Telco Customer Churn

Julius Hermanto

Workflow

Business Problem

Objectives

Data Understanding

Descriptive Statistics

Exploratory Data Analysis

Deep Dive Questions

Data Preprocessing

Churn Prediction Modeling

Recommendations

Business Problem

26,53% of 7043 customers stopped using Telco Services which had a hit to the company's revenue.

Objectives

01

Find the cause of churn using Exploratory Data Analysis.

02

Predict customer churn.

03

Present recommendations for minimizing churn rate.

Data Understanding

Data Understanding

The Dataset has 7043 data with 21 columns:

- 4 columns contain numerical data
- 17 columns contain non-numerical data

The target variable is **Churn** which contains either **0** or **1** indicating it is non-numerical data.

Variable	Description
customerID	Customer identifier number
gender	Customer's gender
SeniorCitizen	Customer Age Range
Partner	Customer has or has no partner
Dependents	Customer have or have no dependents
tenure	How long they've been a customer
PhoneService	Service – Phone Service
MultipleLines	Service – Multiple Lines
InternetService	Service – Internet Service
OnlineSecurity	Service – Online Security
OnlineBackup	Service – Online Backup
DeviceProtection	Service – Device Protection
TechSupport	Service – Tech Support
StreamingTV	Service – Streaming TV
StreamingMovies	Service – Streaming Movies
Contract	Customer signed contract
PaperlessBilling	Paper / Paperless Billing
PaymentMethod	Customer's payment method of choice
MonthlyCharges	Customer's monthly charges
TotalCharges	Customers total charges
Churn	Customer who left within the last month

Descriptive Statistics

Descriptive Statistics

Based on non-numerical data, most customers are **Male** with **no partner** and **no dependents**. They mostly subscribe **only** to a single line of **Phone Service**. Even if they do subscribe to an internet service, the top subscribed internet service is **Fiber Optic**, and they don't subscribe to additional services that we offer. Customers top contract use is **month-to-month** and they prefer **using paperless billing** with a payment method of **electronic check**.

Most of the customers **do not churn** but there it raises a concern that **26.53%** of the total data **churned**.

According to numerical data:

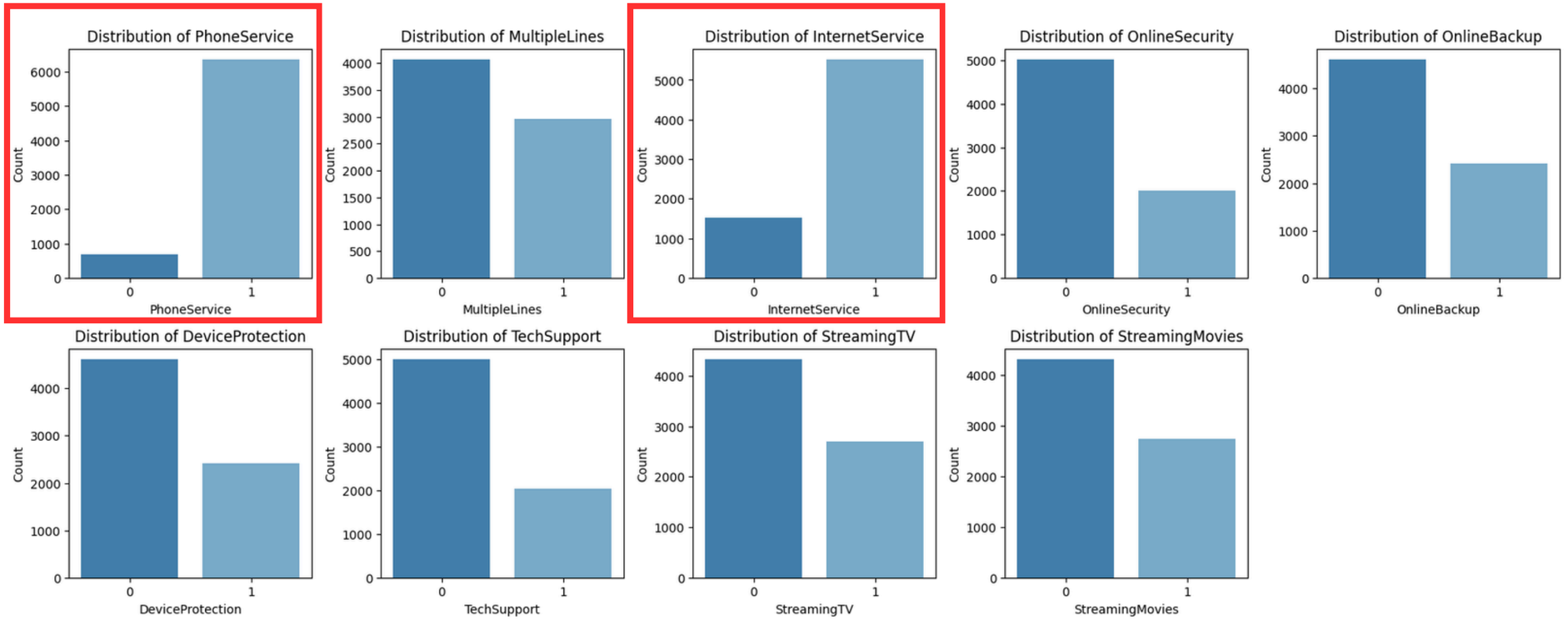
- Customers spend a median tenure of **29**
- Customers spend median monthly charges of **$70.35**
- Customers spend a median total charge of **$1397.48**

Exploratory Data Analysis

Univariate Analysis

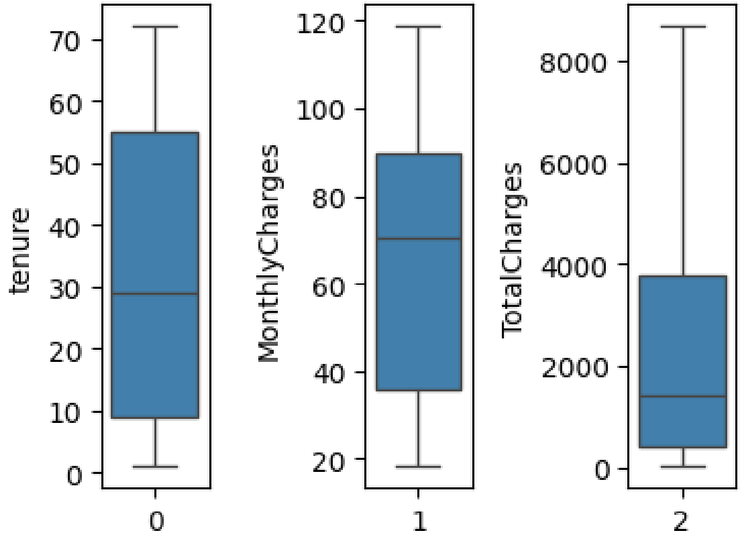
Univariate Analysis

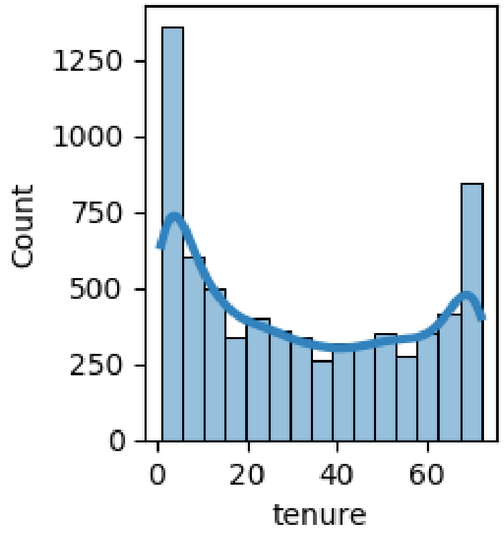
Most customers only subscribe to **Phone Service** and **Internet Service** without considering subscribing to our other services. Based on the distribution, **InternetService** may provide deep insight into why customers may churn, since it's upselling packages like **OnlineSecurity**, **TechSupport**, etc may correlate to Internet Service.

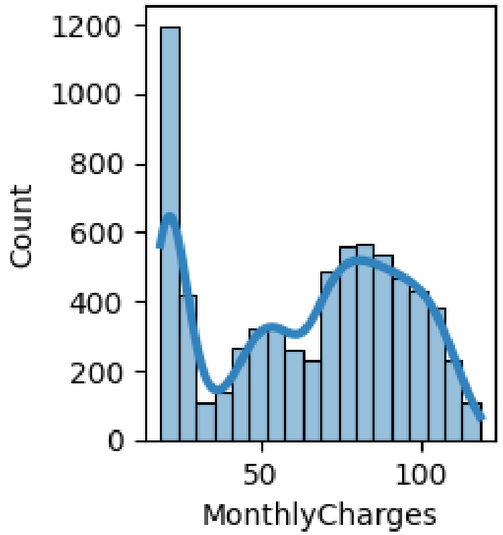


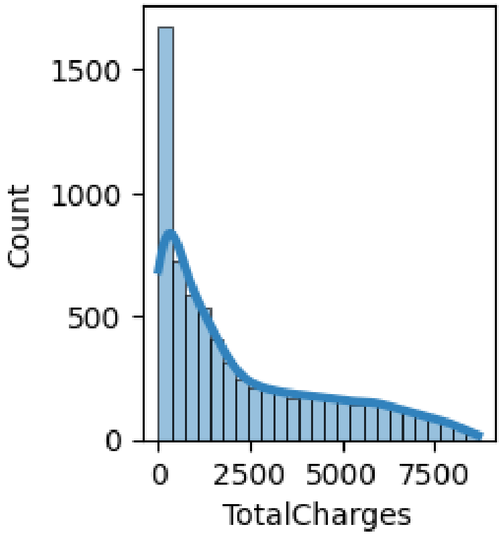
Univariate Analysis

- **Tenure and Monthly Charges**: Both features exhibit bimodal skew, indicating **possible gaps in feature pricing**. This disparity might influence the customer churn rate.
- **Total Charges**: The positive skewness suggests that **most customers prefer smaller packages** or subscribe to fewer features.



