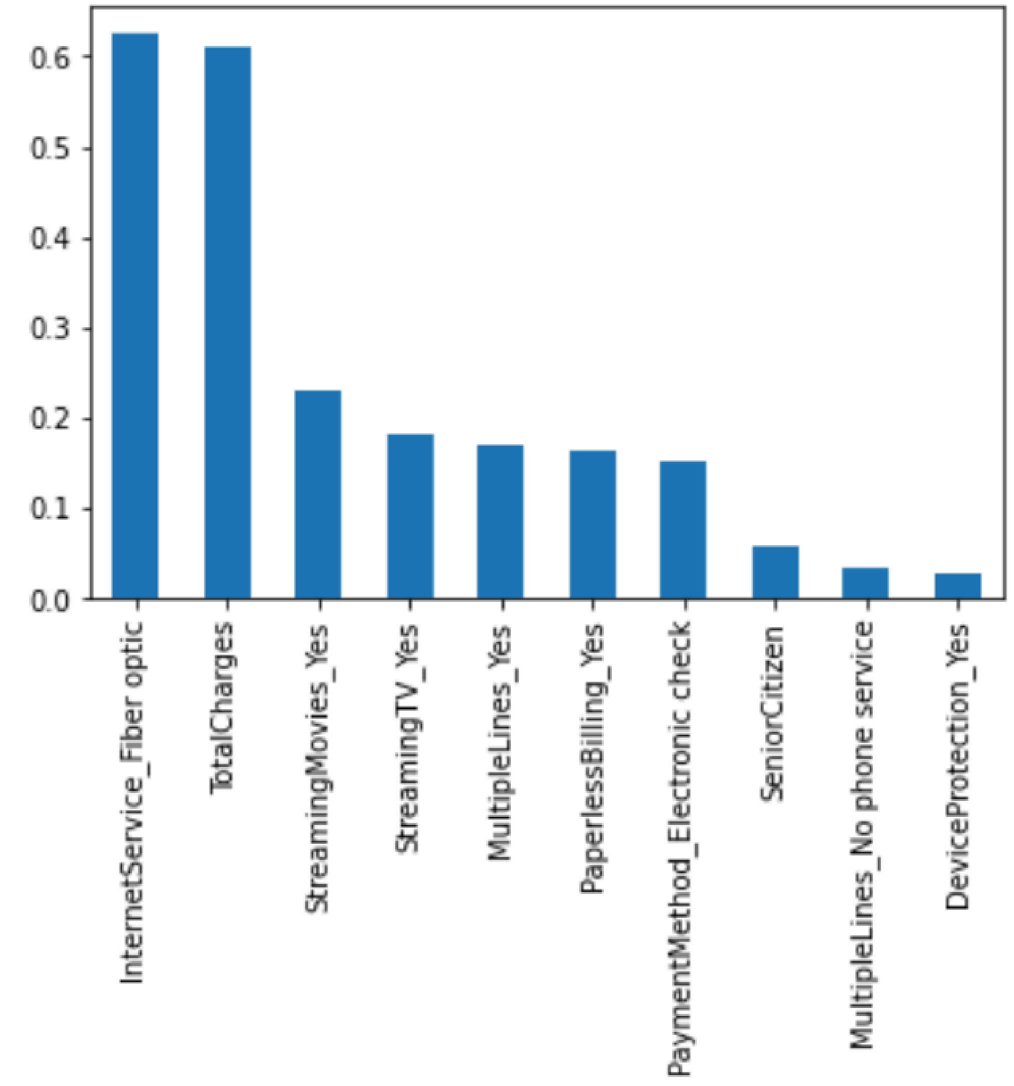
02 Decision Tree

