Syriatel Customer churn Prediction

Author: Lynette Mwiti
lynette.mwiti@student.moringaschool.com

Overview

- Customer churn is a significant challenge for businesses, particularly the telecom industry.
- The company seeks to understand the factors leading to churn and develop a predictive model to identify at-risk customers.
- We're going to use the CRISP_DM method to be able to analyse and come up with predictions to reduce churn and increase customer retention.

Project Workflow

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation
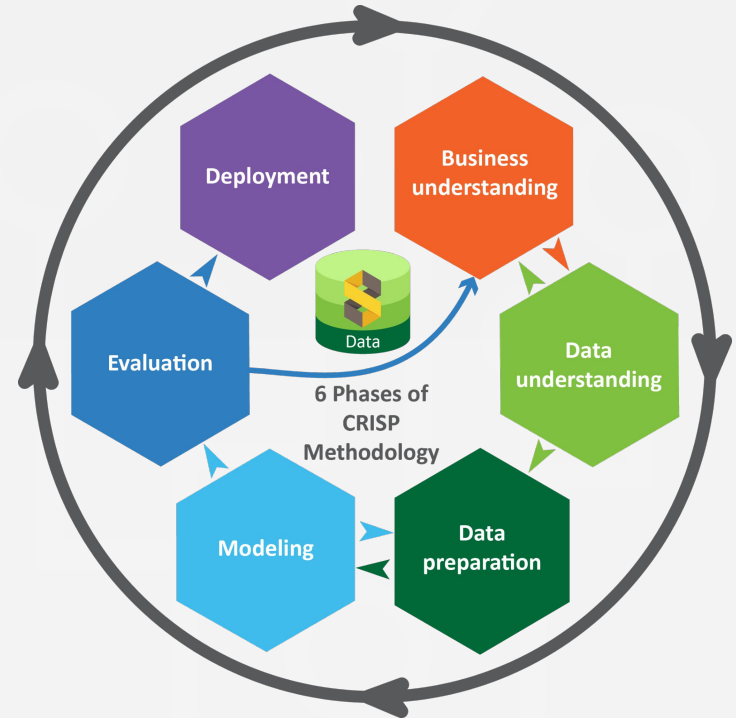
Business understanding

SyriaTel is a telecommunications company experiencing a high churn rate, meaning many customers are leaving and switching to competitors. The company seeks to understand the factors leading to churn and develop a predictive model to identify at-risk customers. The ultimate goal is to reduce churn, increase customer retention, and enhance profitability.

Specific Objectives

- Identify factors leading to customer churn.
- Develop an accurate predictive model for customer churn.
- Implement strategies to retain customers identified as at risk of churning.
- Developing a robust churn prediction model with high recall score of 0.8
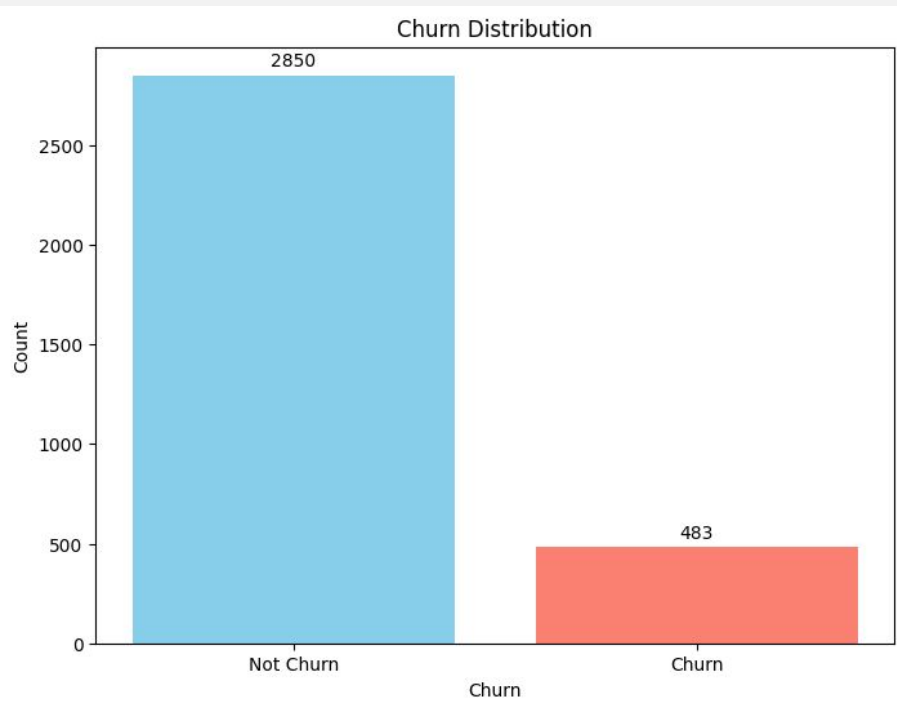
Data understanding

