

✓ Análise de Dados do Airbnb em Nova York: Tendências, Preços e Disponibilidade

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

✓ Carregando conjunto de dados.

```
data = pd.read_csv('AB_NYC_2019.csv') # criando o atributo "data" para gerar os dados da tabela
data.shape # comando para conferir a quantidade de linhas e colunas
```

```
(48895, 16)
```

✓ Limpeza de dados

```
data_clean = data.dropna()
data_clean.shape #conferindo quantidade de dados após limpeza
```

```
(38821, 16)
```

-> Conferido que sem a limpeza de dados, a tabela havia 48895 linhas e 16 colunas. -> Após a limpeza, a tabela possui 38821 linhas e 16 colunas.

com estes dados podemos concluir que aproximadamente 10.079 dados nulos foram retirados do dataset.

✓ Dados de preço por região

```
#aplicando método "agg()" para realizar mais de uma operação
estatisticas_por_regiao = data_clean.groupby('neighbourhood_group').agg(
    max_preco=('price', 'max'),
    media_preco=('price', 'mean'),
    min_preco=('price', 'min'),
    total_listagens=('calculated_host_listings_count', 'sum'),
    avaliacao_media=('reviews_per_month', 'mean')
).reset_index()
```

```
# Exibindo a tabela resultante
display(estatisticas_por_regiao)
```

	neighbourhood_group	max_preco	media_preco	min_preco	total_listagens	avaliacao_media
0	Bronx	800	79.558857	0	1930	1.838549
1	Brooklyn	10000	121.463289	0	36717	1.283356
2	Manhattan	9999	180.071596	10	145835	1.271756
3	Queens	10000	95.783683	10	15341	1.941923
4	Staten Island	625	89.964968	13	750	1.872580

Próximas etapas: [Gerar código com estatisticas_por_regiao](#) [Ver gráficos recomendados](#) [New interactive sheet](#)

Nesta tabela podemos analisar os dados por região pelo preço máximo, mínimo, médio e total.

