Capstone Project – DTH / E-Commerce Customer Churn

- NIMESH MARFATIA

Capstone Journey & Assessment:

Capstone Journey:

1st Mentor Connect – 20 Nov

PN1 Submission

2nd Mentor Connect – 4 Dec

PN2 Submission

3rd Mentor Connect – 18 Dec

Capstone Presentation

Final Report Submission

Assessment:

STAGES	GRADES
Project Notes – I	20
Project Notes – II	20
Capstone Presentation	20
Final report submission	40
Total Marks	100

Today’s Agenda

1) Discuss the business context.

2) Understanding Data Dictionary and Data Set

3) Data cleaning and pre - processing (like outlier treatment, missing value treatment etc.)

4) How to generate insights from EDA?

5) Discuss about any finer nuances that could be used to generate insights.

Project Notes -1

Criteria	Pts
1. Problem Understanding Defining problem statement, Need of the study/project, Understanding business/social opportunity	4
2. Data Report Understanding how data was collected in terms of time, frequency and methodology, Visual inspection of data (rows, columns, descriptive details), Understanding of attributes (variable info)	2
3. Exploratory Data Analysis Univariate, Bivariate Analysis, Missing Value, Outlier Treatment, Correlation, Multi collinearity, VIF, etc	10
4. Insights from EDA Data is Balanced, What to do?, Clustering or any other insights about data	4

What do we mean by Customer Churn

- Customer churn is when the customers either switch from their service provider to its competitor or stop using the service altogether.
- As per existing research, it costs five to six times more to acquire customers than to retain existing customers.
- This emphasizes the importance of managing churn by any organization.
- Since churn is the antithesis of retention, it not only affects the size of your customer base, but directly impacts your customer lifetime value.

Effects of Customer Churn

- What are the reasons for customer churn?
  - Customer Service is not good
  - Better Subscription Rates
  - Better Offers – Value Added
  - Product is not stable
  - Cannot reach Customer Service, etc
- What are the effects of customer churn?
  - Affects Profitability
  - Future Business
  - Brand Image
  - Higher cost of Acquisition
  - Loss in Market Share, etc

Problem Statement Given:

- The Company is facing a lot of competition in the current market and it has become a challenge to retain the 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 thing because 1 account can have multiple customers. hence by losing one account the company might be losing more than one customer.
- 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 be very clear on the campaign offer because your recommendation will go through the revenue assurance team.
- If they find that you are giving a lot of free (or subsidized) stuff thereby making a loss to the company; they are not going to approve your recommendation. Hence be very careful while providing campaign recommendation.

Business Understanding, Objective & Scope

- We have a data set from leading online E-Commerce company which have multiple attributes and are deciding factor to know the churn rate of the customers so that the company can provide promotional offers to build a good relation with the customers and in turn increase their revenue.

- Here the company's motive is to decrease the Customer's churn away rate by offering good deals and cashbacks.

- The dataset requires thorough analysis so that we can find out the essential attributes driving the churn rate of customers.

- Identify similar profile and behavior of customers who churn and create offers to retain them.

- The study of data will help to provide proper insights to the E-commerce / DTH company so that they can retain the right customers and accordingly provide some luring promos to make them their permanent customers.

What type of problem is it?

