Predicting Customer Churn: Unlocking Insights for Retention at SyriaTel

Author: Stanley Macharia

Introduction

Author: Stanley Macharia

▶ Customer churn is a key problem for telecommunications firms since losing customers has a direct influence on revenues and growth. This project seeks to create a machine learning classifier that can predict if a client will churn (leave SyriaTel). By studying customer usage habits, plan subscriptions, and contacts with the organization, we identify significant churn contributors and give practical ideas for reducing it.

Project Goals

- Predict customer churn.
- Compare different machine-learning models.
- Provide Business Insights and Recommendations.
- Enhance Business Decision-Making.
- Increase Revenue and Growth by reducing customer churn.

Author: Stanley Macharia

Data Understanding

- The data has 3,333 entries, each representing a consumer. The idea is to evaluate these records for trends that suggest client turnover. Here's a full analysis of the data:

- **Target Variable**:

- Churn: A binary variable that indicates if a client has churned. True: Customer churn occurred. False: The customer was kept.

- Features: The number of features is 21 columns, including the target variable (churn).

- **Data Types**:

- *Categorical*: state, phone number, international plan, voice mail plan, churn.

- ***Numerical***:

- *Integer*: account length, area code, amount of vmail messages, total day calls, total evening calls, total night calls, total international calls, and customer service calls.

- *Float*: total day minutes, total day charge, total evening minutes, total evening charge, total night minutes, total night charge, total international minutes, total international charge.
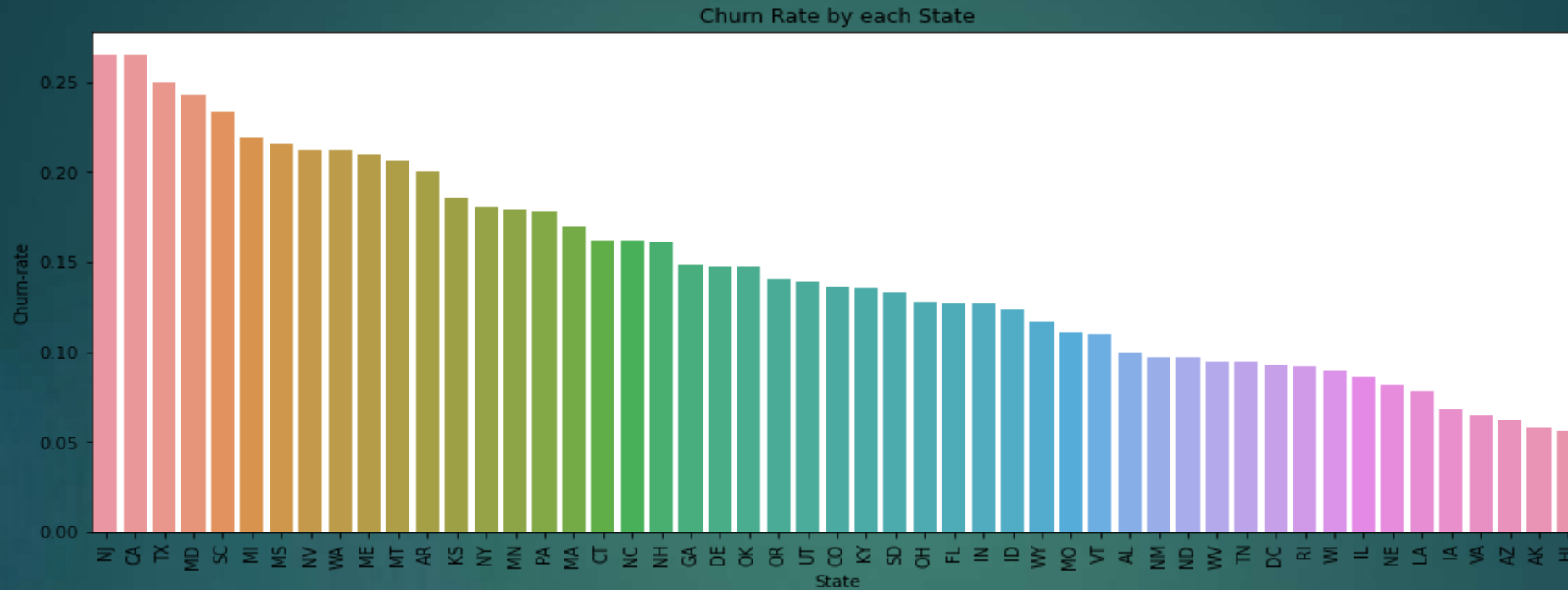
Author: Stanley Macharia

Techniques and methods used in the project

Author: Stanley Macharia

▶ **Data Acquisition and preprocessing**: This involves identifying the data and then cleaning, dealing with missing and duplicate data preparing it for analysis.

