**Trabalho Final Estatística - MBA em Machine Learning & Artificial Intelligence**

Andrey Ferreira de Almeida - RM: 344545

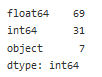
Diego Cohen - RM 344154

Vitor Silva - RM 344669

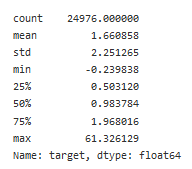
* **O método**
* Foi utilizado um Jupyter Notebook com Python 3.9 e as bibliotecas:
  + Pandas
  + Numpy
  + Matplotlib
  + Train\_test\_split
  + Seaborn
  + Stats
  + StandardScaler
  + Pyplot
  + RandomForestRegressor
  + Permutation\_importance
  + LinearRegression
  + DecisionTreeRegressor
* Foram realizadas as seguintes etapas:
  + Exploração dos dados
    - Linhas e colunas:



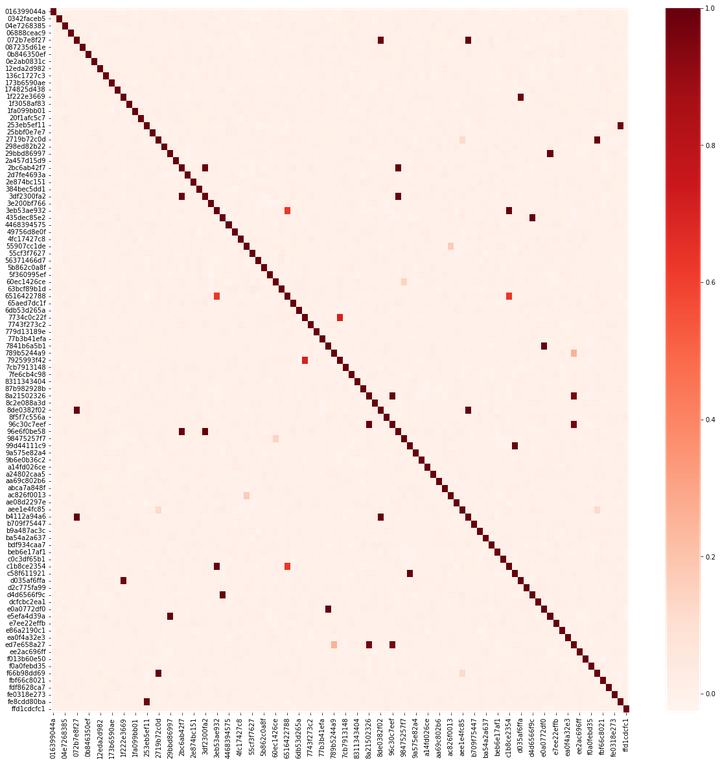
* + - Datatypes:



* + - Target:



* + Separação em x e y
  + Separação em treino e teste
  + Verificação de correlação das variáveis:



* + Dropada uma a cada duas colunas com alta correlação (>0.5)
  + Removidos outliers em 3 desvios-padrão
  + Testada (e não aplicada) dummerização
  + Normalização z-score
  + Classificação de variáveis por RandomForest
  + Treino e teste de modelos:
    - Regração Linear
    - Decision Trees
  + Obtenção do melhor modelo
  + Observação sobre a qualidade do modelo
  + Escrita da variável dependente ‘target’ no dataframe de teste
* **Arquivos no ZIP de entrega**
  + Trabalho Final Estatística - MBA em Machine Learning & Artificial Intelligence - Descritivo
    - Este arquivo
  + train.csv
    - Arquivo de treino fornecido
  + test.csv
    - Arquivo de teste fornecido (sem variavel resposta)
  + entrega.csv
    - Arquivo de teste com variável resposta predita por Regressão Linear
  + Trabalho Final Estatística HTML.html
    - Versão em HTML do Jupyter Notebook, para simples conferência
  + Trabalho Final Estatística.ipynb
    - Script do trabalho