

PUC Minas Virtual

Pós-Graduação Latu Sensu em
Ciência de Dados e Big Data

Rio de Janeiro
2021



Trabalho de Conclusão de Curso

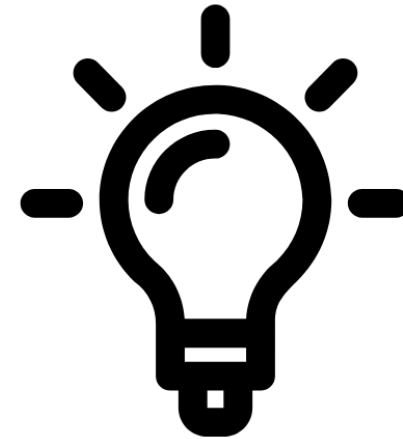
Big Deal: Encontrando hospedagens
com os melhores custos-benefícios.

Fabio Rafael Ladislei de Souza



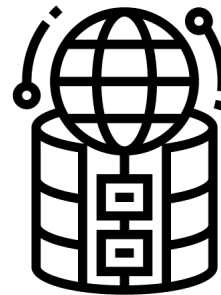
Definição do Problema

Ordenar anúncios do AirBNB por um ranking de custo-benefício, que será calculado com base em modelos preditivos de regressão.

