## Report on telecommunication churn dataset

The telecommunication dataset consisted of 13 columns below:

* State
* Account length
* Area code
* International plan
* Voicemail plan
* Number of Vmail messages
* Total day minutes
* Total day calls
* Total day charge
* Total evening minutes
* Total evening calls
* Total evening charge
* Total international minutes
* Total international calls
* Total international charge
* Customer service calls
* Churn

# Preprocessing analysis

During the preprocessing stage I got some insights from the data.

An average of 591 minutes is spent on calls per day.

An average 0f 305 calls are made per day.

Only about 9% of customers have international plans.

Only 27% of customers have voice mail plans.

A total of 1017022 number of calls were made

A total charge on calls = 168017.75

WY has the highest call per day with 32523 numbers of calls per day hence the highest revenue generated with 6079.48 charges on calls per day

# Feature engineering

After the preprocessing stage I noticed the summing up the call to total calls, charge and minutes spent per day, would me more meaningful than using calls during the day, night and evening, so I created some new features ;

* Total calls per day
* Total charge and
* Total minutes spent on calls per day

These additional features improved the model score and the roc\_score by 0.05 %

# Model

I used decision tree classifier Algorithm to predict whether customers are going to churn or not and got a model score of 0.9700 and a roc\_score of 0.929.