# **Customer Segmentation using RFM method.**

The entire exercise will be a combination of programming, data analysis, and machine learning. The solution will include,

1. **EDA (Exploratory Data Analysis) to understand the key Metrics in the dataset with visualization.**
   1. Month wise Revenue (Sum of Quantity \* MRP for each invoice SKU)
   2. Monthly Revenue Growth Rate (Rate of change in revenue each month)
   3. Monthly Active Customers (number of customers billed at least once in a month)
   4. Monthly Invoice Count (number of invoices with at least 1 positive quantity sold)
   5. Average Revenue per Invoice
   6. New Customer Ratio (customer that billed for the first time in a month, and has not done in past)
   7. Monthly Retention Rate (Ratio of retained customers from previous month vs total active customers)
2. **Customer Segmentation**

Customer segmentation is the tagging and grouping of customers with shared characteristics like age, industry, gender, etc. With customer segmentation, you can easily personalize your marketing, service, and sales efforts to the needs of specific groups. The result is a potential boost to customer loyalty and conversions.

You can also segment customers based on how much they spend, how often, and what products. This is more behavior focused.

If you want to increase retention rate, you can do a segmentation based on churn probability and take actions. But there are very common and useful segmentation methods as well. We are going to implement one of them to our business: **RFM**.

**RFM stands for Recency - Frequency - Monetary Value.**

Theoretically we will have segments like below:

* **Low Value**: Customers who are less active than others, not very frequent buyer/visitor and generate very low - zero - maybe negative revenue.
* **Mid Value**: In the middle of everything. Often using our platform (but not as much as our High Values), fairly frequent and generates moderate revenue.
* **High Value**: The group we don’t want to lose. High Revenue, Frequency, and low Inactivity.

The Python code, will include following sections,

* 1. Recency
     1. We need to find out the most recent purchase date of each customer.
     2. See how many days they are inactive for.
     3. Apply K-means clustering to assign customers a recency score.
  2. Frequency
     1. We need to find the total number of orders for each customer.
     2. See how frequency looks like in our customer dataset.
     3. Apply K-means clustering to assign customers a frequency score.
  3. Revenue
     1. We need to find the total revenue for each customer.
     2. See how revenue looks like in our customer dataset.
     3. Apply K-means clustering to assign customers a revenue score.
  4. Overall Score
     1. Mean of the above 3 against the group by of sets.
     2. Divide the overall score in 3 sets,
        1. 0 to 2 : Low Value (meaning, Increase Frequency)
        2. 3 to 4 : Mid Value (meaning, Improve Retention + Increase Frequency)
        3. 5+ : High Value (meaning, Improve Retention)

1. **Customer Lifetime Value Prediction**

Customer’s Lifetime Value Prediction is done to sustain and improve on the revenue and profitability. Naturally, these actions make some customers super valuable in terms of lifetime value but there are always some customers who pull down the profitability. We need to identify these behavior patterns, segment customers and act accordingly.

Calculating Lifetime Value is the easy part. First, we need to select a time window. It can be anything like 3, 6, 12, 24 months. By the equation below, we can have Lifetime Value for each customer in that specific time window:

**Lifetime Value: Total Gross Revenue - Total Cost**

This equation now gives us the historical lifetime value. If we see some customers having a very high negative lifetime value historically, it could be too late to act.

* Define an appropriate time frame for Customer Lifetime Value calculation. **Take 6 months for this exercise.**
* Identify the features we are going to use to predict the future and create them. **We will use XGBoost algorithm for this.**
* Calculate lifetime value (LTV) for training the machine learning model.
* Build and run the machine learning model.
* Check if the model is useful.

1. **Churn Prediction**

Till here, we have identified our best customers, their lifetime value predictions, but we also need to retain our best customers. That makes Retention Rate one of the most important metrics here.

Retention Rate is an indication of how good is your product market fit (PMF). If your PMF is not satisfactory, you should see your customers churning very soon. One of the powerful tools to improve Retention Rate (hence the PMF) is Churn Prediction. By using this technique, you can easily find out who is likely to churn in the given period.

Steps:

1. Feature Engineering
2. Effect of Features on Retention Rate using Logistic Regression
3. Classification model using XGBoost