▶ **Exploratory Data Analysis**: Examining data trends and patterns using summary statistics and infographics to comprehend consumer behavior.

▶ **Feature Engineering & Selection:** Determining the key elements that affect retention of customers and choosing the most suitable ones to model.

Cont…

Author: Stanley Macharia

- **Machine Learning Models**: Using various models, such as Decision Trees, K-Nearest Neighbors (KNN), and Logistic Regression, to forecast customer churn.

- **Business Recommendations**: Analyzing data and offering practical insights to assist optimize company plans and lower customer churn.

Findings:

Author: Stanley Macharia

Churn Rate by each State

The states on the left side of the graph notice the greatest churn rates. These states might be locations where consumers are unsatisfied or where there is greater competition, increasing the probability that customers would leave SyriaTel. The progressive fall from greatest to lowest churn rates illustrates that turnover varies significantly between states. This reinforces the prior result that state has a major impact on turnover.
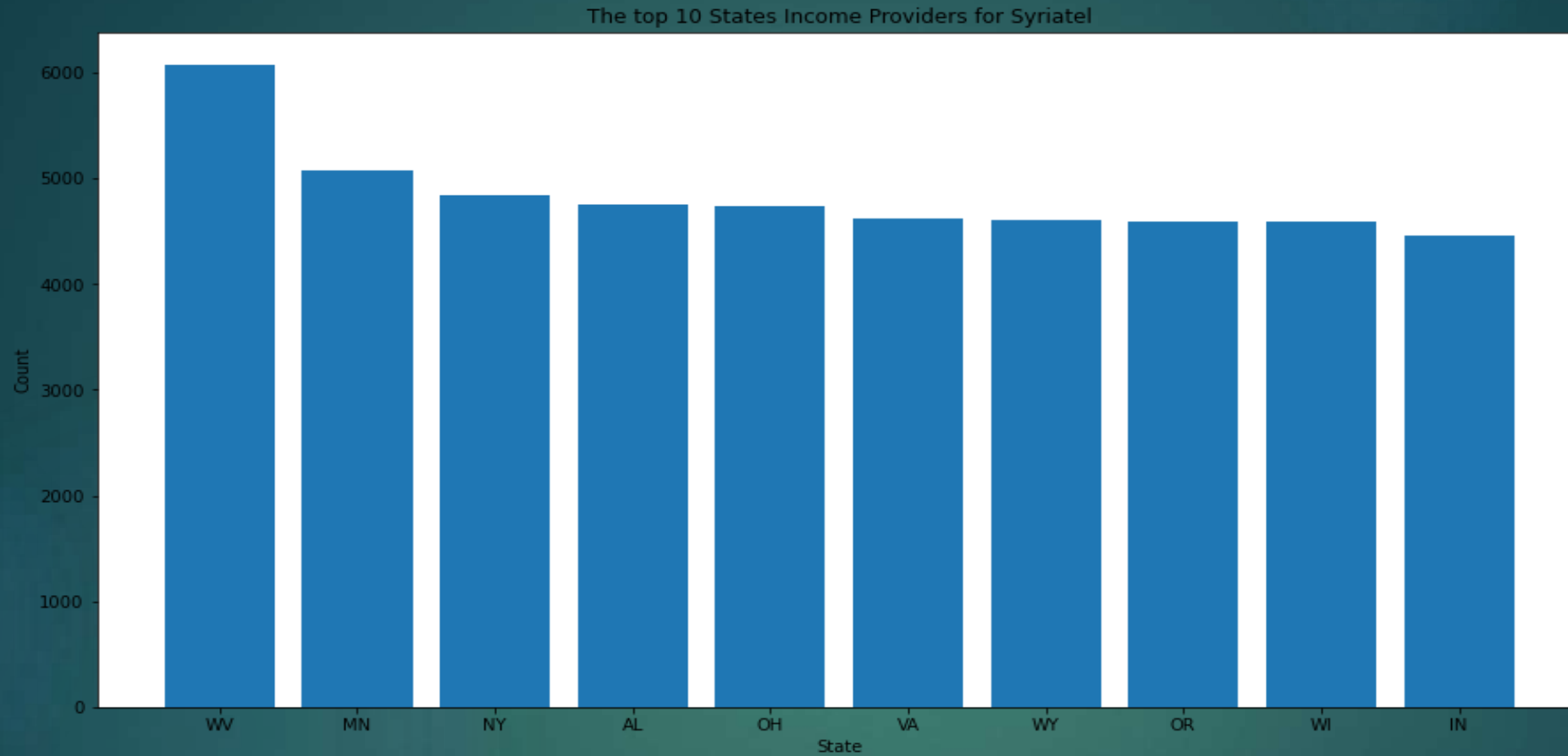
We can see that West Virginia generates the highest income for Syriatel

