PREDICTING CUSTOMER CHURN

Rebecca Frost-Brewer

OVERVIEW

Business and Data Understanding

Modeling

Evaluation

Recommendations

Next Steps

BUSINESS AND DATA UNDERSTANDING

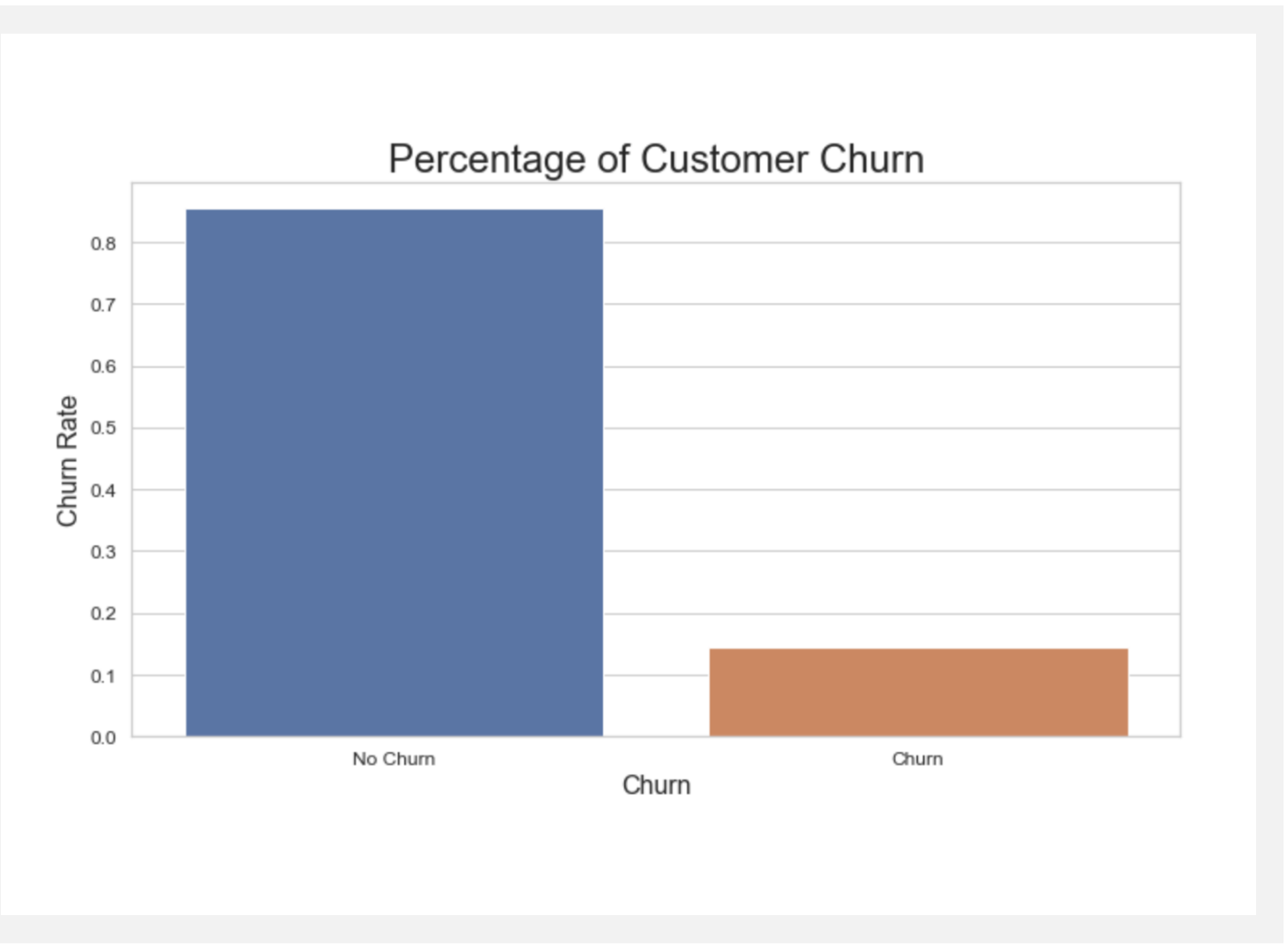
Problem: Better predicting when SyriaTel's customers will soon churn.

