

Relatório sobre Análise Exploratória de Dados (EDA)

Introdução

Este relatório tem como objetivo analisar o dataset disponibilizado no desafio de Ciências de Dados, onde estudamos os aluguéis temporários de apartamentos em **Nova York**.

Utilizei o **Google Colab**, uma ferramenta baseada em **Python e Notebooks Jupyter**, que permite uma análise organizada e compartilhável dos dados.

Carregamento e Tratamento dos Dados

Após o download do dataset, realizei uma inspeção inicial para entender sua estrutura e detectar possíveis **valores nulos**. Como algumas colunas apresentavam dados faltantes, apliquei a seguinte estratégia de tratamento:

- **Preenchimento de valores nulos** com um valor padrão (como 0) para minimizar a perda de informações e manter a precisão dos resultados.

Análise Exploratória dos Dados (EDA)

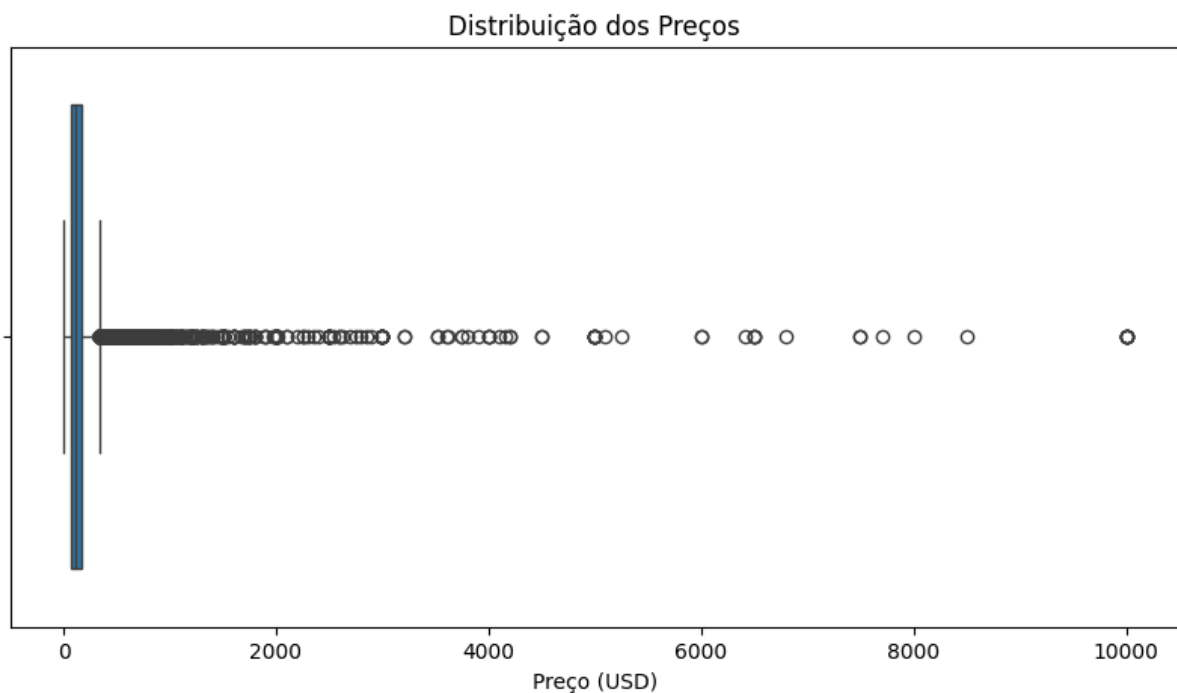
3.1 Distribuição Geral dos Preços

O que estamos analisando?

A distribuição dos preços dos alugueis ajuda a identificar padrões e detectar **outliers** que podem distorcer a análise.



