**Business Problem**

An E-commerce company is expanding aggressively and the leadership team has asked us to come up with a plan to increase revenue by 25% in the next quarter without having to cut down on any existing operations.

**1) Problem Definition**

Data Problem:

Identify existing customers, who are likely to buy more products and who are likely to switch to other E- commerce websites, target them directly with specific campaigns.

**2) Hypothesis Generation**

1. Customers buying behaviour related to

* Customer clustering
* Geography
* Demographics
* Channel of first buy
* Time since first buy
* Higher Customer experience
* Product clustering

2. Revenue from customers related to

* + - High Return rate ~ customer experience
    - High conversion rate ~offers, Upsell, cross sell
    - APV

3. Customer Retention related to

* + - Product and delivery quality
    - Competitive pricing
    - Service quality - customer support
    - Transaction security
    - Personalised content

**3) Data Collection/Extraction**

What kind of data do you need based on the above hypotheses?

* Geographic data
* Demographic data
* Channel of first buy
* Time since first buy
* Items purchased
* Price of items
* Customer Ratings
* Personalised Return rate
* Personalised conversion rate
* Personalised APV

Which variables do you require and how would you collect them?

* customer ID, ~ Customer profile
* age ~ Customer profile
* Item description ~ Purchase details
* Location ~ Purchase details
* timestamp of transaction ~ Purchase details
* Quantity purchased ~ Purchase details
* unit price ~ Purchase details
* Channel of first buy ~ Purchase details
* Ratings ~ website

**4) Data Transformation and Exploration**

a) What kind of visualization techniques will you use to explore the data?

* For Univariate Analysis- Line Chart, histogram

We can also use a box plot or violin plot to compare the spread of the variables and provides an insight into outliers.

* Bi Variate Analysis - Bar plot, scatter plot

b) Do you need to transform any variables before proceeding with the analysis?

* Average order Value
* No of repeat order
* Seasonality - product
* CLTV
* Customer Acquisition cost
* APV ranking

**5) Model Building**

1. What is the evaluation metric for your problem?

* email click-through and Return rate
* customer acquisition costs
* customer lifetime value
* Sales conversion rate ~ related to sales target
* Average Purchase value
* Customer Retention rate
* Repeat customer rate
* Cart abandonment rate
* Refund and return rate
* Product revenue by categories, regions
* Profit margins by categories

b) What kind of models will you build?

# **Predictive Analysis for- email marketing, Personalised Product Recommendations, offers, Upsell, Cross sell.**

# **Market Basket Analysis – Product clustering**

# Classification (supervised) and clustering (unsupervised) algorithms are useful for this segmentation

## **Price Optimization** – competitive pricing

# regression modelling- regression model on the various data points around sales, conversion rates, seasonality, product attributes, marketing channels, etc. helps to fix optimal prices for the products.

c) What is your model validation strategy?

Predictive: To test the predictive analysis model, splitting dataset into two sets: training and test datasets. These datasets should be selected at random and should be a good representation of the actual population and Cross-validation of target variable

**6) Model Implementation**

a) Which model, based on the ones you have built, is best suited to your business problem? Is there any trade-off between the accuracy and the interpretability?

**Predictive Analysis seems to be best suited as it will have large application on marketing, Personalised Product Recommendations, offers, Upsell, Cross sell etc.**

I will go for more accurate campaigns rather than focusing on wide interpretability of variables

b) Any specific steps you’ll follow for monitoring your model’s performance?

* Check the accuracy and efficiency of Model created.
* Update training dataset
* Refine model