Telecom retention and churn prediction are important aspects of managing customer relationships in the telecommunications industry. Churn occurs when customers stop using the services provided by a telecom company and switch to a competitor. To prevent churn and keep customers, telecom companies use churn prediction models. These models analyze different factors like customer behavior, how they use the services, their personal details, and the quality of the services they receive. Advanced techniques like machine learning and predictive modeling to identify customers who are likely to stop using the services. By identifying possible churners in advance, telecom companies can take action to keep these customers.

The following are the reasons we need to manage churn

1. Churn is a key driver of EBITDA margin and an industry wide challenge
2. A churned customer provides less or no revenue and increases competitor market share.
3. It costs up to 5 times as much for a service provider to acquire a new subscriber as to retain an existing one.

Different type of churn in telecom industry

1. Tariff plan churn - Example 50 dollars to 30 dollars monthly plan change
2. Service churn - Weekly/Monthly subscription
3. Product churn - Postpaid to prepaid
4. Usage churn - Inactivity or zero usage
5. Subscriber Churn - port out to competition

Decision cycle of a subscriber:-

I am a mobile Customer - and I haven’t thought about churning because

- Scenario 1 - Because it’s too complex or don’t have time or it’s not worth it - inert subscriber
- Scenario 2 - Because my operator is the best - unconditionally loyal customer
- Scenario 3 - Locked in my contract - Locked in Subscriber

I am a mobile Customer - and I have thought about churning because

- Scenario 1 - I found a better offer - Conditional churner
- Scenario 2- My needs have changed - Lifestyle Migrator
- Scenario 3 - Because I am not satisfied - Unsatisfied Churner

Key drivers that influence churn

1. Handset Loss/Upgrade
2. Cost of Service/Competitor pricing
3. Network Quality

4. Customer Care Quality

Key drivers for subscriber loyalty

1. Offer and services
2. Price
3. Quality of products and services
4. Quality of customer service
5. Length of contract period
6. Perception of telecom brand
7. Marketing programmes and campaigns

Data science led approach to manage churn

1. Capture and analyze
    a. Business understanding
    b. Identify data requirements and explore data availability
    c. Request and extract data required to build a model
    d. Aggregate, clean and standardize data in desired format for model
2. Report and Predict
    a. Business analysis of standardized data
    b. Business model design
    c. Development and implementation of predictive model
3. Engage and Act
    a. List of churn drivers/ KPI's for tracking and monitoring
    b. A generated list of recommended subscribers for targeted churn campaigns
    c. Recommendations on monthly churn initiatives

We use the telco customer dataset from kaggle, with the following columns

1. customerID: Unique identifier for each customer.
2. gender: The gender of the customer.
3. SeniorCitizen: A binary indicator (1 or 0) representing whether the customer is a senior citizen or not.
4. Partner: A binary indicator (1 or 0) indicating whether the customer has a partner or not.
5. Dependents: A binary indicator (1 or 0) showing whether the customer has dependents or not.
6. tenure: The number of months the customer has been with the mobile service provider.
7. PhoneService: A binary indicator (1 or 0) representing whether the customer has phone service or not.
8. MultipleLines: A categorical variable indicating the customer's multiple lines subscription status (e.g., 'Yes,' 'No,' or 'No phone service').

9. InternetService: The type of internet service the customer has (e.g., 'DSL,' 'Fiber optic,' or 'No internet service').
10. OnlineSecurity: A categorical variable indicating the customer's online security subscription status (e.g., 'Yes,' 'No,' or 'No internet service').
11. OnlineBackup: A categorical variable representing the customer's online backup subscription status (e.g., 'Yes,' 'No,' or 'No internet service').
12. DeviceProtection: A categorical variable indicating the customer's device protection subscription status (e.g., 'Yes,' 'No,' or 'No internet service').
13. TechSupport: A categorical variable representing the customer's tech support subscription status (e.g., 'Yes,' 'No,' or 'No internet service').
14. StreamingTV: A categorical variable indicating the customer's streaming TV subscription status (e.g., 'Yes,' 'No,' or 'No internet service').
15. StreamingMovies: A categorical variable representing the customer's streaming movie subscription status (e.g., 'Yes,' 'No,' or 'No internet service').
16. Contract: The type of contract the customer has (e.g., 'Month-to-month,' 'One year,' or 'Two year').
17. PaperlessBilling: A binary indicator (1 or 0) showing whether the customer has opted for paperless billing or not.
18. PaymentMethod: The customer's preferred payment method.
19. MonthlyCharges: The amount charged to the customer on a monthly basis.
20. TotalCharges: The total amount charged to the customer over their tenure.
21. Churn: A binary variable (1 or 0) indicating whether the customer has churned or not.

EXPLORATORY DATA ANALYSIS

Step 1: Identify the ratio of churners

Based on our EDA we have approx 27% churners and 73% are active
On further analysis we see that the number of churners and non churners is -

Non churner 5174
Churner 1869

This is a highly imbalanced dataset

We then analyze for missing values.

As a general thumb rule

1. For features with less missing values - we can use different imputation technique to predict the missing values
2. For features with very high number of missing values - it is better to drop those columns as they give very less insight on analysis

Step 2: Data Cleaning

Total charges should be a numeric amount. Lets convert it to numerical data type and then analyze the total number of null values. We now see that total charges has 11 null values, but since the null values are so less, we can drop it

Divide customers into bins based on tenure e.g., for tenure < 12 months: assign a tenure group if 1-12, for tenure between 1 to 2 yrs, tenure group of 13-24 and so on. This will help simplify our analysis.

Subsequently we can drop customerID and tenure.

Step 3: Data Exploration

1. Plot distribution of individual predictors by churn
2. Convert Target variable ‘Churn’ in a binary numeric variable i.e., Yes=1; No=0
3. Convert all the categorical variables into dummy variables
4. On plotting a relationship between Monthly Charges and Total Charges we see that Total charges increase as Monthly Charges increase - as expected
5. We get an **insight : Churn is high when Monthly charges are high**
6. We get a **surprising insight: Higher Churn at lower total charges**
7. Build a corelation of all predictors with churn
   Derived Insight:

   a. **HIGH** Churn seen in case of **Month to month contracts**, **No online security**, **No Tech support**, **First year of subscription** and **Fibre Optics Internet**
   b. **LOW** Churn is seens in case of **Long term contracts**, **Subscriptions without internet service** and **The customers engaged for 5+ years**
   c. Factors like **Gender**, **Availability of PhoneService** and **# of multiple lines** have almost **NO** impact on Churn