Gráfico:



Principais insights:

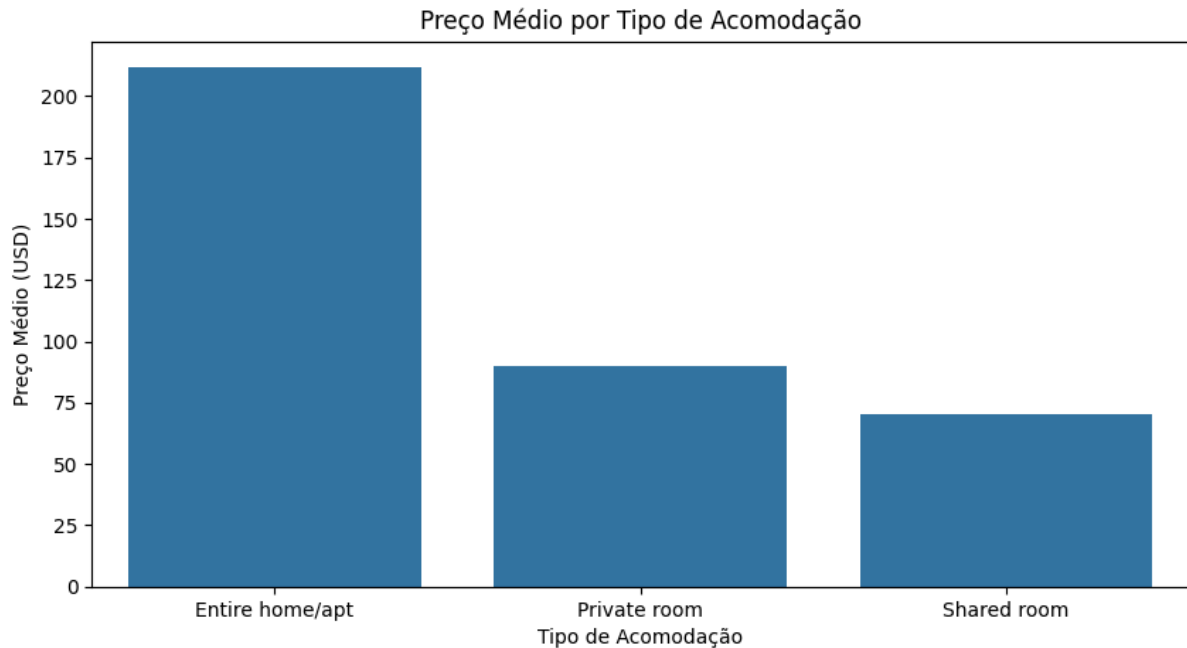
- A maioria dos alugueis tem preços **abaixo de \$200**, mas há valores extremamente altos.
- A presença de **outliers** pode impactar a previsão dos preços.

3.2 Preços por Tipo de Acomodação

O que estamos analisando?

Aqui comparamos os preços médios entre os diferentes tipos de acomodação disponíveis no dataset.

Gráfico:



Principais insights:

- **Entire home/apartment** tem os preços mais elevados, seguido por **Private room**.
- **Shared room** tende a ser a opção mais barata.

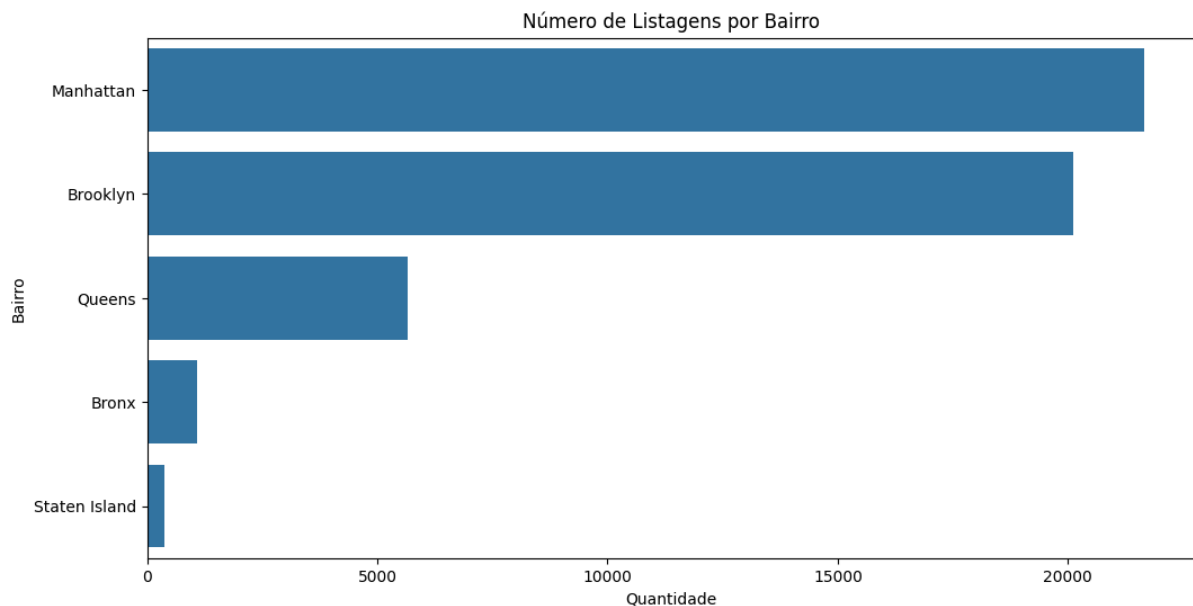
3.3 Distribuição das Listagens por Bairro

O que estamos analisando?