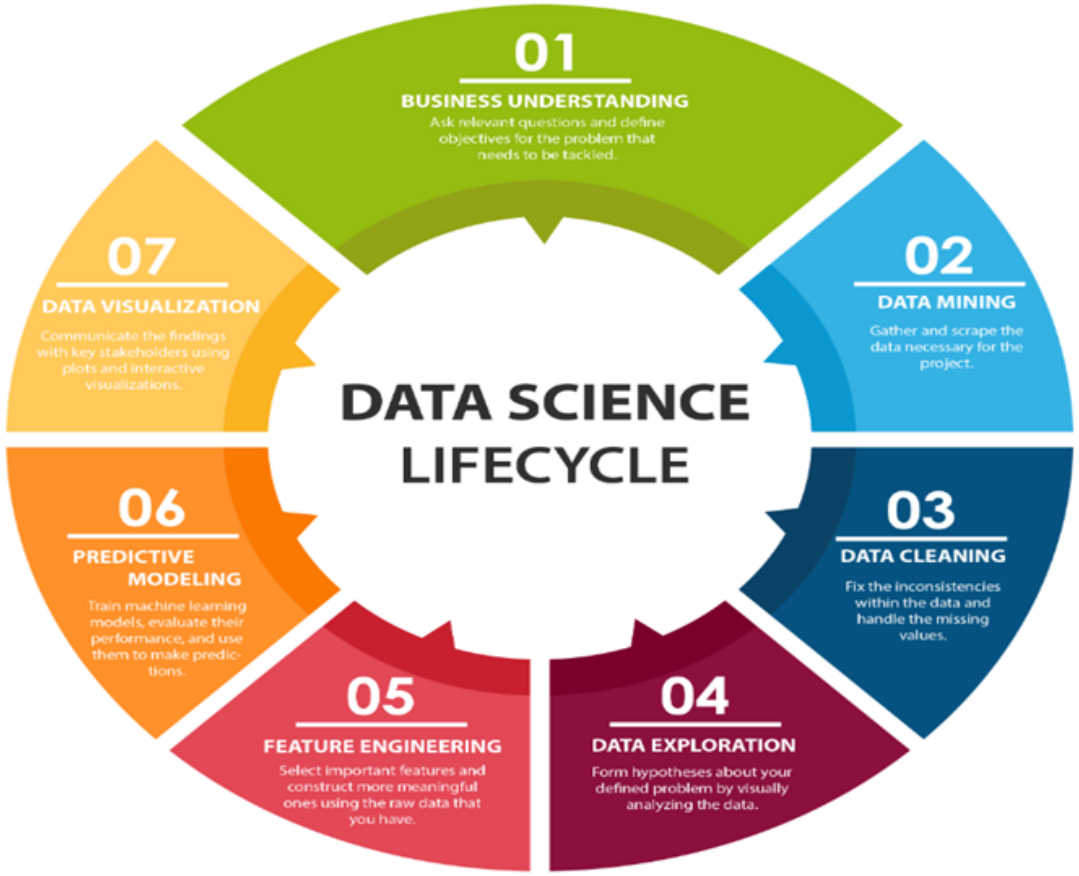
Supervised learning:

- You train the **model** using data which is well labelled.
- Accomplishing a task by providing **training**, input and output patterns to the systems.

Classification type:

- It specifies the class to which data elements belong to and is best used when the output has finite and discrete values.

Modelling Steps



Data Gathering & Variable Type

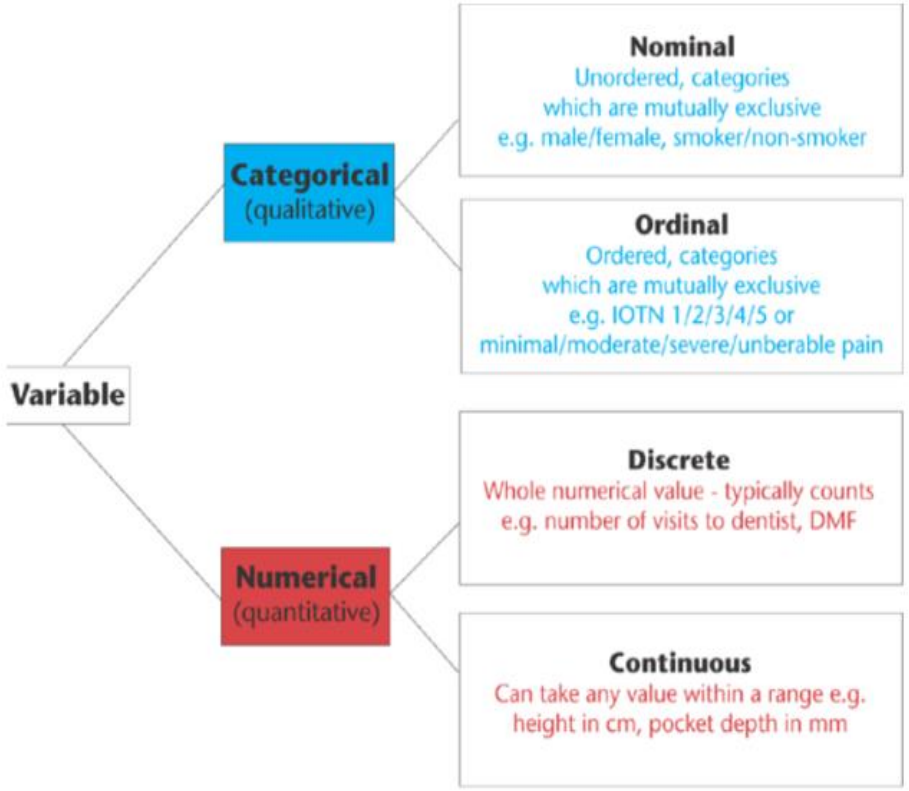
We have 19 variables & 11260 rows of data for building a model which can describe the churn rate for the company.