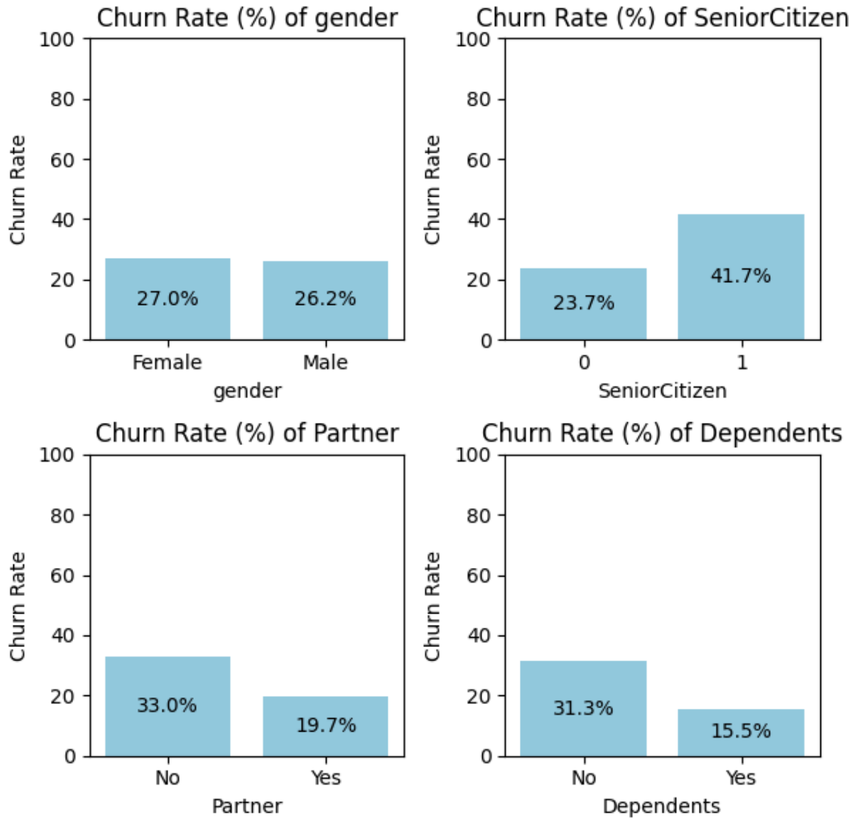






Multivariate Analysis

Customer Demographic Analysis

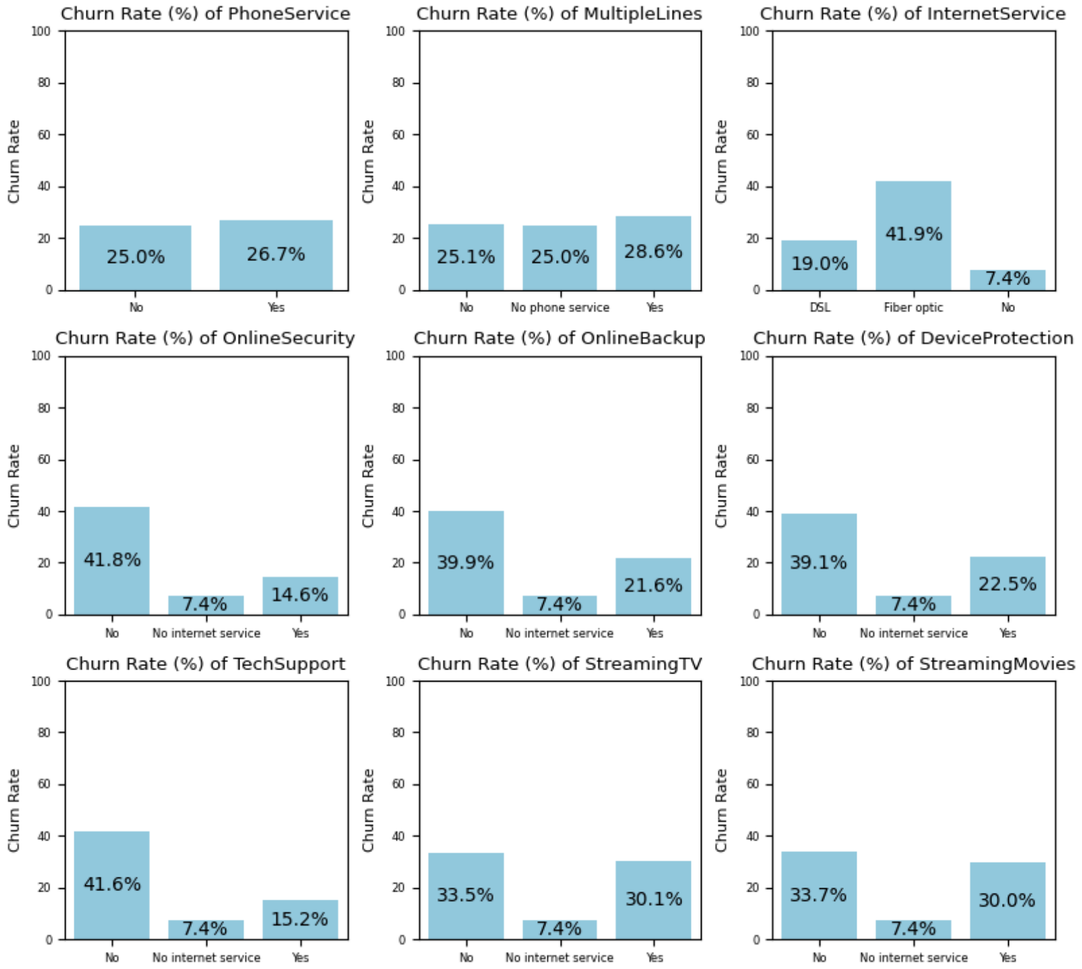


- After running the dependency test using chi2_dependency, **Gender** does not correlate to **Churn** while **SeniorCitizen**, **Partner**, and **Dependents** correlate to **Churn**.

- Customers who are **senior citizens do tend to churn**.

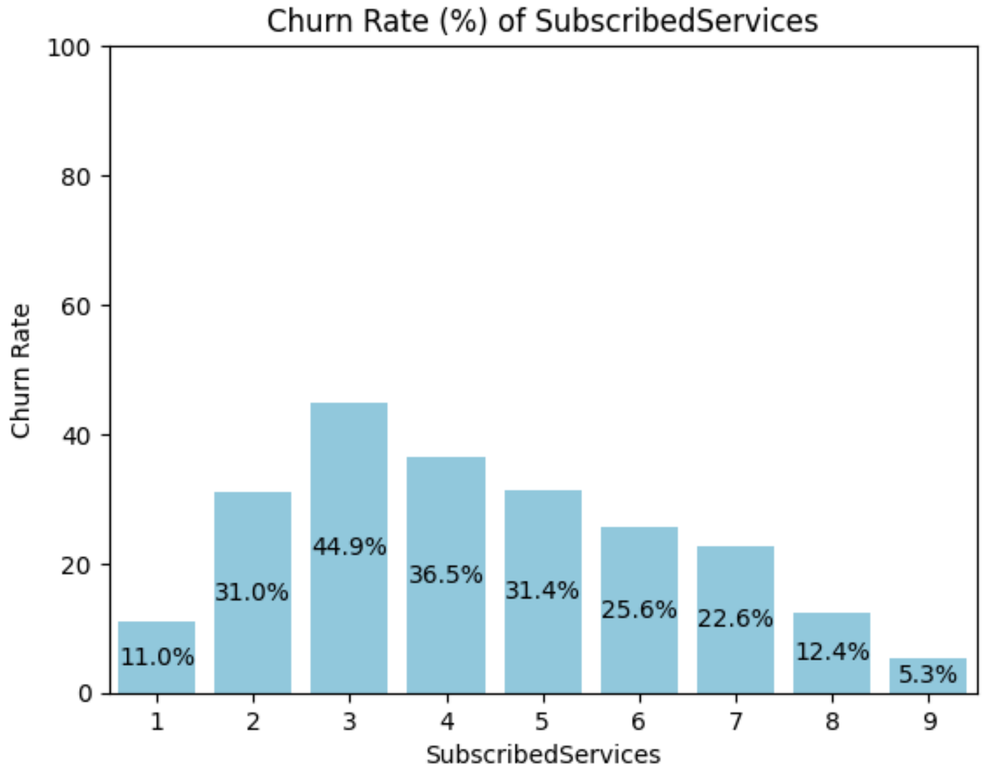
- Customers who have **no partner** (either single, divorced, deceased, etc) and have **no dependents do tend to churn**.

