<u>Telco-Customer-Churn</u>

<u>EDA and Decision Tree Analysis Report</u>

Submitted to:

Prof. Faiz Hamid

Submitted by:

Akash SK

2336304

Introduction

Using provided Telco-Customer-Churn dataset find insights on what prompted customers to change Telecom provider and what worked and helped in retaining customers using Exploratory Data Analysis (EDA) and Decision Tree analysis. Comparing results from 2 analysis and record observations.

Exploratory Data Analysis

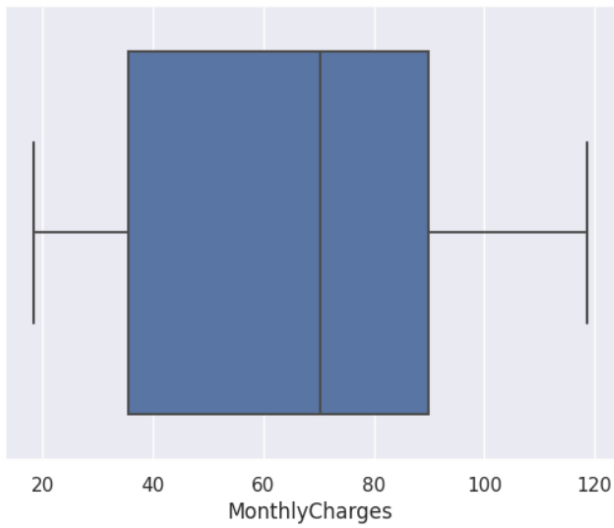
In the EDA phase of the analysis, the dataset "Telco-Customer-Churn" was thoroughly examined to uncover its inherent characteristics and uncover potential trends and patterns. The dataset contains a total of 7043 entries and 21 columns. The key features that were explored during the EDA process include:

- customerID: A unique identifier for each customer.
- gender: The gender of the customer.
- SeniorCitizen: A binary flag indicating whether the customer is a senior citizen.
- Partner: Whether the customer has a partner.
- Dependents: Whether the customer has dependents.
- tenure: The duration of the customer's relationship with the service.
- PhoneService: Whether the customer has phone service.
- MultipleLines: Whether the customer has multiple phone lines.
- InternetService: The type of internet service subscribed to.
- OnlineSecurity: Whether the customer has online security features.
- OnlineBackup: Whether the customer has online backup features.
- DeviceProtection: Whether the customer has device protection features.
- TechSupport: Whether the customer has tech support features.
- StreamingTV: Whether the customer has streaming TV services.
- StreamingMovies: Whether the customer has streaming movie services.
- Contract: The type of contract between the customer and the service provider.
- PaperlessBilling: Whether the customer uses paperless billing.
- PaymentMethod: The method of payment used by the customer.
- MonthlyCharges: The monthly subscription cost for customers.
- TotalCharges: The total amount charged to customers over their subscription period.
- Churn: Whether the customer has churned (left the service).