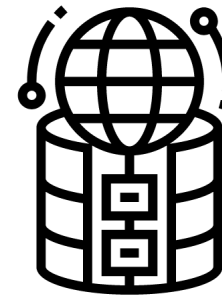


Coleta dos Dados

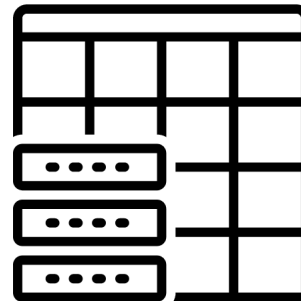
- Leitura dos dados



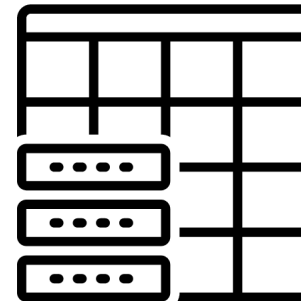
Listings



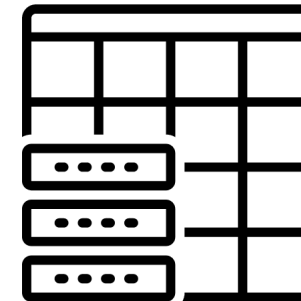
Metro
Quadrado



Anúncios

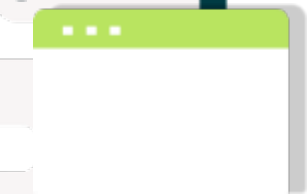


Amenities



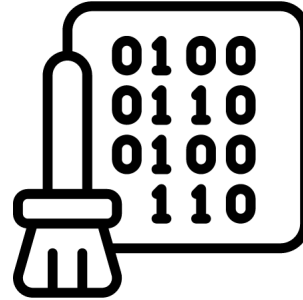
Metro
Quadrado

- Geração dos
dataframes iniciais

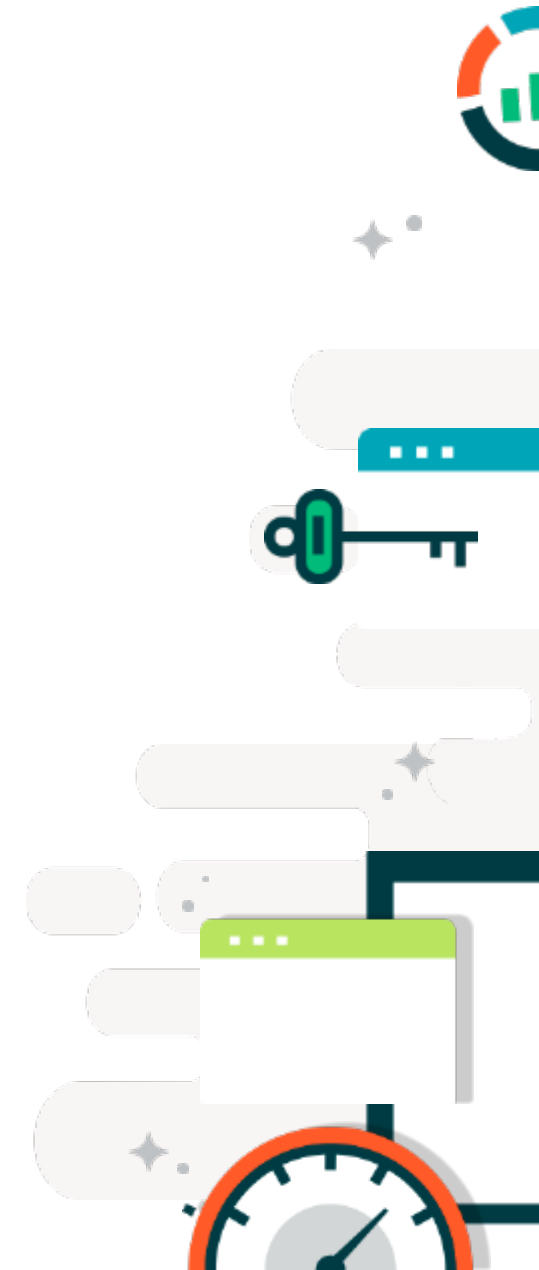
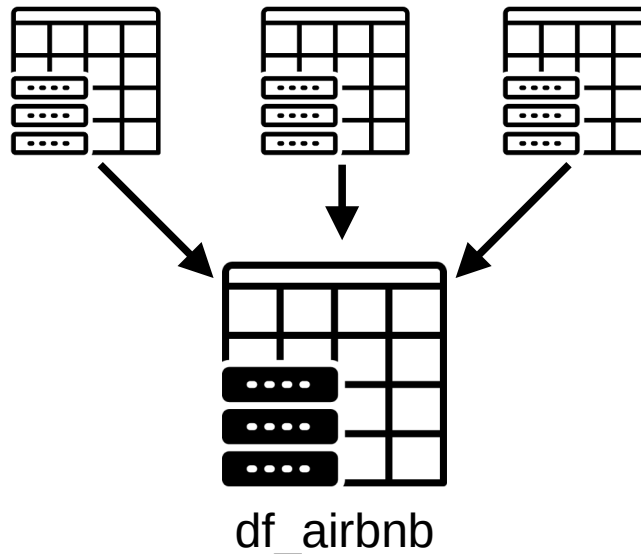


Processamento dos Dados

- Exclusão de registros incompletos

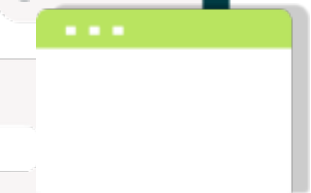
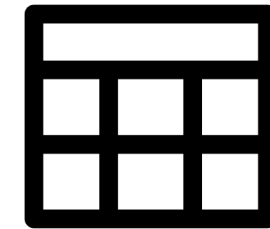
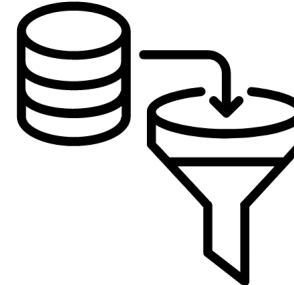
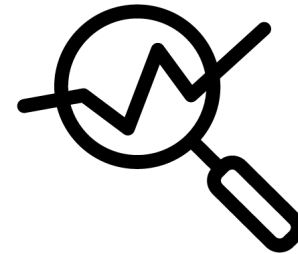