Telco Service Analysis



- After running the dependency test using chi2_dependency, only **PhoneService** does not correlate to **Churn**.
- Customers who subscribe to **MultipleLines** tend to churn.
- Customers who subscribe to **Fiber Optic** Internet Service have a higher chance of churning than other options.
- Customers who subscribe to an **InternetService** that do not use protection nor support like **OnlineSecurity**, **TechSupport`**, **OnlineBackup**, **DeviceProtection** tend to be more likely to churn.

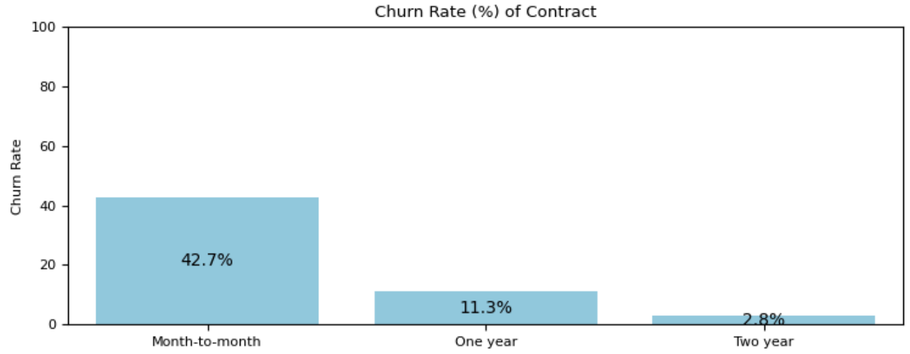
Telco Service Analysis (2)

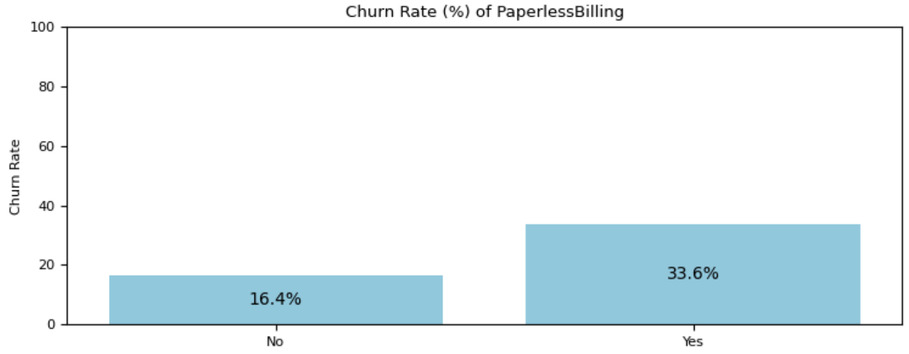


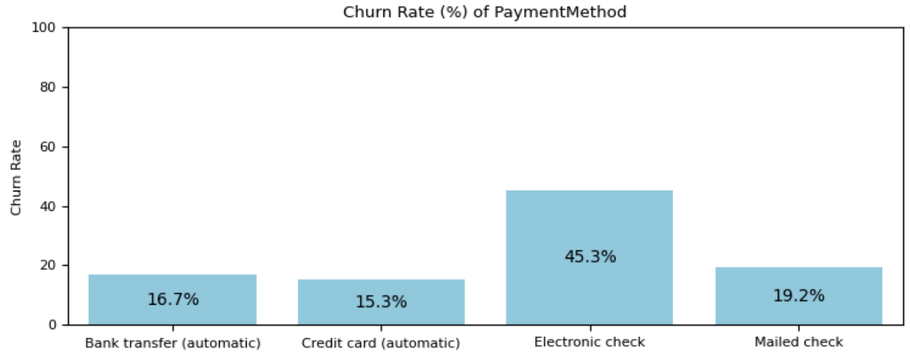
The more customer subscribes to our services, the less likely they churn.

Customer Account Info Analysis

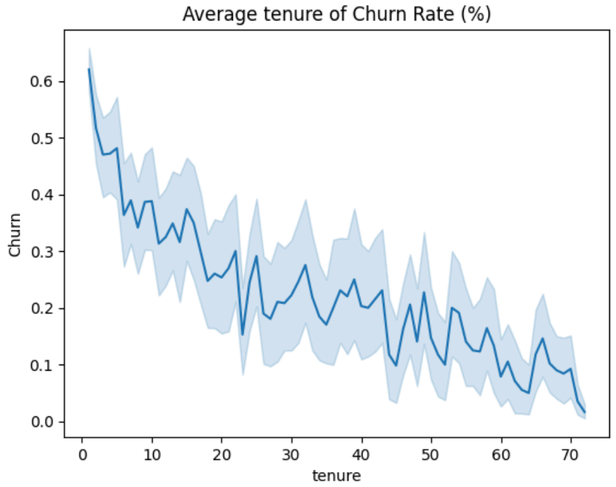
- After running the dependency test using chi2_dependency, **all of the customer account info correlates to Churn.**
- Customers with month-to-month contracts, with paperless billing, and using electronic checks tend to churn.

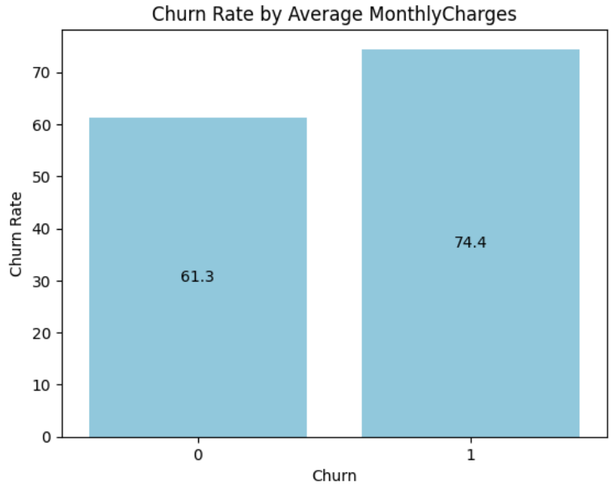


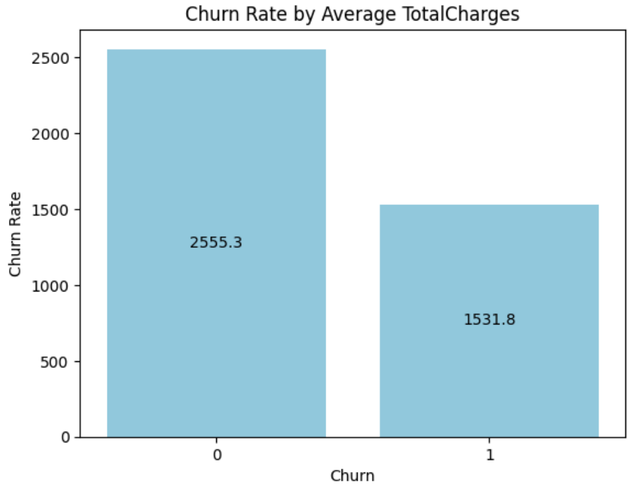




Customer Spending Analysis







- Customers with low **tenure** tend to churn.
- Customers who tend to churn have higher than average **MonthlyCharges** than those who do not churn.
- A combination of low tenure with high monthly charge and high churn rate affects a low total charge and presents us with a low churn rate on above average totalCharges.

Deep Dive Questions

Deep Dive Questions

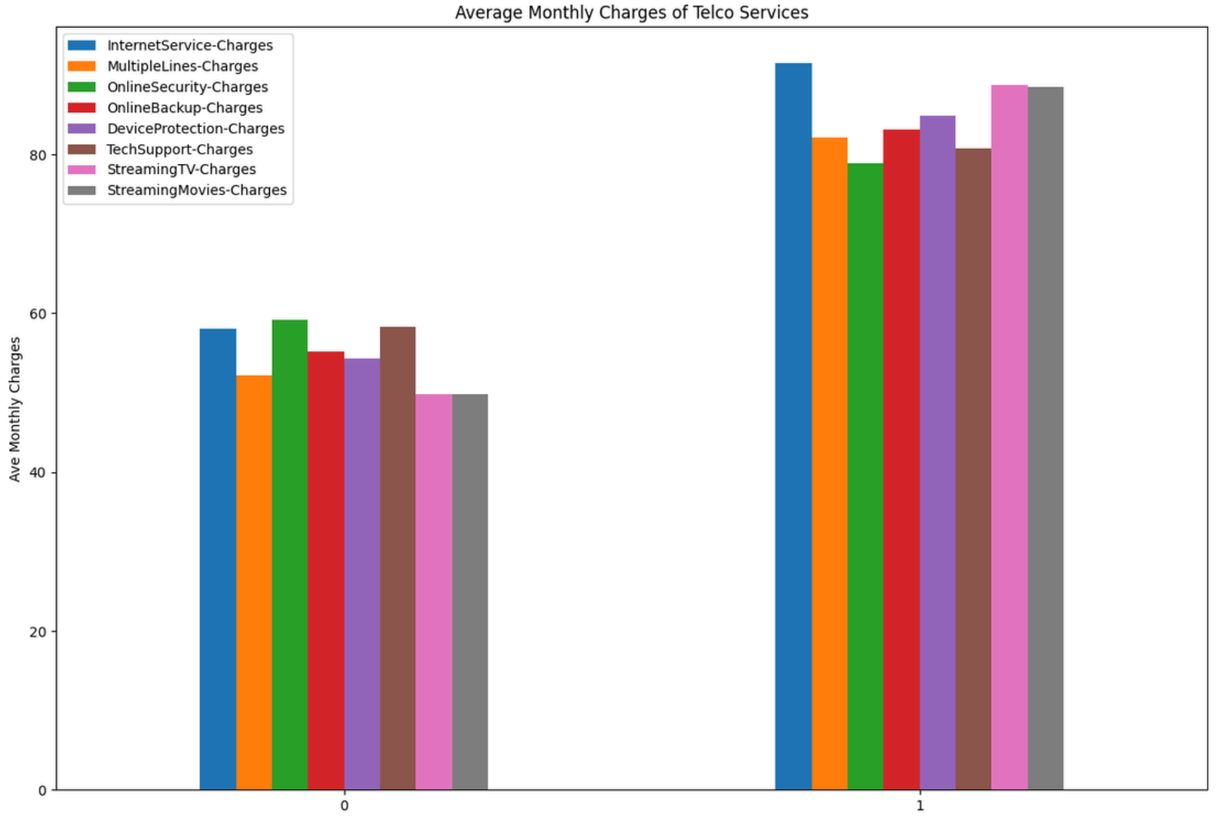
1. What service price affects customer's churn rate?