The data we are going to use is called **customer_churn.csv** from syriatel company found in syria , sourced from kaggle.
it has 3333 rows and 21 columns.

17 numerical features and 4 categorical features.

⊙ EDA Results

Distribution of Churn feature

Out of the 3,333 customers in the dataset, 483 have terminated their contract(churn) while 2850 have not terminated their contract (no churn) That is 14.5% of customers lost.
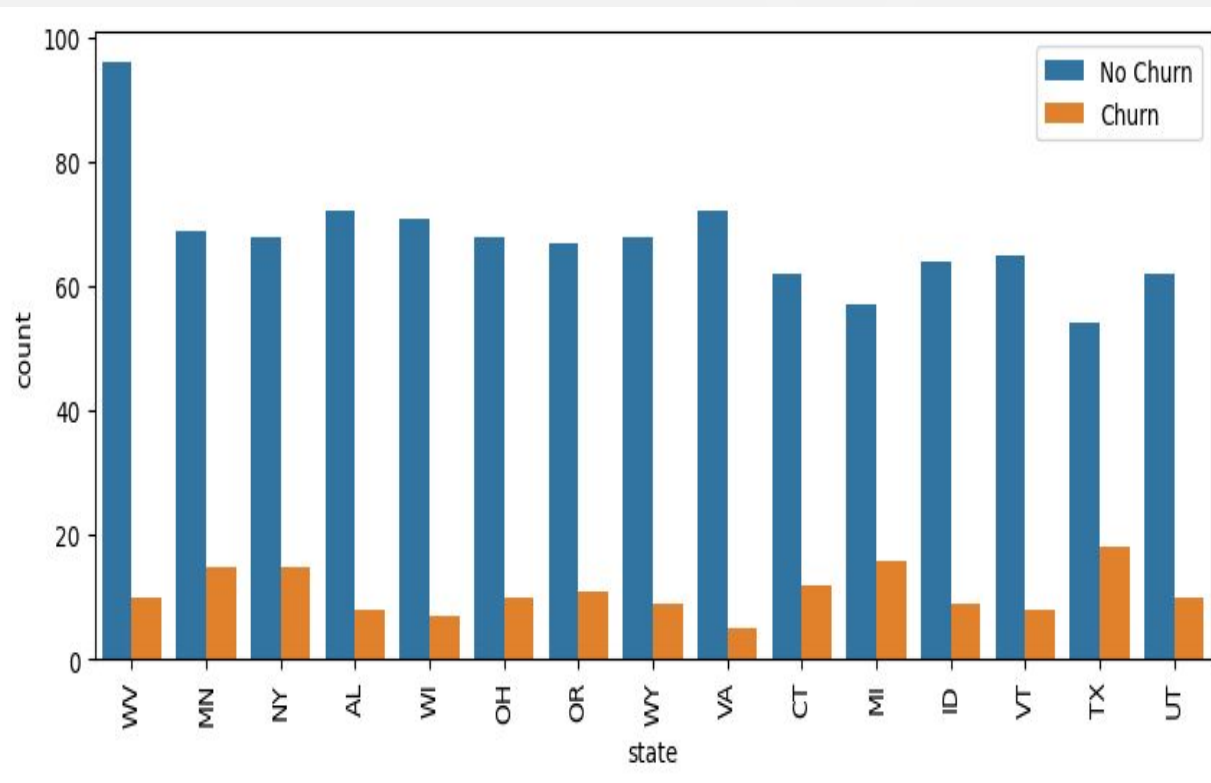
Distribution of Area code data