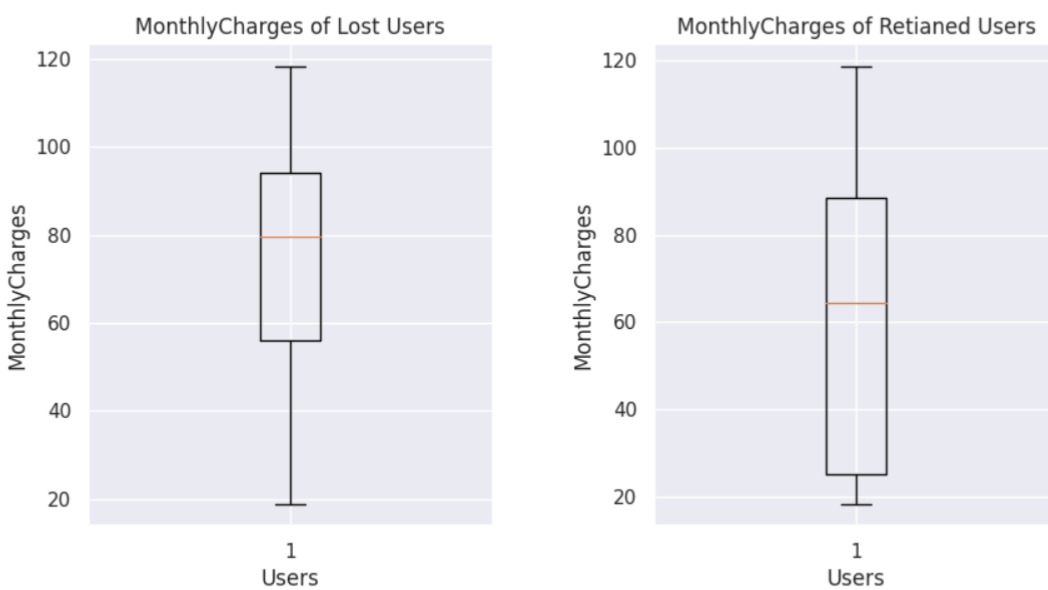
There were not null entries in the dataset, provided the data is clean data was further expressed in visual form using box plots, QQ plots, bar plots etc.

1. Monthly Charges

   Plotting monthly charges using a box plot it was observed that there is no outliers in the data.

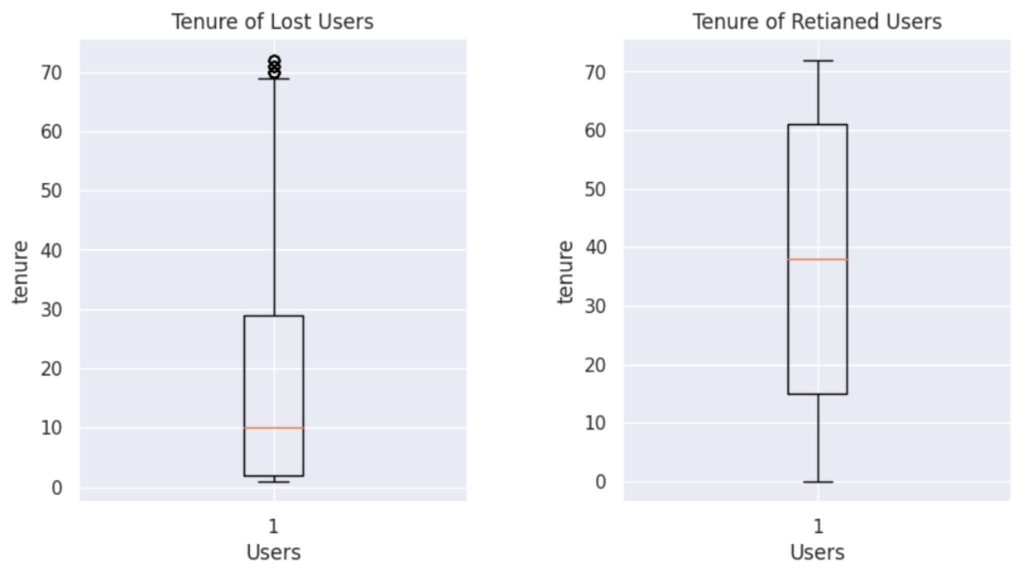


Creating comparative plots for box plot between churned and Un-churned customers monthly charges gives a better insight.



On average, lost users had higher monthly charges than retained users. Additionally, the range of values for lost users is larger than that for retained users, indicating more variability in monthly charges among lost users.