2. Is the price of **InternetService** fiber optic the reason for the high churn rate?

3. Do customers with higher **tenure** tend to choose longer-term **Contracts**, and how does this impact their churn rate?

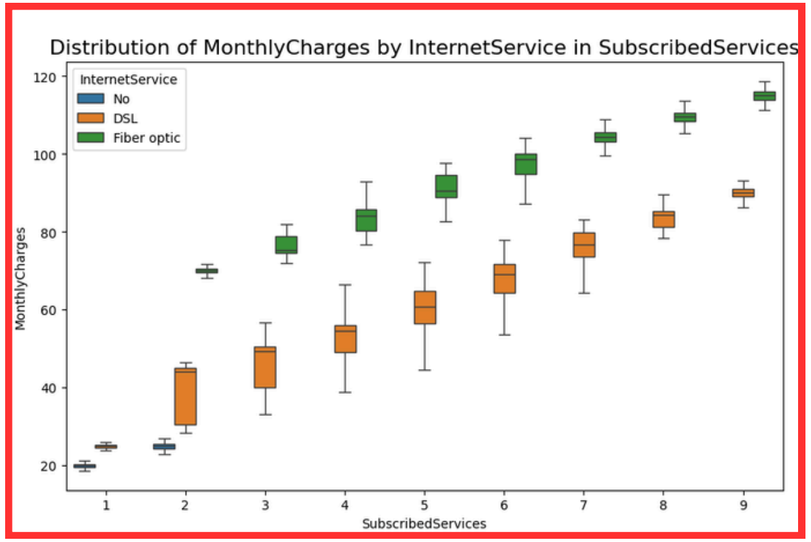
4. What services do customer choose based on the length of their tenure?

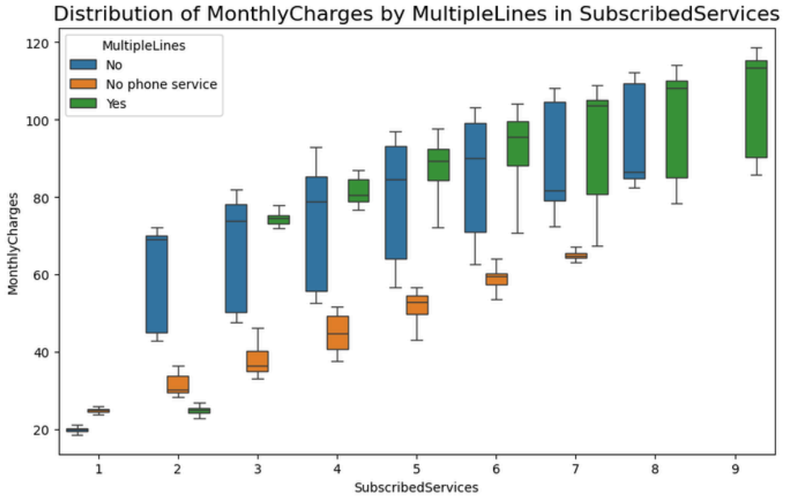
1.What service price affects customer's churn rate?

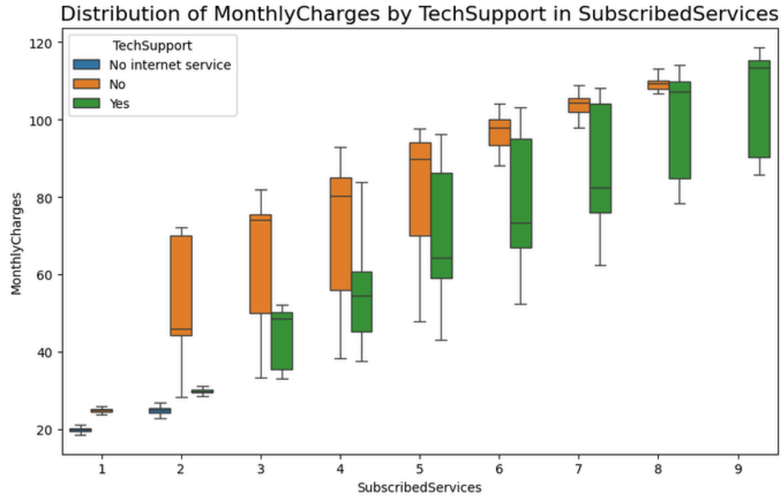


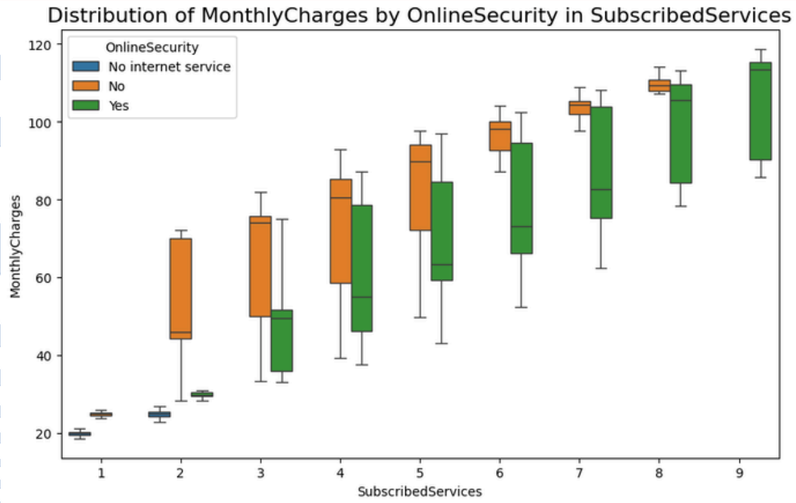
Internet Service ranks the top of the most expensive average monthly charge

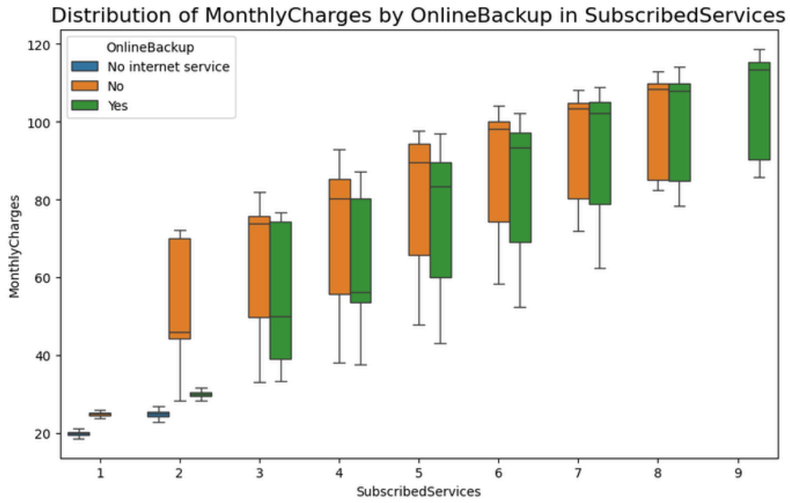
1.What service price affects customer's churn rate?

