

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Contributor Roles:

1. Sai Krishna Vamshi Devarasetty - krishnavamshidevarasetty@gmail.com
 1. Data Wrangling
 2. Comparison of Churn vs Non-Churn Data
 3. Analysis of Churn vs Non-Churn Data
 4. Area code-based Analysis
2. Gangadhar Palle - pallegangadhar156@gmail.com
 1. Data Wrangling
 2. Comparison of Day, Evening and Night Calls (No. of Calls)
 3. Analysis of Day, Evening and Night Calls (No. of Calls)
 4. State based Analysis
3. Abhishek Sharma - abhisheksharmatrio@gmail.com
 1. Data Wrangling
 2. Comparison of Day, Evening and Night Call minutes (No. of minutes)
 3. Analysis of Day, Evening and Night Call minutes (No. of minutes)
 4. International Plan based Analysis
4. Nitish Rao - nitishrao1896@gmail.com
 1. Data Wrangling
 2. Comparison of Day, Evening and Night Call charges (Total charges for calls)
 3. Analysis of Day, Evening and Night Call charges (Total charges for calls)
 4. Voicemail plan-based Analysis

Complete Jupyter Notebook is prepared by combined Collaboration.

Please paste the GitHub Repo link.

Github Link: <https://github.com/codes-by-vamshi/EDA-TelecomChurn>

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Orange S.A., formerly France Telecom S.A., is a French multinational telecommunications corporation. The Orange Telecom's Churn Dataset, consists of cleaned customer activity data (features), along with a churn label specifying whether a customer canceled the subscription.

As first step we explored data and figured out of all columns except 4(State, International Plan, Voicemail Plan, Churn) remaining all are of numerical types i.e, either int64 or float64.

Even though the Area Code is numerical it is not ordinal that means there is no mathematical significance to that number other than for referring. State is the only variable which is string. Although International Plan and Voicemail plan are given as strings(Yes/No) we need to consider them as Boolean for analysis.

Dataset consists of 3333 data points, out of which 483 are Churned and 2850 didn't churned. To do further analysis divided given dataset into two pandas data frames one with Churn being True and the other being False.

First we tried to figure out if users using service at any specific duration(day, evening or night) is resulting in Churning? And we figured out that among the people who use telecom service more than avg. has high Churn rate compared to users using less service.

Secondly we want to figure out if there is any correlation among duration of the day for users using the telecom service. But it resulted in almost 0 correlation implying that the duration during which user using service didn't had any impact.

Next when we did area code based analysis to see if any area has significant impact on Churning we figured out the IQR(Inter Quartile Range) is more for Churned Users indicating the service usage is more widely spread for Churned users when compared to Non Churned.

In State based analysis we figured out 11 states out of total 51 states have more than 20% churn rate. NJ and CA having highest churn rate of around 26.47%.

Further when we did analysis based on opting of International plan we have seen the churn rate being 40% for users opting for international plan, which clearly indicates that users are not at all satisfied with their International Plan.

On the other hand for Voicemail plan the churn rate is less than half of those who didn't opt voice mail plan depicting that their voicemail service is very good.

And finally we figured out users with more than or equal to 4 customer service calls are likely to churn 4 times more than others.