### **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

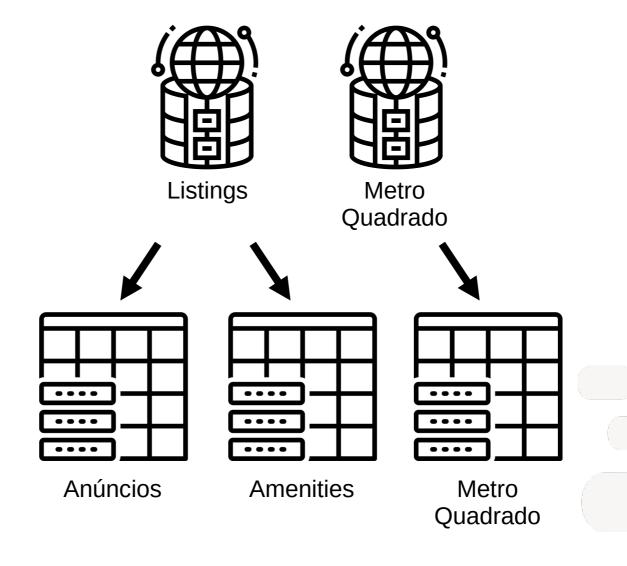
Ordenar anúncios do AirBNB por um ranking de custo-benefício, que será calculado com base em modelos preditivos de regressão.



### Coleta dos Dados

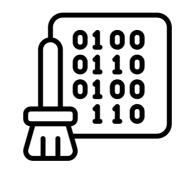
Leitura dos dados

 Geração dos dataframes iniciais

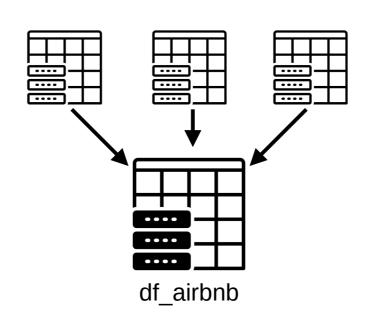


### **Processamento dos Dados**

 Exclusão de registros incompletos



Agregação
dos dataframes





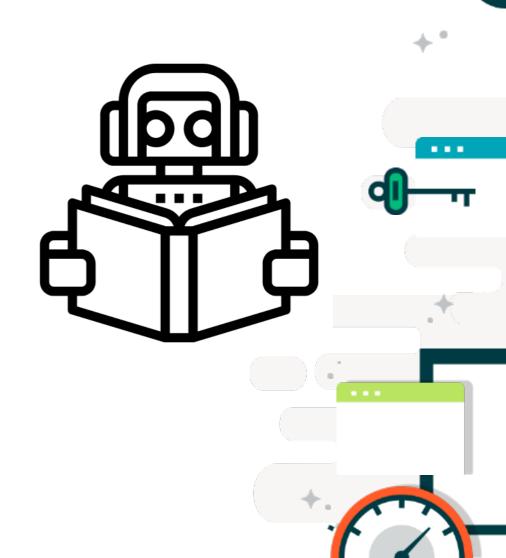
# Análise Exploratória

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



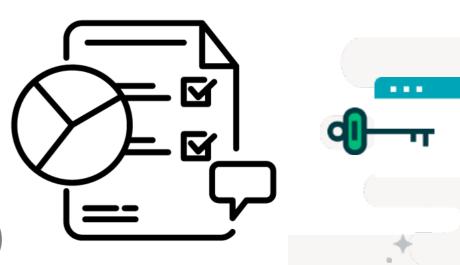
# **Modelagem Preditiva**

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



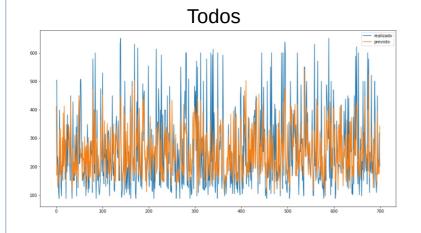
### Interpretação dos Resultados

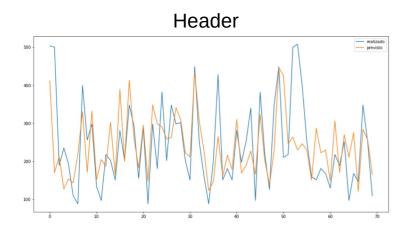
- Análise da performance do modelo
- Teste do modelo
- Visualização das predições
- Análise dos resultados (previsto x realizado)
- Criação do ranking de custo-benefício
- Visualização dos anúncios ordenados pelo ranking