- Agregação dos dataframes



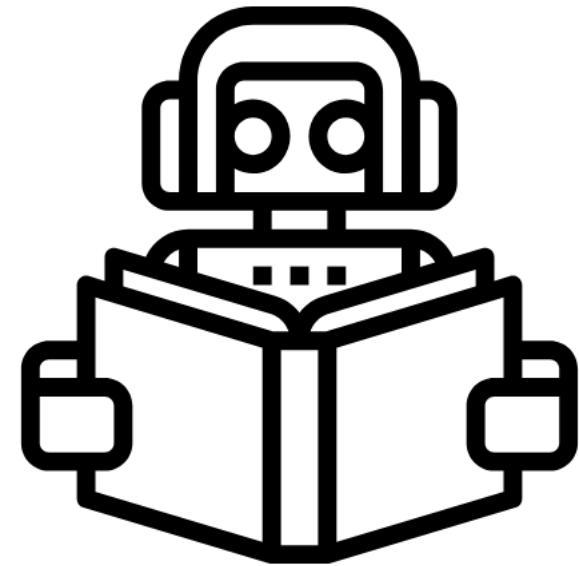
Análise Exploratória

- Exclusão de valores discrepantes
- Análise de correlação
- Seleção de atributos
- Plotagens diversas
- Filtragem dos dados
- Adequação dos dados ao aprendizado de máquina
- Matriz de correlação dos atributos



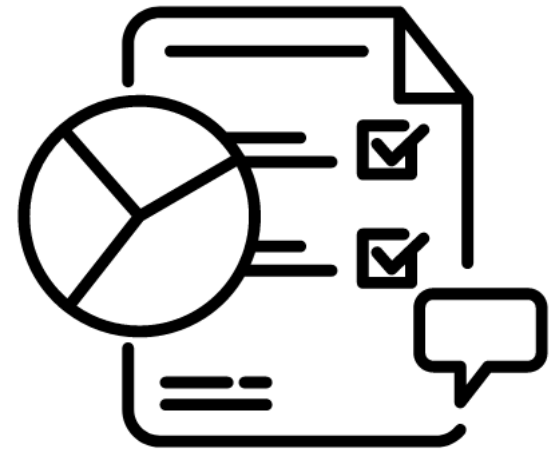
Modelagem Preditiva

- Particionamento dos dados
- Preparação de modelos preditivos
- Definição do algoritmo de base
- Preparação da validação cruzada
- Otimização dos modelos
- Treinamento do modelo



Interpretação dos Resultados

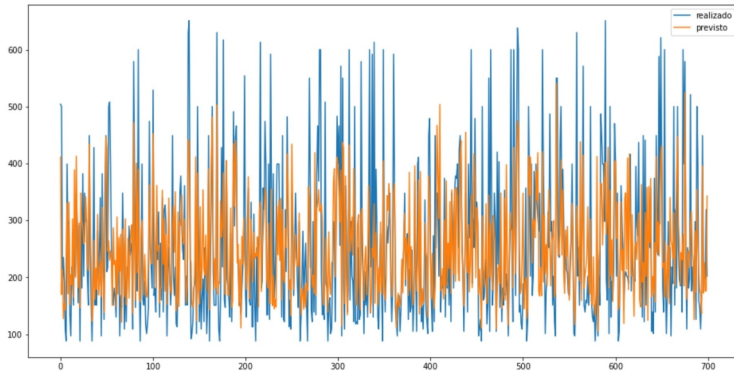
- Análise da performance do modelo
- Teste do modelo
- Visualização das previsões
- Análise dos resultados (previsto x realizado)
- Criação do ranking de custo-benefício
- Visualização dos anúncios ordenados pelo ranking



