Telecom Churn Analysis
Case Study

Introduction

- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate.
- There are two main models of payment in the telecom industry - postpaid (customers pay a monthly/annual bill after using the services) and prepaid (customers pay/recharge with a certain amount in advance and then use the services).
- Thus, churn prediction is usually more critical (and non-trivial) for prepaid customers, and the term ‘churn’ should be defined carefully.

Objectives

- In this presentation, we will walk through the process of analyzing telecom churn data. Churn analysis is crucial for businesses to identify and retain high-value customers. We'll cover the steps from data understanding to preparation, exploration, and modeling.
- To predict Customer Churn.
- Highlighting the main variables/factors influencing Customer Churn.
- Finding out the best model for our business case and providing executive summary.

Dataset Overview

Load and Explore the Dataset

- Load the telecom churn dataset.
- Check initial and last rows of the data.
- Examine the columns, data types, and basic statistics.
- Check for missing values.

Data Cleaning and Imputing Missing Values

- Identify columns with more than 70% missing values.
- Impute missing values in recharge-related columns.
- Drop unnecessary columns.
- Impute or drop remaining missing values.

Data Preparation

Feature Engineering

- Create new variables like 'tenure' and derive insights.
- Visualize the distribution of the 'tenure' variable.

Exploratory Data Analysis (EDA)

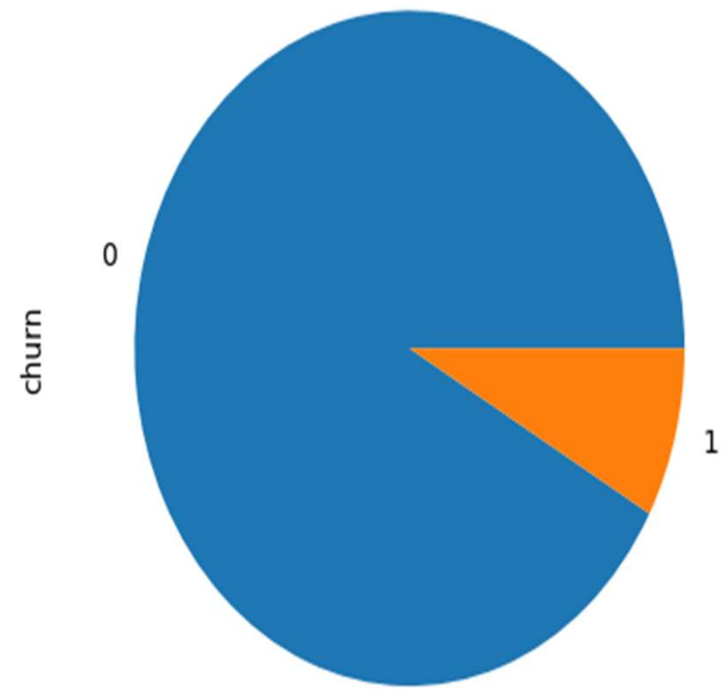
- Explore the relationship between tenure and churn.
- Examine the correlation between variables using a heatmap.

Feature Selection

- Drop highly correlated variables.
- Visualize the correlation of features with the target variable (churn).

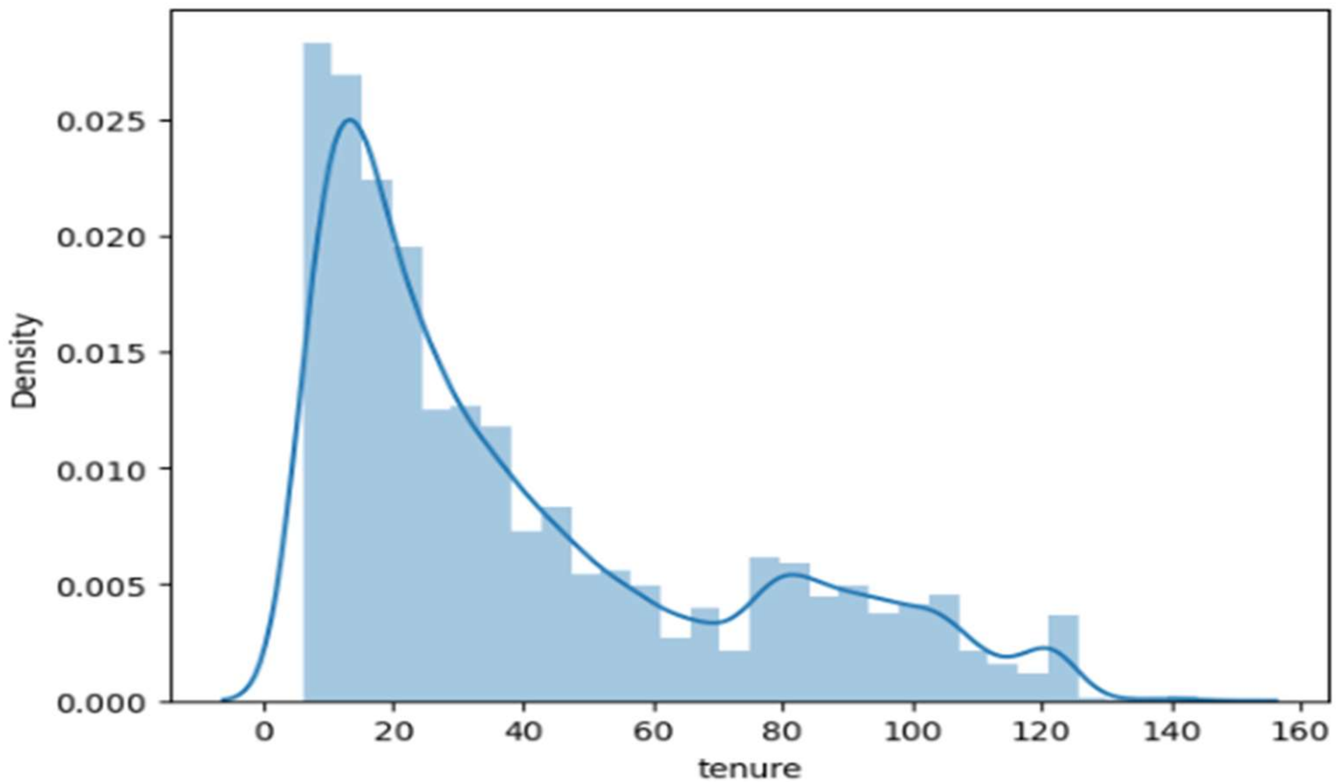
Churn/Non-Churn Percentage:

As we can see that 97% of the customers do not churn, there is a possibility of class imbalance

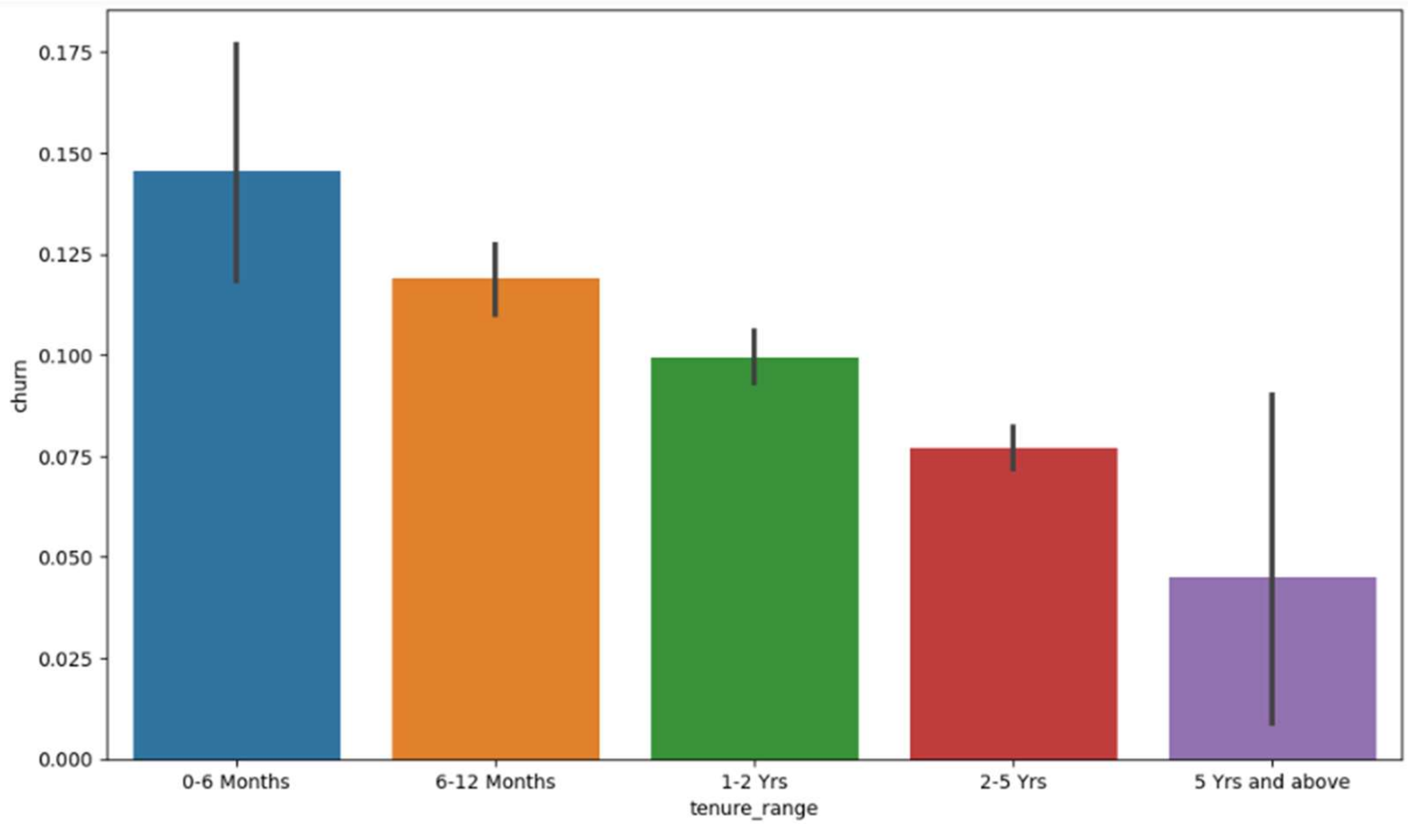


```
# Checking the distribution of he tenure variable

sns.distplot(telecom_df_filtered['tenure'],bins=30)
plt.show()
```



Plotting a bar plot for tenure range:



Conclusion

- The presentation covers the entire process from understanding and cleaning the data to preparing it for modeling.
- Feature engineering, EDA, and feature selection are essential steps in building an effective predictive model.
- The dataset is now ready for training machine learning models to predict customer churn.