

PUC Minas Virtual

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

Rio de Janeiro
2021



Trabalho de Conclusão de Curso

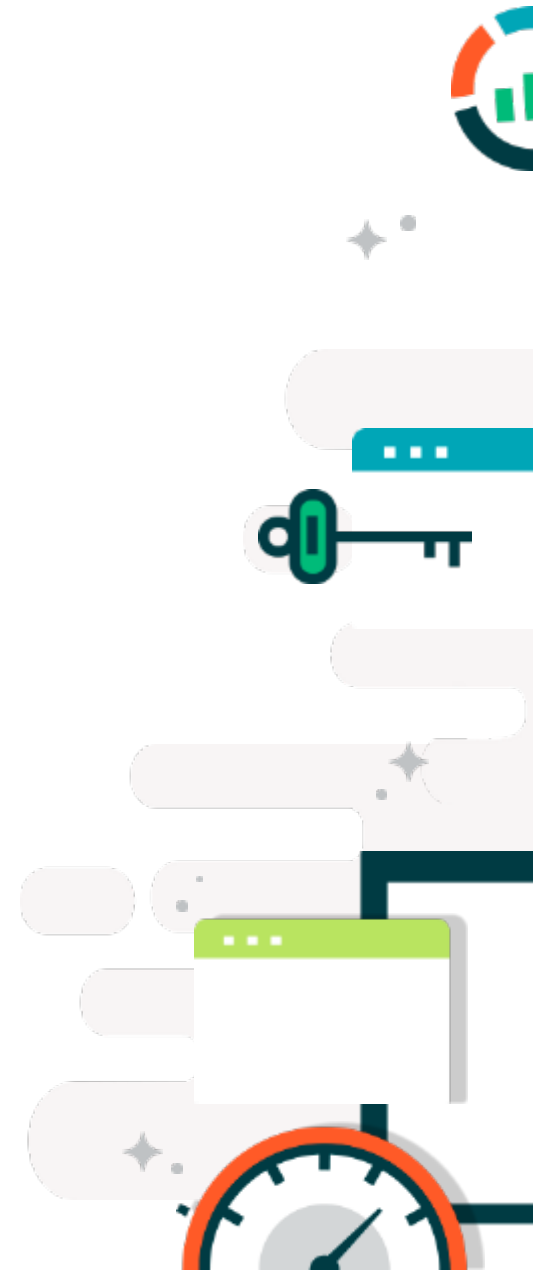
Big Deal: Encontrando hospedagens
com os melhores custos-benefícios.

