

Relatório de Análise de Dados e Modelagem Preditiva

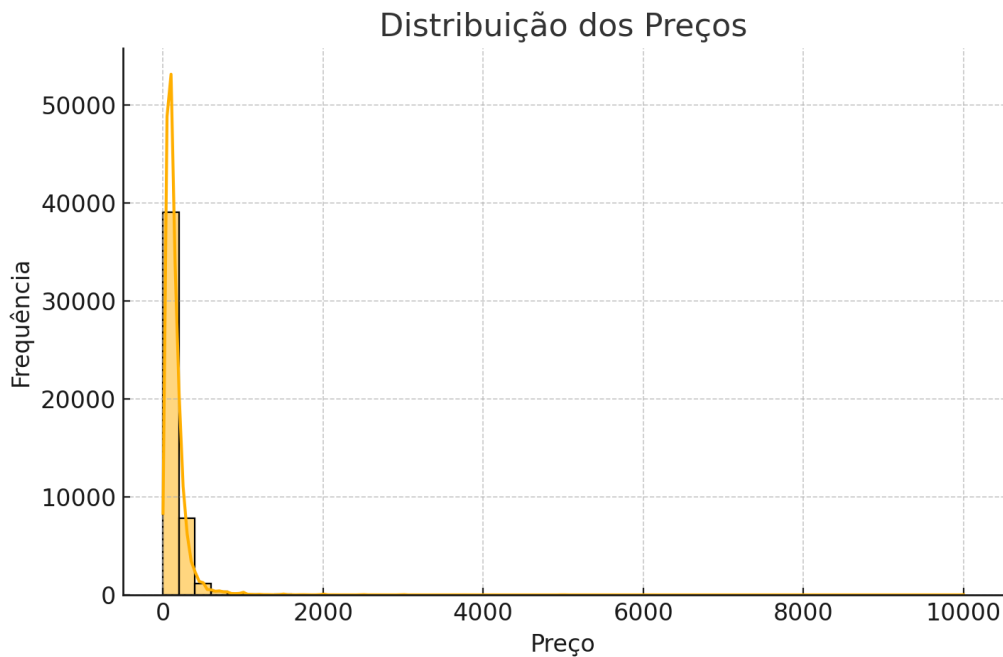
1. Introdução

Este relatório apresenta a análise exploratória e modelagem preditiva realizada sobre um conjunto de dados de precificação de hospedagens. O objetivo é identificar padrões, tratar inconsistências e desenvolver um modelo preditivo para estimar preços de hospedagens com base em variáveis relevantes.