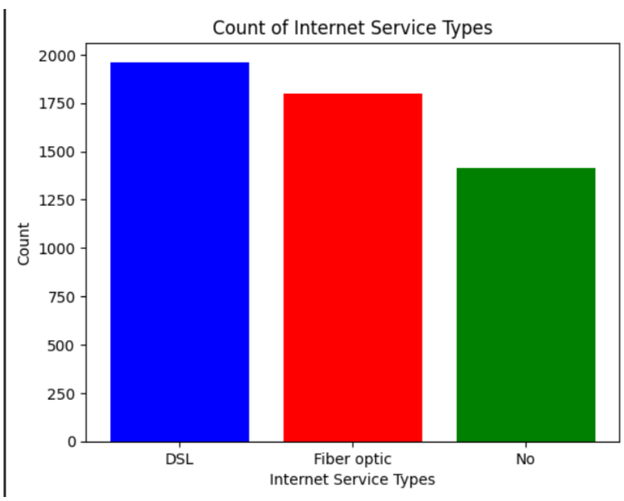
2. Tenure



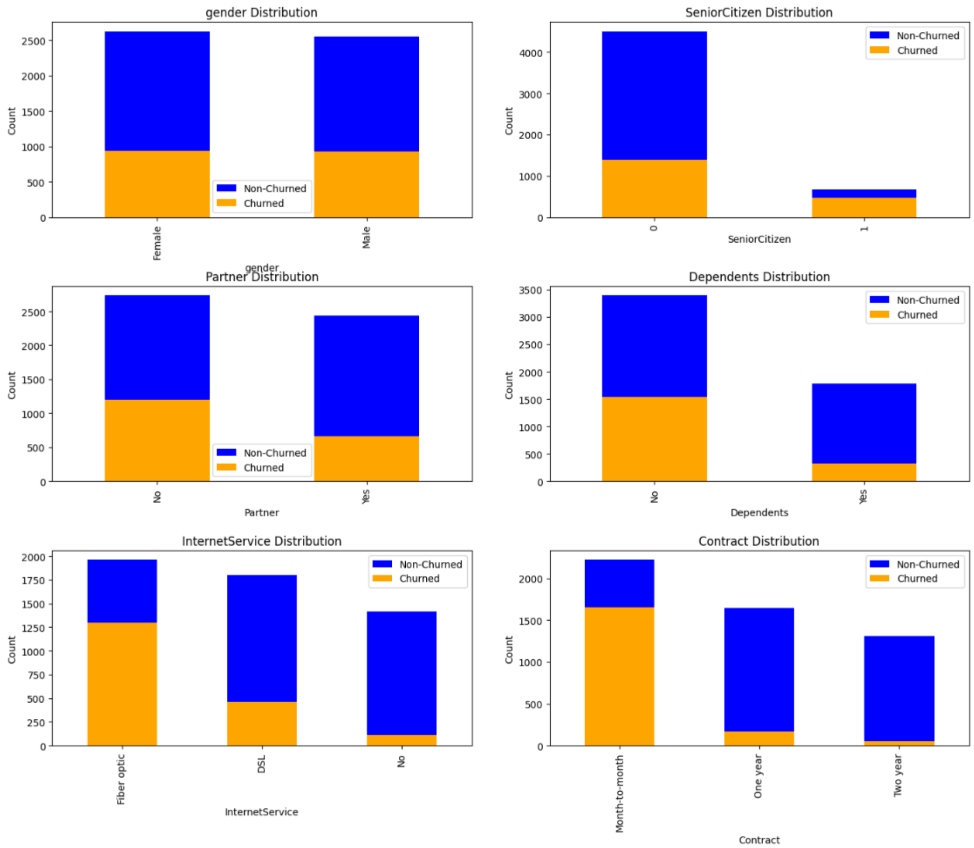
The median value for the "Tenure of Lost Users" is around 20, while the median value for the "Tenure of Retained Users" is around 40. This means that, on average, lost users had a shorter tenure than retained users. Additionally, both plots have a similar range of values, indicating a similar level of variability in tenure among both lost and retained users