Author: Stanley Macharia

How Income varies on different times of the day
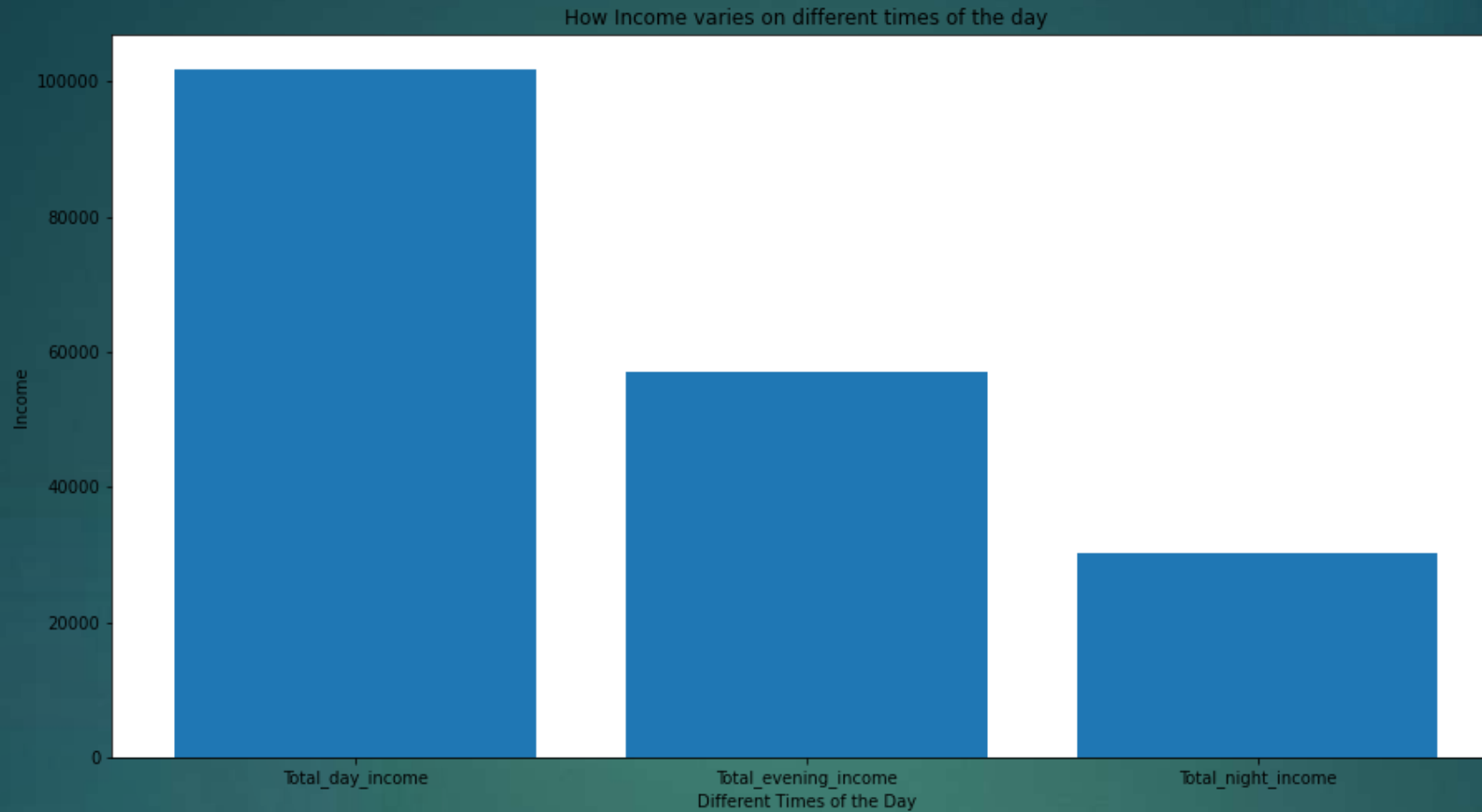
Author: Stanley Macharia

