Seriatel Customer Churn Prediction Project

CONTENT

01 OVERVIEW

02 BUSINESS UNDERSTANDING

03 DATA UNDERSTANDING

04 MODELING

05 EVALUATION

06 RECOMMENDATIONS

07 NEXT STEPS

OVERVIEW

- Increased competition in the dynamic telecom market has raised concerns about customer churn rate, as customers have the option to switch between different companies.
- By leveraging big data in the telecom industry, it becomes possible to predict customer churn and implement measures to address this issue.
- Predicting customer churn enables telecommunication firms to proactively meet customer needs and improve their retention rates.
- Utilizing big data analytics can help telecommunication companies stay competitive by identifying potential churners and taking proactive actions to retain customers.

BUSINESS UNDERSTANDIG

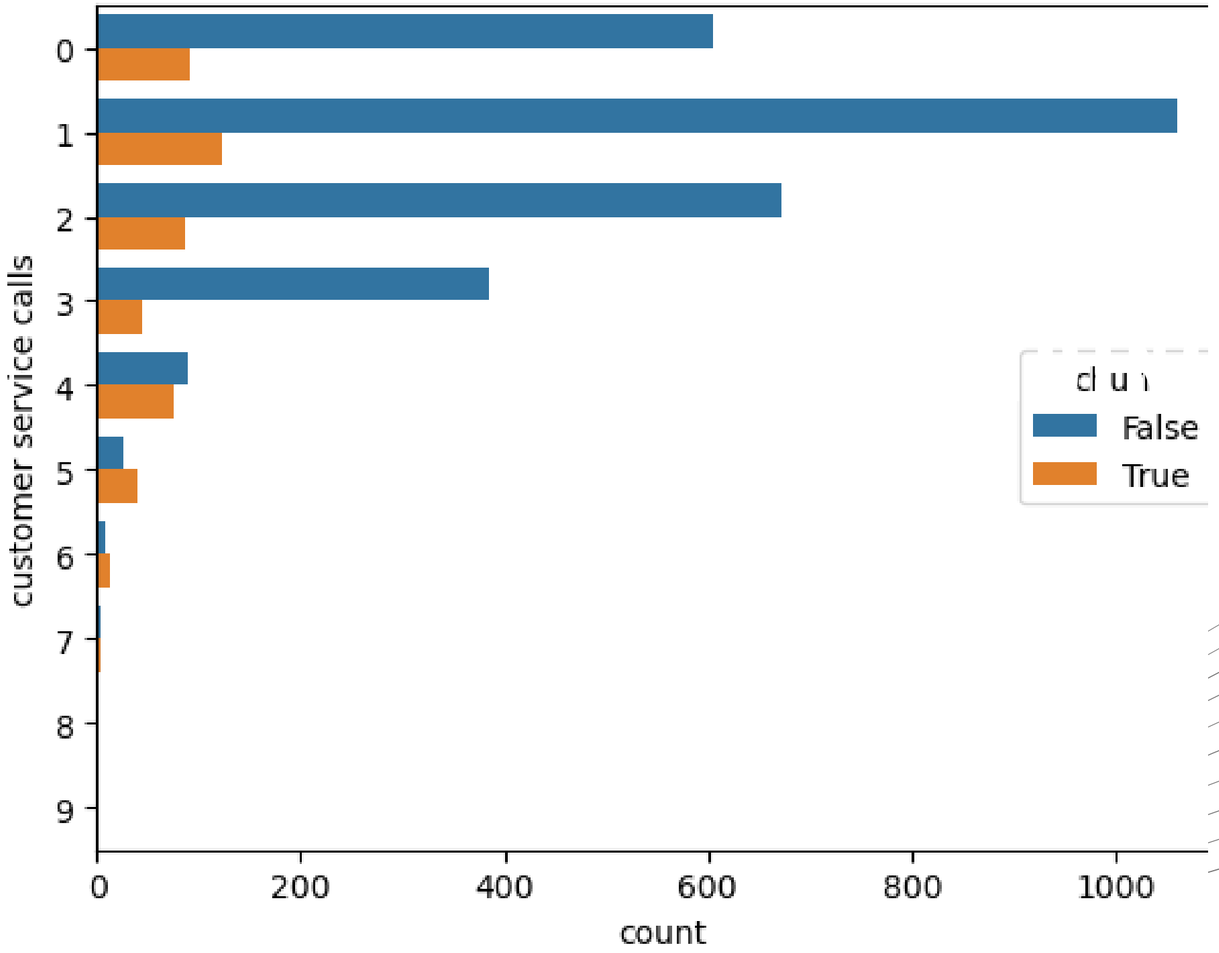
1. Increased competition in the telecom industry necessitates predicting customer churn to retain customers effectively.
2. SyriaTel recognizes the importance of customer retention for competitive parity and cost-efficiency.
3. Analyzing data and using predictive analytics enables SyriaTel to anticipate customer behaviour and take proactive measures to reduce churn rates.
4. Demographic and customer usage data are leveraged to build predictive machine learning models for forecasting churn, allowing SyriaTel to adapt and optimize retention strategies.