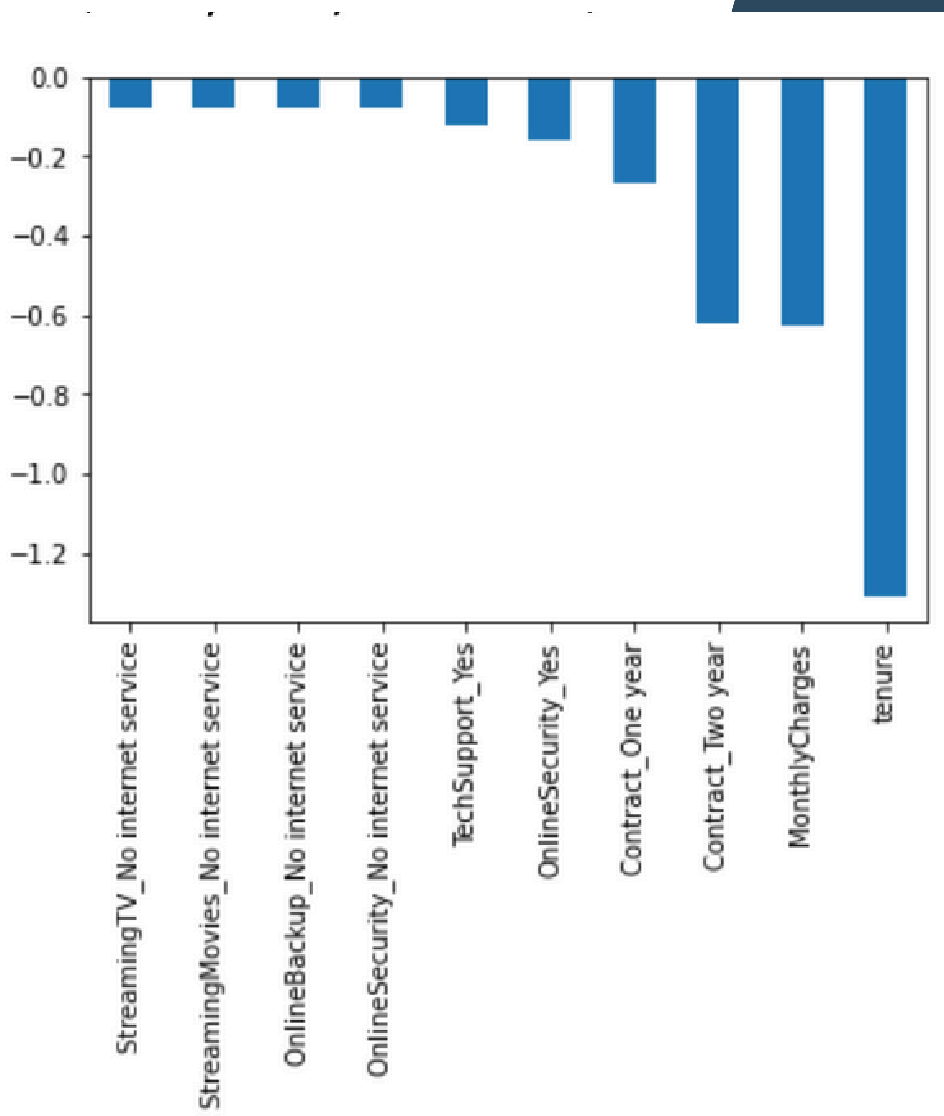
- precision(1):0.58
- recall(1):0.47
- accuracy:0.77