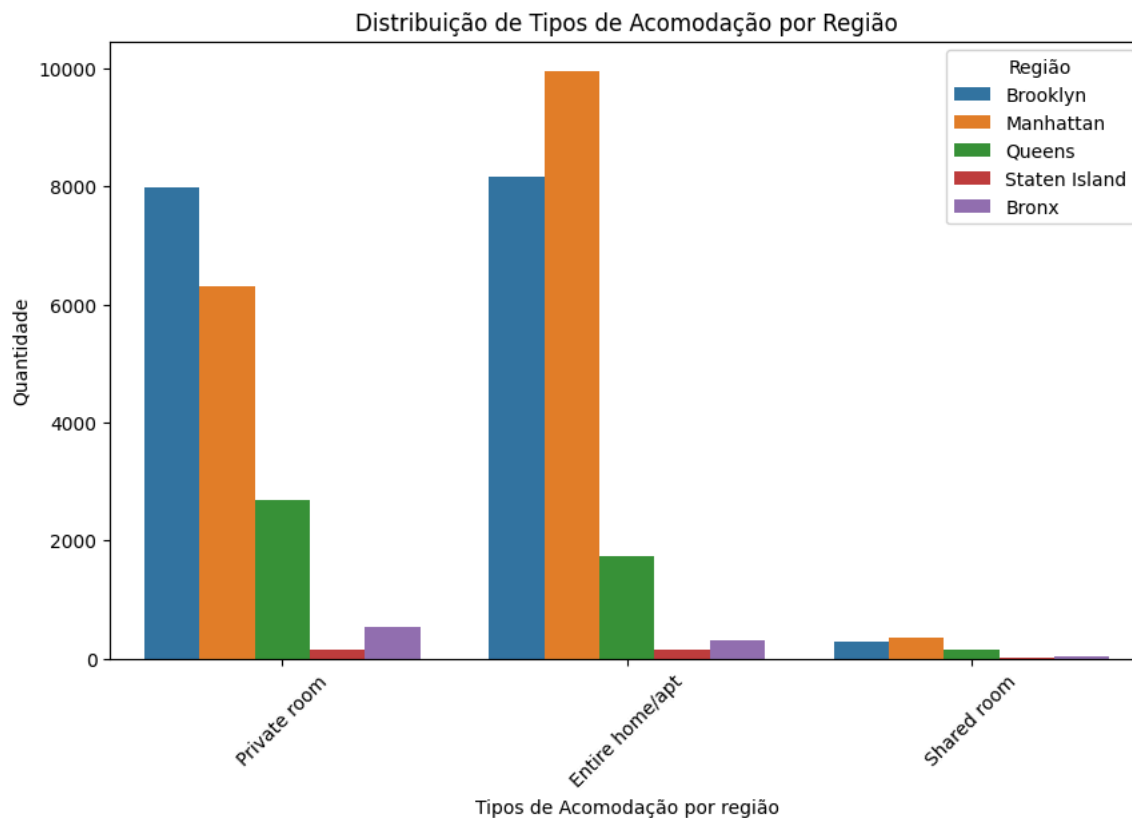
✓ Análise de Tipos negrito de Acomodação

```
acomodacao = data_clean['room_type']
regiao = data_clean['neighbourhood_group']
```

```
# VISUALIZAÇÃO DE DADOS
```

```
plt.figure(figsize=(10, 6))
sns.countplot(x=acomodacao, hue=regiao)
plt.title('Distribuição de Tipos de Acomodação por Região')
plt.xlabel('Tipos de Acomodação por região')
plt.ylabel('Quantidade')
plt.xticks(rotation=45)
```

```
plt.legend(title='Região')
plt.show()
```



```
# Iterar sobre cada região e criar gráficos de pizza
regioes = data_clean['neighbourhood_group'].unique()

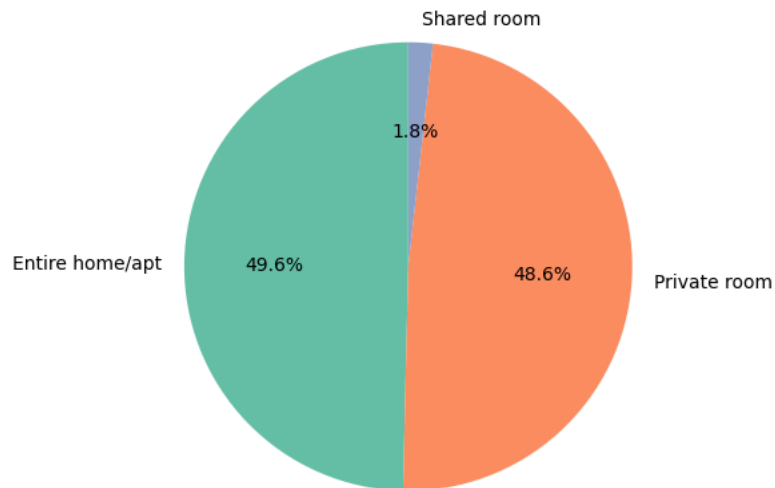
for regioao in regioes:
    # Filtrar os dados por região
    dados_regiao = data_clean[data_clean['neighbourhood_group'] == regioao]

    # Contar os tipos de acomodação
    tipos_acomodacao = dados_regiao['room_type'].value_counts()

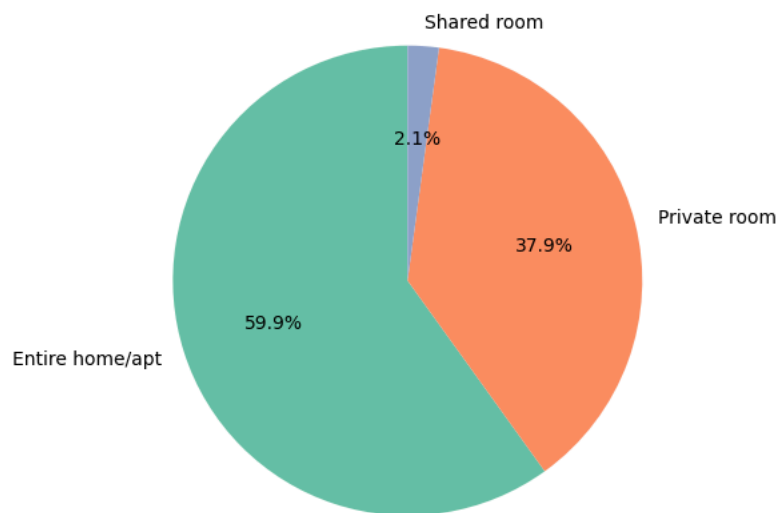
    # Plotar o gráfico de pizza
    plt.figure(figsize=(6, 6))
    plt.pie(
        tipos_acomodacao,
        labels=tipos_acomodacao.index,
        autopct='%1.1f%%',
        startangle=90,
        colors=sns.color_palette('Set2', n_colors=len(tipos_acomodacao))
    )
    plt.title(f'Distribuição de Tipos de Acomodação - {regiao}', fontsize=14)
    plt.axis('equal') # Assegura que o gráfico seja um círculo
    plt.tight_layout()
    plt.show()
```



