










	protigen34 Add files via upload ...	now	History
..			
	Classification Stuff.txt	2 hours ago	
	Flatiron Spectrum.pdf	now	
	Flatiron Spectrum.pptx	2 hours ago	
	Main Project Notebook - Jupyter Notebook.pdf	now	
	Main Project Notebook.ipynb	2 hours ago	
	README.md	2 hours ago	
	bigml.csv	2 hours ago	
	download (1).png	2 hours ago	
	download.png	2 hours ago	

 README.md



# SyriaTel



## Overview

This project analyzes the telephone/internet company SyriaTel - and their 3300 users listed in this study among their mobile telephone userbase. The primary focus will be the churn amount, or the amount of users who have reported to have left the service. This presents the business problem at hand - at which we'll attempt to solve.

## Business Problem

A large amount of users have left the service. We need to retain further users to prevent them from leaving.

## Data

<https://www.kaggle.com/datasets/becksddef/churn-in-telecoms-dataset>

## Methodology

1. Initial Trials - We will use multiple algorithms and models to predict user churn rate. The model will predict who it believes is most likely to leave the service. The goal will be to target those users to prevent them from leaving.

2. States and Area Codes - Through data manipulating, we can find out what states and area codes produces the highest churn amount. The users in those specific states and area codes or state/area code combinations will additionally be our target.
3. Decision Tree Method - The Decision Tree Method we believe will be the best method of approach and yield the most accurate predictions. However, we will first attempt to try out other methods to make sure.
4. Final Model - We will settle on a final model. Likely the model with the highest recall or precision score. Believing that to be the most accurate prediction model. The False Positive prediction becomes the item of main importance after the model produces its final predictions, and therefore the False Positive amount. False Positives are users who the model predicts a high probability of leaving, but has not yet left. Those are the users specifically we need to prevent from leaving.

## Project and Trials

1. Initial trials ran. Models such as KNearest Neighbors, Logistic Regression, GridSearch, and Cross Validation are some of the methods used.
2. States and Area Codes are combined to find the most frequent churns by state and/or area codes.
3. Decision Tree Model found to be the best model.
4. Final Model is a form of Decision Tree Model.

## Results

1. Final recall score is .7625
2. Final False Positive count is 58
3. States of New York, Texas, New Jersey, Maryland, and Michigan are reported to have left the service at the highest rate.
4. Additionally Area Code 415 in the States of New York, Connecticut, Nevada, Arkansas, and Texas are reported to have left the service at the highest.

Chart 1:

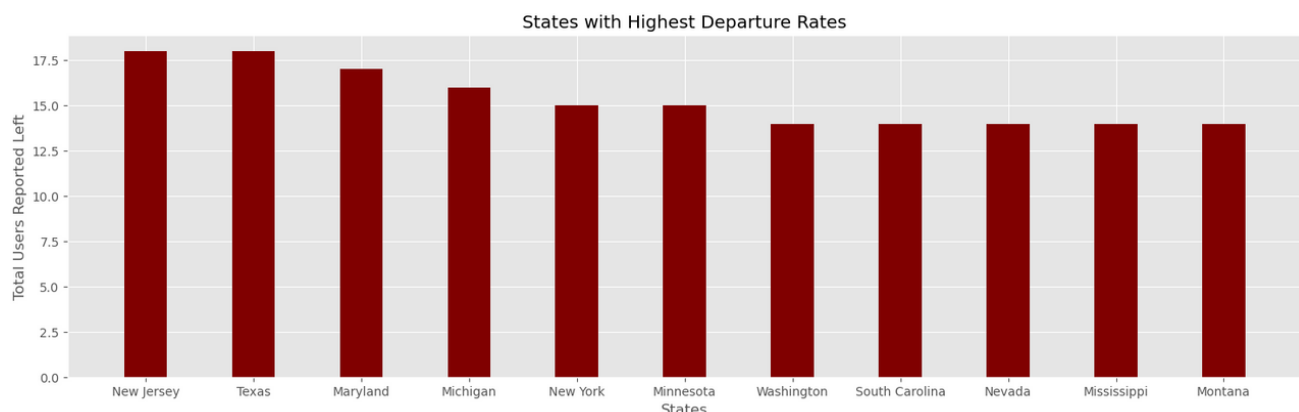
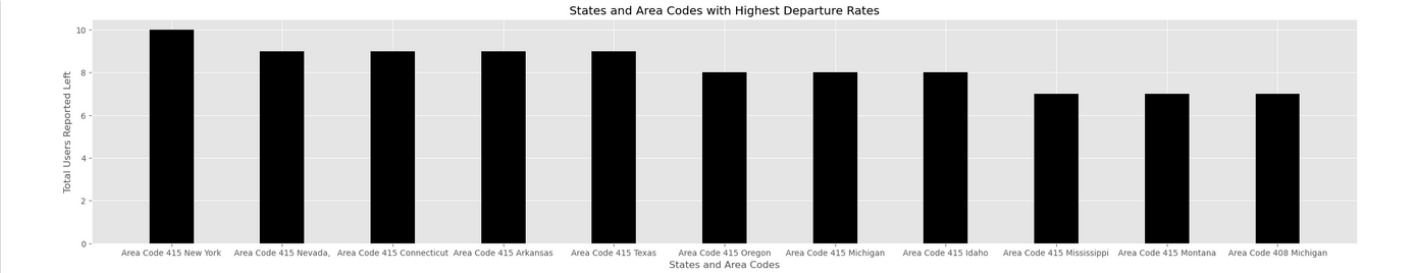


Chart 2:



## Conclusion and Final Goals

We are successful in finding the most important results to move further with our goals. Our final goal is to prevent further users from leaving our service. We have narrowed down our search to target more specifically what state and area code they might be in. We have also an approximate prediction of just exactly how many users might leave the service. It is now much easier to move forward with our plans to retain these users. The discount, bundles, and coupons offered in high probability of leaving states in those area codes will be a more generous offer than normally.

Give feedback