**Python II Group Assignment Group D, GMBD 2022**

* **Ahmed Aljeshi**
* **Juan Francisco Balbi**
* **Paula Escusol Entío**
* **Isobel Rae Impas**
* **Paliz Mungkaladung**

## 1) Import basic set of libraries for data loading/EDA

## 2) Import files to be used & display column key

## 3) Initial EDA & scope transformations

### 3.0 Initial EDA

### 3.1 Reduce dataset from loan to account level by removing duplicates and keeping only last loan per account per customer

### 3.2 Preliminary EDA on customer level dataset

## 4) DATA CLEANING AND DATA PREPARATION

### 4.1 Group columns for transformations/cleaning and present transformation/cleaning plan

### 4.2 bin bucket column

### 4.3 remove unecessary columns

### 4.4 tarnsform date columns from string to datetime to DAYS /AGE

## 5) Analyze missingness in customer level data for MNAR/MAR patterns/dataframe split by applying **missing\_not\_at\_random method from** RiskDataFrame class

## 6) Split dataset into corporate/individual segments based on MNAR by applying **missing\_not\_at\_random method from** RiskDataFrame class

## 7) Corporate dataset analysis, transformations, modelling, baseline GINIi and Variable by Variable Risk Based Segmentation Analysis

### 7.1 Corporate: Preliminary EDA

### 7.2 Corporate: Impute MAR values using most frequent value

### 7.3 Corporate: Outlier detection/removal and Scaling

In [25]:

### 7.4 Corporate: dummy encode categoricals for logistic regression

### 7.5 Corporate: EDA

### 7.6 Corporate: Logit regression and Baseline metrics

### 7.7 Corporate Baseline Gini

### 7.8 Corporate Variable by Variable Risk Based Segmentation Analysis

### 7.8.1 Corporate: Analysis Model 1 - Gender M and F

### 7.8.2 Corporate: Full Model vs Seg 1 on Seg 1

### 7.8.3 Corporate: Model vs Seg 2 on Seg 2

### …

### 7.9 Corporate Execution Report

8) Individual dataset analysis, transformations, modelling, baseline GINIi and Variable by Variable Risk Based Segmentation Analysis

### 8.1) Individual: Preliminary EDA

### 8.2) Individual: Impute MAR values using most frequent value

### 8.3) Individual: Outlier detection/removal and Scaling

In [25]:

### 8.4) Individual: dummy encode categoricals for logistic regression

### 8.5 Individual: EDA

### 8.6 Individual: Logit regression and Baseline metrics

### 8.7 Individual Baseline Gini

### 8.8 Individual Variable by Variable Risk Based Segmentation Analysis

### 8.8.1 Individual: Analysis Model 1 - Gender M and F

### 8.8.2 Individual: Full Model vs Seg 1 on Seg 1

### 8.8.3 Individual: Model vs Seg 2 on Seg 2

### …

### 8.9 Individual Execution Report

# 9) Final Corporate/ Individual Execution Summary Repport