OBJECTIVES

- Create machine learning models that can predict customer churn.
- Comparing the build machine learning models and determining the most accurate model for prediction.
- The analysis aims to identify the specific features that have a significant impact on the customer churn rate in SyriaTel and provide valuable recommendations based on the findings hence helping to mitigate churn rates in the company and improve customer retention.

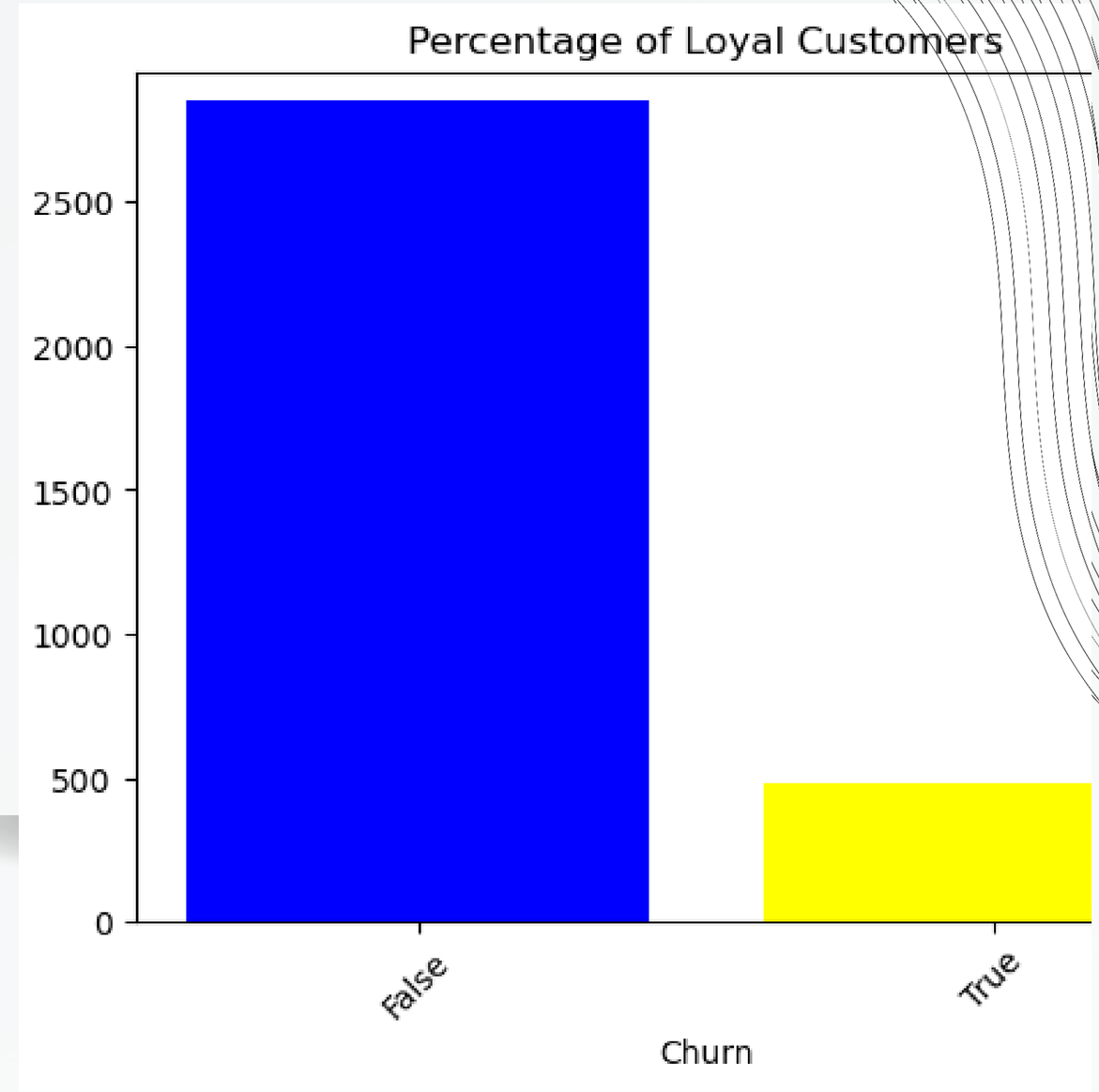
DATA EXPLORATION: CUSTOMER CALL RELATIONSHIP

- There is strong relationship between calls and loyalty as most people making the calls are loyal to Syriatel company
- This shows that there is lower probalility of switching among customers.