<u>What Worked?</u>

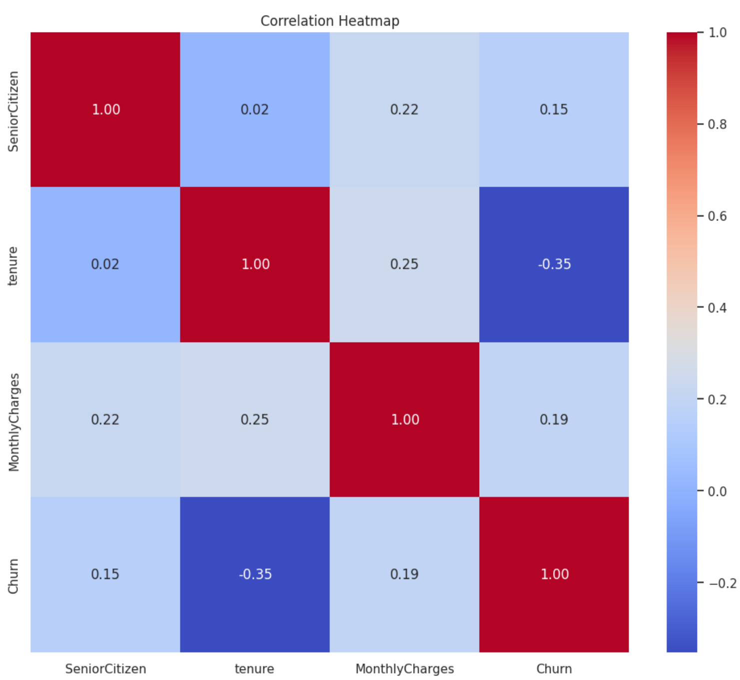
Customers who have already left the company won't come back, so we need to figure out what worked for the costumers who were retained, so we can apply that make sure the retained users are not churned further.



The graph above shows the internet service types opted by Un-churned users.



The graphs above show various factor as comparison between churned and non-churned users.



The correlation heatmap gives an insight how different factors influence churn of customers. From the matrix we can see than churn rate has high negative correlation with tenure.

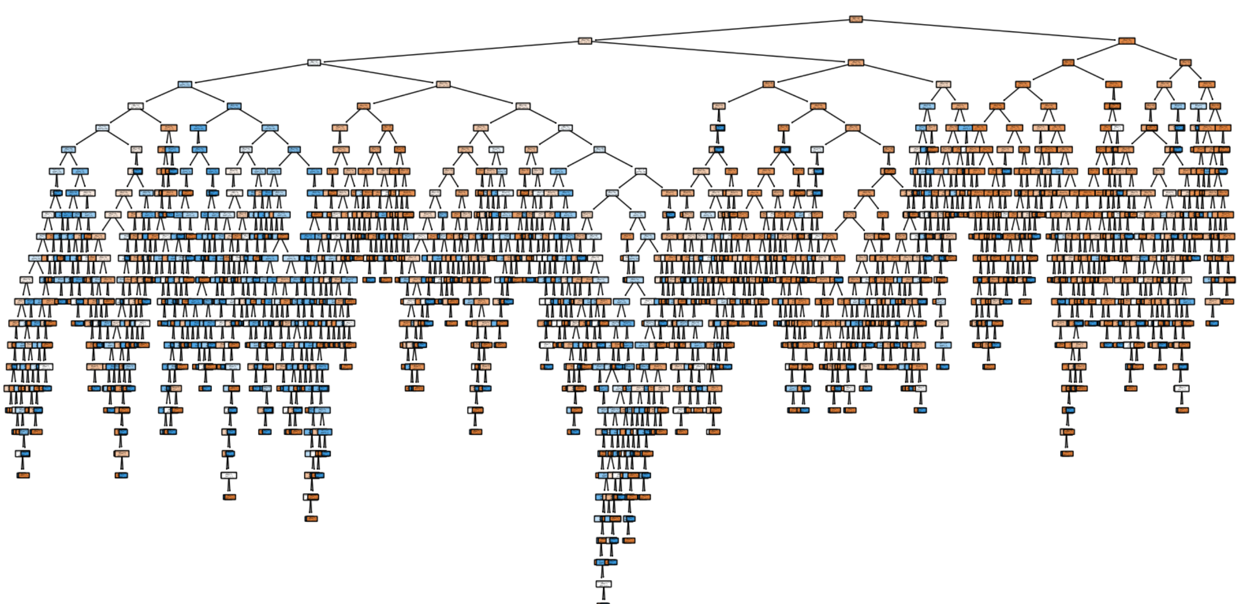
Decision Tree

The data set was feed into a decision tree prediction model and subsequent tree was generated with rules that would help to identify and predict if a user is at the risk of churning or not. The data was feed into a decision tree but before that the costumer id was dropped cause its unique to user.

