Dataset Exploration and Analysis

1) How many rows and attributes in the dataset?

Answer: The data set we have chosen have nearly 1000 rows and 23 columns.

2) How many missing data and outliers?

Answer: The data chosen doesn't have any missing data but has few outliers which needs to be taken care of

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

Answer: The dataset chosen doesn't have any duplicate or incomplete data

4) Are the variables correlated to each other?

Answer: Yes, we see that many variables are correlated to each other such as attributes total transaction amount and total transaction count and many more as described in the Literature Review.

5) Are any of the pre-processing techniques needed?

Answer: The dataset did not require any pre-processing for the EDA

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

Answer: Yes, plots such as histograms, bar plots etc helped in visualization of the data and to find the relationship between different attributes and their distributions

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.