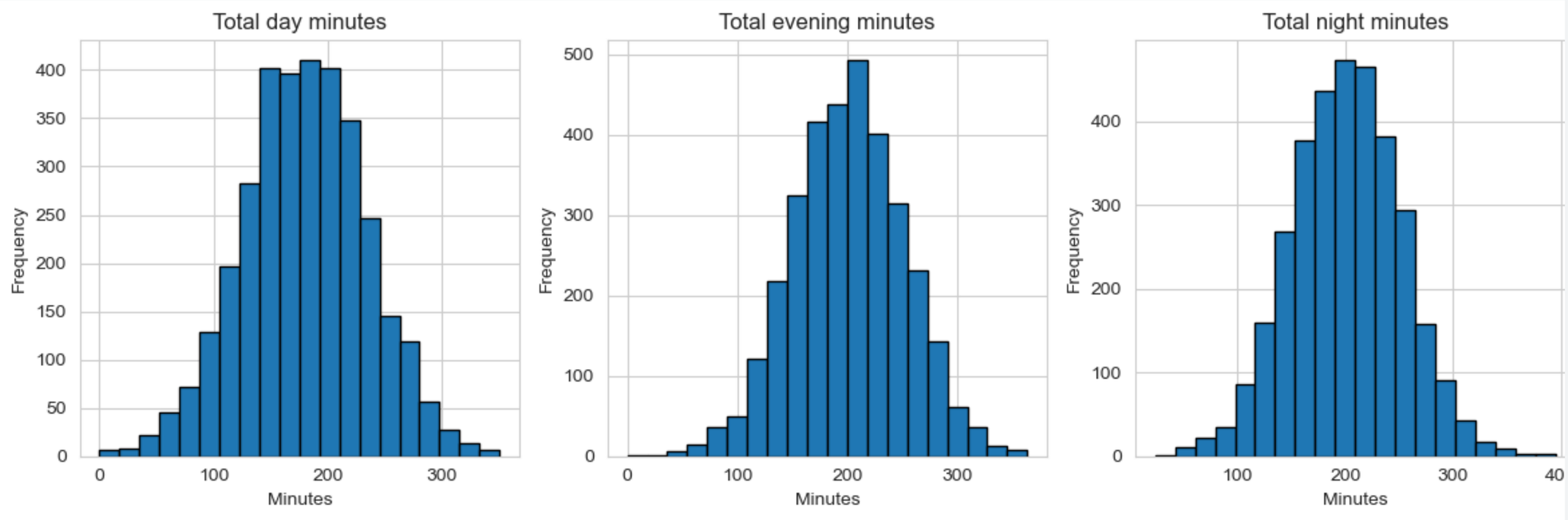


DATA EXPLORATION: CUSTOMER CHURN

- It can be seen that West virginia has highest calls of 32523 and california the lowest with 10582
- In average the calls from the chart above is 15,000.