Majority or half (49.7%) of the customers are in area code 415. while area code 510 and 408 are 25.2% and 25.1% respectively.
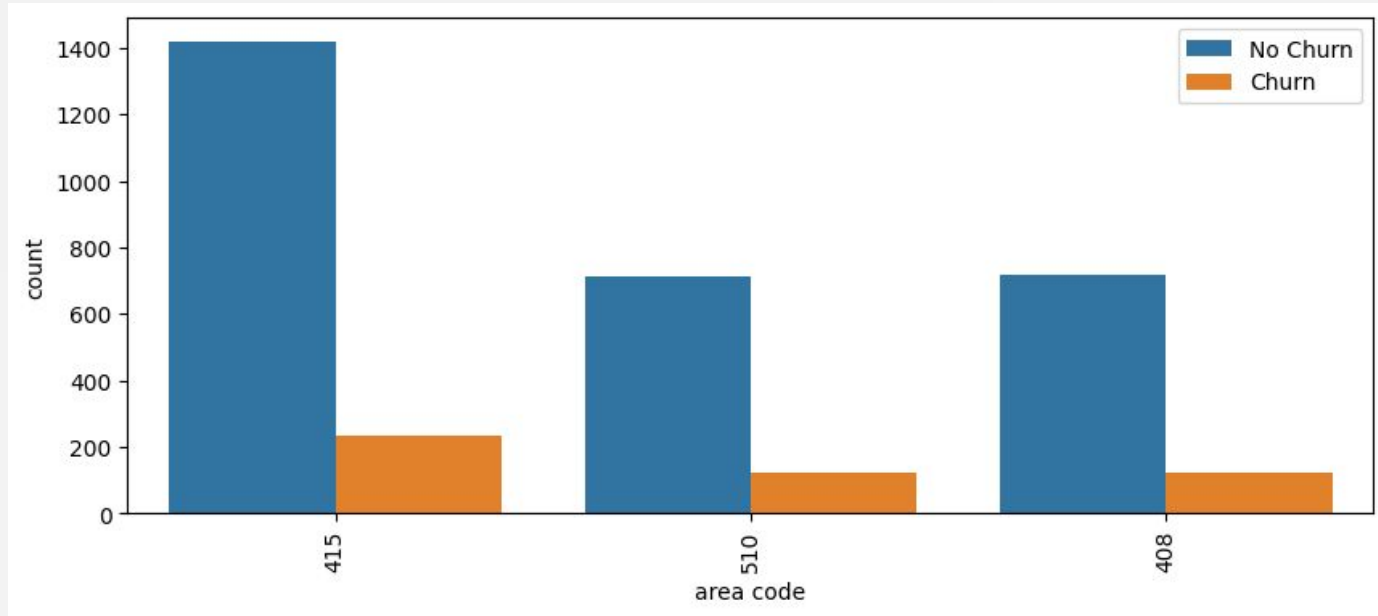

Distribution of Area Code Feature
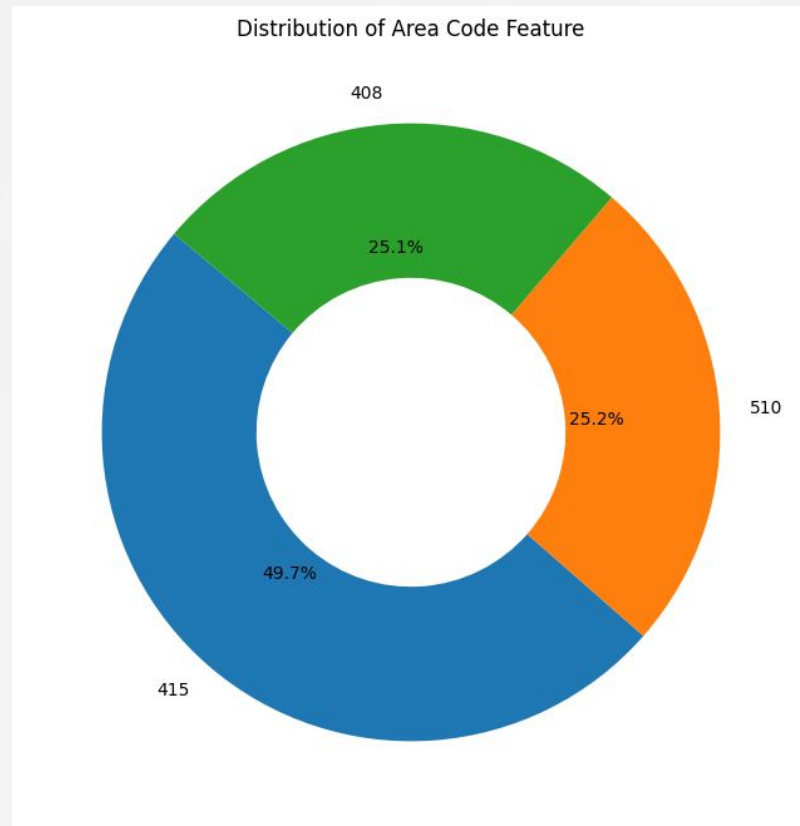
Distribution of state against churn

