**Documentation for Credit Card Analysis**

**Introduction:**

Credit card churn prediction is a vital application of data analytics and machine learning in the financial industry. Churn, in this context, refers to the phenomenon where credit card customers close their accounts or stop using them regularly. Predicting credit card churn is essential for financial institutions as it can help them retain profitable customers, reduce customer acquisition costs, and improve overall customer satisfaction

**Before we begin, let me introduce you to business problem.**

A bank manager was disappointed as time to time many people have shut down their credit card services from that bank. and the bank manager wants to find out who is going to leave the bank in future so he can give better offers to those customers to stick with them. So the bank manager collected the data for credit card services and wants to find answers.

**Features:**

**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.

**Customer\_Age :** Demographic variable - Customer's Age in Years.

**Gender :** M=Male, F=Female.

**Education\_Level :** Educational Qualification of the account holder (example: high school, college graduate, etc.)

**Marital\_status :** Married, Single, Divorced, Unknown.

**Income\_Category** : Annual Income Category of the account holder (< $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown)

**Card\_Category :** Type of Card (Blue, Silver, Gold, Platinum)

**Months\_on\_book :** Period of relationship with bank.

**Credit\_limit :**The maximum amount of money a lender will allow you to spend using a particular credit card or revolving line of credit.

**Total\_Revolving\_Bal :** The balance that carries over from one month to next.

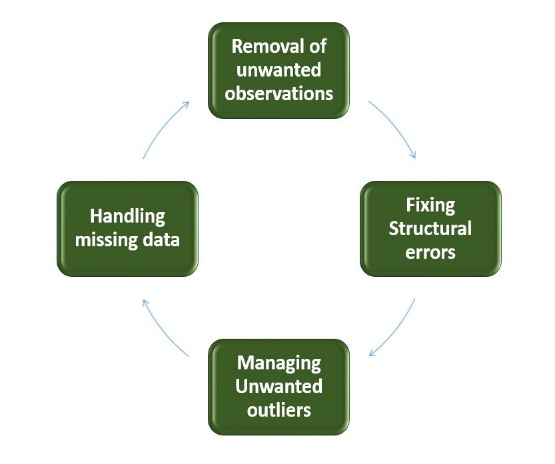
**Avg\_Open\_To\_Buy :** The difference between credit limit and present balance in the account.

**Total\_Trans\_Amt :**The amount that is debited from the balance in connection with your use of card.

**Total\_Trans\_Ct :** Total transactions count by (Last 12 months)

**Avg\_Utilization\_Ratio** : Num Average Card Utilization Ratio.

**Data Cleaning…..**

Data cleaning is a crucial step in the Analysis pipeline, as it involves**identifying and removing any missing, duplicate, or irrelevant data.**

Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_1

Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_2

* As already mentioned in the problem statement that the last two columns are not useful at all. so we are going to **drop** it before proceeding to next step.
* **The dataset is free from missing values and duplicate rows.**
* **Unknown values are removed from Income\_category column.**