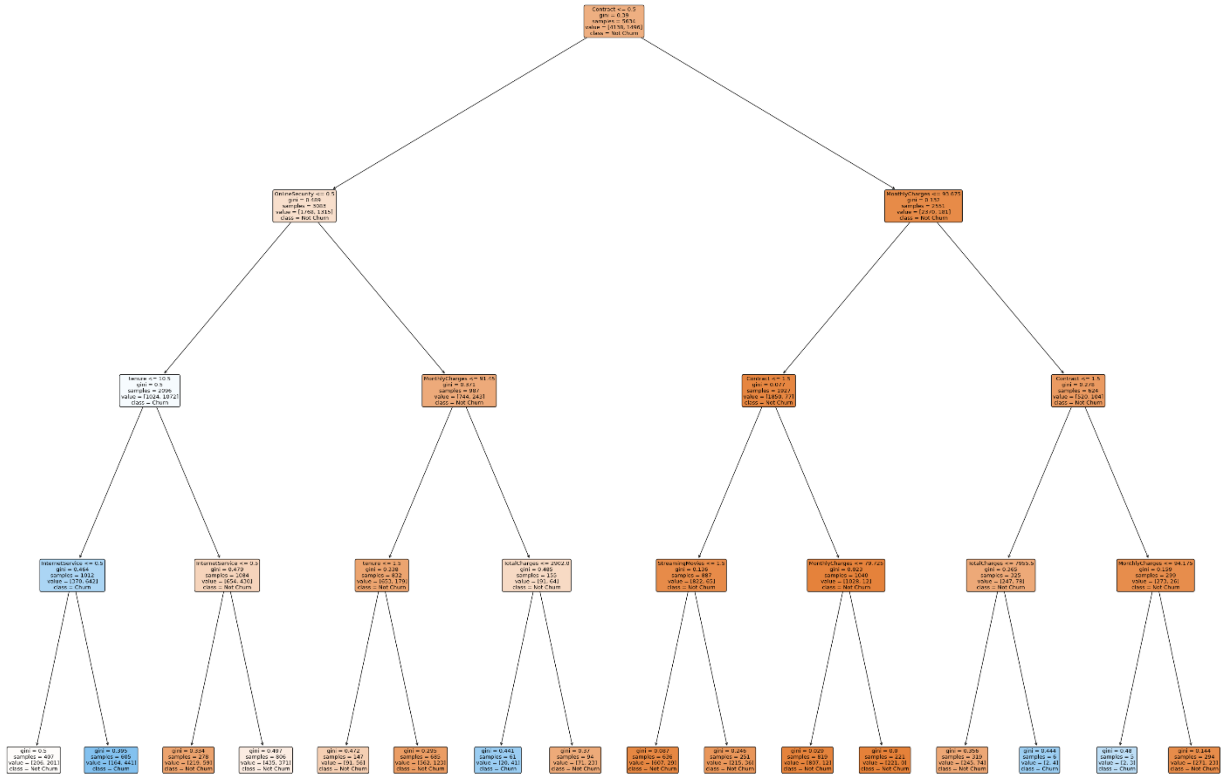
The dataset was split up into train and test data and provided an acceptable accuracy.

```
Accuracy: 0.7253371185237757
Confusion Matrix:
 [[833 203]
 [184 189]]
Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.80      0.81      1036
           1       0.48      0.51      0.49       373

    accuracy                           0.73      1409
   macro avg       0.65      0.66      0.65      1409
weighted avg       0.73      0.73      0.73      1409
```

