

1. Faça uma análise exploratória dos dados (EDA), demonstrando as principais características entre as variáveis e apresentando algumas hipóteses de negócio relacionadas. Seja criativo!

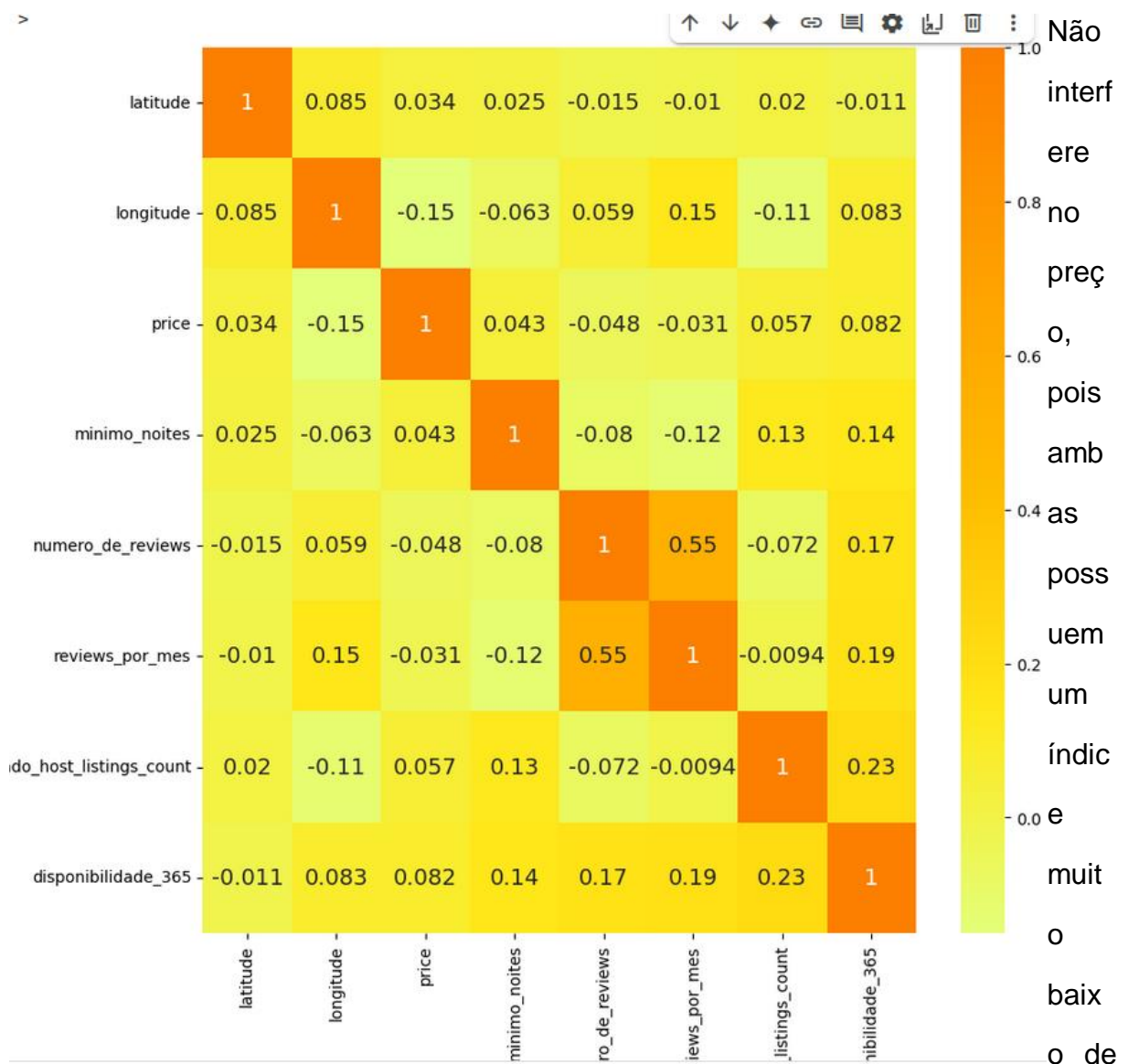
Os dados apresentados possuem colunas String e Numéricas. Isso não facilita um bom treino. Nosso problema exige tratamento prévio excluindo colunas desnecessárias e também normalizando outras no pré-processamento.

Entendo que colunas como: nome, host\_name, avaliações, dentre outras não são necessárias. No entanto, se tivéssemos a nota dada nas avaliações isso geraria valor ao cliente e ao vendedor.

Outras mais poderiam entrar na avaliação como índice de criminalidade, proximidade com colégios, shoppings e templos religiosos. Estes fatores costumam influenciar nas decisões de compra.

2. Responda também às seguintes perguntas:

1. Supondo que uma pessoa esteja pensando em investir em um apartamento para alugar na plataforma, onde seria mais indicada a compra?
2. O número mínimo de noites e a disponibilidade ao longo do ano interferem no preço?



correlação sendo 0.043 na primeira e 0.082 na segunda.

- Existe algum padrão no texto do nome do local para lugares de mais alto valor?

Aparentemente não, pois os nomes estão muito misturados com números e/ou características, dificultando o uso de uma classificação.

Explique como você faria a previsão do **preço** a partir dos dados. Quais variáveis e/ou suas transformações você utilizou e por quê? Qual tipo de problema estamos resolvendo (regressão, classificação)? Qual modelo melhor se aproxima dos dados e quais seus prós e contras? Qual medida de performance do modelo foi escolhida e por quê?

, como todo modelo

- A partir dos dados , os passos a se tomar levam em consideração: Colunas descartáveis inicialmente como: 'id', 'nome', 'host\_name', 'ultima\_review', 'room\_type', 'bairro\_group', 'bairro';
- Tratar as colunas string, transformando-as em números classificatórios e utilizáveis através do One Hot Encoding;
- Normalizar as colunas numéricas excessivamente grandes;
- O modelo de Regressão Linear pareceu o melhor, pois queremos uma previsão final de preço.
- É de fácil implementação, porém, como todo modelo, precisa ser treinado mais constantemente, pois não acompanha de forma alguma fatores que podem influenciar enormemente nos preços como criminalidade, mudança de Leis locais e/ou federais, projetos que surgem na região, desastres, ou mesmo um famoso que se mudou praquela bairro. Exemplo claro disso é a especulação imobiliária que está ocorrendo no entorno de Jericoacoara como Preá, Cruz, Caiçara.
- Escolhi como medidas: Média de Erro Absoluta e Quadrática e R2 Score para avaliar o modelo.