

Customer Intelligence and Big Data

Alan Rijnders and Lorenzo Severi

11/4/2021

We start reading in the data to perform the analysis

```
#read in data
data <- read.csv("ch.csv")
```

We now start exploring the dataset

```
#summary of the dataset
summary(data)
```

```
##           X           CLIENTNUM      Attrition_Flag      Customer_Age
##  Min.      :    1      Min.      :708082083      Length:10127      Min.      :26.00
##  1st Qu.: 2532      1st Qu.:713036770      Class :character      1st Qu.:41.00
##  Median : 5064      Median :717926358      Mode  :character      Median :46.00
##  Mean   : 5064      Mean   :739177606                        Mean   :46.33
##  3rd Qu.: 7596      3rd Qu.:773143533                        3rd Qu.:52.00
##  Max.    :10127      Max.    :828343083                        Max.    :73.00
##      Gender      Dependent_count      Education_Level      Marital_Status
##  Length:10127      Min.      :0.000      Length:10127      Length:10127
##  Class :character      1st Qu.:1.000      Class :character      Class :character
##  Mode  :character      Median :2.000      Mode  :character      Mode  :character
##                               Mean   :2.346
##                               3rd Qu.:3.000
##                               Max.    :5.000
##  Income_Category      Card_Category      Months_on_book      Total_Relationship_Count
##  Length:10127      Length:10127      Min.      :13.00      Min.      :1.000
##  Class :character      Class :character      1st Qu.:31.00      1st Qu.:3.000
##  Mode  :character      Mode  :character      Median :36.00      Median :4.000
##                               Mean   :35.93      Mean   :3.813
##                               3rd Qu.:40.00      3rd Qu.:5.000
##                               Max.    :56.00      Max.    :6.000
##  Months_Inactive_12_mon      Contacts_Count_12_mon      Credit_Limit      Total_Trans_Amt
##  Min.      :0.000      Min.      :0.000      Min.      : 1438      Min.      : 510
##  1st Qu.:2.000      1st Qu.:2.000      1st Qu.: 2555      1st Qu.: 2156
##  Median :2.000      Median :2.000      Median : 4549      Median : 3899
##  Mean   :2.341      Mean   :2.455      Mean   : 8632      Mean   : 4404
##  3rd Qu.:3.000      3rd Qu.:3.000      3rd Qu.:11068      3rd Qu.: 4741
##  Max.    :6.000      Max.    :6.000      Max.    :34516      Max.    :18484
##  Total_Trans_Ct      Avg_Utilization_Ratio
##  Min.      : 10.00      Min.      :0.0000
##  1st Qu.: 45.00      1st Qu.:0.0230
##  Median : 67.00      Median :0.1760
##  Mean   : 64.86      Mean   :0.2749
##  3rd Qu.: 81.00      3rd Qu.:0.5030
```

Max. :139.00 Max. :0.9990