Customer who churned, majority are from *Texas, New Jersey, Maryland, Miami and New York*.

Distribution of area code by churn

Area code 415 and are code 510 had the highest number of customers that had churned.

Churn distribution by day charges

The plot shows the distribution of day charges against churn

We can see that the churn customers have a peak at 40 which suggests that while day charges are somewhat indicative of churn behavior, they are not the only factor.
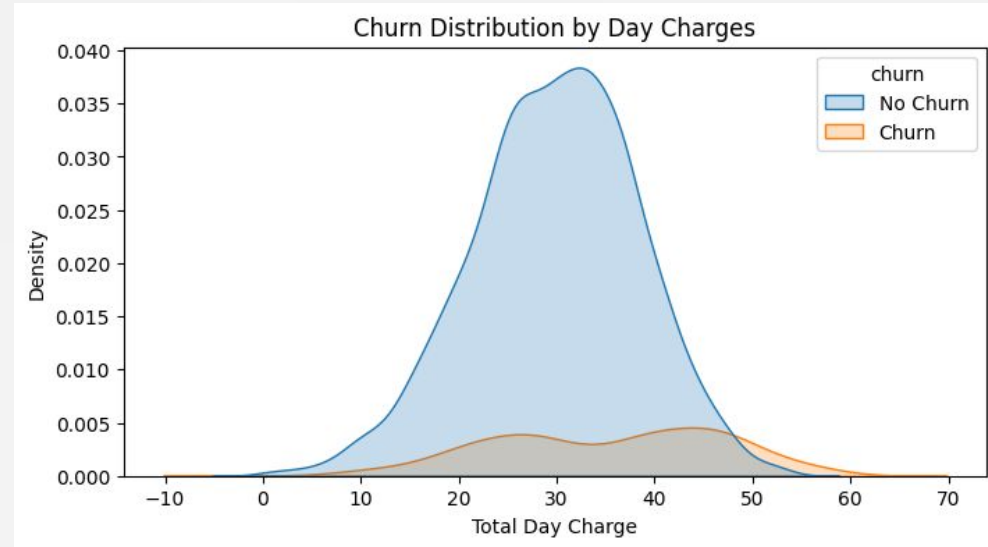
Data Modeling

we will be using the following algorithms for data modeling

- **Logistic Regression**
- **Decision Tree**
- **Random Forest**
- **XGBoost**

We will also proceed to use the **ROC_AUC** metric to evaluate the performance of our models.