Here the target variable is Churn which is dependent on other 18 variables.

Scope is fixed as to identify the customers who are going to churn by developing the end to end models on lead indicators.

Quantitative: Numerical **Variables** that have are measured on a numeric or **quantitative** scale.

A **qualitative variable**, also called a categorical **variable**, are variables that are not numerical.



Problem Understanding

Defining problem statement

Need of the study/project

Understanding business/social opportunity

Please Note:

- Cover each point, do not miss any point in the report.
- Write your observations / understandings on each point.

Data Report

- Understanding how data was collected in terms of time, Frequency and methodology
- Visual inspection of data (rows, columns, descriptive details)
- Number of Rows & Columns (Variables)
- Understanding of attributes (variable info, renaming if required)

Data Dictionary – Variable Description:

Variable Description

Variable	Description
AccountID	account unique identifier
Churn	account churn flag (Target)
Tenure	Tenure of account
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
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
Login_device	Preferred login device of the customers in the account

EDA – Exploratory Data Analysis

This is an important step in the complete Data Analysis and Model Development cycle. Let's look at some of the important activities to be performed in this phase

Univariate analysis

- Describe the data and find patterns that exist within it.

Bivariate analysis (relationship between different variables , correlations)

- Find out if there is a relationship between two different variables

Removal of unwanted variables

- Drop the features you do not find related to the Target variable.

EDA – Exploratory Data Analysis

Data Balancing

- Churn Variable: **Yes 1896 (16.83%), No 9364 (83.16%)**.
- Instances of one of the two classes is higher than the other, in another way, the number of observations is not the same for all the classes in a classification dataset.
- Oversampling using SMOTE (Synthetic Minority Over-sampling Technique).
- Under sampling using K-means algorithm

Correlation Test

- Check Multi Collinearity
- As there are multiple features in this data set which have a high correlation, it is important to handle multicollinearity.
- Dropping the features because of high correlation should be used as a last option if nothing really works out.

EDA – Exploratory Data Analysis

Missing Values:

- Numeric Variable – Mean / Median (if outliers are present)
- Categorical Variable – Most Common Class / Unknown Class
- If a feature has more than 15-20% missing value (15% is used as a norm, imputation of more than 15% is not recommended) and such features can be removed from the model building dataset.

Profiling of Continuous and Categorical Columns.

- Objective of this step is to
- Identify how variability in the continuous features is observed with respect to dependent feature
- For categorical feature what is the distribution of dependent variable class across each level.

EDA – Exploratory Data Analysis

Outlier treatment

There are multiple ways in which outliers can be treated. We can either use “Capping” technique or we can remove the outlier values.

Identify the outlier through Box Plot.

Dummy Variables:

Where a categorical variable has more than two categories, it can be represented by a set of dummy variables, with one variable for each category.

Business insights from EDA

Is the data unbalanced? If so, what can be done? Please explain in the context of the business

- Checking data balance is very critical for Classification problems.

Any business insights using clustering

- Clustering is generally done on Unsupervised data.
- If you wish to try you can try and build separate model for each cluster.

Any other business insights

- Add any other insights then you can add the same in this section.

Points to remember

- Understand and make clear the problem statement
- Explain the data dictionary in detail
- EDA should been done thoroughly.
- Do not paste only codes in your report.
- Focus should be on business report/notes.

Thank You