# Introduction:

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**

# Business ****Objective**:**

The objectives are:

* Analyze customer-level data of a leading telecom firm
* Build predictive models to identify customers at high risk of churn
* Identify the main indicators of churn.

# Approach taken:

The dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively.

In churn prediction, we assume that there are **three phases** of customer lifecycle:

The ‘good’ phase:

In this phase, the customer is happy with the service and behaves as usual.

The ‘action’ phase:

The customer experience starts to sore in this phase, for e.g. he/she gets a compelling offer from a competitor, faces unjust charges, becomes unhappy with service quality etc. In this phase, the customer usually shows different behavior than the ‘good’ months. Also, it is crucial to identify high-churn-risk customers in this phase, since some corrective actions can be taken at this point (such as matching the competitor’s offer/improving the service quality etc.)

The ‘churn’ phase:

In this phase, the customer is said to have churned. We **define churn based on this phase**. At the time of prediction (i.e. the action months), this data is not available to you for prediction. Thus, after tagging churn as 1/0 based on this phase, you discard all data corresponding to this phase.

After defining the churn phase, we filter the high value customers which are defined as below

## Filter high-value customers

As we need to predict churn only for the high-value customers. Define high-value customers as follows: Those who have recharged with an amount more than or equal to X, where X is the **70th percentile** of the average recharge amount in the first two months (the good phase).

# Hypothesis:

We are formulating certain hypotheses and trying to validate them using the EDA steps. Some of these hypotheses are:

1. The average, maximum and total amount of recharge is declining as we are moving from good phase to Churn phase for churn customers
2. The consumption of 2g and 3g data is declining for churn customers as we are moving from good phase to churn phase
3. Average Revenue per user from Good phase to action phase is decreasing from good phase to churn phase
4. Since we are more interested in identifying the customers who are likely to churn, we will take Recall/True Positive Rate/ Sensitivity to be good

# Recommendation

* Customers having less than 4 years of tenure are more likely to churn. The company needs to concentrate more on that segment.
* Roaming on Incoming and Outgoing Calls for 8th month are strong indicators of churn behavior
* Outgoing local calls to landline, fixed line, mobile and call center provides a strong indicator of churn behavior.

# Financial benefits:

After taking the steps mentioned in the recommendations, it is likely that the company will generate more revenues from the high value customers and retain most of those.