**EDA and Visualization**

1)How many rows and attributes in the dataset?

The data set we have choosen have nearly 1000 rows and 23 columns.

2) How many missing data and outliers?

The data choosen doesn’t have any outliers but has few outliers which has to be taken care of

3) Any inconsistent, incomplete, duplicate or incorrect data?

The dataset choosen doesn’t have any duplicate data

4) Are the variables correlated to each other?

Yes we many variables that are correlated to each other like attribute total transaction amount and total transaction count and many more as described in the report 1.

5)Are any of the preprocessing techniques needed?

The dataset did not require any preprocessing

6) Do we get any insights from histograms/ bar charts/ line plots, etc.?

Yes it helped in visualization and to find the relationship between different attributes

What do the columns mean?

**CLIENTNUM** : Unique identifier for the customer holding the account.

**Attrition\_Flag**: Internal event (customer activity) variable - if the account is closed then 1 else 0.

***Demographic variables***:

* + **Customer\_Age**: Customer's Age in Years.
  + **Gender**:M=Male, F=Female.
  + **Dependent\_count**:Number of dependents.
  + **Education\_Level**: Educational Qualification of the customer (example: high school, college graduate, etc.).
  + **Marital\_Status**: Married, Single, Divorced, Unknown.
  + **Income\_Category**: Annual Income Category of the customer (< 40K,40K,40K - 60K, 60K−60K−80K, 80K−80K−120K, >).

***Product variables***:

* + **Card\_Category**: Type of Card (Blue, Silver, Gold, Platinum).
  + **Months\_on\_book**: Period of relationship with bank.
  + **Total\_Relationship\_Count**: Total no. of products held by the customer.
  + **Months\_Inactive\_12\_mon**: No. of Months in the last 12 months.
  + **Contacts\_Count\_12\_mon**: No. of Contacts in the last 12 months.
  + **Credit\_Limit**: Credit Limit on the Credit Card.
  + **Total\_Revolving\_Bal**: Total Revolving Balance on the Credit Card.
  + **Avg\_Open\_To\_Buy**: Open to Buy Credit Line (Average of last 12 months
  + **Total\_Amt\_Chng\_Q4\_Q1**: Change in Transaction Amount (Q4 over Q1).
  + **Total\_Trans\_Amt**: Total Transaction Amount (Last 12 months).
  + **Total\_Trans\_Ct**: Total Transaction Count (Last 12 months).
  + **Total\_Ct\_Chng\_Q4\_Q1**: Change in Transaction Count (Q4 over Q1).
  + **Avg\_Utilization\_Ratio**: Average Card Utilization Ratio.