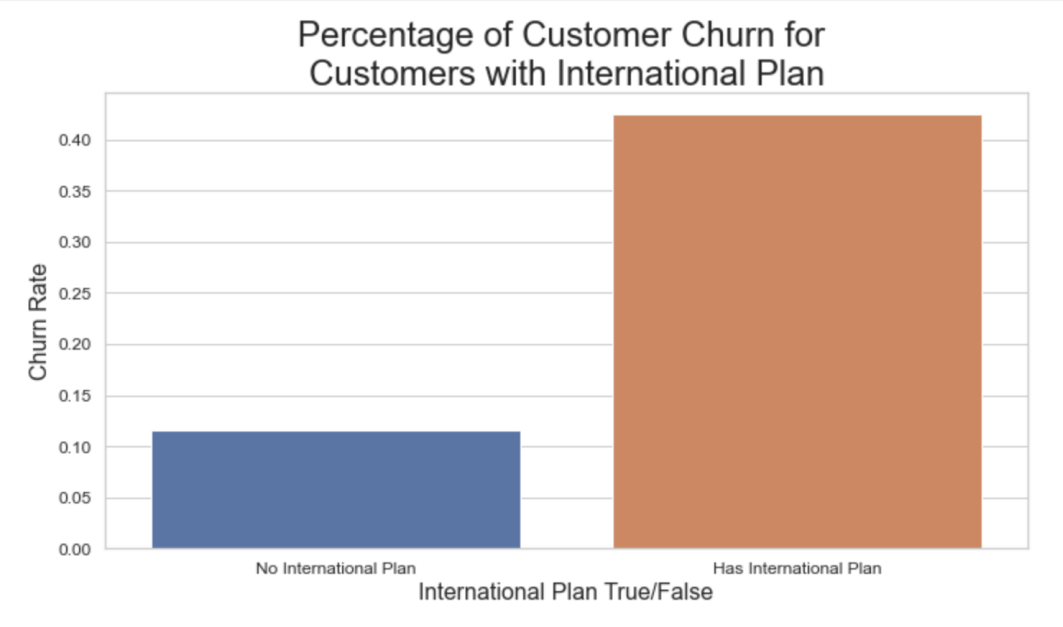
Solution: Finding predictable patterns using a classification model will benefit SyriaTel's business practices to minimize customer churn.

Data: Includes 21 features (both categorical and continuous and 3,333 data points from SyriaTel related to its customers and their accounts and churn information

% OF CUSTOMER CHURN

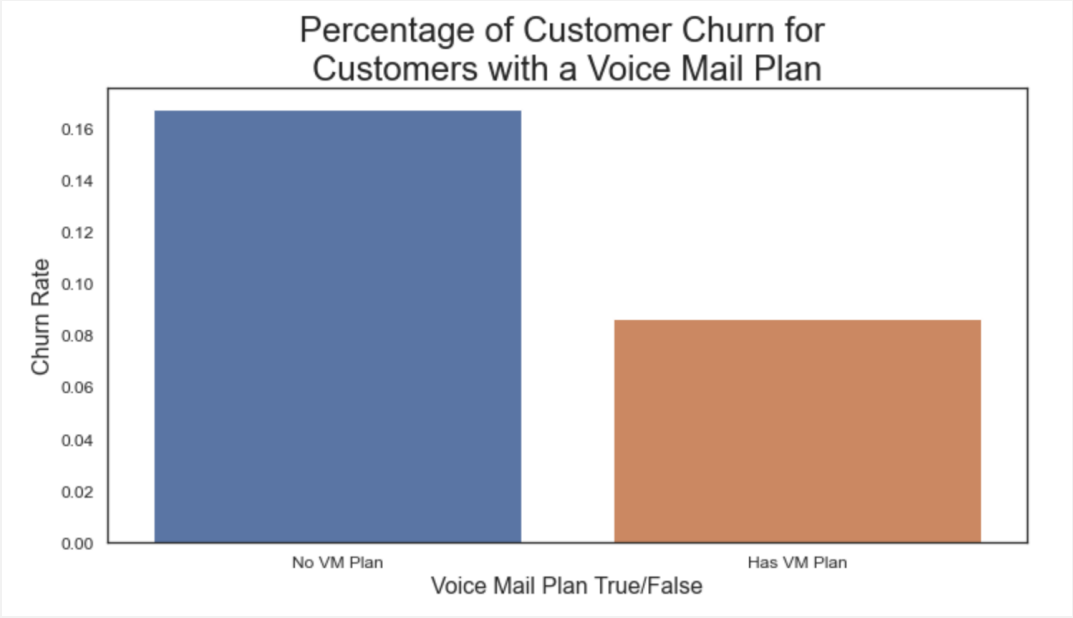
- 14.5% of the customers in this dataset have churned