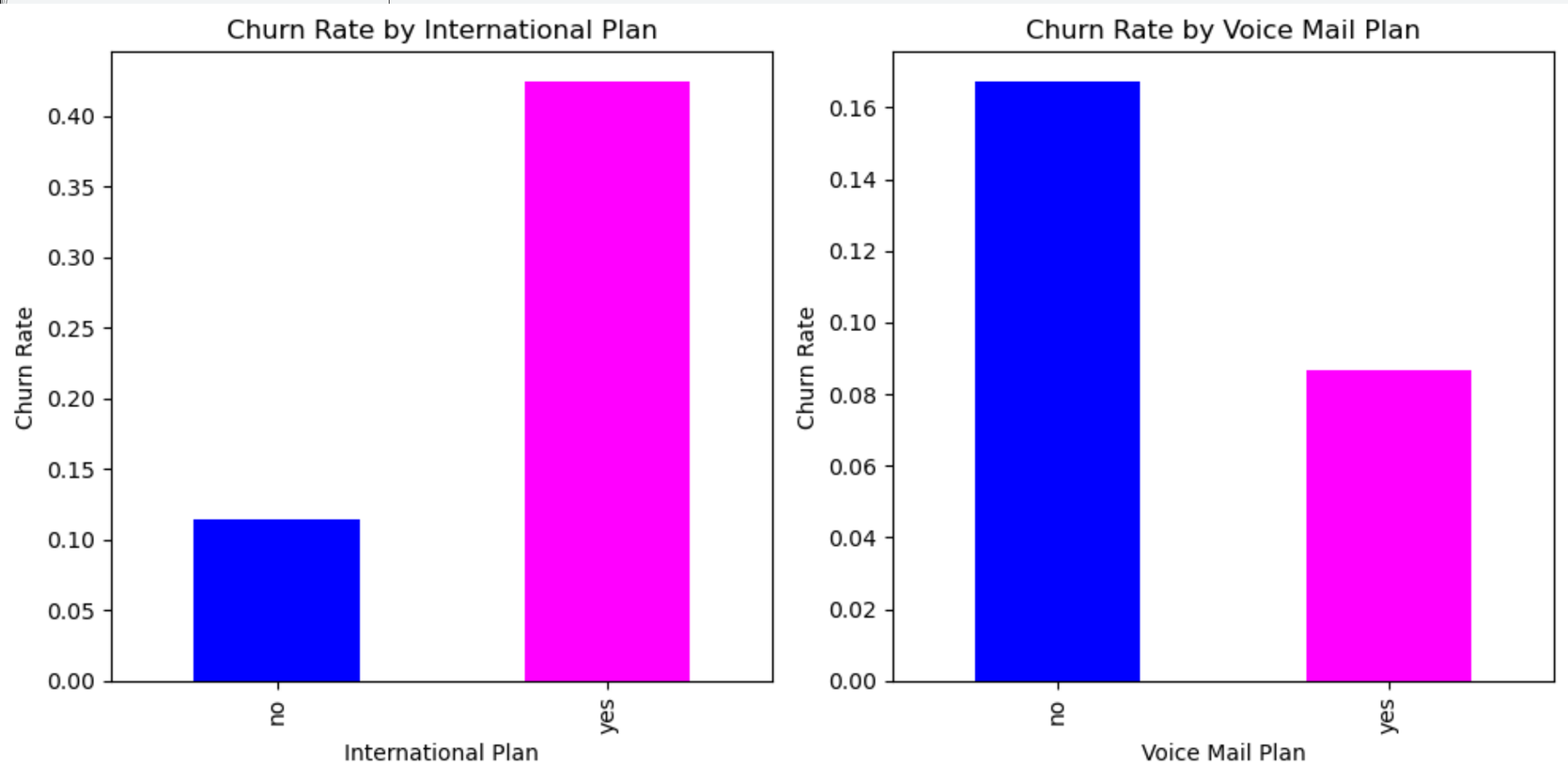


DATA EXPLORATION: TOTAL MINUTES



- The distribution of total ninutes a day, in the evening and at night is normal.
- All the total calls made on diffrent times of the day have almost same mean.