Fabio Rafael Ladislei de Souza



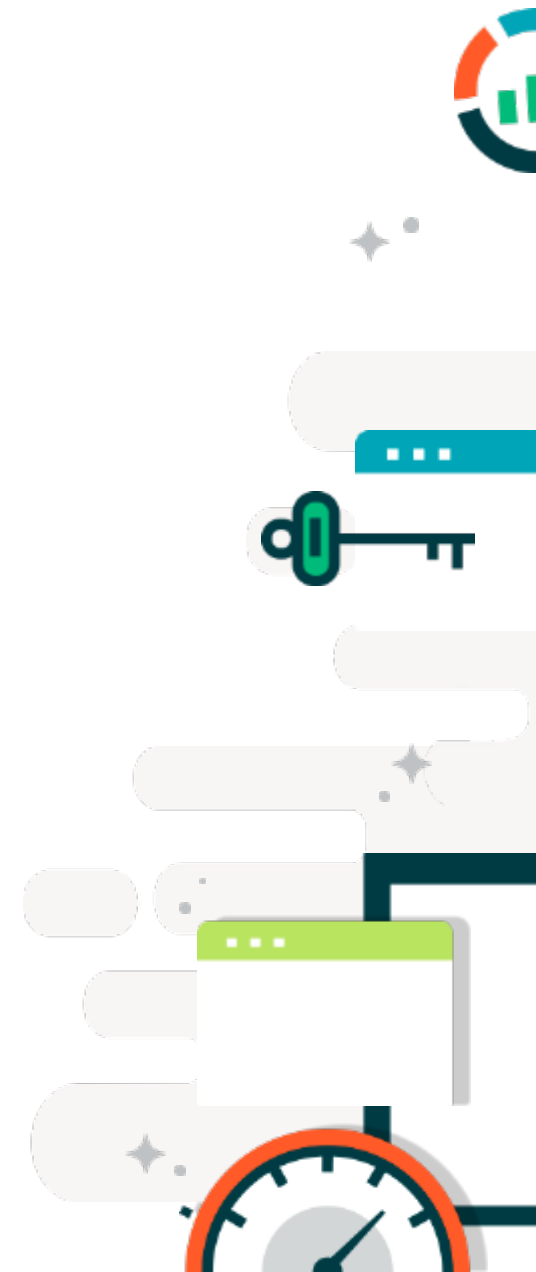
Definição do Problema

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



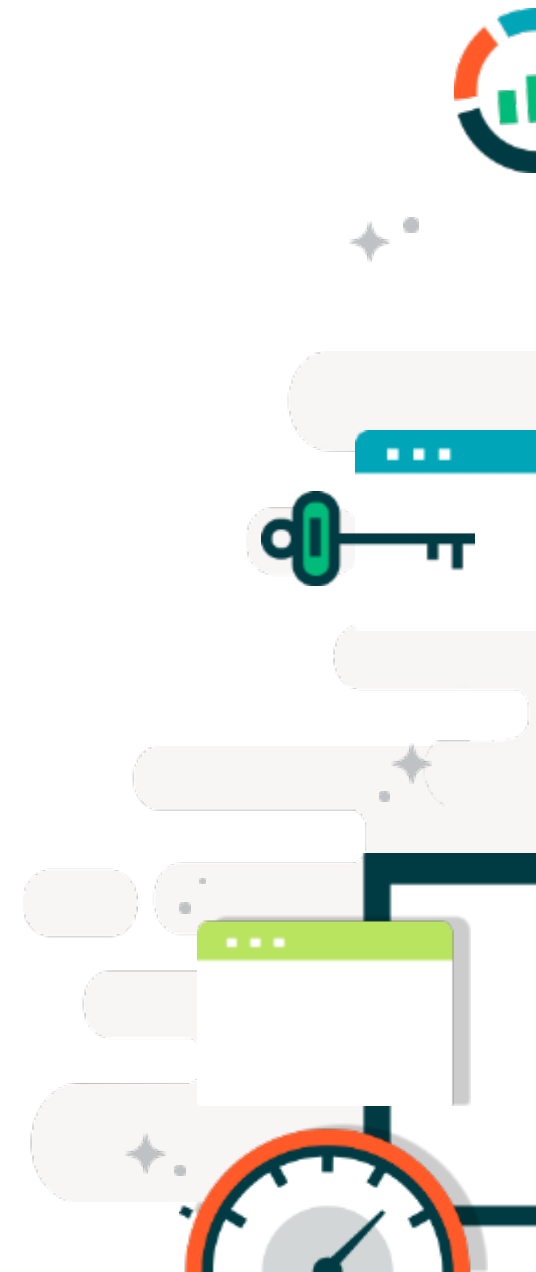
Coleta dos Dados

- Leitura dos dados
- Geração dos dataframes iniciais



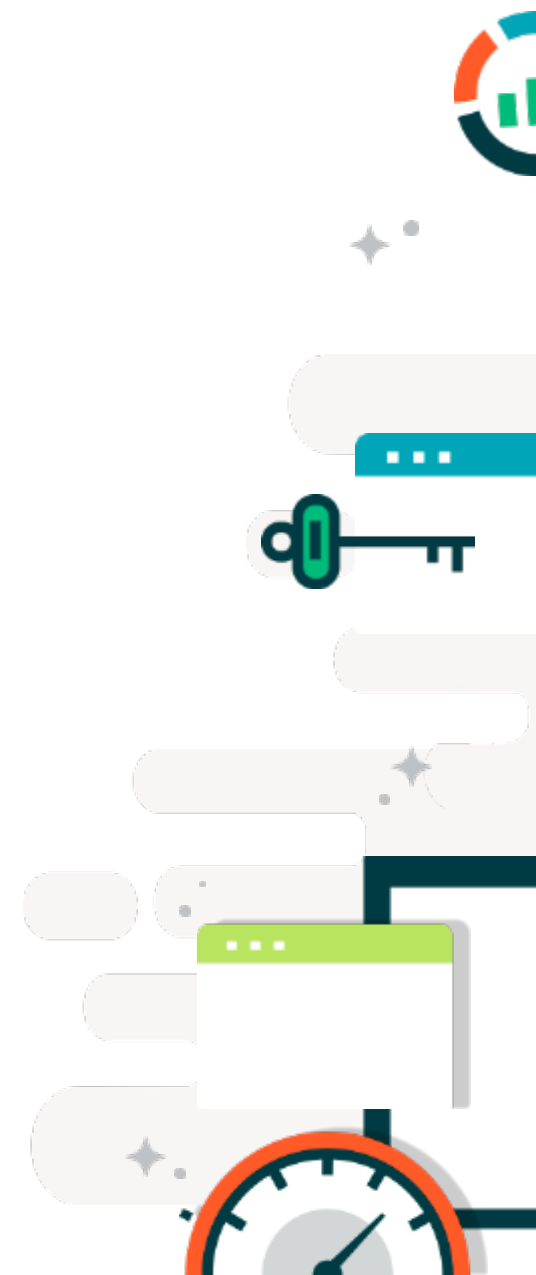
Processamento dos Dados

- Exclusão de registros incompletos