According to the bar chart above, we can see that during the day Syriatel makes the most money and during night-time, Syriatel makes less money.

Models and model selection

Author: Stanley Macharia

▶ In this project I created 4 models to determine which is better when it comes to churn prediction.

▶ Logistic Regression model

▶ Logistic Regression with SMOTE model

▶ KNN model

▶ Decision Tree model

▶ All these are used to deal with classification models where the data is labeled.

Model Selection
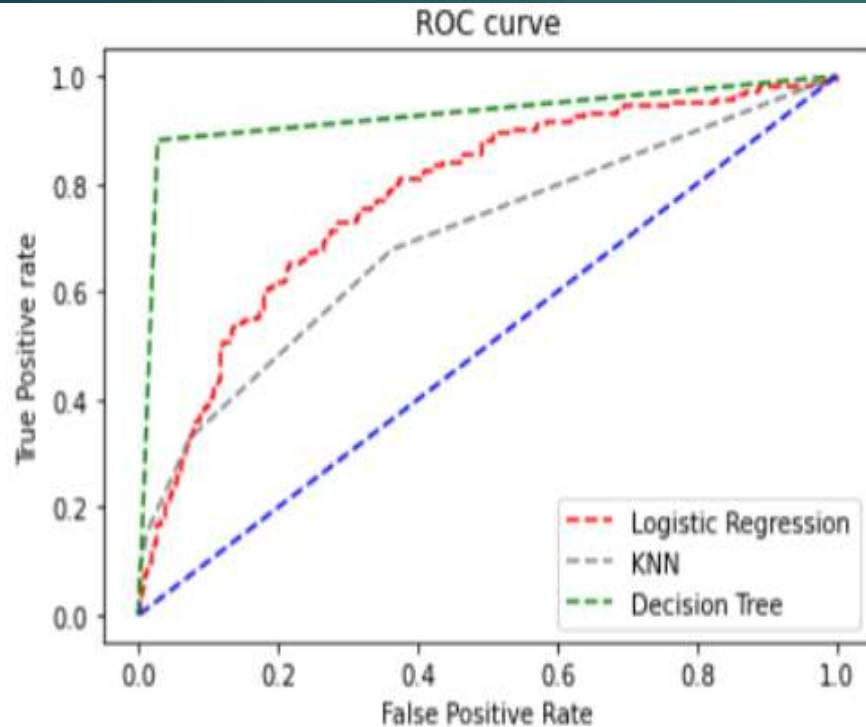
ROC (Receiver Operating Characteristic) Curve is a graphical representation used to evaluate the performance of a classification model, especially in binary classification problems.

Author Stanley Macharia

The ROC curve clearly shows that the **Decision Tree model** outperforms the others. Its green curve is the steepest and spans the biggest area under the curve (AUC), indicating that it performs the best at differentiating between positive and negative classes. Essentially, the Decision Tree model finds a good mix between properly recognizing true positives and reducing false positives, making it the best option for this dataset.

Recommendations

Author: Stanley Macharia

- **Check and Refine the Model**: To guarantee that the model reacts to changing consumer behavior, regularly check its performance over time and retrain it with fresh data. Consider including customer feedback and service interaction data to improve the model's predictive ability.

- **Investigate Ensemble approaches**: While the decision tree model worked well, more investigation into ensemble approaches such as Random Forest or Gradient Boosting might produce even better results. These strategies may increase recall and general resilience, resulting in a more complete solution to the churn prediction problem.

- **Customer Feedback Loop**: Create a feedback loop to collect data from clients who were expected to leave but opted to stay. Use this information to improve retention techniques and the prediction model.

- **Proactive Retention Strategies**: Create focused retention efforts for high-risk customers identified using the model. These might include targeted offers, better customer service, and loyalty programs.