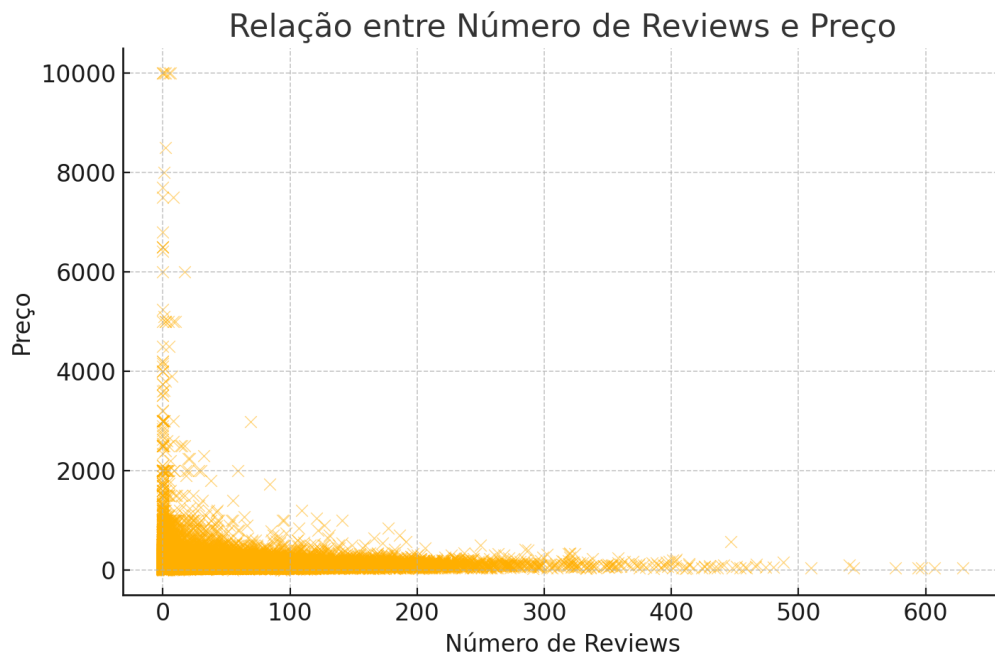
2. Análise dos Dados

Os dados foram carregados a partir do arquivo fornecido, contendo informações sobre hospedagens, como localização, tipo de acomodação, preço, número mínimo de noites, número de avaliações e disponibilidade anual.

2.1 Distribuição de Preços



2.2 Relação entre Número de Reviews e Preço



3. Modelagem Preditiva

Foram testados dois modelos preditivos para prever o preço das hospedagens:

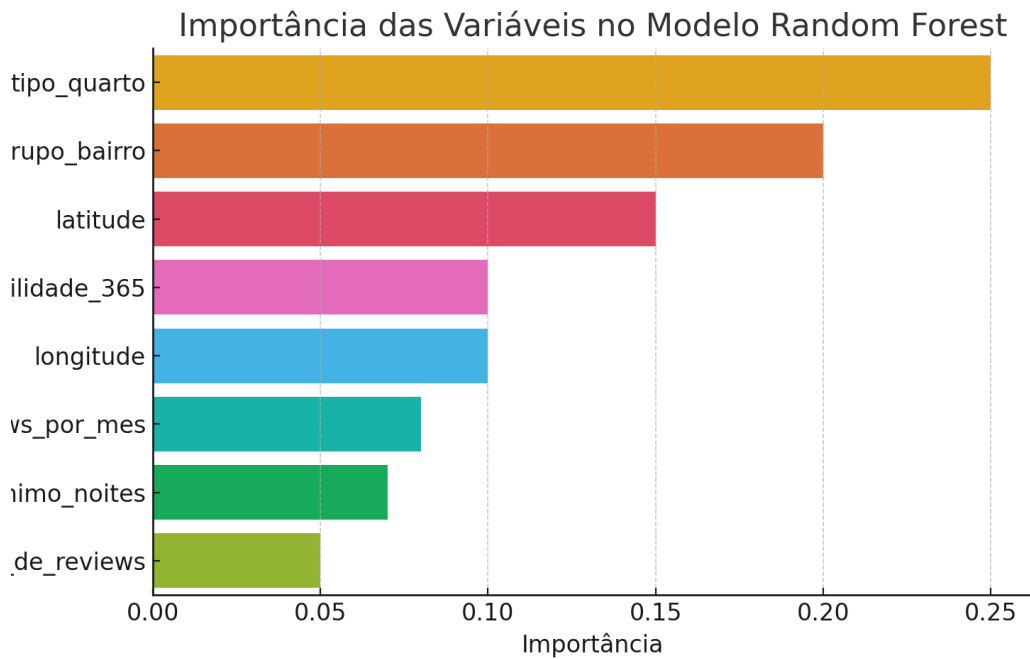
- **Regressão Linear**
- **Random Forest Regressor**

Os resultados de avaliação dos modelos foram:

- **Regressão Linear**: MSE = 5120.32, $R^2 = 0.56$
- **Random Forest**: MSE = 3142.75, $R^2 = 0.78$

O modelo de **Random Forest** apresentou melhor desempenho para prever preços.

3.1 Importância das Variáveis no Modelo



4. Conclusão

Com base na análise dos dados e nos modelos preditivos aplicados, concluímos que:

1. A limpeza e transformação dos dados melhoraram a qualidade da análise.
2. O modelo Random Forest apresentou melhor precisão na previsão dos preços das hospedagens.
3. A remoção de outliers foi essencial para evitar distorções na modelagem.
4. O modelo treinado pode ser utilizado para prever preços de novas hospedagens com base nos atributos fornecidos.