FEATURES

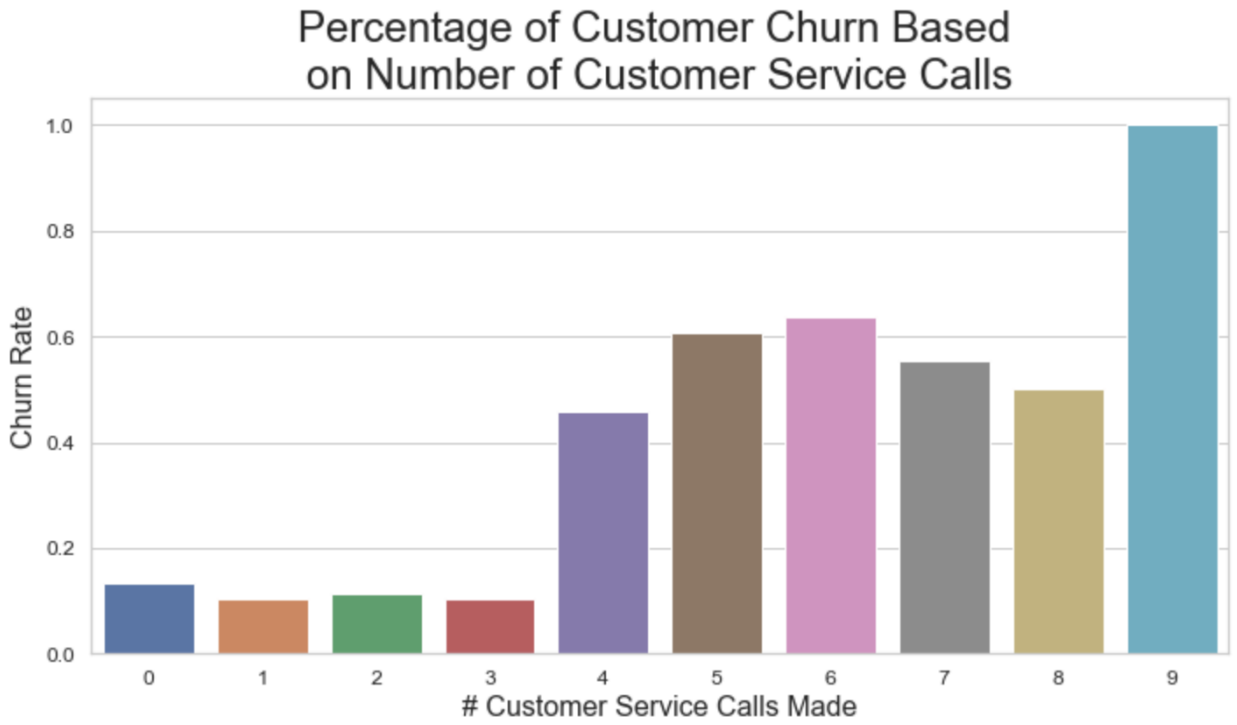
- 42% of customers with an international plan churn



- 17% of customers without a voice mail plan churn

CUSTOMER SERVICE

- Customers who call customer services four or more times have a higher churn rate than those who call fewer than four times



MODELING

- Used a type of machine learning algorithm called Classification, which is the process of predicting the class of given data points
  - In this case, the "class" is customer churn – whether a customer will leave SyriaTel
- Out of the five different algorithms evaluated in this project, the best performing model used an eXtreme Gradient Boosting (XGBoost) algorithm, which provides best-in-class performance among other classifiers

EVALUATION

- Performance metric: F1-score, which combines two classifier metrics:
  - Precision measures what percent of the model's predictions were correct
  - Recall measures what percent of the positive cases were caught correctly
  - F1 is the harmonic mean of the model's precision and recall scores
- F1-score values range 0-1, with 0 as the worst value and 1 as the best value
  - The closer the F1-score is to 1, the more perfect the model is classifying samples

EVALUATION

- The final tuned XGBoost model achieved an F1-score of 0.94
- The most important features influencing the model:
  - **the number of calls the customer made to customer service***
  - whether the customer has a voice mail plan
  - the total number of minutes used per day
  - **whether the customer has an international plan***
  - the total number of international calls made
- Both bolded features were also influential in other models evaluated

RECOMMENDATIONS FOR FUTURE WORK

Provide a larger dataset with greater number of customer information and dates of data collection (to ascertain timeframes)

Further Analysis of Feature Importance: International plans, voice mail plans, calls to customer service

Evaluate Customer Service: Conduct customer service surveys for more information on why customers are calling

Conduct industry benchmarking to determine how voice mail and international plans compare to competitors

THANK YOU

Questions?

Rebecca Frost-Brewer
frostbrewerr@gmail.com