

Correlação e Regressão Simples/Multivariada

Desafio Final

Sugestões de bibliotecas para serem utilizadas:

Pandas, NumPy, Seaborn, Matplotlib, Statsmodels, Scipy

Considerando a base de dados *kc_house_data.csv* responda as questões de 1 a 6 abaixo:

1. Codifique e exiba a matriz de correlação entre as variáveis numéricas. Além disso, qual é a correlação entre o preço *price* e o número de quartos *bedrooms*? Existe alguma diferença na correlação quando consideramos apenas casas com uma área total *sqft_living* superior a 2000 pés quadrados?
2. Existe alguma correlação entre o preço *price* e a área total da casa *sqft_living*, considerando apenas casas com pelo menos dois banheiros *bathrooms*?
3. Como a quantidade de banheiros *bathrooms* influencia na correlação entre a área total da casa *sqft_living* e o preço *price*?
4. Qual é a relação entre a condição da casa *condition* e o preço *price*, considerando apenas casas com uma área total *sqft_living* superior a 3000 pés quadrados?
5. Existe alguma correlação entre a localização geográfica *lat* e *long* e o preço *price* para casas com pelo menos três quartos *bedrooms*?

6. Calcule a correlação entre uma variável categórica(waterfront) e uma variável numérica(price) usando ANOVA.

Considerando a base de dados `car_price.csv` responda as questões de 7 a 11 abaixo:

7. Codifique e exiba a Matriz de correlação para as variáveis numéricas e dê exemplos de correlações positivas, negativas e neutras.
8. Codifique e exiba Gráficos de Dispersão para cada uma das variáveis numéricas em relação à variável de interesse *Price*.
9. Crie um modelo de Regressão Linear Simples, exiba a Tabela de Regressão e exiba o *plot* da Reta Estimada.
10. Codifique e exiba o gráfico dos resíduos do modelo de Regressão Simples.
11. Crie um modelo de Regressão Multivariada, exiba a Tabela de Regressão e exiba o gráfico dos resíduos do modelo.