## Interpretação dos Resultados

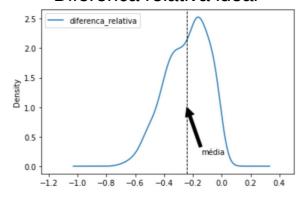
### Dados sem ordenação





#### Ranking de custo-benefício

#### Diferenca relativa ideal

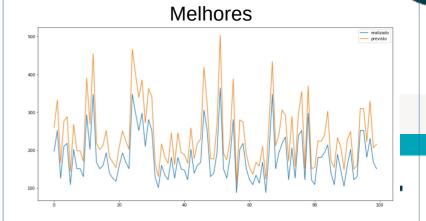


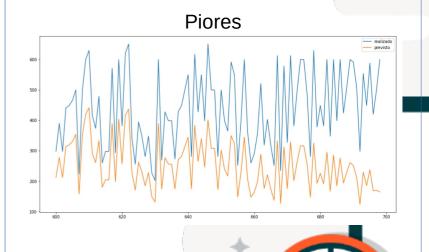
#### Definição do ranking

	realizado	previsto	diferenca_relativa	ranking
id				
25953323	197.0	260.190190	-0.242862	0.999825
7626969	252.0	332.931869	-0.243088	0.999752
5325370	126.0	166.661286	-0.243976	0.999465
32671767	210.0	276.459231	-0.240394	0.999378
34564320	218.0	288.653801	-0.244770	0.999208
3197436	588.0	237.648787	1.474239	0.445172
31960056	420.0	168.947873	1.485974	0.441379
12863588	500.0	169.973076	1.941642	0.294097
30744239	600.0	164.995170	2.636470	0.069514
32655506	592.0	153.704861	2.851537	0.000000
700 nous	× 4 colum	nc.		

700 rows × 4 columns

#### Dados ordenados

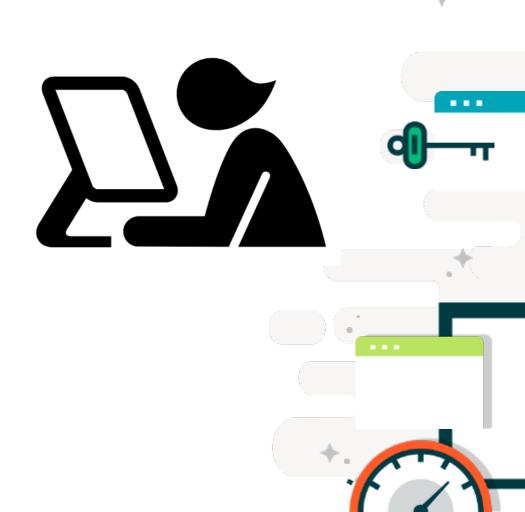




## Apresentação dos Resultados

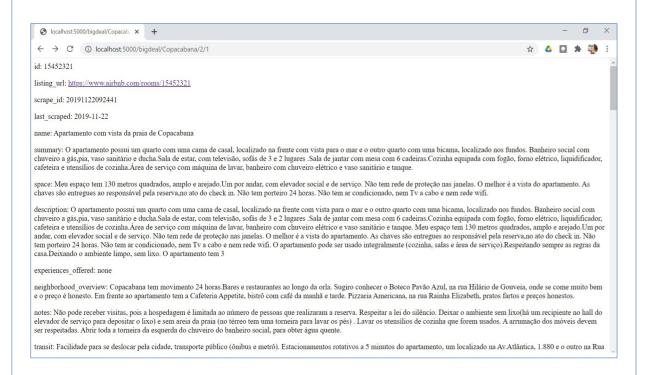
Criação da aplicação web

• Exemplos de utilização da API

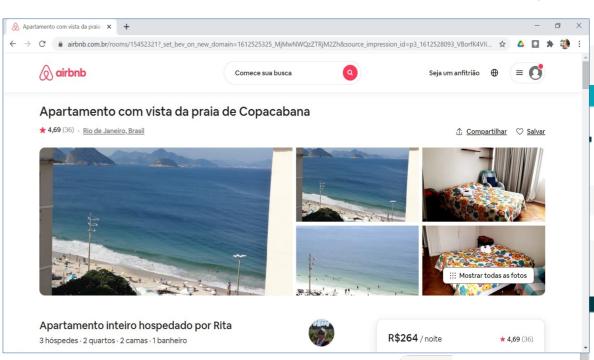


## Apresentação dos Resultados

### Utilização da API



#### Melhor custo-benefício







Obrigado