BUSINESS UNDERSTANDING

**DATA UNDERSTANDING**

Todas as variáveis estavam com valores em falta excepto o year\_birth

1.1 Tamanho do data frame

1.2 Descrever as variáveis/data frame individualmente

Verificar variáveis

No nome também está implícito o género

Depois criar a nova feature

1.2.1 Variable Analysis Congruence Check

Fazer para todas as variáveis (Longevity identificar que tem valores diferentes)

Ver idade e ver as pessoas com menos de 10 anos

Cada Variável Categórica

Countplot

Cada variável numérica

Boxplot

Kurtosis e Skewness

Histogram

Pairplot APENAS PARA CORRELACIONADAS

Bar Service – Staff

Onlinebooking – bar service

Wifi – onlinebooking

Wifi – bar service

Staff – onlinebooking

foodDrink – confort

Location – Reception schedule

Cleanliness – Pricequality

Cleanliness – Checkout

Resumo do Outlier detection (DBSCAN ou Anomaly Detection)

1.3 Ver os missing values

Dummy variables

Normalização

Usar a moda para substituir ou o KNN imputer

1.4 Verificar duplicados

Verificamos que existe

Justificamos que é improvável pessoas com o mesmo nome avaliar tudo igual e retiramos

Vamos eliminar 3

Verificar pessoas com o mesmo nome mas avaliações diferentes

1.5 Multicolinearity

Data correlation

**DATA PREPARATION**

Prepara análise de variáveis (kurt e skew e correlação)

**MODELLING**

PCA