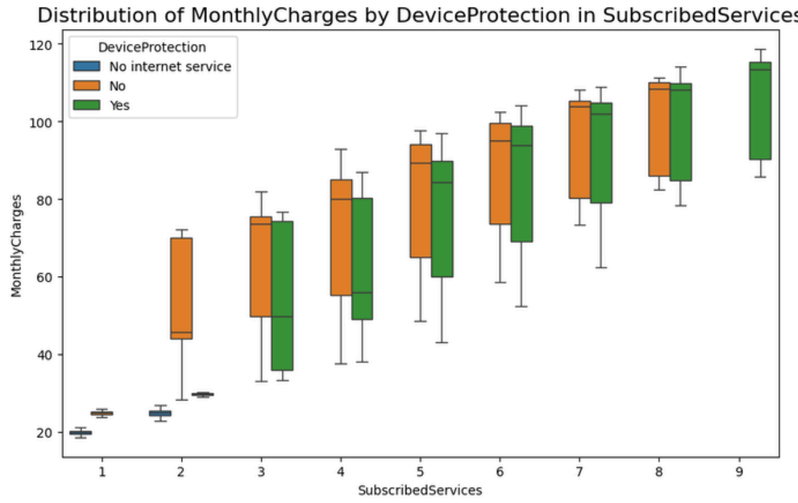




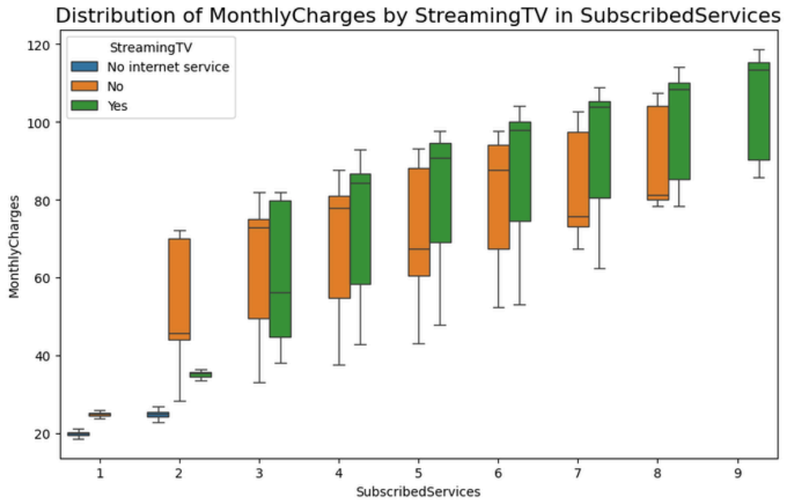


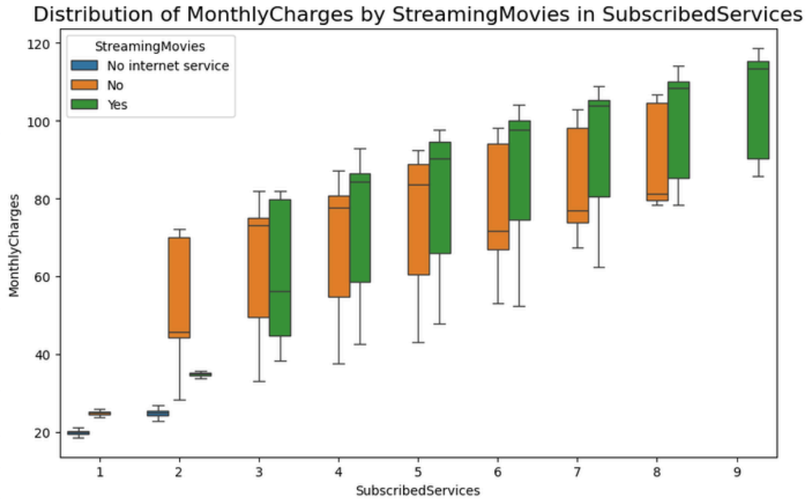






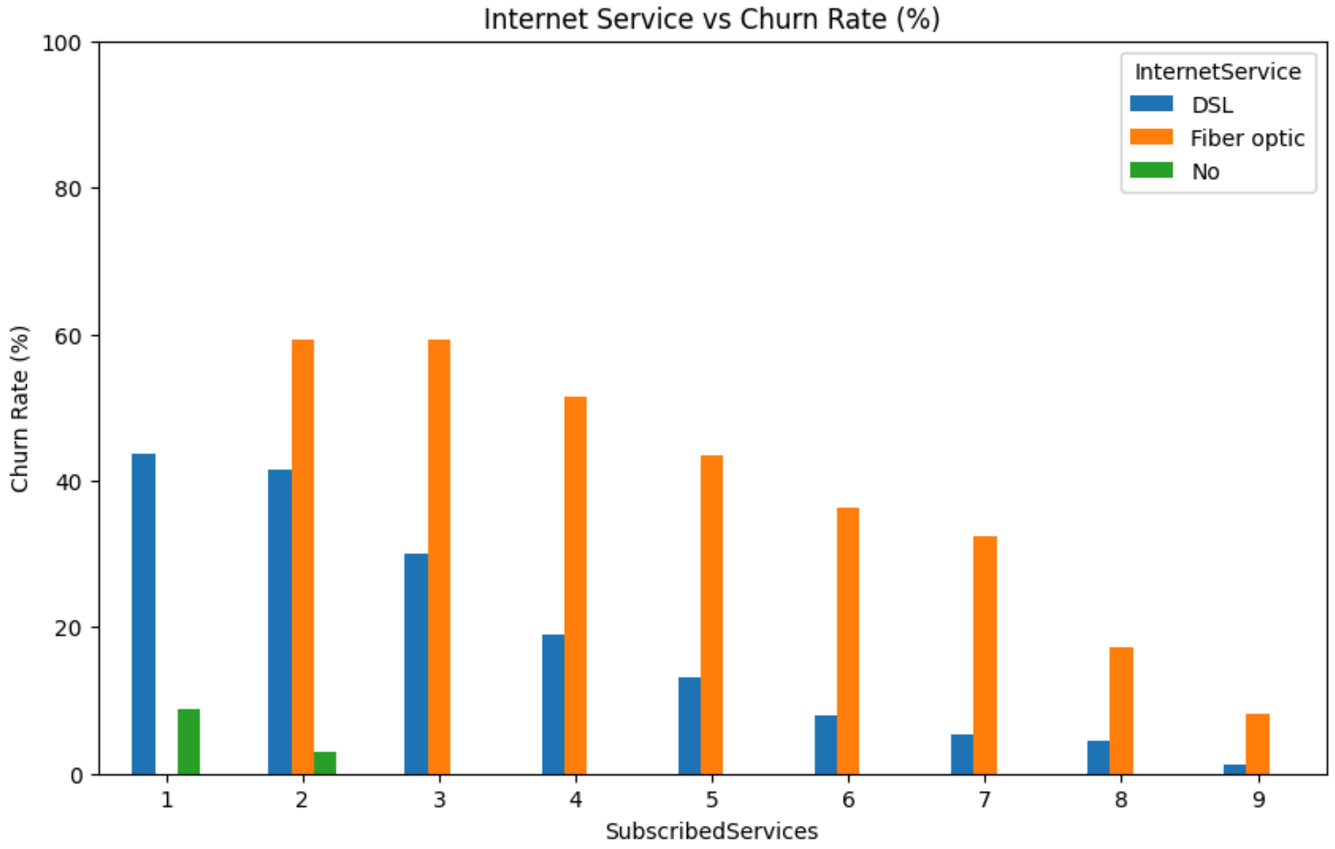
1.What service price affects customer's churn rate?





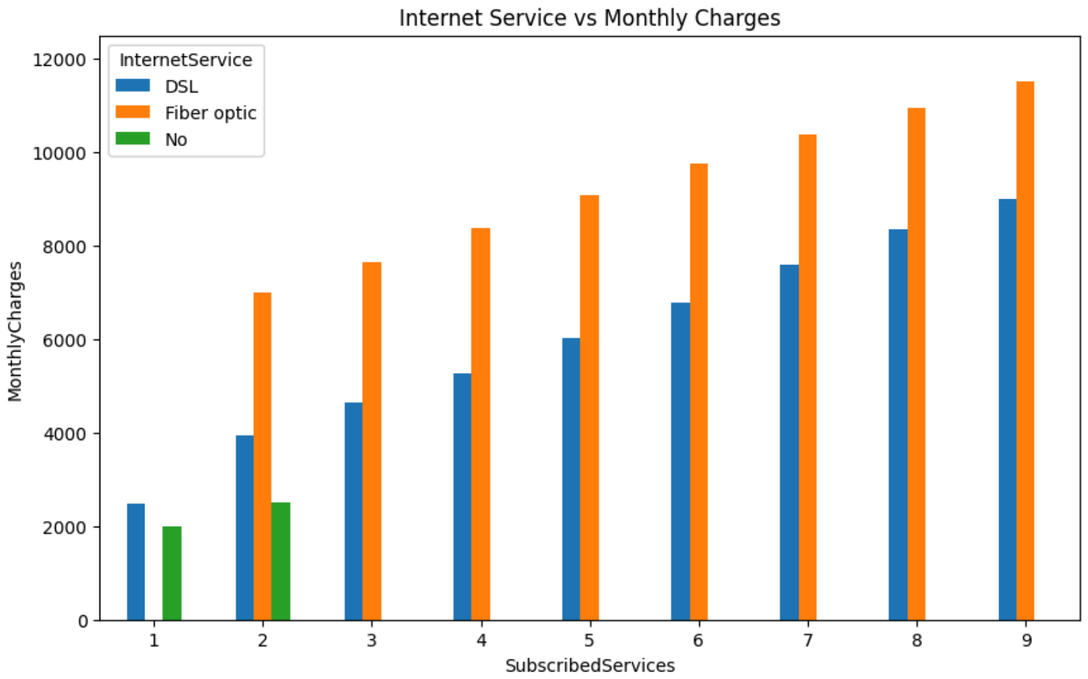
- Customers are willing to pay more the longer their tenure is.
- **TechSupport**, **OnlineBackup**, **OnlineSecurity**, **DeviceProtection**, **StreamingTV**, and **StreamingMovies** tend to have similar distribution in price, even though they correlate highly to churn rate, they do not specifically impact churn rates much.
- Distribution on InternetService **Fiber Optic is higher significantly** compared to **DSL**.
- **InternetService Fiber Optic** may be the reason for customer churn, and other services do not.