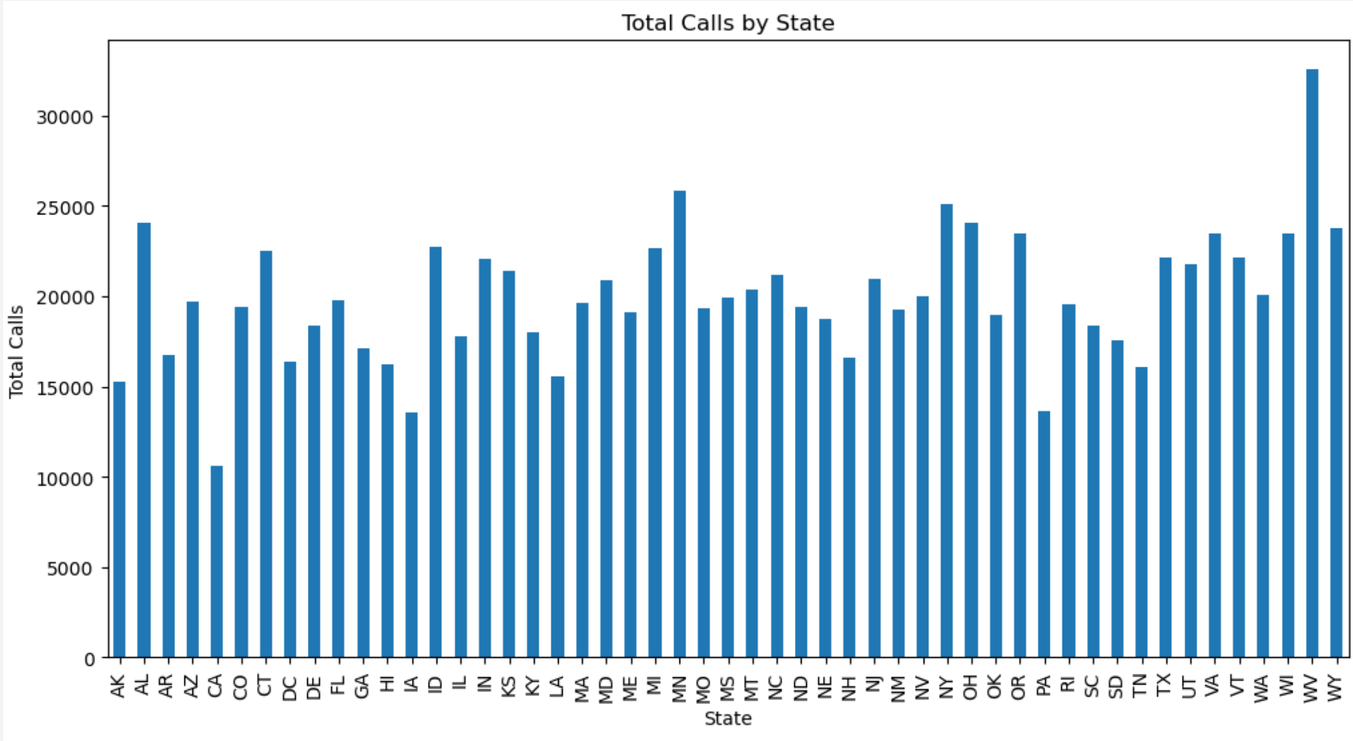
DATA EXPLORATION: CUSTOMER PLAN



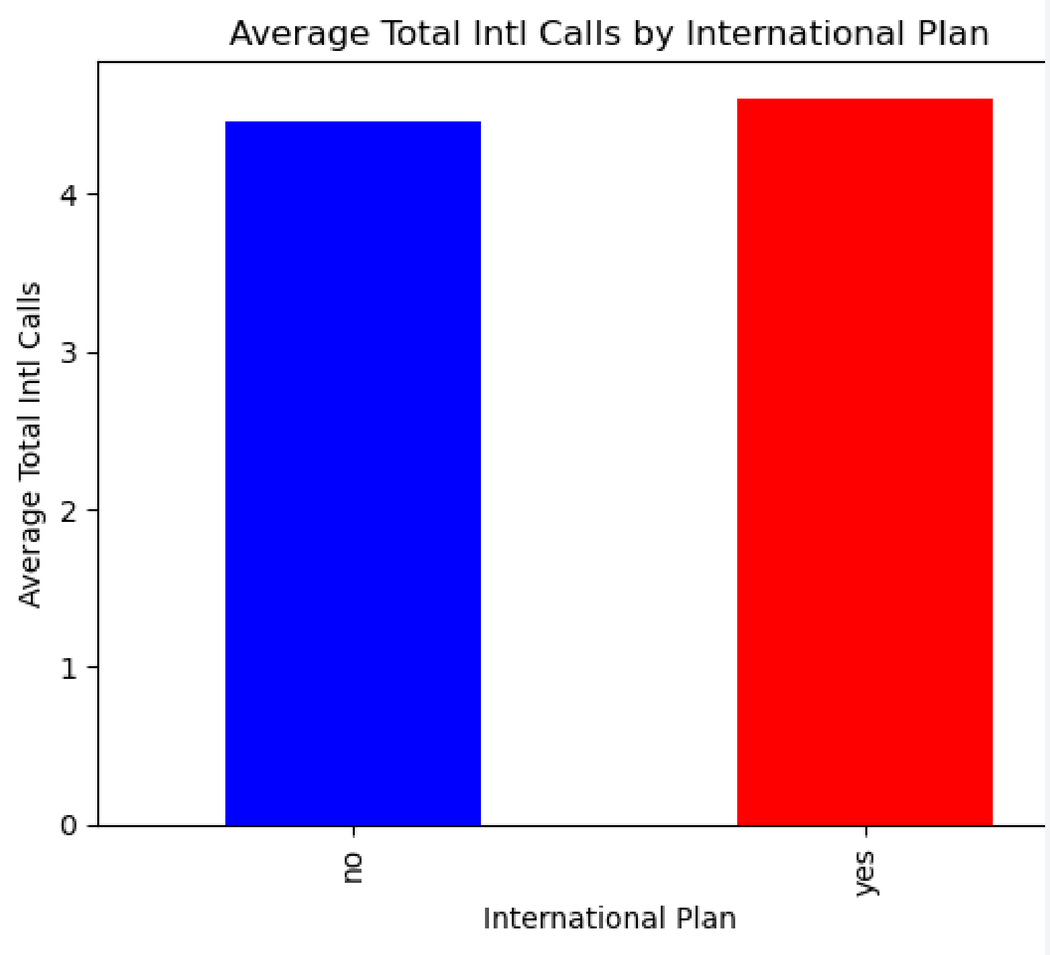
- There is more customers loyalty by Voice mail plan subscribers due to lower churn rate
- Customers have higher rate of churn and a higher probability to switch from Seriatel in terms of international plan.

DATA EXPLORATION: AVERAGE CALLS

- It can be seen that West virginia has highest calls of 32523 and california the lowest with 10582
- In average the calls from the chart above is 15,000.