Accuracy:0.73



The visual representation of the tree however was too big and unoptimized for observation, thus we pruned the tree to a depth of 4 resulting in the tree given below



Exporting the rules from the final tree is as follows:

Decision Tree Rules:

|--- Contract <= 0.50

| |--- OnlineSecurity <= 0.50

| | |--- tenure <= 10.50

| | | |--- InternetService <= 0.50

| | | | |--- class: 0

| | | |--- InternetService > 0.50

| | | | |--- class: 1

| | |--- tenure > 10.50

| | | |--- InternetService <= 0.50

| | | | |--- class: 0

| | | |--- InternetService > 0.50

| | | | |--- class: 0

```
|   |--- OnlineSecurity >  0.50
|   |   |--- MonthlyCharges <= 91.45
|   |   |   |--- tenure <= 1.50
|   |   |   |   |--- class: 0
|   |   |   |--- tenure >  1.50
|   |   |   |   |--- class: 0
|   |   |--- MonthlyCharges >  91.45
|   |   |   |--- TotalCharges <= 2902.00
|   |   |   |   |--- class: 1
|   |   |   |--- TotalCharges >  2902.00
|   |   |   |   |--- class: 0
|--- Contract >  0.50
|   |--- MonthlyCharges <= 93.67
|   |   |--- Contract <= 1.50
|   |   |   |--- StreamingMovies <= 1.50
|   |   |   |   |--- class: 0
|   |   |   |--- StreamingMovies >  1.50
|   |   |   |   |--- class: 0
|   |   |--- Contract >  1.50
|   |   |   |--- MonthlyCharges <= 79.72
|   |   |   |   |--- class: 0
|   |   |   |--- MonthlyCharges >  79.72
|   |   |   |   |--- class: 0
|   |--- MonthlyCharges >  93.67
|   |   |--- Contract <= 1.50
|   |   |   |--- TotalCharges <= 7955.50
|   |   |   |   |--- class: 0
|   |   |   |--- TotalCharges >  7955.50
|   |   |   |   |--- class: 1
|   |   |--- Contract >  1.50
|   |   |   |--- MonthlyCharges <= 94.17
|   |   |   |   |--- class: 1
|   |   |   |--- MonthlyCharges >  94.17
|   |   |   |   |--- class: 0
```

Rule 1: Contract <= 0.50

If the value of the "Contract" feature is less than or equal to 0.50, move to the next condition.

This condition seems to suggest that customers with shorter contract durations are being considered.

Rule 2: Online Security <= 0.50 (Inside Rule 1)

If the value of the "Online Security" feature is less than or equal to 0.50, move to the next condition.

This indicates that the availability of online security might influence the decision for customers with shorter contracts.

Rule 3: Tenure <= 10.50 (Inside Rule 2)

If the value of the "tenure" feature is less than or equal to 10.50, move to the next condition.

This suggests that the tenure of customers with shorter contracts and no online security is being considered.

Rule 4: Internet Service <= 0.50 (Inside Rule 3)

If the value of the "Internet Service" feature is less than or equal to 0.50, the predicted class is 0 (non-churn).

Customers with shorter contracts, no online security, shorter tenure, and a specific type of internet service are predicted to not churn.

Rule 5: Internet Service > 0.50 (Inside Rule 3)

If the value of the "Internet Service" feature is greater than 0.50, the predicted class is 1 (churn).

Customers with shorter contracts, no online security, shorter tenure, and a different type of internet service are predicted to churn.

CONCLUSION

- Shorter contract durations, lack of online security, and higher monthly charges are key factors linked to higher churn rates.
- Specific internet service types, combined with contract and monthly charges, influence churn predictions.
- Recommendations include offering longer contracts with benefits, improving online security, and optimizing pricing strategies.
- Tailoring marketing efforts to streaming preferences and contract types can enhance customer retention strategies.