**Assignment 2**

In Today’s time, there is a fierce competition in retail industry in understanding their customers better, especially their most profitable customer groups and the groups that have the biggest potential to become such and how to retain these groups.

Retail data is growing exponentially in variety, volume, velocity & value with each passing year.

Smart retailers are well aware that this data can be utilized and eventually holds the prospective for profit. As a result, these retailers are becoming more conscious about utilization of data and information kept in their repositories, so they can integrate and analyze these large volumes of data to come up with results that can support the quality of their decision making, in order to stay at a competitive advantage and to increase profits.

In this Assignment, you have to address the issue of churn prediction and customer retention in retail industry. This would not help the retailers to plan out retention strategies for their most profitable group of customers but to also design effective marketing programs for the future. In this way companies can retain their valued customers and can also make cost effective marketing strategies which are solely targeted to right customer base.

Consumer churn occurs when a buyer discontinues his or her dealings with an organization. In retail, a buyer is considered as churned once his/her last transaction outdates a particular amount of time. Once a profitable consumer becomes a churn, the loss experienced by the organization is not just the lost profits but also the cost which would be required to make new marketing plans and strategies to attract a new customer base.

**Data summary:**

A food store in Pakistan has provided data set for this project dated between September 17, 2014, and October 26, 2014 (6 weeks approx.). The dataset involves only three columns. Total rows are 1058198 and each row represents the purchasing done by customers in each day.

You have to make a new dataset which contains features which are given in the excel file:

**\*** **Reference date** is the last date of 5th week.

**Task:**

▪ Preprocess the dataset and make a feature set (provided above).

▪ Calculate the week with the highest earning?

▪ Identify the most valued customer?

▪ Categorize the customers into 3 groups (i.e., Poor, Mediocre, Rich)

▪ Make 3 more graphs/insights other than these.

You also have to submit original jupyter notebook file and generated features.csv file into the Google classroom