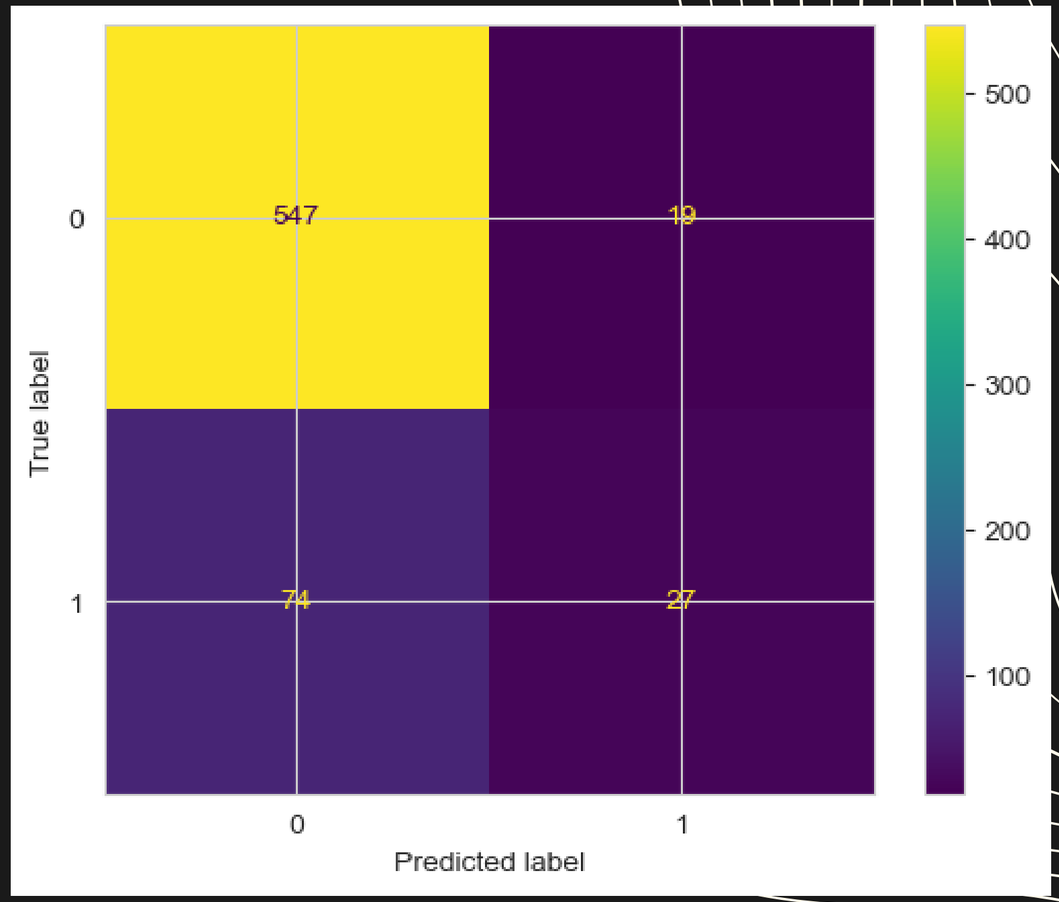
DATA EXPLORATION: AVERAGE CALLS BY INTERNATIONAL PLAN



- In average there almost same number of people getting international plan from the most international calls and from the ones who do not make oftenly.