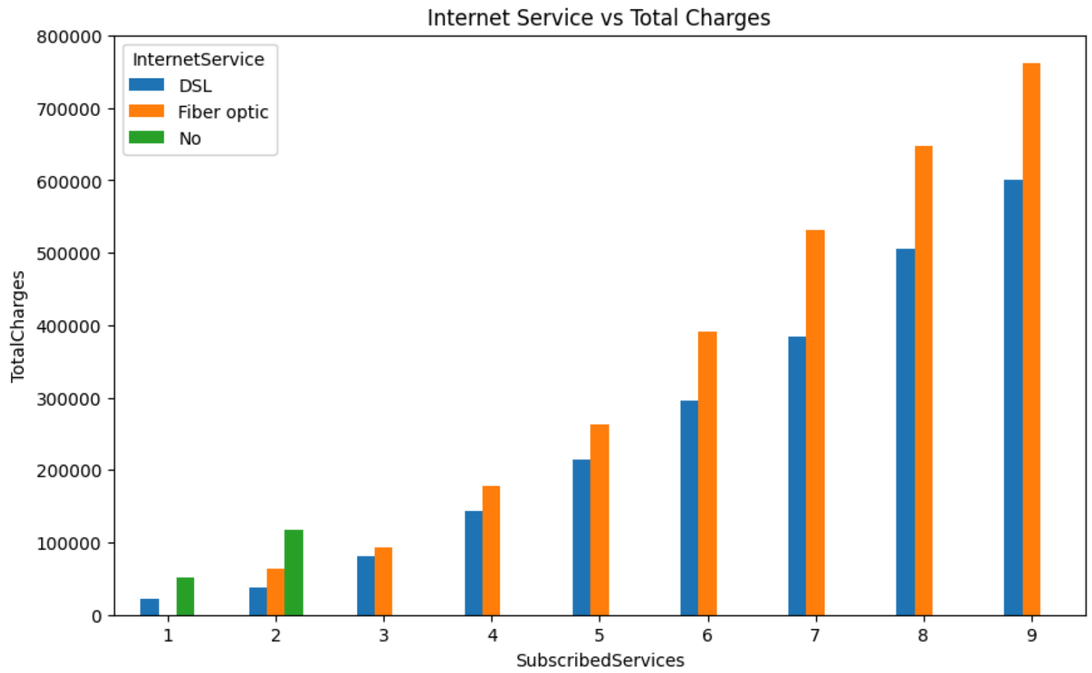
2.Is the price of InternetService fiber optic the reason for the high churn rate?



The difference on **Churn rate** is significant with approximately **~15-20%** between DSL & Fiber Optic. Even though as customers subscribe to more services, the churn rate drops, but there is still a significant difference in churn rate.

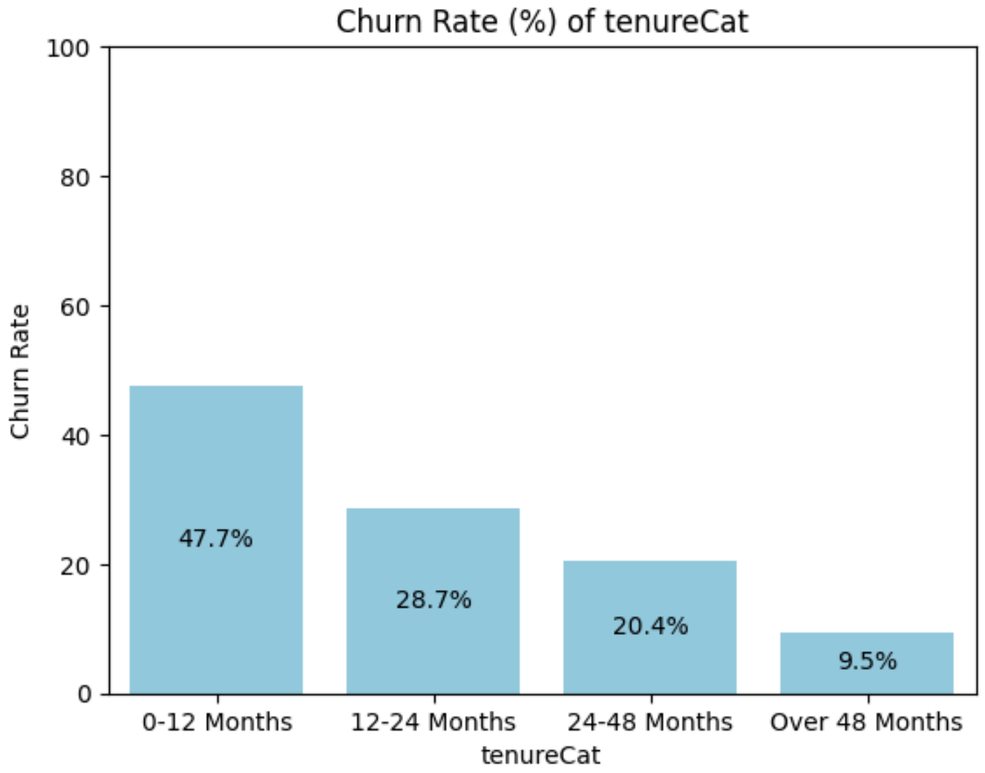
2.Is the price of InternetService fiber optic the reason for the high churn rate?





Based on MonthlyCharges and TotalCharges, even when adding more subscribed services, the price difference is still at a significance. Therefore: **Internet Service Fiber Optic prices is the reason of high churn rate.**

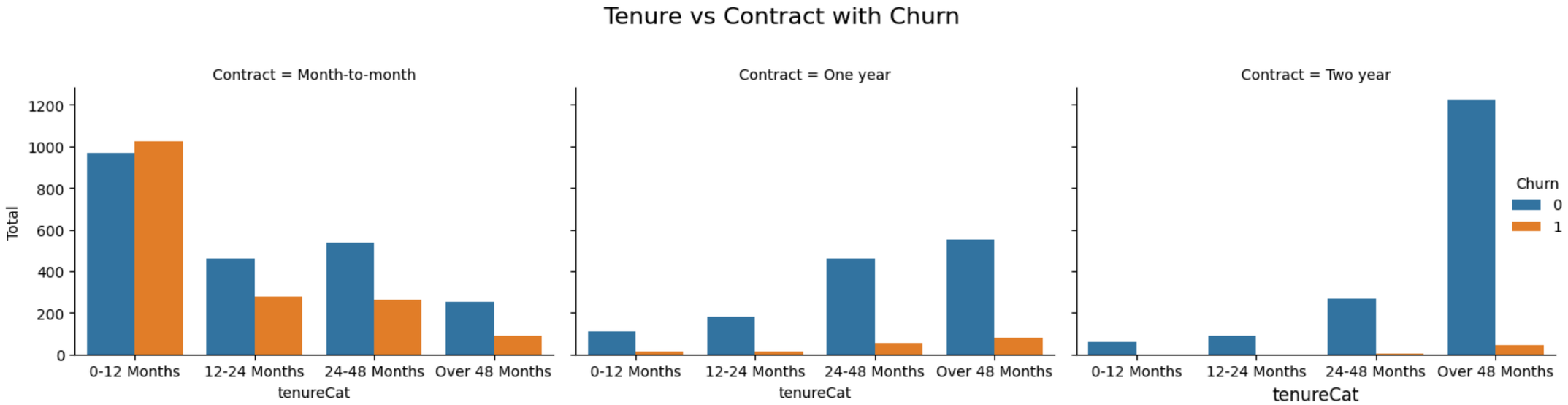
3. Do customers with higher tenure tend to choose longer-term Contracts, and how does this impact their churn rate?



We first split tenure into categories according to contract:

1. tenure < 13 = **0 - 12 Months**
2. tenure < 25 = **12 - 24 Months**
3. tenure < 49 = **24 - 48 Months**
4. tenure >= 49 = **Over 48 Months**