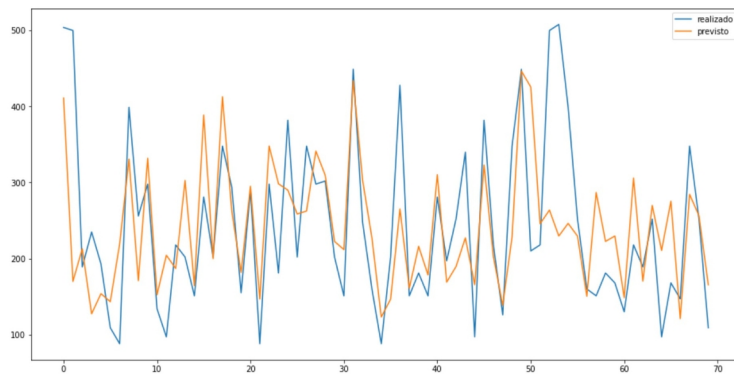
Interpretação dos Resultados

Dados sem ordenação

Todos

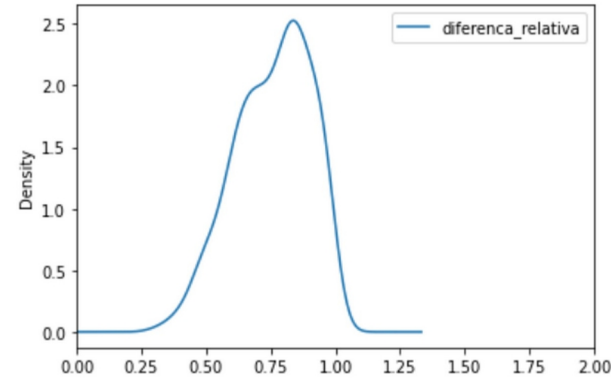


Header



Ranking de custo-benefício

Diferença ideal



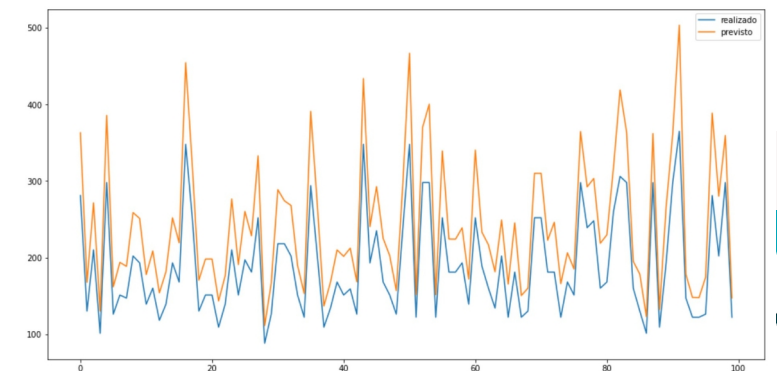
Definição do ranking

	realizado	previsto	diferenca_relative	ranking
id				
10151402	281.0	363.051861	0.773994	0.999747
3532589	130.0	167.622461	0.775552	0.999747
2361521	210.0	271.432251	0.773674	0.999643
22602640	101.0	130.088369	0.776395	0.999473
9903460	298.0	385.705412	0.772610	0.999297
...
3197436	588.0	237.648787	2.474239	0.447645
31960056	420.0	168.947873	2.485974	0.443831
12863588	500.0	169.973076	2.941642	0.295731
30744239	600.0	164.995170	3.636470	0.069900
32655506	592.0	153.704861	3.851537	0.000000

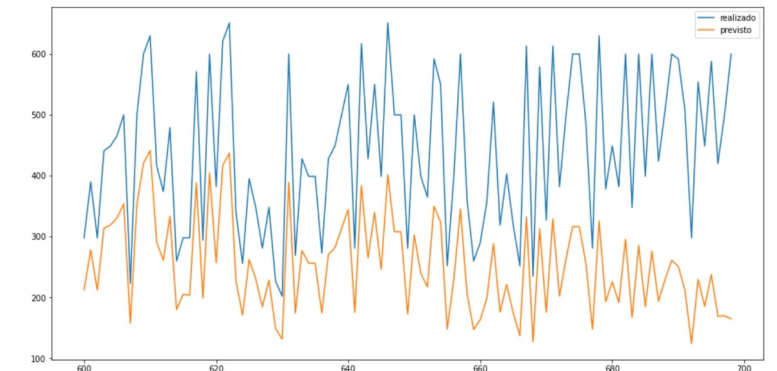
700 rows × 4 columns

Dados ordenados

Melhores



Piores



Apresentação dos Resultados

- Exibição da lista de anúncios ordenada pelo ranking
- Exibição dos dados do anúncio selecionado





Obrigado