03 Logistic Regression

- precision(1):0.69
- recall(1):0.60
- accuracy:0.82

Finding and recommendation

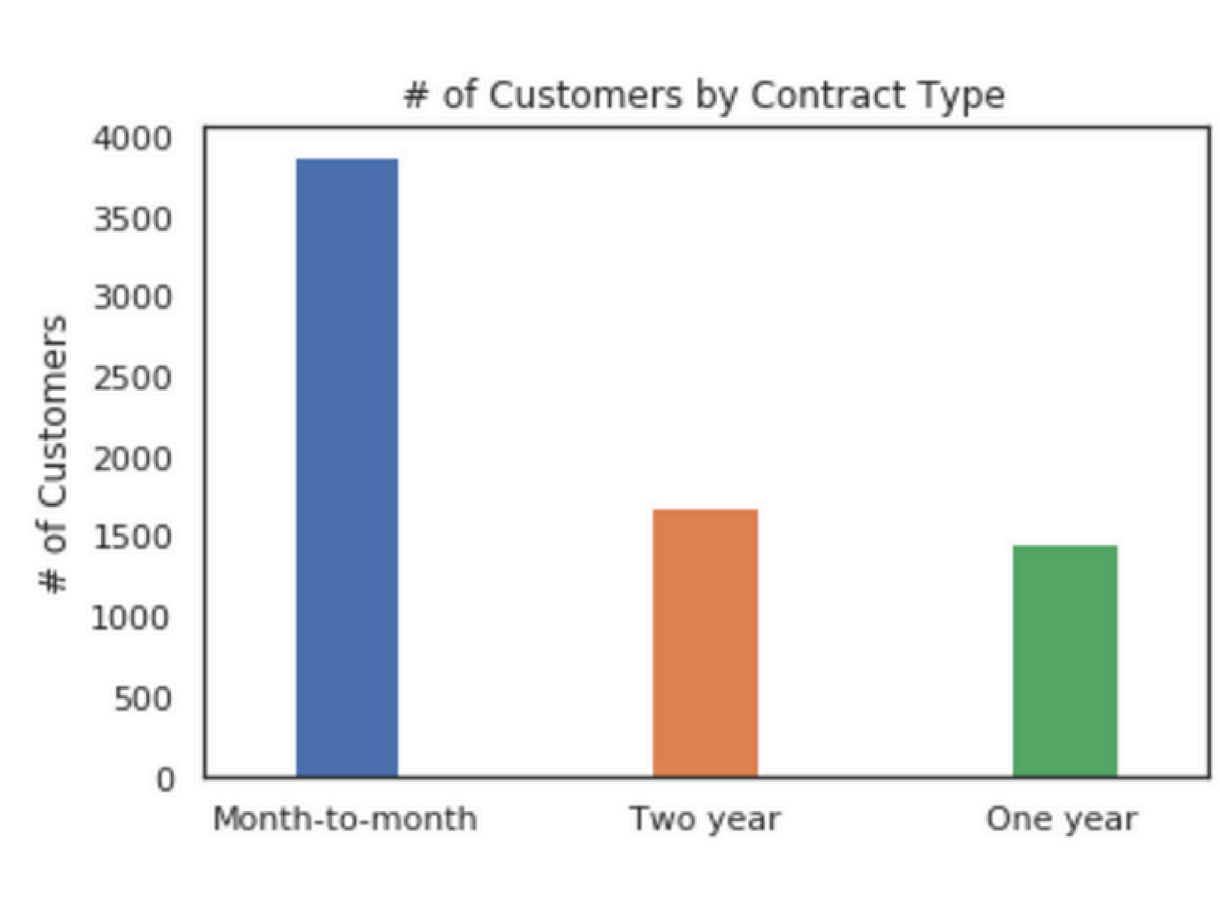




- higher fibre optic internet services, total charges, can lead to higher churn rates.
- monthly charges , longer contract and tenure can lead to a lower churn rates

Recommendation

1. Encourage Longer Contracts
   - Correlation: Offering longer contract terms leads to lower churn rates. Customers who commit for an extended period may feel a sense of obligation to stay with the service.Longer contracts can provide customers with predictable pricing, reducing anxiety over potential rate increases.



Recommendation

2. Lower the Total Charges
    - Correlation: Higher churn rates may result from higher total fees. Many consumers are sensitive to price, especially in competitive markets. When faced with high costs, they are more likely to evaluate other options.

Thank you!

References

https://www.kaggle.com/code/bandiatindra/telecom-churn-prediction#After-going-through-the-above-EDA-we-will-develop-some-predictive-models-and-compare-them.