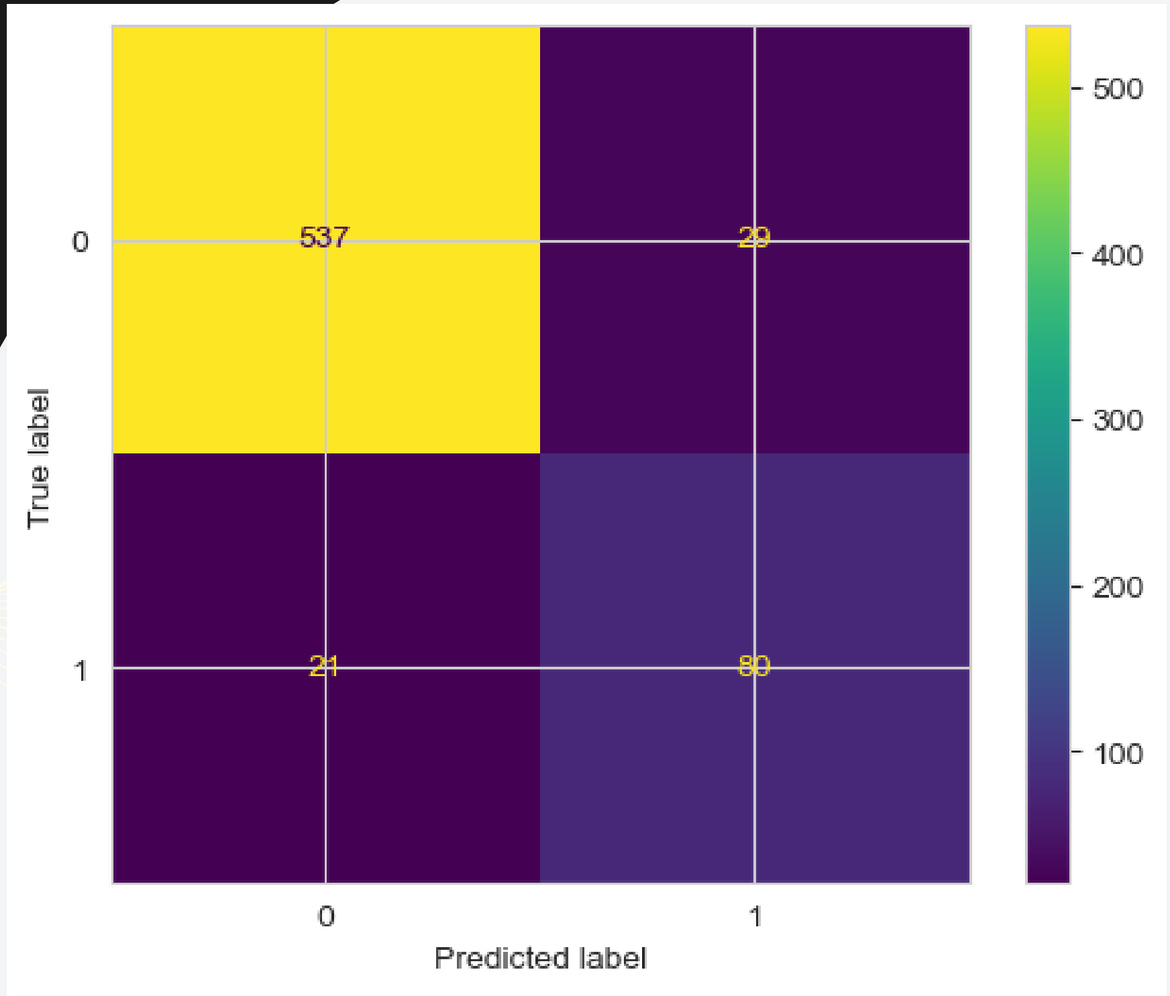
MODEL 1: LOGISTIC REGRESSION

- Logistic regression had an accuracy rate of 89.6% on training data and 86.0% on testing data.
- There was overfit hence cross-validation with 5 folds was done.
- This improved testing data accuracy to 86.2% and testing data dropped to 89.2%.



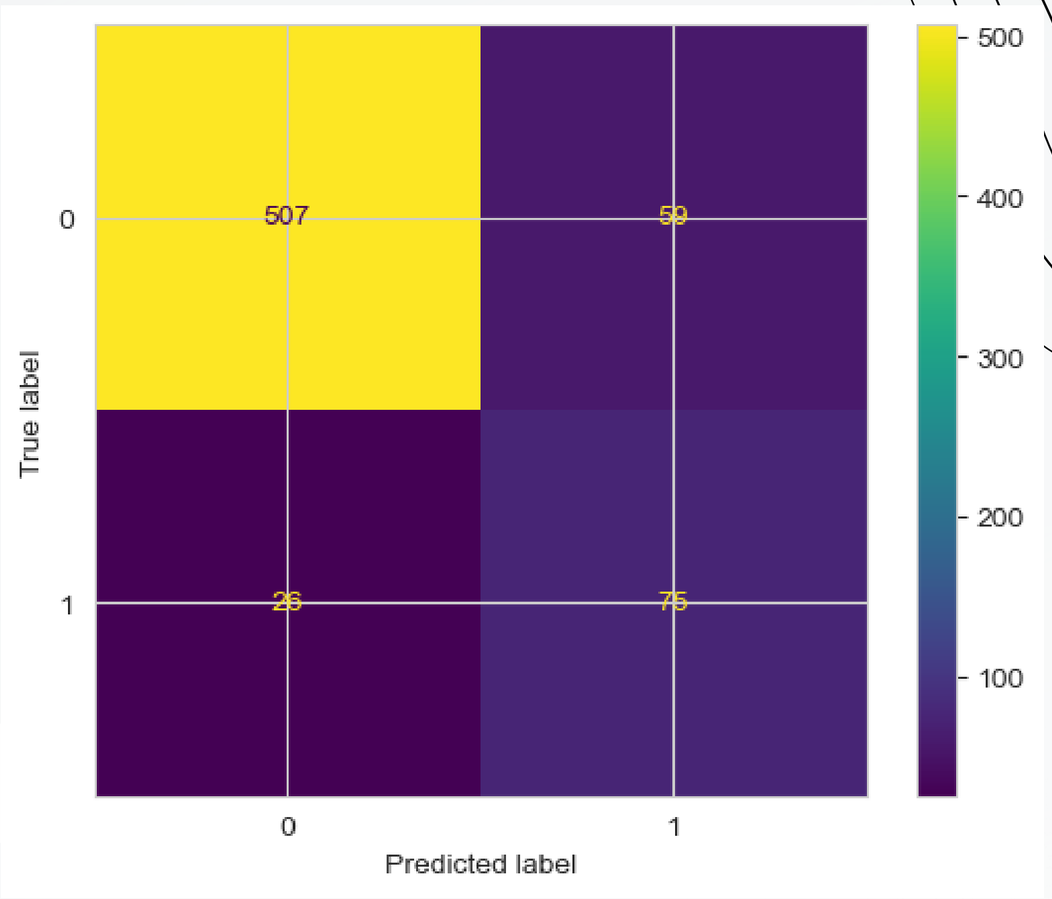
MODEL 2: DECISION TREE

- The decision tree classifier model has an 87.1% accuracy level on training data and 87.2% on testing data in predicting customer churn and not churn.