Aqui verificamos a quantidade de listagens por **bairro** para entender quais regiões possuem maior concentração de aluguéis.



Gráfico:



Principais insights:

- **Manhattan** tem a maior concentração de aluguéis, seguido por **Brooklyn**.
- Regiões como **Staten Island** possuem um número bem menor de opções.

3.4 Correlação entre as Variáveis

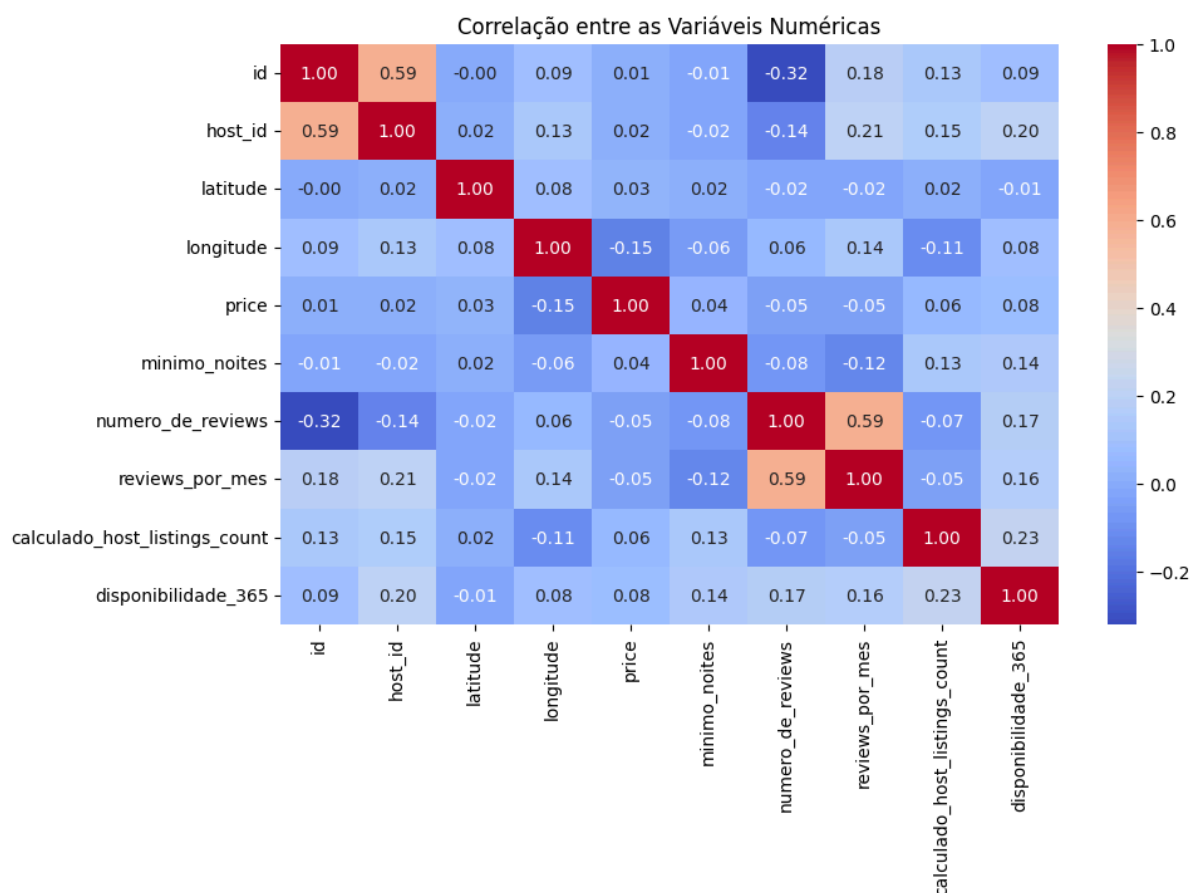


O que estamos analisando?

