Telecom churn analysis - Insight Report

### **HYPOTHESIS**

* **First**, identify high value customer from the last three months based on the Average Revenue per unit (ARPU) in the ‘Good Phase’, in June & July
* **Second**, identify the customers who are likely to churn in September based on the usage of services.
* **Third**, come up with appropriate measure to retain high value customers by looking into their attributes of last three months. This will ensure their permanence with Operator which will in turn increase revenue.

### **EXPLORATORY DATA ANALYSIS**

* **Extraction of dataset**

The csv file is downloaded from internal drive of the database.

* **Data cleaning/Missing data**
* The features with 85% missing values, the columns are dropped
  + The features with 5% - 25% missing values, we imputed the missing values using median or mode method
* The column having single unique value like mobile\_number, circle\_id, loc\_og\_t2o\_mou, lstd\_og\_t2o\_mou, loc\_ic\_t2o\_mou, last\_date\_of\_month are dropped
* The columns with Boolean values are night\_pack, fb\_users. These are filled by the most occurring value i.e. is 1 or 0.
* Columns with dates are date\_of\_last\_recharge, date\_of\_last\_recharge\_data. These columns are filled by first day of every month.
* **Initial insights**

****

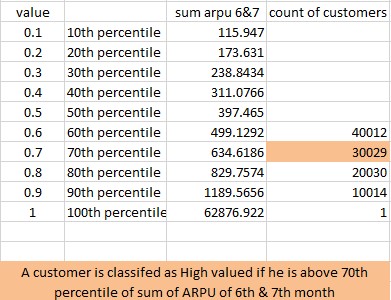
The initial insight is to obtain an overview about the data and which features are required and how they affect other features.

* **Data preparation**

As we are dealing with high dimensional dataset, determining important features based on our problem statement is crucial.

The first step is to classify the **High value** and **Low value** customers using the service network.

* As June and July is considered as “**Good Phase**”, most of the customers are happy with the services and continuing to use the operator
* The type of customer is distinguished by considering **70th percentile of the sum of ARPU of June & July** as the benchmark. Customers above 70th percentile are high valued and below 70th percentile are Low valued customers.



The second step is to classify the **Churn** and **Un-churn** customers in the month of September.

* For this classification we chose to go with Usage based analysis. The variables for consideration are Total outgoing calls, Total incoming calls, 2G data in MB & 3G data in MB.
* To classify if a customer will churn or not can only be decided in the “**Action phase**” i.e. in the month of August.
* It was considered if the customer makes and incoming or outgoing call of less than 0.1 min (6sec) and not use the internet service at all is likely to churn in September.
* This benchmark of limiting the call duration to 0.1mins (6sec) is decided by considering the least duration a customer makes or receives a call as per the provided dataset.
* Also as this data was collected in 2014, internet was highly used by everyone in their daily life. Taking this into account we arrived at the benchmark of whoever doesn’t use the internet services at all may likely to churn in the next month.

