**Customer Churn Prediction**

**Context**

A DTH (Direct to Home) service provider named Magix is facing a lot of competition in the current market, and it has become a challenge to retain existing customers in the current situation. Hence, the company wants to develop a model through which they can do churn prediction of the accounts and provide segmented offers to the potential churners. In this company, account churn is a major problem because one account can have multiple customers. Hence, by losing one account, the company might be losing more than one customer.

Thus, being a data scientist, you have been assigned to develop a churn prediction model for this company and provide business recommendations on the campaign. Your campaign suggestion should be unique and very clear on the campaign offer because your recommendation will go through the revenue assurance team.

**Objective**

The objective of the problem statement is to predict the customers who are going to churn based on usage and other demographic factors and provide recommendations to the company to minimize the churn.

**Business Opportunity**

Customers are a fundamental part of a DHT's growth, given that, their satisfaction plays an important role in the pursuit of service success. An effective churn prediction model will enable the company to identify potential churn customers and dissuade them from leaving by improving the offers tailored to them specifically. Also, when this model is proven and tested effectively, the Direct-To-Home service provider company can sell this proprietary model to other companies facing the same challenge. This will further boost the company's revenue.

**Data Report**

***Data Collection***

* The data was collected on various accounts using their account IDs to know if they were still operating or if they had been closed or abandoned (churn).
* The campaign survey was taken at different categories of cities in different locations ranked by city tiers.
* Over the course of 12 months, the number of times customers tied to an account contacted customer care was recorded.
* Details tied to each account ID were collected such as the number of users tied to each account, gender and marital status of the primary customer, account segment on the basis of spend, average monthly revenue over the last 12 months, preferred login device of customer, etc.

***Visual Inspection of Data***

* The data has 11260 rows and 19 columns
* Each Account ID is unique without repetition
* There are many null values with empty spaces, while there are many special characters such as @, +, #, &&&&, regular + instead of regular plus, F instead of Female which has to be replaced with NaN & then missing value imputation to be performed.

***Data Dictionary***

This dataset has 11260 rows and 19 features, with churn being the target variable.

| **Variable** | **Description** |
| --- | --- |
| AccountID | account unique identifier |
| Churn | account churn flag (Target) 1 - Churn, 0 - No Churn |
| Tenure | Tenure of account in months |
| City\_Tier | Tier of primary customer's city |
| CC\_Contacted\_L12m | How many times all the customers of the account has contacted customer care in last 12months |
| Payment | Preferred Payment mode of the customers in the account |
| Gender | Gender of the primary customer of the account |

| Service\_Score | Satisfaction score given by customers of the account on service provided by company |
| --- | --- |
| Account\_user\_count | Number of customers tagged with this account |
| account\_segment | Account segmentation on the basis of spend |
| CC\_Agent\_Score | Satisfaction score given by customers of the account on customer care service provided by company |
| Marital\_Status | Marital status of the primary customer of the account |
| rev\_per\_month | Monthly average revenue generated by account in last 12 months (Its a encoded ordinal values) |
| Complain\_l12m | Any complaints has been raised by account in last 12 months |
| rev\_growth\_yoy | revenue growth percentage of the account (last 12 months vs last 24 to 13 month) |
| coupon\_used\_l12m | How many times customers have used coupons to do the payment in last 12 months |
| Day\_Since\_CC\_connect | Number of days since no customers in the account has contacted the customer care |
| cashback\_l12m | Monthly average cashback generated by account in last 12 months ( these values are in points,but monetary value not disclosed) |
| Login\_device | Preferred login device of the customers in the account |

***Detailed Analysis of Customer Churn Prediction***