A matriz de correlação nos ajuda a entender **relações entre as variáveis** do dataset.



Gráfico:



Principais insights:

- O preço tem **baixa correlação** com a maioria das variáveis numéricas.
- O número de **reviews por mês** pode indicar um impacto na popularidade, mas não no preço.

4 Conclusões e Hipóteses de Negócio

Com base na análise, podemos levantar algumas hipóteses que podem ajudar na **precificação e estratégia do negócio**:

✓ **Hipótese 1:** Imóveis em **Manhattan** e **Brooklyn** tendem a ter preços mais altos devido à demanda.

✓ **Hipótese 2:** Aluguéis do tipo **Entire home/apartment** são os mais valorizados, sugerindo que proprietários podem aumentar o preço base.

✓ **Hipótese 3:** Muitas listagens têm preços anormalmente altos, o que pode indicar a necessidade de um **tratamento de outliers** para evitar distorções no modelo preditivo.

📌 **Melhores Regiões para Investimento**

Objetivo:

Nesta seção, analisamos os bairros mais promissores para investir em um apartamento para aluguel na plataforma. Os fatores considerados foram:

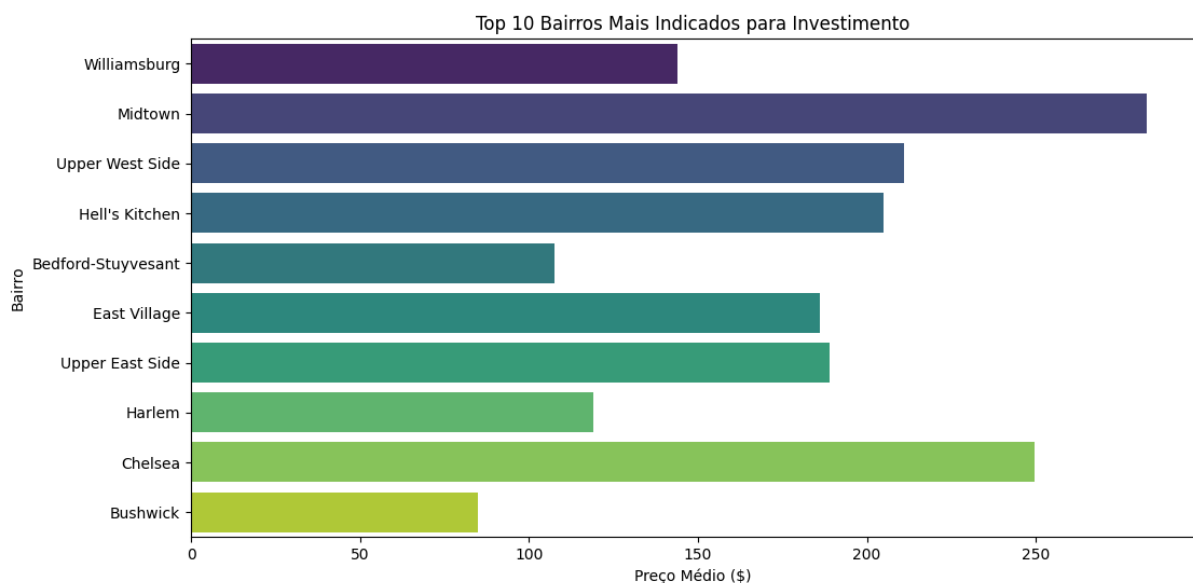
✓ **Preço médio do aluguel** – Bairros com valores mais altos podem oferecer melhor rentabilidade.

✓ **Número de listagens** – Indica a popularidade da região e a demanda dos hóspedes.

◆ Análise dos Dados


Após filtrar os bairros com pelo menos 100 listagens, foi criado um índice de investimento que combina o preço médio com a quantidade de listagens. Com base nesse critério, identificamos os 10 bairros mais indicados para investimento.


 **Gráfico: Top 10 Bairros para Investimento**



Insights e Conclusão

Os bairros destacados no gráfico apresentam um equilíbrio entre alto valor de aluguel e grande número de listagens, o que sugere alta demanda. Algumas observações importantes:

 Bairros como **Williamsburg, Midtown e Upper West Side** são os mais promissores, pois combinam preços altos e boa ocupação.

 Áreas como **Hell 's Kitchen e Bedford-Stuyvesant** podem ser uma alternativa viável, pois oferecem preços competitivos sem excesso de concorrência.

