Telecom Churn Case Study

Problem Statement

Business Problem Overview:

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 to 25% annual churn rate. Given the fact that it costs 5 to 10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.

Main Goals:

- For many incumbent operators, retaining high profitable customers is the number one business goal.
- To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.
- In this project, you will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

Understanding a Customer's Behaviour During Churn:

Customers generally do not decide to switch to competitor network instantly, it is rather done over a period of time (this is mainly applicable to high-value customers). In churn prediction, we assume that there are three phases of customer lifecycle:

1) The good phase: Here in this scenario the customer is quite happy with the service and no behavioural changes can be observed.
2) The action phase: In this phase, the customer experience starts to sore like maybe customer gets a compelling offer from a competitor, faces unjust charges, becomes unhappy with service quality etc. In this phase, the customer usually shows different behaviour than during the good months phase. Now is the crucial time to identify high-churn-risk customers, so that we could rectify the problems (such as matching the competitor's offer/improving the service quality etc.) and retain the customer.
3) The churn phase: By the time customer reaches this phase, he/she is said to have churned. We define churn based on this phase. Note that at the time of prediction (i.e., the action phase), this data will not be available to us for prediction. Therefore, after tagging churn as 1/0 based on this phase, we discard all data corresponding to this phase.

As we are working over a four-month window, the first two being the good phase, the third being action phase and the fourth month being the churn phase.

Analysis Approach:

- The industry of telecommunications experiences an average of 15 - 25% annual churn rate. Considering the fact that it costs 5 to 10 times more to acquire a new customer than to retain an existing one, customer retention has become comparatively more important than customer acquisition.
- Here, provided data is of 4 months data related to customer usage. In this case study, we are going to analyse customer-level data of a leading telecom firm, build few predictive models to identify customers that are at high risk of churn and identify the main indicators of churn so that the company can act on it beforehand.
- Churn is predicted using two approaches which are Usage based churn and Revenue based churn.
- Usage based churn: customers whose usage of services like incoming or outgoing calls, internet etc. over a period of time is zero.
- In this case study we are going to consider only usage-based churn.

- In Indian and southeast Asian markets, approximately 80% of revenue comes from the top 20% customers (high-value customers). So, if we could reduce churn of these high-value customers, then we will be able to reduce significant revenue leakage. Therefore, in this case study our main focus will be on high value customers.
- The dataset contains customer-level information for a span of four consecutive months June, July, August and September which are encoded as 6, 7, 8 and 9 respectively.
- The business objective is to predict the churn in the last (September) month using the data from the first three months.
- This is a classification problem, where we need to predict whether the customers are about to churn or not.

Analysis Steps

Data Cleaning:

- We started with importing necessary packages and libraries.
- We loaded the dataset into a data-frame.
- We checked the number of columns, their data types, null count and unique value count so that we could get better knowledge about the data in hand and to check if the columns are under correct data-type.
- No duplicates were detected when checked for duplicate values (rows) in the given data.
- Since 'mobile number' is the unique identifier in the available data, we made it our index to retain the identity.
- Renamed some columns that do not follow the naming standard to make sure all the variables follow the same naming convention.
- After column renaming, we converted the columns into their respective data types.
- Converted the date columns which have 'object' as their data type to the proper date-time format.
- Since our analysis is mostly focused on the high value customers, we filtered for high value customers to carry out the further analysis.
- Checked for missing values.
- Dropped all the columns that has missing values of more than 50%.

- As we have 4 months data and each month's revenue & usage data is not related to each other, we did month-wise drill down on missing values.
- Since some of the columns had similar range of missing values, we compared them with their related columns to check if they might be imputed with 0.
- We found that 'last_date_of_the_month' had some missing values which are very important so, we imputed the last date basing on the month.
- Found some columns with only one unique value, so it is not useful for the analysis. So, we dropped those columns.
- After checking all the data preparation tasks, we tagged the Churn variable (which is our target variable).
- After imputing, we dropped churn phase columns (Columns belonging to 9th month).

Exploratory Data Analysis:

- The users that are likely to churn are the users with negative average revenue for the telecom company in both phases.
- Customers with low aon are comparatively more likely to churn than customers with higher aon.
- Revenue generated from the customers that are likely to churn is highly unstable.
- Customers that regularly use 2g network in 6th and 7th months are less likely to churn.
- Customers with fall in usage of 2g network in 7th month are very likely to churn.
- Customers with stable consumption of 3g network during 6th and 7th months are less likely to churn.
- Customers with fall in usage of 3g network in 7th month are more likely to churn.
- Correlation analysis was done.
- We created the derived variables and removed the variables that were used in the process of deriving new ones.
- Outlier were treated accordingly.
- We checked categorical variables and the contribution of classes in those variables.
- Created dummy variables wherever they were needed.

Recommendations:

Following are the strongest indicators of churn:

- Customers who churn show lower average monthly local incoming calls from fixed line in the action phase compared to other customers.
- Customers who churn show lower number of recharges done in action period.
- Customers who churn have done higher recharge than non-churn customers. This factor is useful when coupled with above factors.
- Customers who churn are more likely to be users of 'monthly 2g package-0 / monthly 3g package-0' in action period.

Based on the above indicators the recommendations to the telecom company are:

- Offer special discounts to customers that are customized to their needs.
- Provide additional data/internet services on high recharges.
- Award personalised rewards based on a customer's recharge like, extra minutes on calls only recharge or STD minutes or extra data etc.,
- Communicate with customers to know their needs and try to accomplish their desires.
- Lower tariffs on data usage, provide a better 2G area coverage where 3G is not available.
- Expansion of 3G network where it is currently not available and strengthen where the network is weak is strongly recommended.
- Concentrate mainly on people who don't recharge as regularly as they are supposed to, they are the customers most likely to churn.