CREDIT-CARD SEGMENTATION

Applied Data Science Capstone Project

author: Zoltan Farkas

# Introduction

Segmentation in marketing is a technique used to divide customers or other entities into groups based on attributes such as behaviour or demographics. It is useful to identify segments of customers who may respond in a similar way to specific marketing techniques such as email subject lines or display advertisements.

As it gives businesses the ability to tailor marketing messages and timing to generate better response rates and provide improved consumer experiences. In this project, the definition of a marketing strategy using a machine learning technique requires the development of a customer segment.

The goal of this analysis report is to discover the Customer Segmentation of a bank, by looking through their behavior/profile while using Credit Card.

Hopefully, I can get a clear segmentation of the customer, so I can deploy effective marketing campaign or sales promotion to the targeted costumer.

# Data acquisition and cleaning

The sample dataset summarizes the usage behavior of about 8000+ active credit card holders during 6 months. The file is at a customer level with 18 behavioral variables.

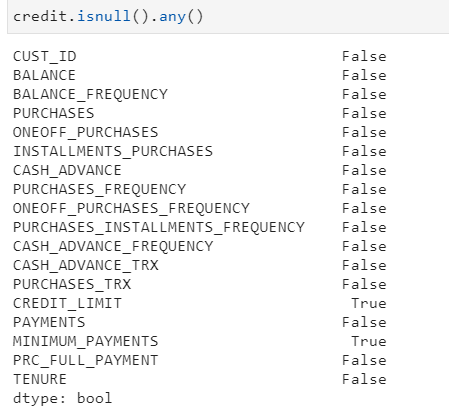
## DATA DICTIONARY:

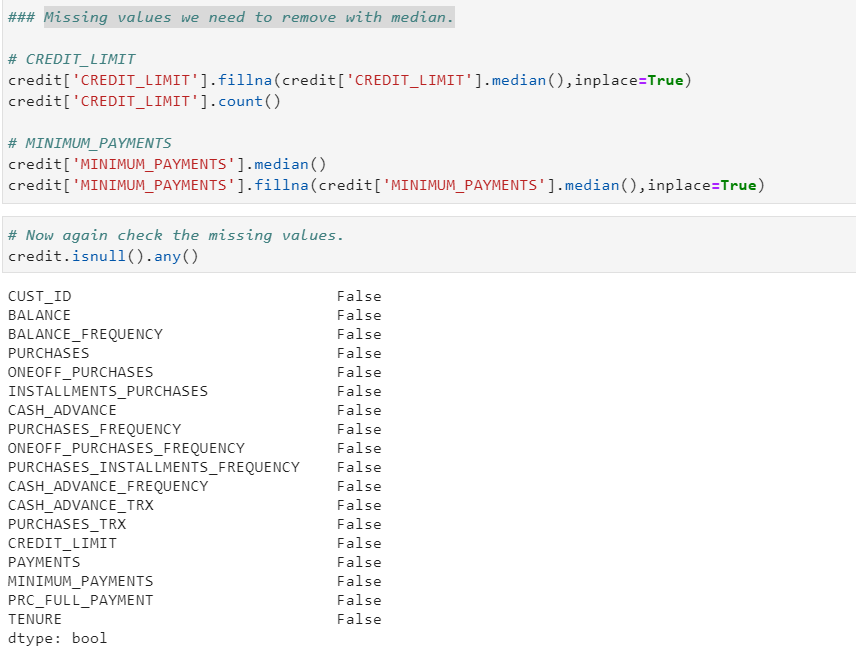
* CUST\_ID: Identification of Credit Card holder (Categorical)
* BALANCE: Balance amount left in their account to make purchases
* BALANCE\_FREQUENCY: Ratio of last 12 months with balance
* PURCHASES: Amount of purchases made from account
* ONEOFF\_PURCHASES: Total amount of one-off purchases
* INSTALLMENTS\_PURCHASES: Total amount of installment purchases
* CASH\_ADVANCE: Cash in advance given by the user
* PURCHASES\_ FREQUENCY: Frequency of purchases (Percent of months with at least one purchase)
* ONEOFF\_PURCHASES\_FREQUENCY: Frequency of one-off-purchases
* PURCHASES\_INSTALLMENTS\_FREQUENCY: Frequency of installment purchases
* CASH\_ADVANCE\_ FREQUENCY: Cash-Advance frequency
* AVERAGE\_PURCHASE\_TRX: Average amount per purchase transaction
* CASH\_ADVANCE\_TRX: Average amount per cash-advance transaction
* PURCHASES\_TRX: Average amount per purchase transaction
* CREDIT\_LIMIT: Credit limit
* PAYMENTS: Total payments (due amount paid by the customer to decrease their statement balance) in the period
* MINIMUM\_PAYMENTS: Total minimum payments due in the period.
* PRC\_FULL\_PAYMEN: Percentage of months with full payment of the due statement balance
* TENURE: Number of months as a customer

## Data cleaning

### Missing values

Missing data is an everyday problem that a data professional need to deal with. Though there are many articles, blogs, videos already available, I found it is difficult to find a concise consolidated information in a single place. That’s why I am putting my effort here, hoping it will be useful to any data practitioner or enthusiast.

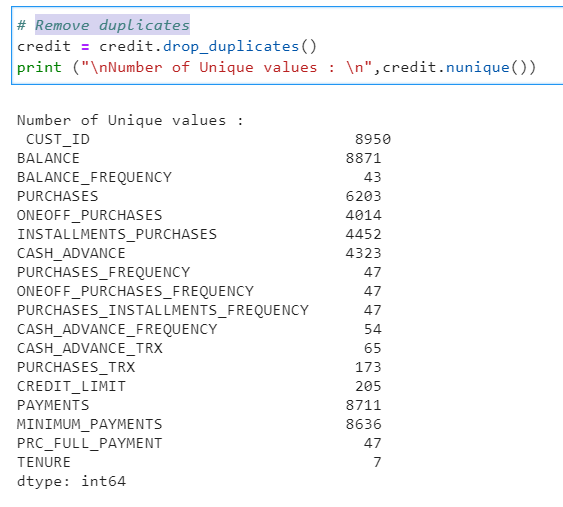
****

**Solution:** Missing values we need to remove with median.

### Duplication

"Duplication" just means that you have repeated data in your dataset. This could be due to things like data entry errors or data collection methods. For example, if you're using a web scraper you may happen to scrape the same webpage more than once, or the same information from two different pages. Whatever the reason, deduplication can lead you to make incorrect conclusions by leading you to believe that some observations are more common than they really are.

**Solution:** Remove duplicates

****

# Exploratory Data Analysis

In statistics, exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.