FINAL PROJECT SUBMISSION

Student Name: Mark Bundi Kimathi

Business Problem

The business problem in this scenario is to provide a Telecom Company a prediction of customer churn do it can effectively focus on a customer retention market program
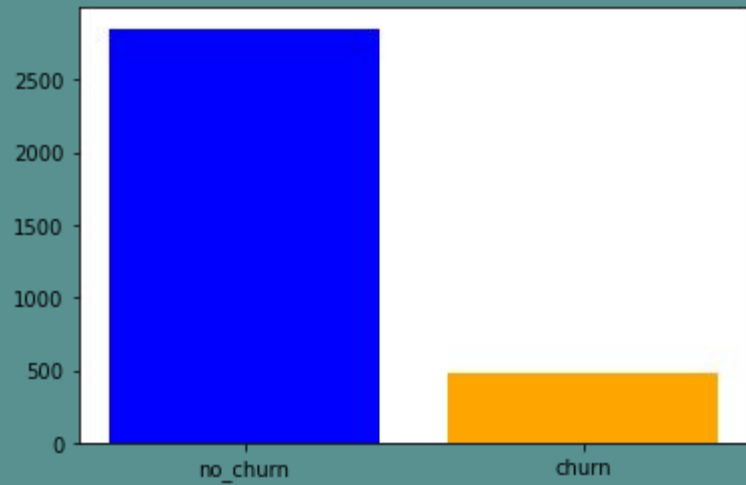
Data Understanding

- The data is from a Telecom Company.
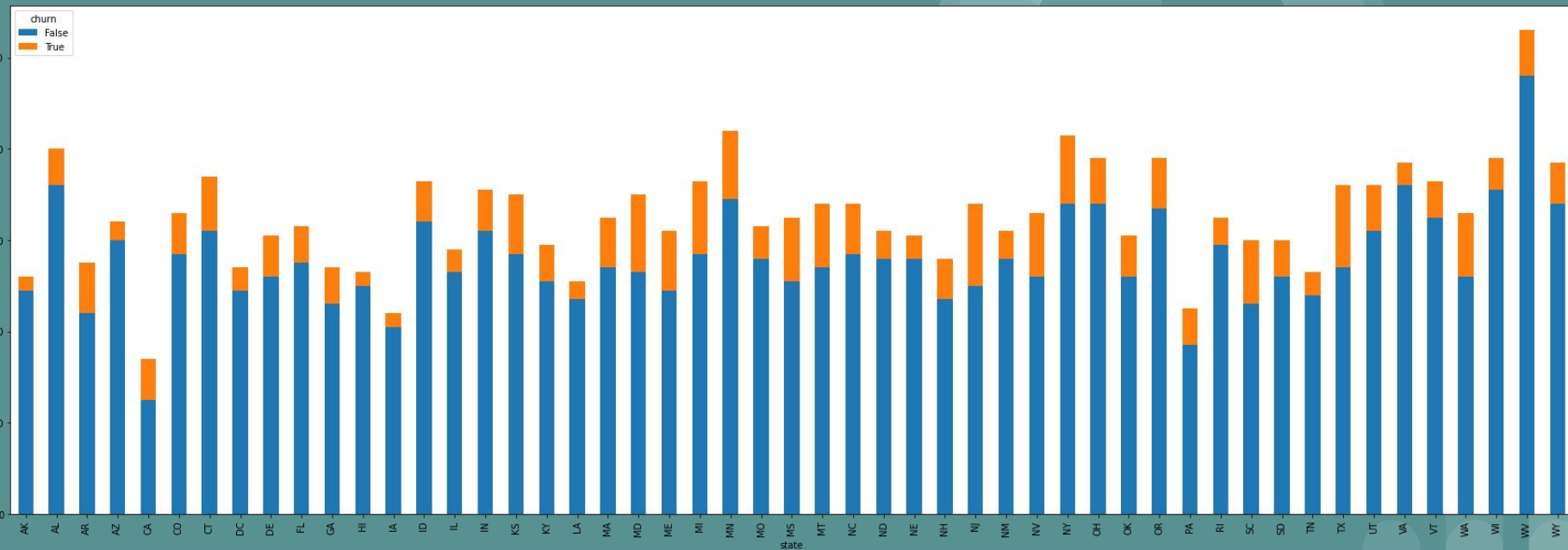- It can be found here: https://www.kaggle.com/becksddf/churn-in-telecoms-dataset.

Customer Churn in total

Customer churn per state

Data Preparation

- The categorical data was transformed into numerical
- The data was scaled to balance the impact of all variables
- The identifier column like phone number were dropped

Modeling

Baseline Model

- The baseline model is a simple model used to contextualize the results of trained models. We create the baseline model to provide a reference point for measuring the performance of other models.
- The baseline model chosen is a Logistical regression model

Observations

The accuracy of the models was:

Logistical regression : 86%

Decision Tree : 91.8%

KNN Classifier: 90.6%

Conclusion

Our goal was to identify clients which are likely to churn, so we can do special-purpose marketing strategies to avoid the churn event. For this we evaluated differently preprocessed datasets and different classifiers.
In the classification chapter we have trained several different classifiers, including a Logistic Regression, a K-Nearest Neighbors Classifier and Decision Tree.