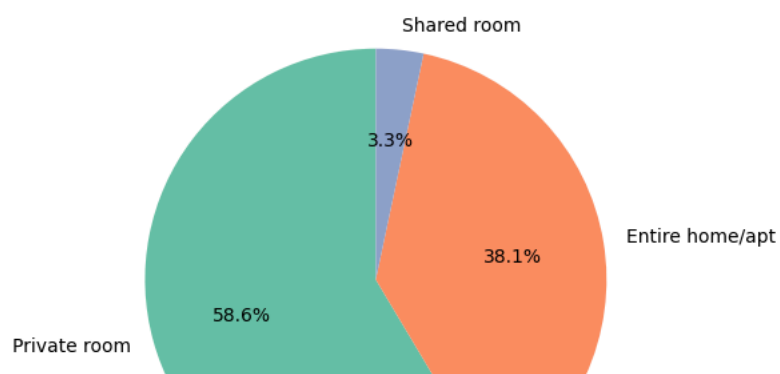
Distribuição de Tipos de Acomodação - Brooklyn



Distribuição de Tipos de Acomodação - Manhattan

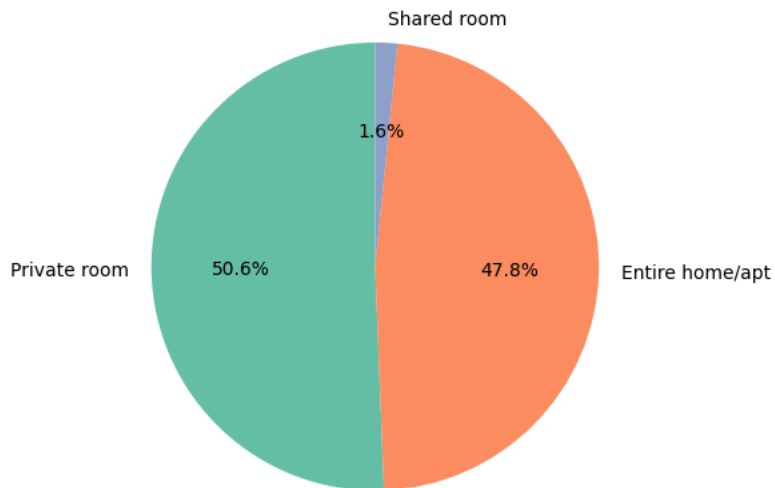


Distribuição de Tipos de Acomodação - Queens

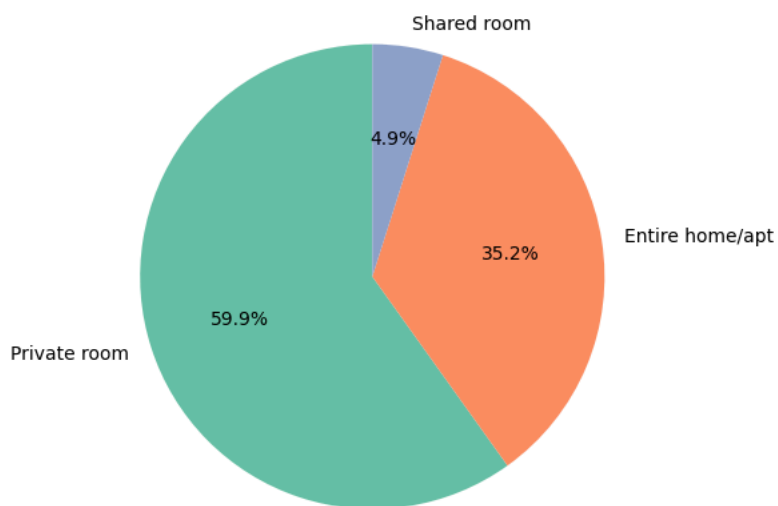




Distribuição de Tipos de Acomodação - Staten Island



Distribuição de Tipos de Acomodação - Bronx



Neste gráfico podemos ver a frequência de cada categoria e sua estatística.