MODEL 3: RANDOM FOREST

- decision tree classifier model which achieved an accuracy level of 88.6% on training data and testing data accuracy of 92.5%.



HYPERPARAMETER TUNING: GRIDSEARCH

RANDOM FOREST

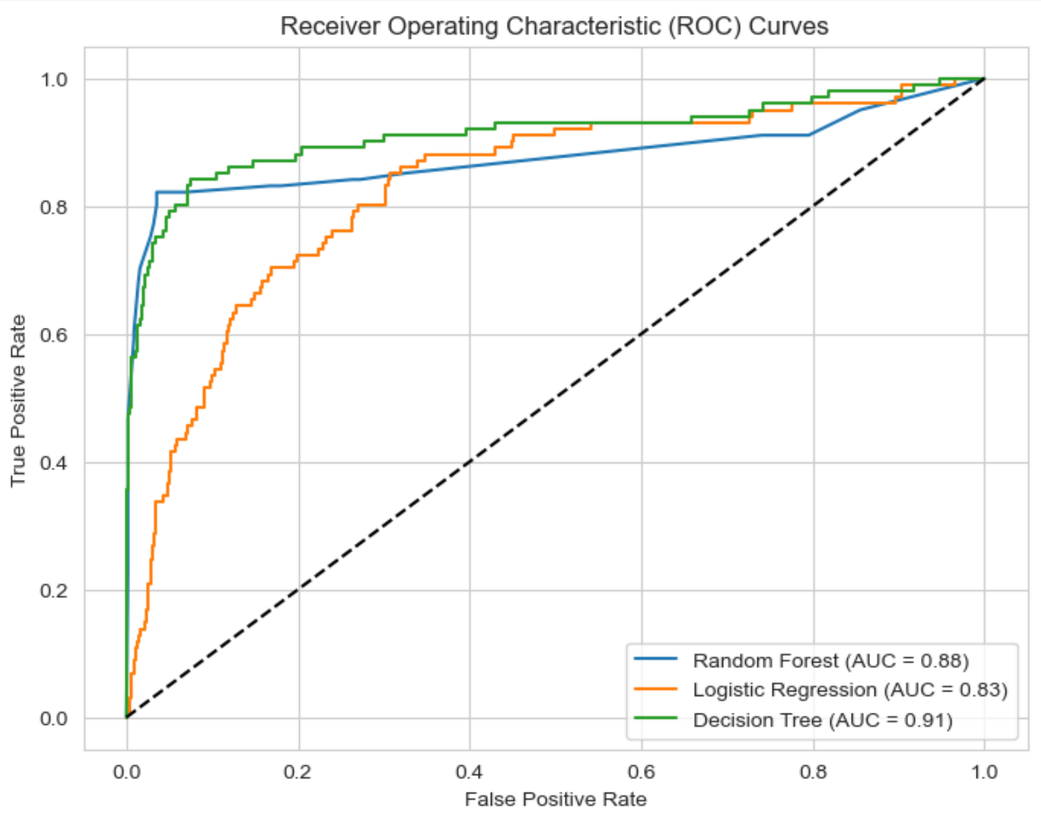
- Accuracy increased to 93.6% on testing data and 92.7% on training data.
-

DECISION TREE

- The accuracy of testing data increased to 93.6% and 90.1% on training data.
-

EVALUATION

- Logistic regression: Lower accuracy (89.2% training, 86.2% testing) indicating limitations in predicting churn.
- Decision tree classifier: Best performance (accuracy 90.4% training, 93.6% testing)afte tuning hence capturing churn patterns well.
- Random forest: Better performance (93.6% training, 92.7% testing) after hyperparamete tuning hence capturing churn fairly.
- ROC scores: Logistic regression (0.83), Decision tree (0.91), Random forest (0.88) showing classification performance.
- Random forest and decision tree training accuracy improved.
- Decision tree had the best overall ability to discriminate and predict customer churn without overfitting.



RECOMMENDATIONS

- 1. Explore alternative algorithms: Consider trying different machine learning algorithms to improve churn prediction accuracy.
- 2. Gather more relevant data: Collect additional data to enhance the accuracy of the churn prediction models.
- 3. Continuously evaluate and refine models: Regularly assess model performance, fine-tune parameters, and retrain models to ensure accuracy and effectiveness in predicting customer churn.

NEXT STEP

- Conduct further analysis: Explore additional factors and features that may contribute to customer churn,
- Implement proactive retention strategies: Utilize the churn prediction model to identify customers at high risk of churn.

MORINGA
Discover · Grow · Transform
MORINGA SCHOOL