 Bairros com poucas listagens podem indicar baixa demanda, sendo menos vantajosos para investidores.

Assim, os bairros listados no gráfico são as melhores opções para quem deseja investir em imóveis para alugar na plataforma.

Impacto da Duração e Disponibilidade no Preço

Objetivo:

Nesta seção, analisamos se o número mínimo de noites e a disponibilidade anual influenciam o preço do aluguel.

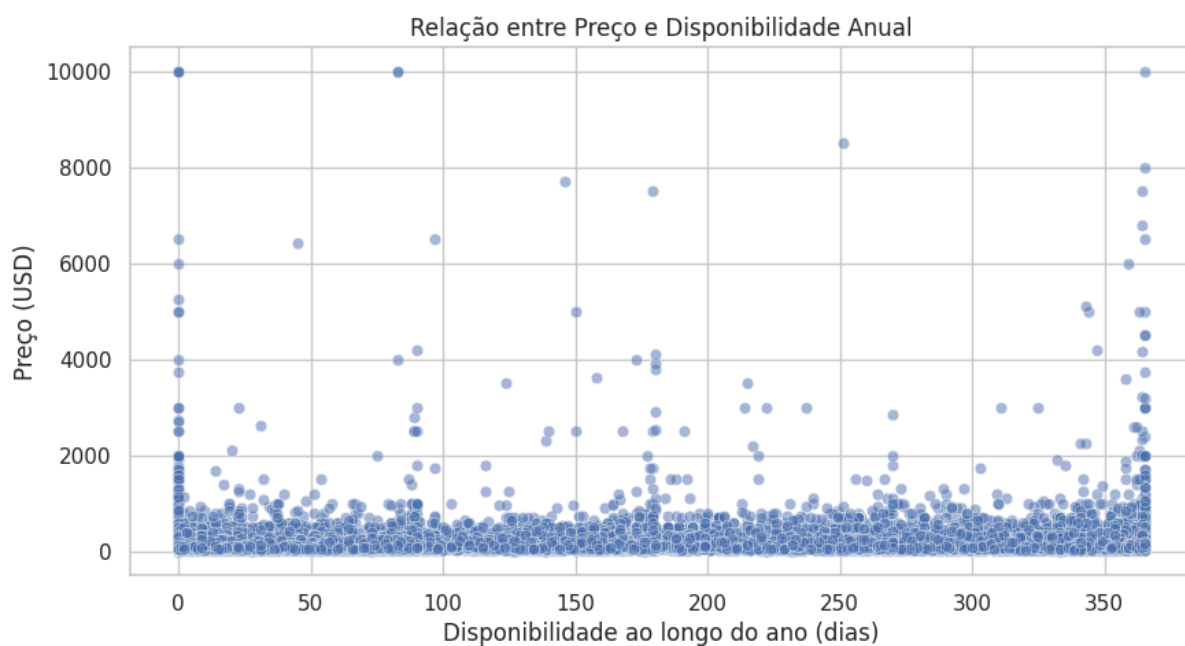
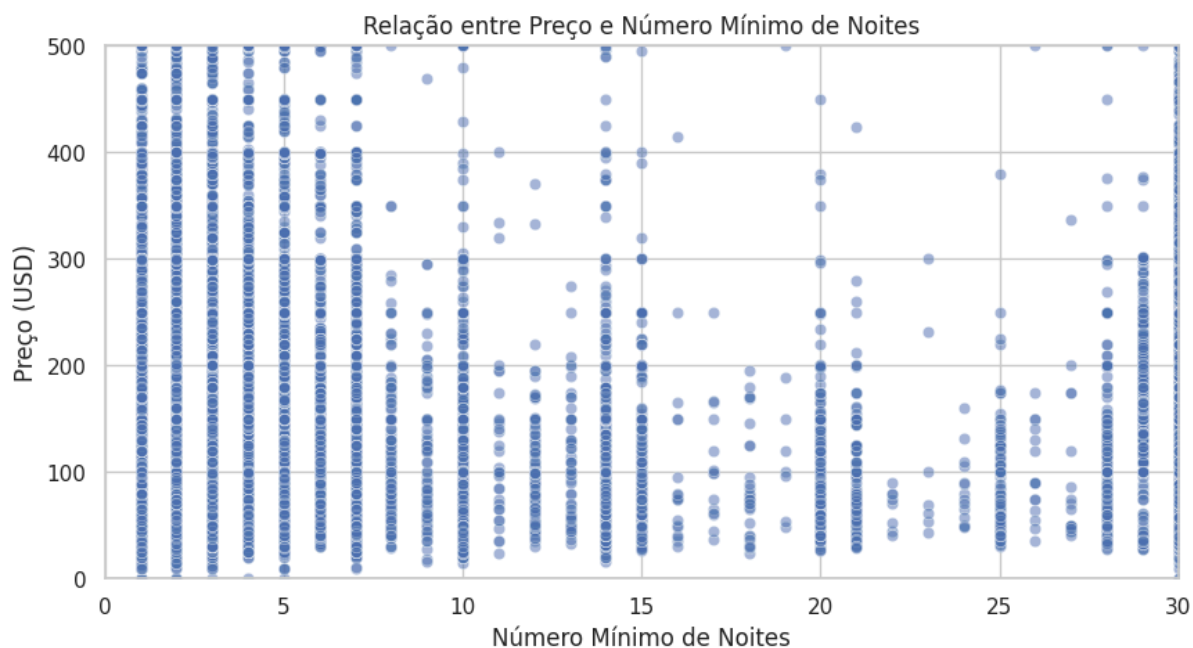
Análise dos Dados