✓ Distribuição de preços

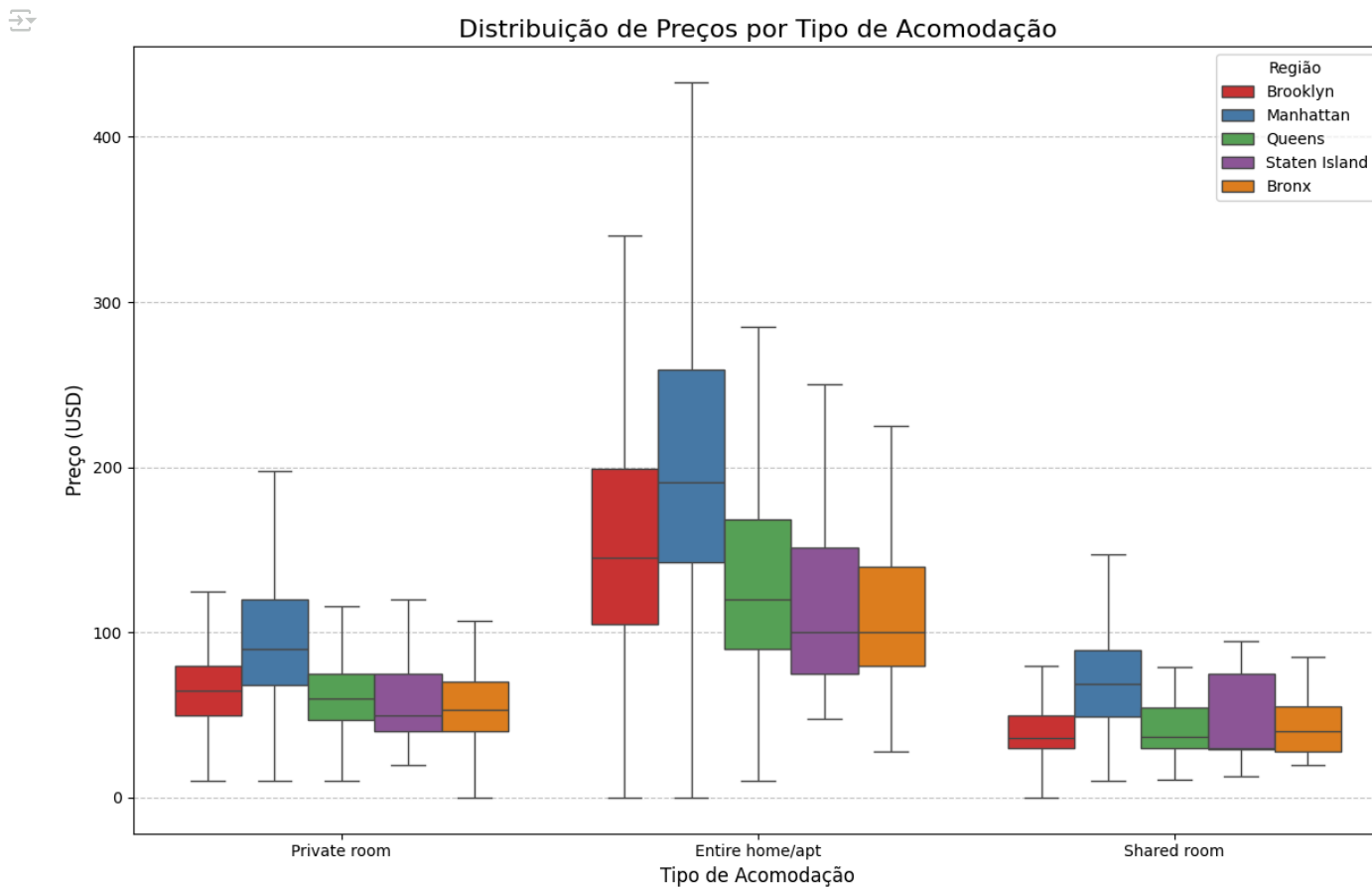
Pergunta: Qual a faixa de preço predominante por tipo de acomodação?

```
tipo_de_acomodacao = data_clean['room_type']
preco = data_clean['price']

# Configurando a figura
plt.figure(figsize=(12, 8))

# Criando o boxplot
sns.boxplot(
    x='room_type',
    y='price',
    data=data,
    palette='Set1',
    hue='neighbourhood_group',
    showfliers=False # Oculta outliers para uma visualização mais limpa
)

# Ajustando o gráfico
plt.title('Distribuição de Preços por Tipo de Acomodação', fontsize=16)
plt.xlabel('Tipo de Acomodação', fontsize=12)
plt.ylabel('Preço (USD)', fontsize=12)
plt.xticks(fontsize=10)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.legend(title='Região', fontsize=10)
plt.tight_layout()
plt.show()
```