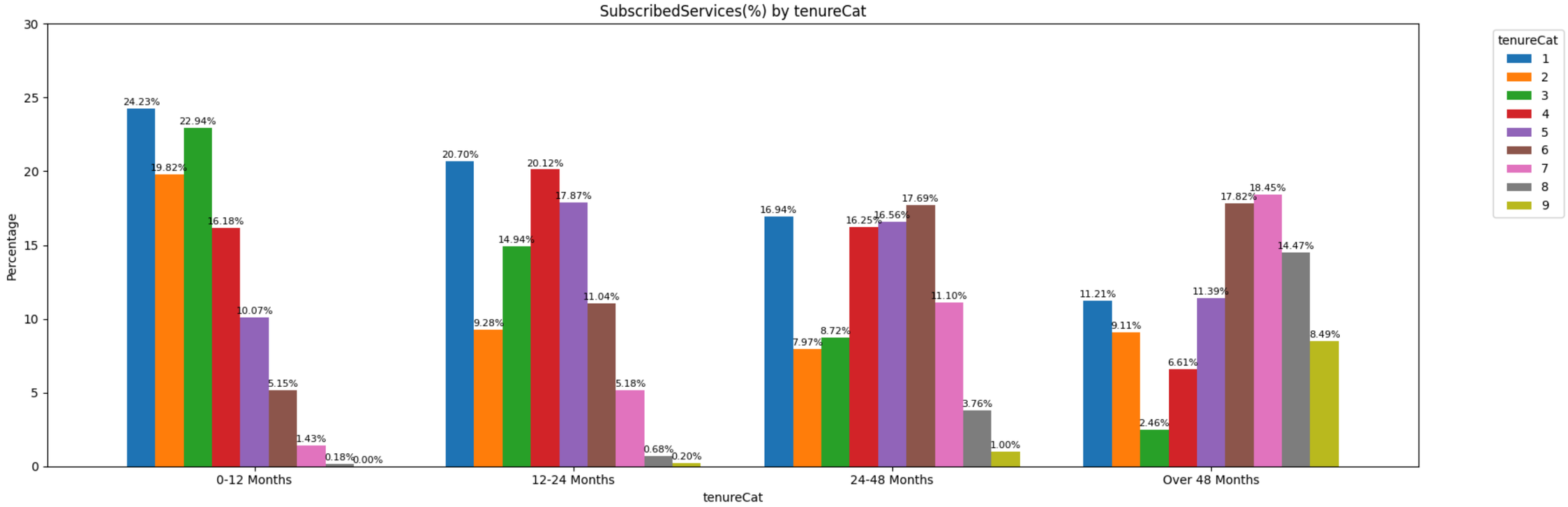
3. Do customers with higher tenure tend to choose longer-term Contracts, and how does this impact their churn rate?



- The **more tenure** a customer has, they tend to choose a longer-term contract and **less likely** to churn.

- Customers with a tenure of <= 12 months (a year) tend to take a month-to-month Contract. This make sense due to previous findings that customers tend to churn significantly when they take a month-to-month Contract.

4. What services do customer choose based on the length of their tenure?



How many services do customers subscribe to based on tenure?

- 0-12 Months: 1 service
- 12-24 Months: 1-2 services
- 24-48 Months: 1 service or 4-6 services
- Over 48 Months: 6-8 services

4. What services do customer choose based on the length of their tenure?

What do customers tend to subscribe to based on tenure excluding PhoneService?

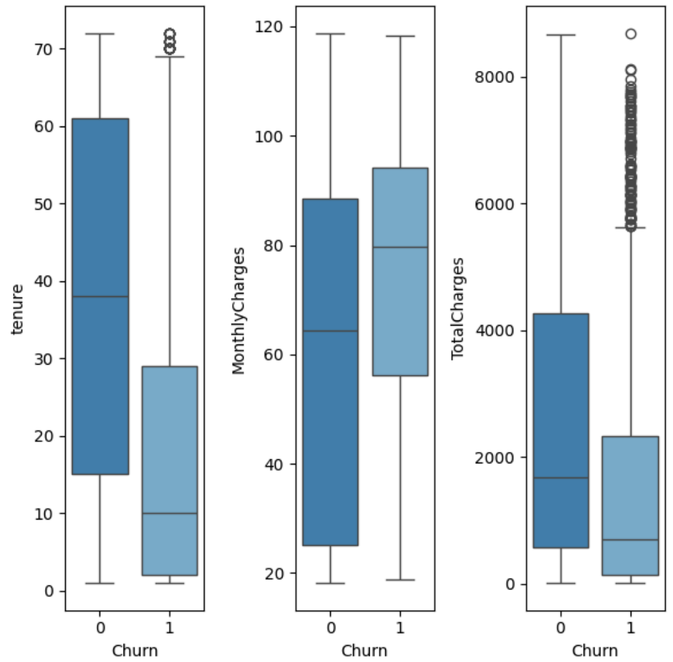
- 0-12 Months: Fiber Optic
- 12-24 Months: Fiber Optic, Multiple Lines
- 24-48 Months: Fiber Optic, Multiple Lines, StreamingTV, StreamingMovies, Device Protection, OnlineBackup
- Over 48 Months: Fiber Optic, Multiple Lines, StreamingTV, StreamingMovies, Device Protection, OnlineBackup, OnlineSecurity, TechSupport

Data Preprocessing

Data Preprocessing

Outlier Handling

Outliers will be kept for further analysis because they provide data where they include high paying customers which include InternetService Fiber Optic, OnlineSecurity, etc, and they tend to stick around for a long tenure period.




```
df_model.drop(
    ['customerID', 'gender', 'PhoneService'],
    axis=1,
    inplace=True
    )
```

Feature Selection

Based on Exploratory Data Analysis previously, we know that **gender** and **PhoneService do not influence Churn** rate. Therefore we drop the two columns so that it doesn't influence our model. We also **remove customerID** since there's too much unique data and it doesn't effect Churn and the model might be overfitting.

After selection, we drop duplicates.

Data Preparation

Label Encoding

We convert categorical data with numerical binary input 0 and 1

```
# Replace 'No' with 0 and 'Yes' with 1 in the 'Dependents' column
df_model['Dependents'] = df_model['Dependents'].replace(['No' ,'Yes'],[0,1])

# Replace 'No' with 0, 'Yes' with 1, and 'No phone service' with 0 in the 'MultipleLines' column
df_model['MultipleLines'] = df_model['MultipleLines'].replace(['No' ,'Yes', 'No phone service'],[0, 1, 0])

# Replace 'No' with 0, 'Fiber optic' with 1, and 'DSL' with 0 in the 'InternetService' column
df_model['InternetService'] = df_model['InternetService'].replace({'No': 0, 'Fiber optic': 2, 'DSL': 1})
```

```
Contract_dict = {
    'Month-to-month' : 0,
    'One year' : 1,
    'Two year' : 2}

df_model['Contract'] = df_model['Contract'].map(Contract_dict)
```

Ordinal Encoding

We convert contract info with hierarchical numerical values

One-Hot Encoding

We convert non hierarchical values in PaymentMethod to its own binary where the category value is 1 and others are 0s

```
df_model = pd.get_dummies(
    df_model,
    columns=['PaymentMethod']
    ).astype(int)
```

Data Sampling

Value Count

We can see that customers who that churn has a lower overall percentage against customers who don’t churn. Here we want to reduce class imbalance and improve model performance by balancing classes.

```
df_model['Churn'].value_counts()
```

✓ 0.0s

```
Churn
0    5069
1    1755
Name: count, dtype: int64
```

```
df_model['Churn'].value_counts(normalize=True)
```

✓ 0.0s

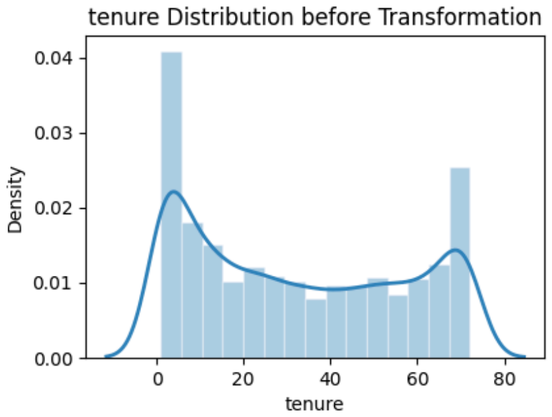
```
Churn
0    0.742819
1    0.257181
Name: proportion, dtype: float64
```

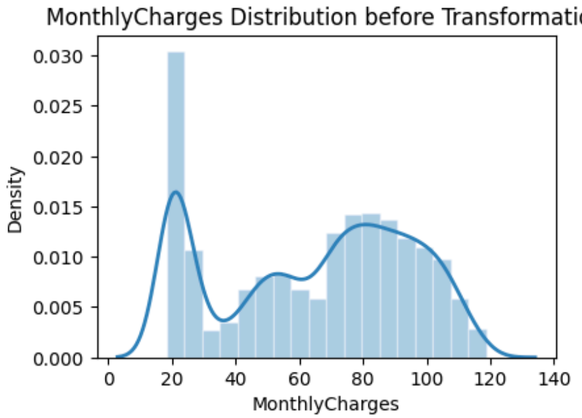
```
X_over_smote, y_over_smote = SMOTE().fit_resample(X, y)
```

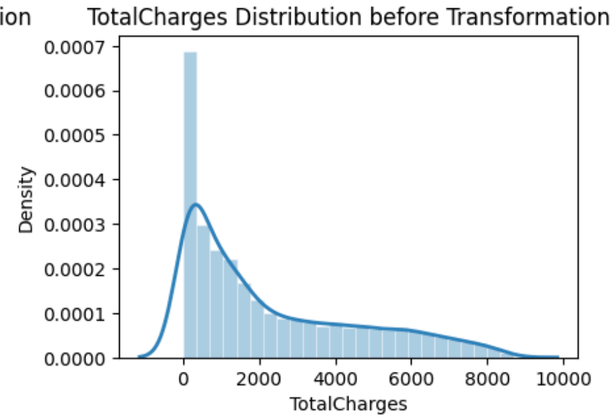
Overpopulate Data

We use SMOTE to overpopulate before splitting the data for modeling.