- Agregação dos dataframes



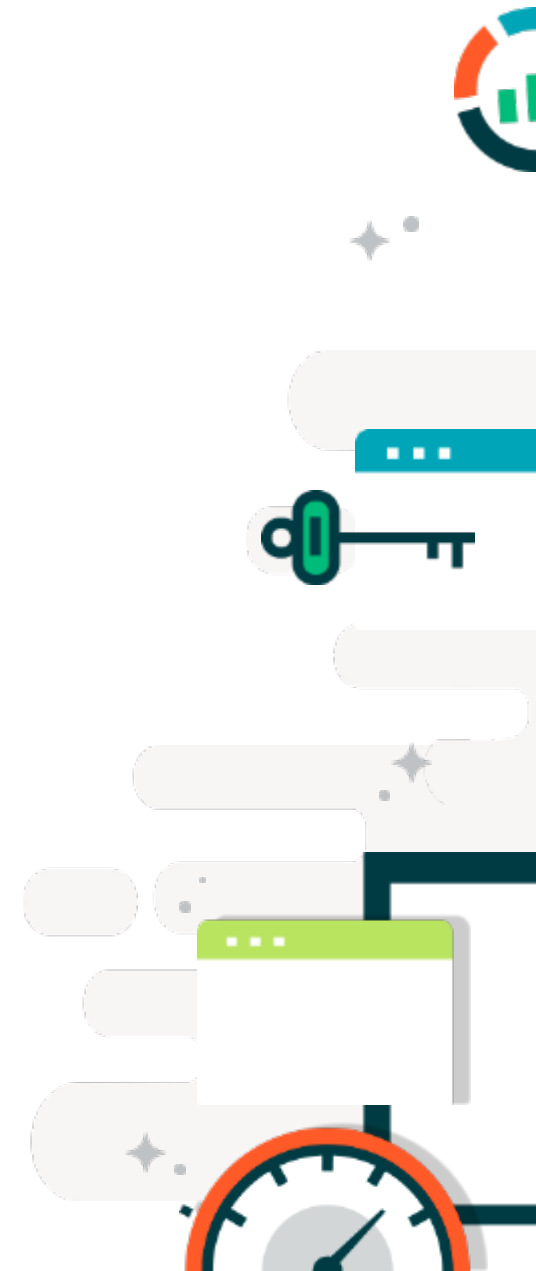
Análise Exploratória

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



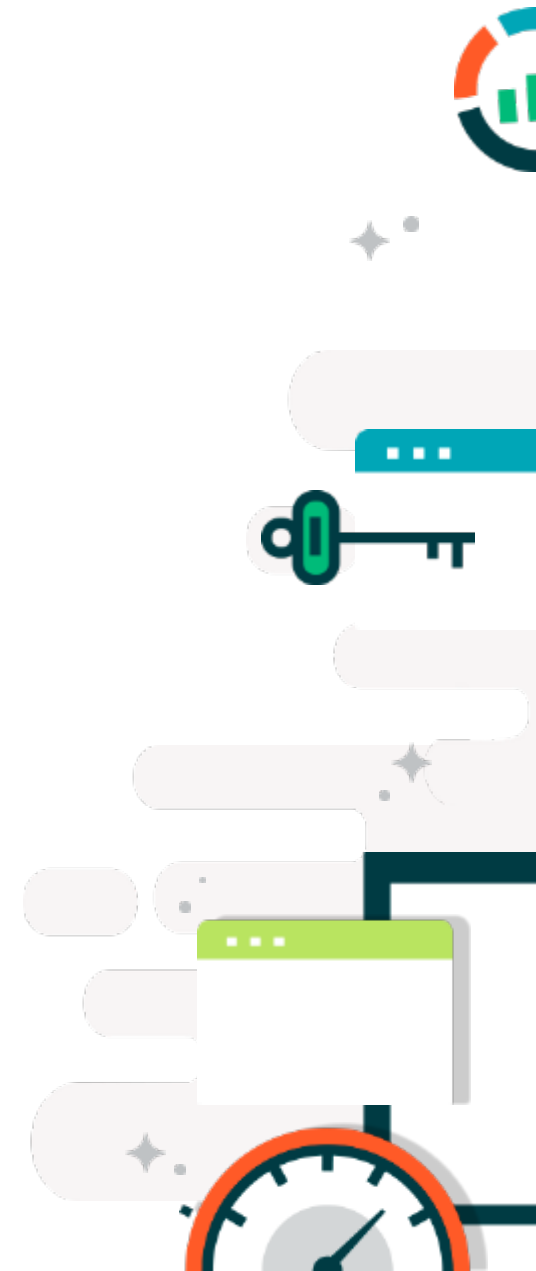
Modelagem Preditiva

- Particionamento dos dados
- Preparação de modelos preditivos
- Definição do algoritmo de base
- Preparação da validação cruzada
- Otimização dos modelos
- Seleção, treinamento e persistência do modelo



Interpretação dos Resultados

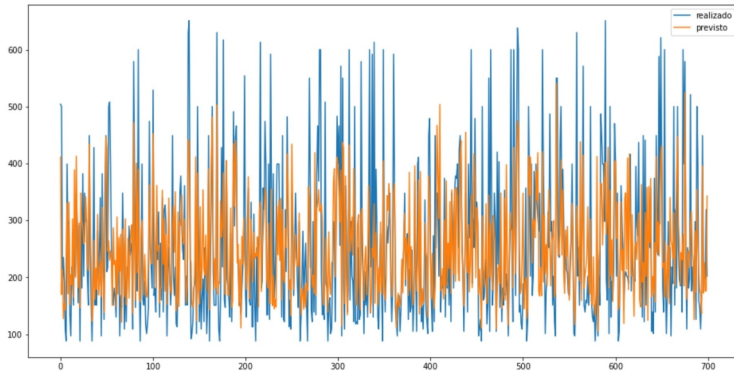
- Análise da performance do modelo
- Teste do modelo
- Visualização das previsões
- Análise dos resultados (previsto x realizado)
- Criação do ranking de custo-benefício
- Visualização dos anúncios ordenados pelo ranking



