



## **Desafio Seazone análise de dados**

**Gráficos, dados e análises para o processo seletivo  
Seazone para a vaga de Estágio em Analista de dados.**

Isis Araujo

Aracaju/SE

Novembro, 2023

## Introdução

Os participantes receberam dois conjuntos de dados em formato CSV, acessíveis através de um aplicativo com uma interface intuitiva. Estes dados incluem informações detalhadas sobre as classificações e características de anúncios de hospedagem no Booking.com. O principal objetivo do desafio é analisar esses dados para extrair insights significativos relacionados à ocupação e ao preço, respondendo a uma série de perguntas estratégicas. O endereço para acesso ao Jupyter Notebook é: [https://github.com/IsisAraujo/desafio\\_pratico\\_Seazone](https://github.com/IsisAraujo/desafio_pratico_Seazone)

## Estrutura dos Dados:

Os dados são divididos em dois arquivos principais:

1. desafio\_rating.csv: Contém dados sobre as avaliações dos imóveis, incluindo identificador do anúncio, título, URL, data, categorias avaliadas e número de avaliações.
2. desafio\_details.csv: Apresenta características específicas de cada imóvel ou anúncio, como identificador, tipo de acomodação, tamanho, recursos disponíveis e localização.

## Perguntas do Desafio:

Os participantes são encorajados a responder perguntas que variam desde a classificação das cidades por número de anúncios e tamanho dos imóveis até análises mais profundas sobre a relação entre as características dos anúncios, suas localizações e as avaliações recebidas. Estas perguntas visam não apenas testar as habilidades técnicas dos candidatos, mas também sua capacidade de gerar insights práticos e estratégicos a partir dos dados.

## Metodologia

A análise de dados foi realizada utilizando Python 3.8 como solicitado e algumas bibliotecas como:

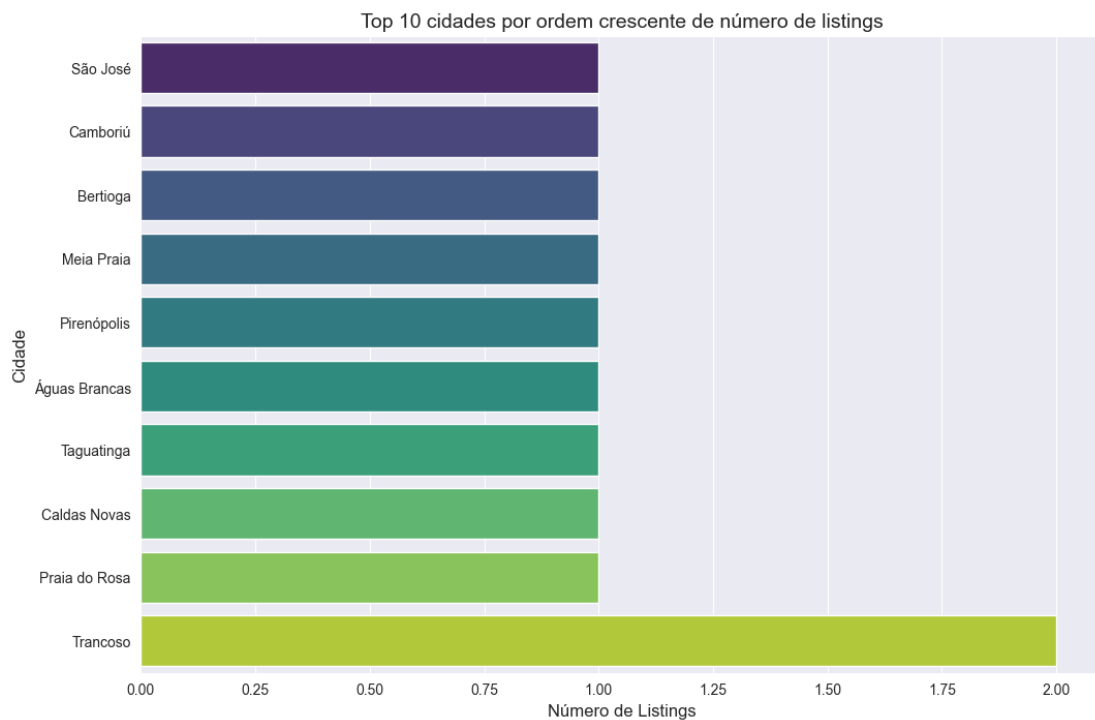
1. **Pandas:** Oferece estruturas de dados e operações para manipular tabelas numéricas.
2. **Matplotlib:** Utilizada para criar visualizações estáticas (gráficos).
3. **Seaborn:** Cria gráficos mais atraentes e informativos.
4. **GeoPy:** Utilizada para geocodificação, que é o processo de converter
5. **Folium:** Criação de mapas interativos.
6. **Scikit-learn (KMeans):** Método de clusterização.

## Pergunta 1

**Ordene as cidades em ordem crescente de número de listings.**

*Resposta:*

As cidades foram ordenadas do menor para o maior número de anúncios. Florianópolis teve o maior número com 768 anúncios, enquanto várias cidades, como São José, Camboriú, Bertioga, entre outras, tiveram apenas 1 anúncio.