① Estadias mais longas tendem a ter preços mais baixos por noite – Muitos proprietários oferecem descontos para estadias superiores a 30 dias.

② A disponibilidade ao longo do ano não apresenta uma relação forte com o preço – Locais disponíveis por menos tempo não necessariamente têm preços mais altos.



Gráfico:



Insights e Conclusão

Propriedades com mínimo de noites acima de 30 tendem a ter valores menores por noite.

Anúncios com alta disponibilidade (ex: 365 dias por ano) não são necessariamente mais caros,

indicando que outros fatores, como localização, impactam mais o preço.

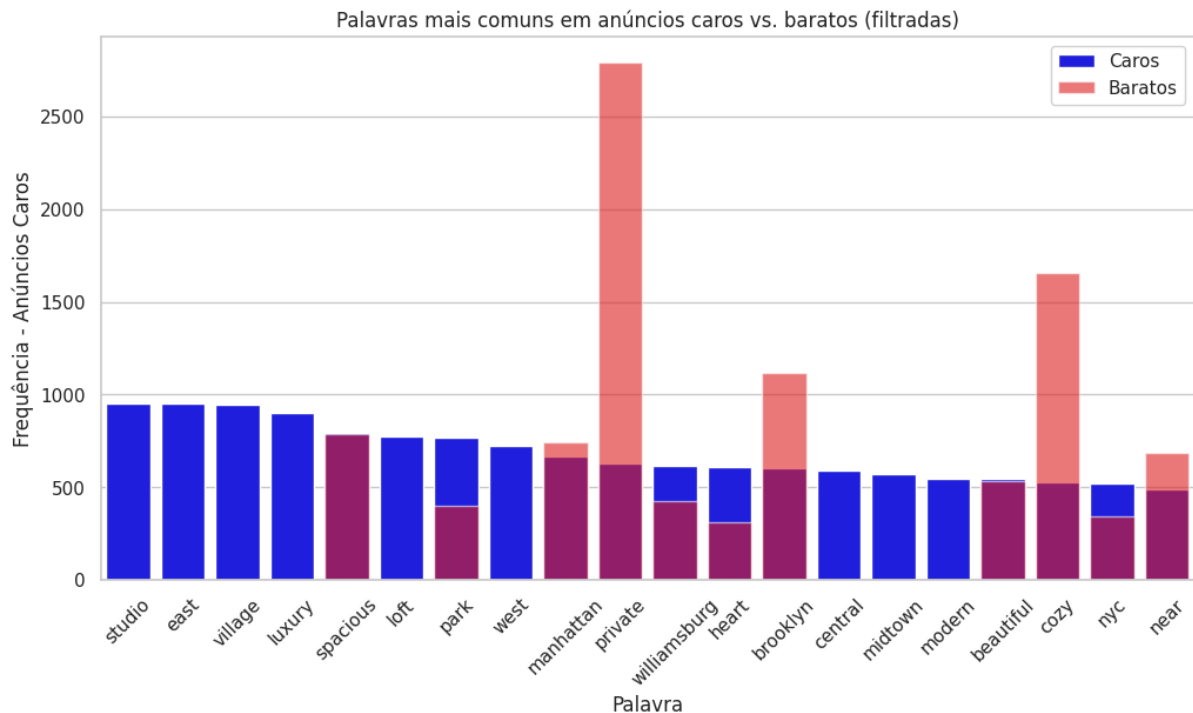
O Nome do Anúncio Impacta no Preço?