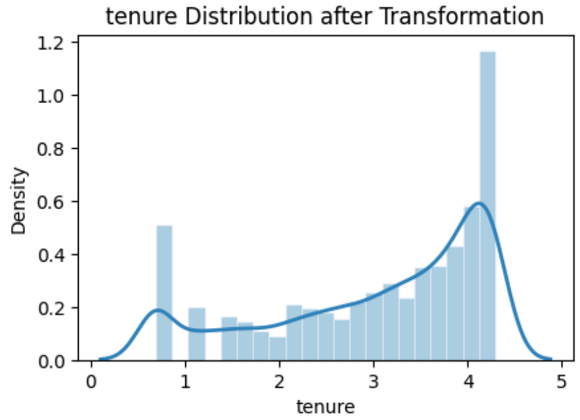
Data Preprocessing

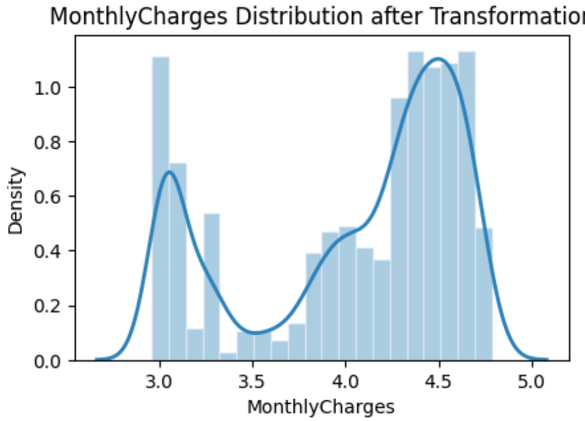


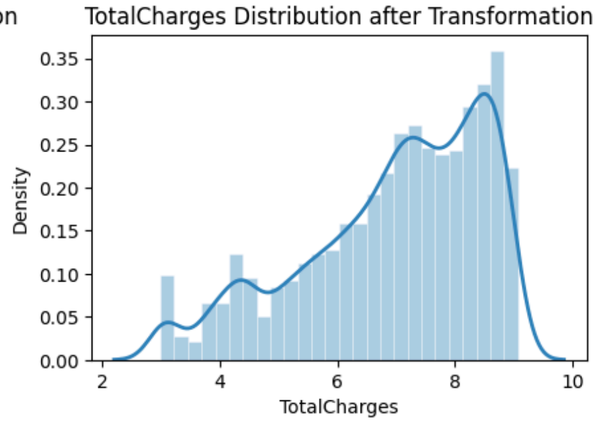




We then transform data to reduce skewness and scale using StandardScaler for our model to understand the limit of the data.



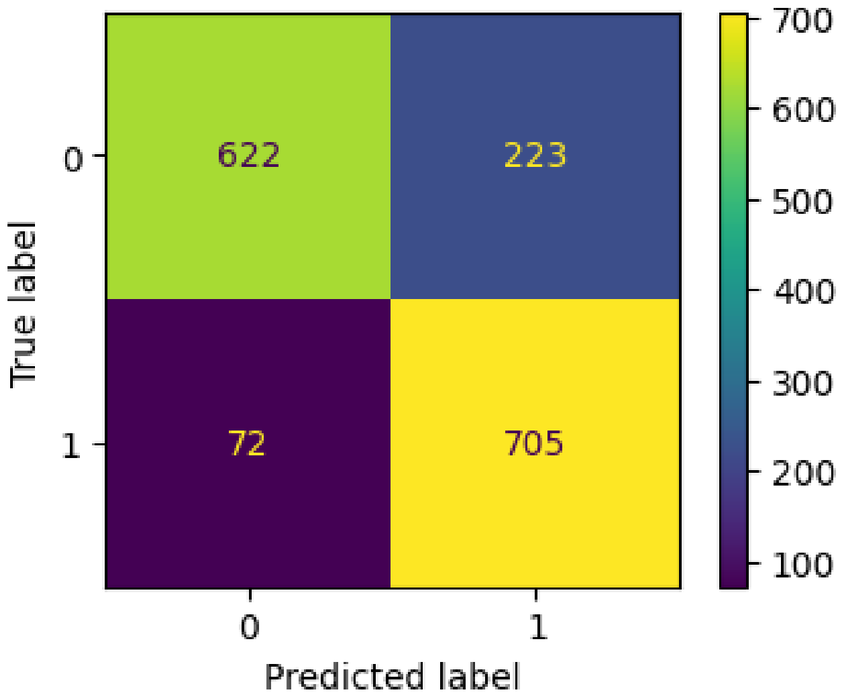




Churn Prediction Modeling

Model we chose for this dataset

```
ACCURACY: 0.8181        PRECISION: 0.7597        RECALL: 0.9073  F1 SCORE: 0.8270        ROC SCORE: 0.8922
```

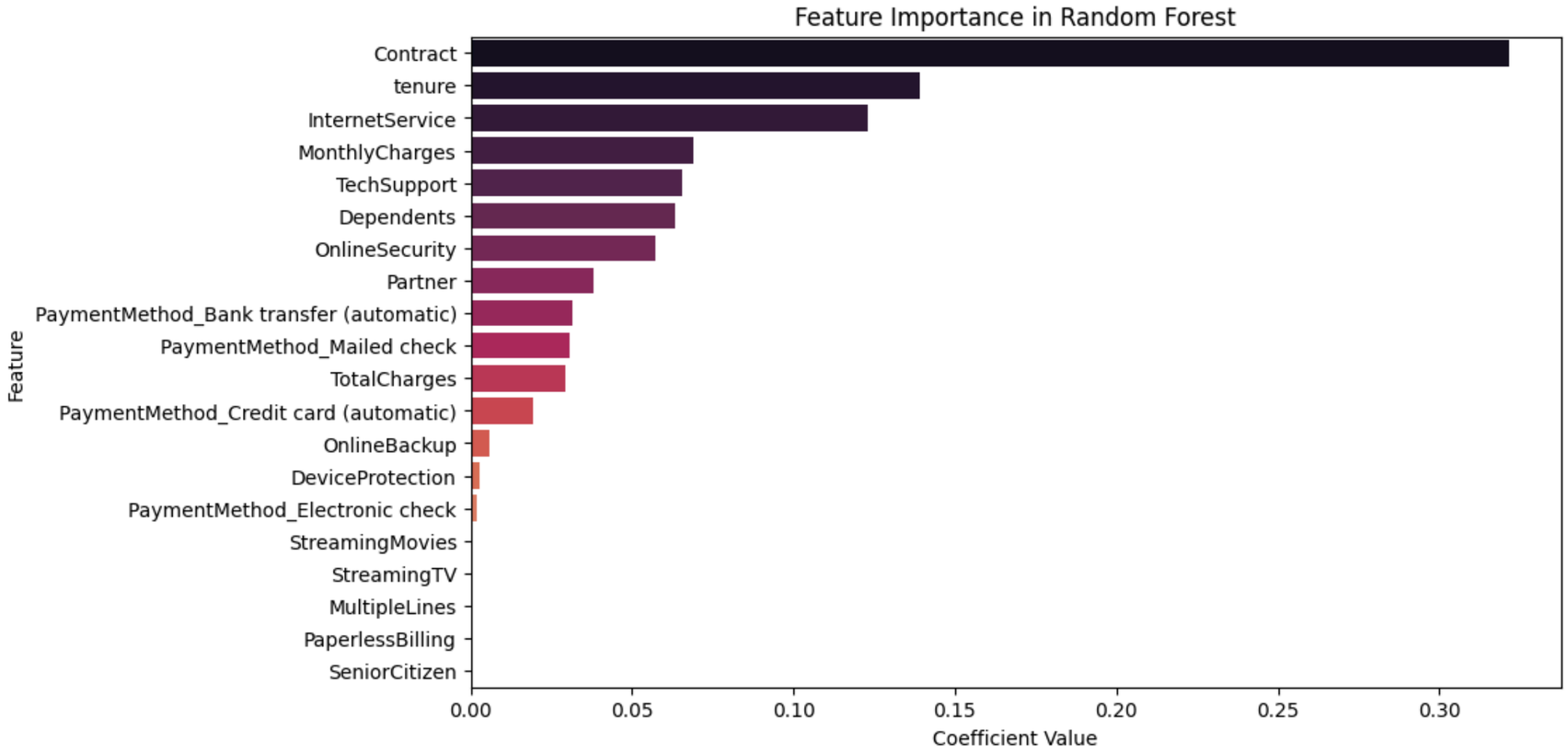


Random Forest with oversampling using the SMOTE method gave us consistent results on training data.

In addition, we tuned hyperparameters to enhance recall further by a few percent resulting in test data that resulted a **recall score of 90.73%** and **ROC score of 89%.**

Feature Importance

According to the feature importance analysis of the model we used, **Contract**, **tenure**, and **InternetService** are key factors influencing customer churn rate.



Recommendations

Recommendations

- **Reevaluate Fiber Optic pricing**; e.g. connection speed, performance rating, after-service, total cost of ownership, future-proofing, etc.

- **Provide package bundling** especially for Fiber Optic subscribers with fast Tech Support and/or Online Security if customers subscribe for a year or over.

- If customers tend to subscribe to month-to-month contracts, contact them after 2 weeks as a **testimony**. If their testimony is positive, upsell products or packages with attractive bonuses (e.g., sign a 1-year contract and get 1 month free), and emphasize the advantages of long-term subscriptions, such as cost savings and added services. If their testimony is negative, provide immediate fixes or compensation and follow up after resolving the issue to ensure satisfaction.

Thank You!