

## PROJECT: GERMAN CREDIT DATA

### Credit Risk Assessment

\*\*\*\*\*

### Feature Engeneering

## Aplicando Engenharia de Atributos em Variáveis Numéricas

### 1. Working Directory

#### Configurando o diretório de trabalho

```
setwd("C:/Users/Utilizador/repos/Formacao_cientista_de_dados/big_data_analytics_R_microsoft_azure_machine_learning")
getwd()
```

**Observação:** Este código contém comandos para filtrar e transformar os dados de aluguel de bikes,

Este código foi criado para executar tanto no Azure, quanto no RStudio;

Para executar no Azure, altere o valor da variavel Azure para TRUE;

Ou seja, se o valor for FALSE, o código sera executado no RStudio;

### 2. Data Collection and Transformation

#### Variável que controla a execução do script

```
Azure <- FALSE
```

```
2.1 Data Collect if(Azure){ source("src/ClassTools.R") Credit <- maml.mapInputPort(1) }else{
source("src/ClassTools.R") Credit <- read.csv("credito.csv", header = F, stringsAsFactors = F )
metaFrame <- data.frame(colNames, isOrdered, I(factOrder)) Credit <- fact.set(Credit, metaFrame)

# Balancear o número de casos positivos e negativos Credit <- equ.Frame(Credit, 2) }
```

#### 2.2 Data Transformation

## Transformando variáveis numéricas em variáveis categóricas

```
toFactors <- c("Duration", "CreditAmount", "Age") maxVals <- c(100, 1000000, 100) facNames <- un-  
list(lapply(toFactors, function(x) paste(x, "__f", sep = " "))) Credit[, facNames] <- Map(function(x, y)  
quantize.num(Credit[, x], maxval = y), toFactors, maxVals)
```

**str(Credit)**

## Output

```
if(Azure) maml.mapOutputPort('Credit')
```