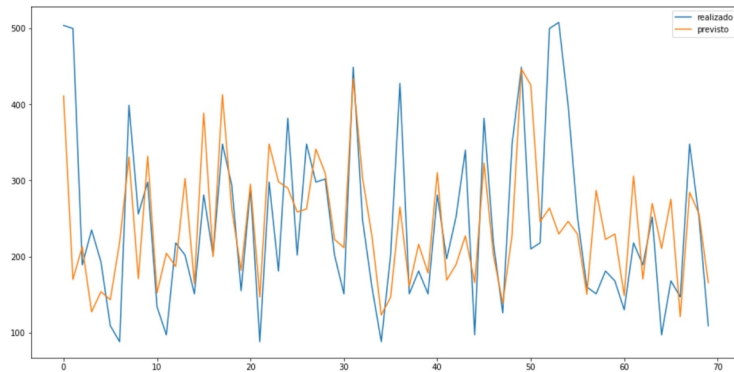
Interpretação dos Resultados

Dados sem ordenação

Todos

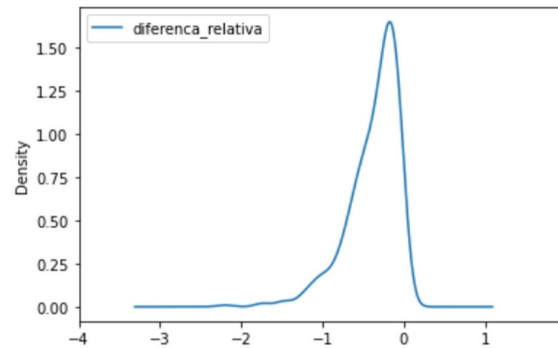


Header



Ranking de custo-benefício

Diferença ideal



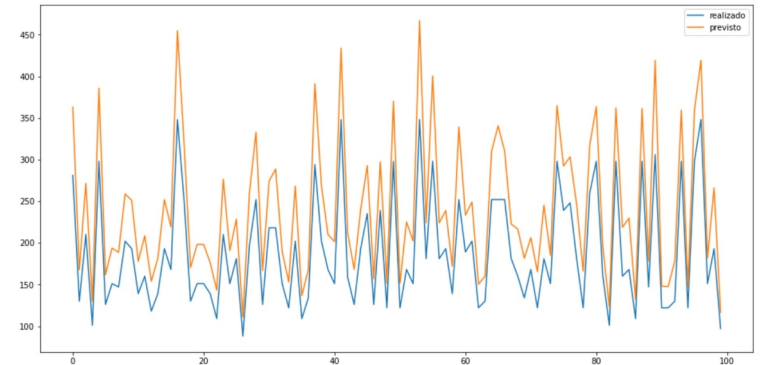
Definição do ranking

	realizado	previsto	diferenca_relativa	ranking
id				
10151402	281.0	363.051861	-0.292000	0.999323
3532589	130.0	167.622461	-0.289404	0.999323
2361521	210.0	271.432251	-0.292535	0.999044
22602640	101.0	130.088369	-0.288004	0.998592
9903460	298.0	385.705412	-0.294313	0.998116
...
34507967	92.0	227.473785	-1.472541	0.383524
23819207	88.0	219.157451	-1.490426	0.374195
2325943	97.0	263.608207	-1.717610	0.255690
22610275	88.0	242.143235	-1.751628	0.237945
13166901	122.0	391.350396	-2.207790	0.000000

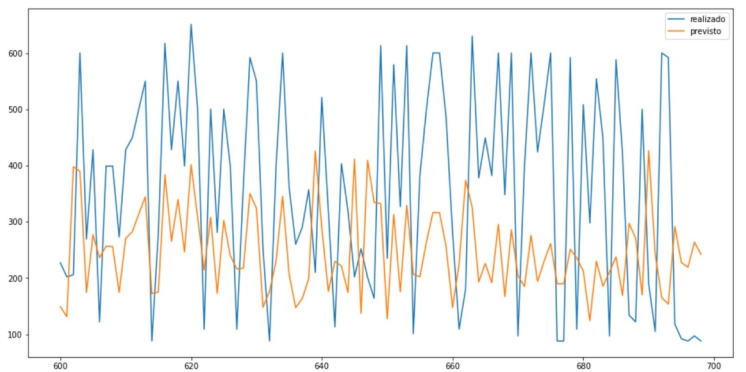
700 rows x 4 columns

Dados ordenados

Melhores

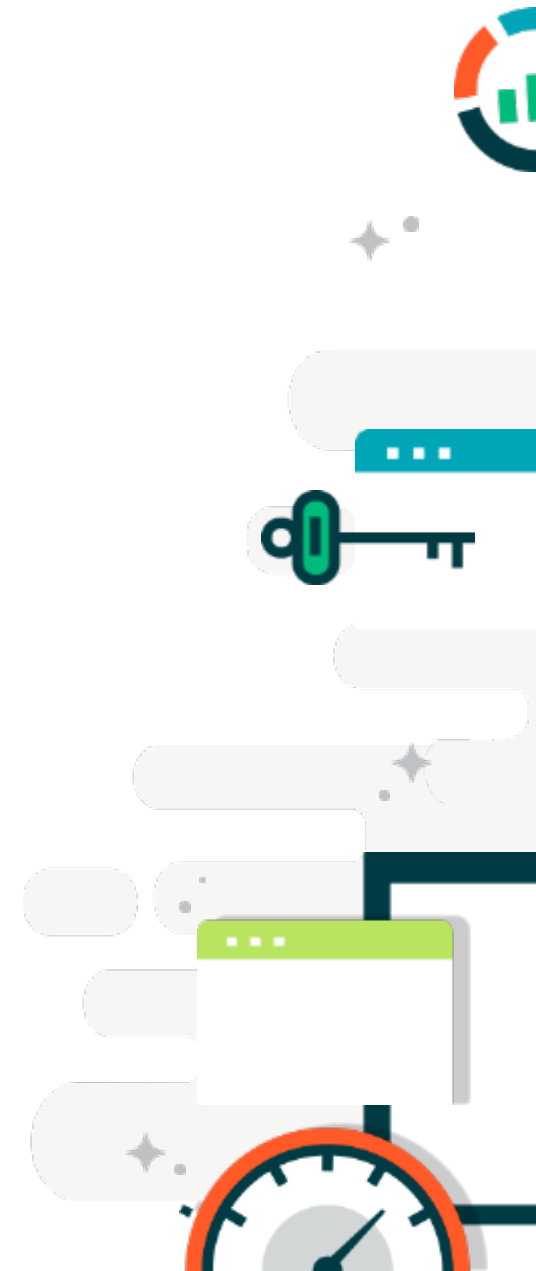


Piores



Apresentação dos Resultados

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





Obrigado