As for dealing with class imbalance, we will use **SMOTE** to generate synthetic examples of the minority class in our dataset
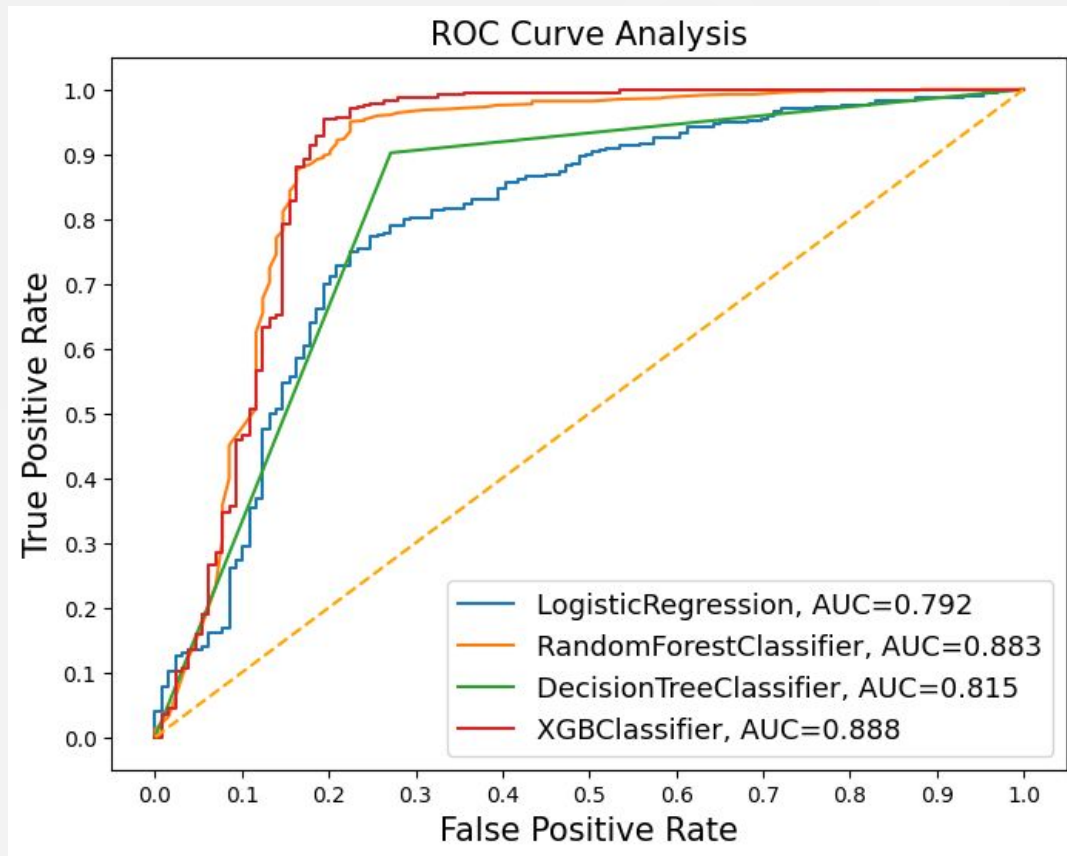
Model Evaluation

Based on recall score

- LogisticRegression   0.772590
- RandomForestClassifier 0.956325
- DecisionTreeClassifier 0.902108
- XGBClassifier   0.974398

Based on ROC_AUC Curve

- Logistic Regression 0.792
- Random Forest Classifier 0.883
- Decision Tree classifier 0.815
- XGB classifier 0.888

Conclusion

The recall score of our XGB classifier was 97%. Which was above our recalls score of 0.80 which shows it is a commendable predictive model, Our primary objective of predicting customer churn with an acceptable recall score has been achieved.

Recommendations

1. State-Specific Retention Strategies:
Focus on customer retention strategies in states with higher churn rates, such as T**exas, New Jersey, Maryland, Miami, and New York.** This can involve targeted marketing campaigns, personalized offers, or improved customer support tailored to the specific needs and preferences of customers in these states.

2. Targeted Promotions:
Offer discounts or promotional offers to customers in area codes 415 and 510, where churn rates are higher. These incentives can help retain customers in these specific regions.

Contd Recommendation

3. Pricing Evaluation:

Review the pricing structure for day, evening, night, and international charges. Consider adjusting pricing plans or introducing discounted packages to address higher charges associated with customer churn.

4. Customer Service Enhancements:

Improve the quality of customer service and reduce the number of service calls. Invest in enhanced training programs for customer service representatives to ensure prompt and effective issue resolution, leading to increased customer satisfaction and lower churn rates.

5. Voicemail Plan Enhancement:

Enhance the value proposition of the voicemail plan to increase adoption among customers. Highlight the benefits and convenience of voicemail services, and consider offering additional features or discounts to encourage customers to sign up.