🔥 Principais Observações "Entire home/apt" tem a maior variação de preços

Os preços são significativamente mais altos e variáveis para esse tipo de acomodação, especialmente em Manhattan e Brooklyn. Há outliers indicando algumas listagens com preços extremamente elevados. "Private room" tem uma faixa de preços mais estável

A maioria dos valores está entre 50 e 150 USD, dependendo da região. Manhattan apresenta preços um pouco mais altos e dispersos. "Shared room" tem os preços mais baixos

Os preços são mais uniformes e bem abaixo das outras categorias. As medianas das regiões são próximas entre si, indicando uma menor variação de preços entre os bairros para este tipo de acomodação. Diferentes regiões apresentam variações distintas

Manhattan tende a ter preços mais elevados e dispersos, principalmente para apartamentos inteiros. Brooklyn também apresenta alta variação de preços, mas com valores um pouco mais baixos do que Manhattan. Queens, Bronx e Staten Island possuem preços mais acessíveis e menos dispersos. 🇺🇸 O que esse boxplot pode indicar? Viajantes com orçamento menor podem preferir Private rooms ou Shared rooms em regiões como Queens ou Bronx. Proprietários podem ajustar os preços com base na região e no tipo de acomodação para competir no mercado. Análises futuras podem verificar se a demanda e as avaliações justificam essas diferenças de preços. Se precisar de mais detalhes ou quiser ajustes no gráfico, me avise! 🚀

✓ Popularidade por Região

```
# Calculando a média de avaliações por bairro
media_avaliacoes = data_clean.groupby('neighbourhood')['number_of_reviews'].mean().reset_index()

# Renomeando as colunas para legibilidade
media_avaliacoes.columns = ['Bairro', 'Média de Avaliações']

# Ordenando os bairros pela média de avaliações em ordem decrescente
media_avaliacoes = media_avaliacoes.sort_values(by='Média de Avaliações', ascending=False)

# Exibindo os 10 bairros com maior média de avaliações
display(media_avaliacoes.head(10))
```



	Bairro	Média de Avaliações
175	Silver Lake	118.500000
69	Eltingville	114.000000
59	East Elmhurst	88.345029
134	Mount Eden	84.000000
163	Richmondtown	79.000000
181	Springfield Gardens	73.412500
192	Tompkinsville	60.000000
32	Castle Hill	59.000000
95	Highbridge	58.227273
119	Manhattan Beach	57.857143

✓ Visualização

```
# Configurando o gráfico de barras
plt.figure(figsize=(12, 8))
sns.barplot(
    data=media_avaliacoes.head(10), # Os 10 bairros mais populares
    x='Média de Avaliações',
    y='Bairro',
    palette='coolwarm'
)

# Ajustando o gráfico
plt.title('Top 10 Bairros com Maior Média de Avaliações', fontsize=16)
plt.xlabel('Média de Avaliações', fontsize=12)
plt.ylabel('Bairro', fontsize=12)
plt.grid(axis='x', linestyle='--', alpha=0.7)
plt.show()
```