



# Bootcamp Data Analytics 2022

Challenge Correlação e Regressão



## Desafio Análise de Correlação:

Suponha que você trabalha em uma empresa do setor imobiliário. Para essa empresa é muito importante precificar corretamente o imóvel, caso contrário, os corretores podem perder a venda. Como parte do time de dados, você é responsável por analisar os principais fatores que podem afetar o preço dos imóveis. Para isso, eles te enviaram a base de dados ([desafio\\_01\\_preco\\_imoveis.csv](#)). Como você poderia ajudá-los?

Para responder esse desafio: A base de dados: [desafio\\_01\\_preco\\_imoveis.csv](#) contém as seguintes informações:

- id : código identificador do imóvel
- date: data em que o imóvel foi cadastrado
- price: preço do imóvel
- bedrooms: número de quartos
- bathrooms: número de banheiros
- sqft\_living: tamanho do imóvel (dentro da casa)
- sqft\_lot: tamanho do lote / terreno
- floors: andares
- waterfront: flag se o imóvel tem vista para o mar. 1 caso positivo, 0 caso contrário.
- view: indica o número de quartos com vista.
- condition: condição do imóvel, escala de 1 a 5
- grade: nota do imóvel
- sqft\_above: tamanho da casa acima do solo (excluindo porão)
- sqft\_basement: tamanho do porão
- yr\_built: ano em que a casa foi construída
- yr\_renovated: ano em que a casa foi renovada
- zipcode: "CEP" do imóvel
- lat: latitude do imóvel
- long: longitude do imóvel

1 - Analise a correlação de todas as variáveis com o preço do imóvel. Quais variáveis são mais correlacionadas? Quais são menos correlacionadas?

2 - Para as top 5 variáveis mais correlacionadas com o preço:

1. plote as suas distribuições (histograma e boxplot)
2. plote o scatterplot (gráfico de dispersão), contendo preço no eixo Y e a variável correlacionada no eixo X.
3. para cada uma das 5 variáveis faça uma regressão linear simples sendo preciso a variável resposta (y) e interprete os resultados.