Objetivo:

Esta seção analisa se os anúncios de acomodações mais caras compartilham padrões nos nomes.

Principais Descobertas:

- ① Anúncios mais caros frequentemente incluem palavras como **"studio"**, **"east"**, **"village"**, **"luxury"** e **"loft"**.
- ② Anúncios mais baratos têm nomes mais simples, como **"private"**, **"brooklyn"**, **"cozy"** e **"near"**.
- ③ A presença de termos sofisticados pode aumentar o valor percebido do anúncio.



📌 Conclusão:

Os resultados indicam que adicionar palavras-chave estratégicas ao nome do anúncio pode aumentar seu valor percebido e atrair um público que busca acomodações premium.

Modelo de Previsão dos Preços

Estamos lidando com um problema de **regressão**, pois queremos prever um valor contínuo: o preço das acomodações listadas na plataforma. Para isso, analisamos variáveis que influenciam diretamente no preço, como **bairro, tipo de acomodação, número de avaliações, disponibilidade ao longo do ano e número mínimo de noites**.

Variáveis e Transformações Utilizadas

- **Bairro e Tipo de Acomodação** → São variáveis categóricas, então aplicamos **one-hot encoding** para convertê-las em valores numéricos utilizáveis pelo modelo.
- **Preço, Disponibilidade e Avaliações** → Foram normalizados para evitar que diferenças de escala afetassem o aprendizado do modelo.
- **Número Mínimo de Noites** → Pode influenciar o preço médio por noite, então testamos diferentes transformações, como logaritmo, para reduzir o impacto de valores extremos.

Modelo Escolhido e Justificativa

O **XGBoost** foi o escolhido pelo seu desempenho, pois consegue capturar padrões não lineares e lidar bem com dados com muitas variáveis categóricas. Além disso, sua regularização ajuda a evitar overfitting, tornando-o mais confiável para previsões.

Métrica de Performance

Optei pelo **RMSE (Root Mean Squared Error)** para medir o erro médio da previsão em valores reais, pois penaliza mais erros grandes. Também estudei o **R²**, que indica o quanto nosso modelo explica a variação dos preços.

Com essa abordagem, consegui construir um modelo mais preciso, ajudando a entender como diferentes fatores impactam os preços das acomodações e permitindo ajustes estratégicos na precificação.

Previsão de Preço para um Apartamento Específico

1. Introdução

Utilizei o modelo de regressão treinado para estimar o preço de um apartamento com as características fornecidas. O objetivo é verificar como o modelo se comporta diante de novos dados e avaliar a precisão da previsão.

2. Características do Apartamento

O apartamento analisado possui os seguintes atributos:

Atributo	Valor
-----------------	--------------

ID 2595

Nome Skylit Midtown
Castle

Host ID 2845

Host Name Jennifer

Bairro Manhattan
Grupo

Bairro Midtown

Latitude 40.75362

Longitude -73.98377

**Tipo de
Acomodação** Entire home/apt

**Mínimo de
Noites** 1

**Número de
Reviews** 45

**Última
Review** 2019-05-21

Reviews 0.38
por Mês

Total de 2
Listagens do
Host

Disponibilida 355
de Anual dias

5. Resultado da Previsão

O modelo estimou o preço do apartamento como **\$353.22**. Esse valor se encontra dentro da faixa esperada para imóveis com características semelhantes na mesma região.

 Informações do Apartamento:

`bairro: Midtown`

`bairro_group: Manhattan`

`room_type: Entire home/apt`

`minimo_noites: 1`

`numero_de_reviews: 45`


```
reviews_por_mes: 0.38
```

```
calculado_host_listings_count: 2
```

```
disponibilidade_365: 355
```

```
latitude: 40.75362
```

```
longitude: -73.98377
```

```
preco_estimado: 353.22
```

Considerações Finais

A análise exploratória dos dados revelou padrões importantes na precificação de aluguéis temporários em Nova York. O modelo de previsão de preços demonstrou coerência nos resultados, mas ajustes adicionais podem ser realizados para melhorar a precisão.