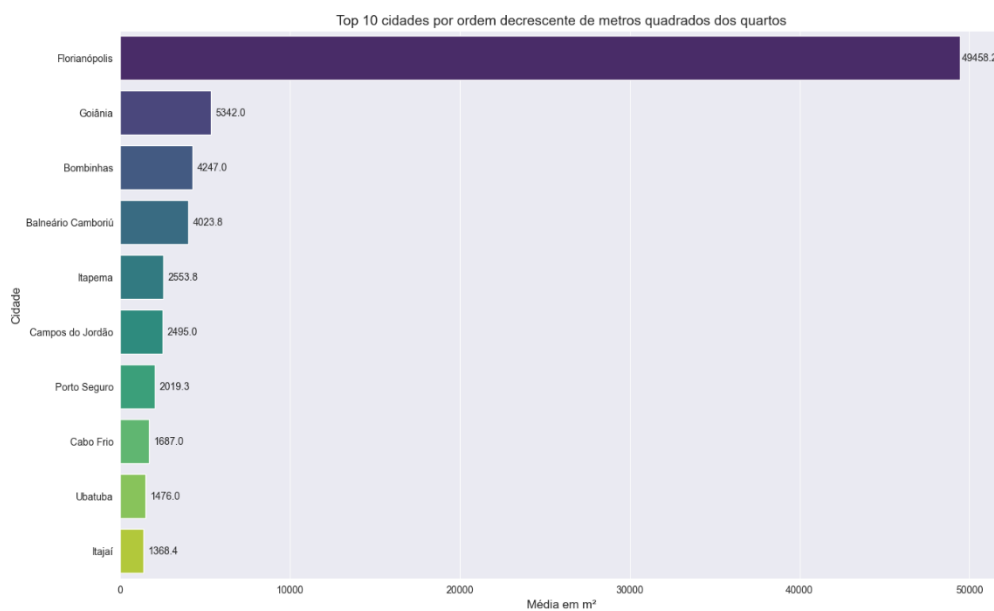


## Pergunta 2

Ordene as cidades em ordem decrescente de metros quadrados.

*Resposta:*

Florianópolis liderou com a maior soma de metros quadrados (49.473,22 m<sup>2</sup>), enquanto cidades como Águas Brancas apresentaram 30 m<sup>2</sup> e Camboriú, 43 m<sup>2</sup>.



## Pergunta 3

Quais cidades têm mais avaliações?

*Resposta:*

Florianópolis teve o maior número de avaliações (87.353), seguida por Anitápolis e Campos do Jordão. Muitas cidades tiveram 0 avaliações

city_name ↕	number_of_ratings ↕
Florianópolis	87353.0
Anitápolis	2106.0
Campos do Jordão	1815.0
Bombinhas	1126.0
Goiânia	531.0
Balneário Camboriú	54.0
Itajaí	23.0
Itapema	19.0

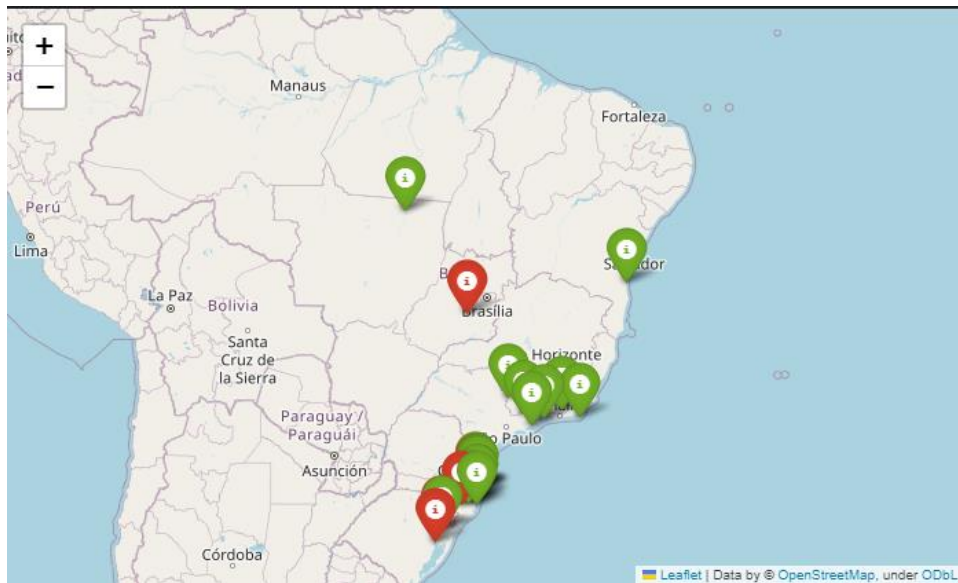
#### Pergunta 4

**Quais cidades têm a maior e a menor média de avaliações? Existe alguma explicação para isso?**

*Resposta:*

Camboriú teve a menor média de avaliações (1.0), e Ilhéus, Angra dos Reis e Brasília as maiores (10.0). A variação nas médias pode ser influenciada por fatores como número de avaliações, qualidade dos serviços e atrações turísticas. Mas parece que destinos mais próximos do litoral têm maiores médias de avaliação.

↕	cidade ↕	media_avaliacao ↕	latitude ↕	longitude ↕
0	Angra dos Reis	10.000000	-23.006397	-44.316326
1	Anitápolis	9.400000	-27.904118	-49.128740
2	Balneário Camboriú	8.150000	-26.992439	-48.633978
3	Bombinhas	8.867647	-27.151916	-48.487569
4	Brasília	10.000000	-10.333333	-53.200000
5	Cabo Frio	9.500000	-22.880437	-42.018923
6	Camboriú	1.000000	-27.027360	-48.653566
7	Campos do Jordão	9.100000	-22.738299	-45.590377
8	Canela	9.000000	-29.363609	-50.809684

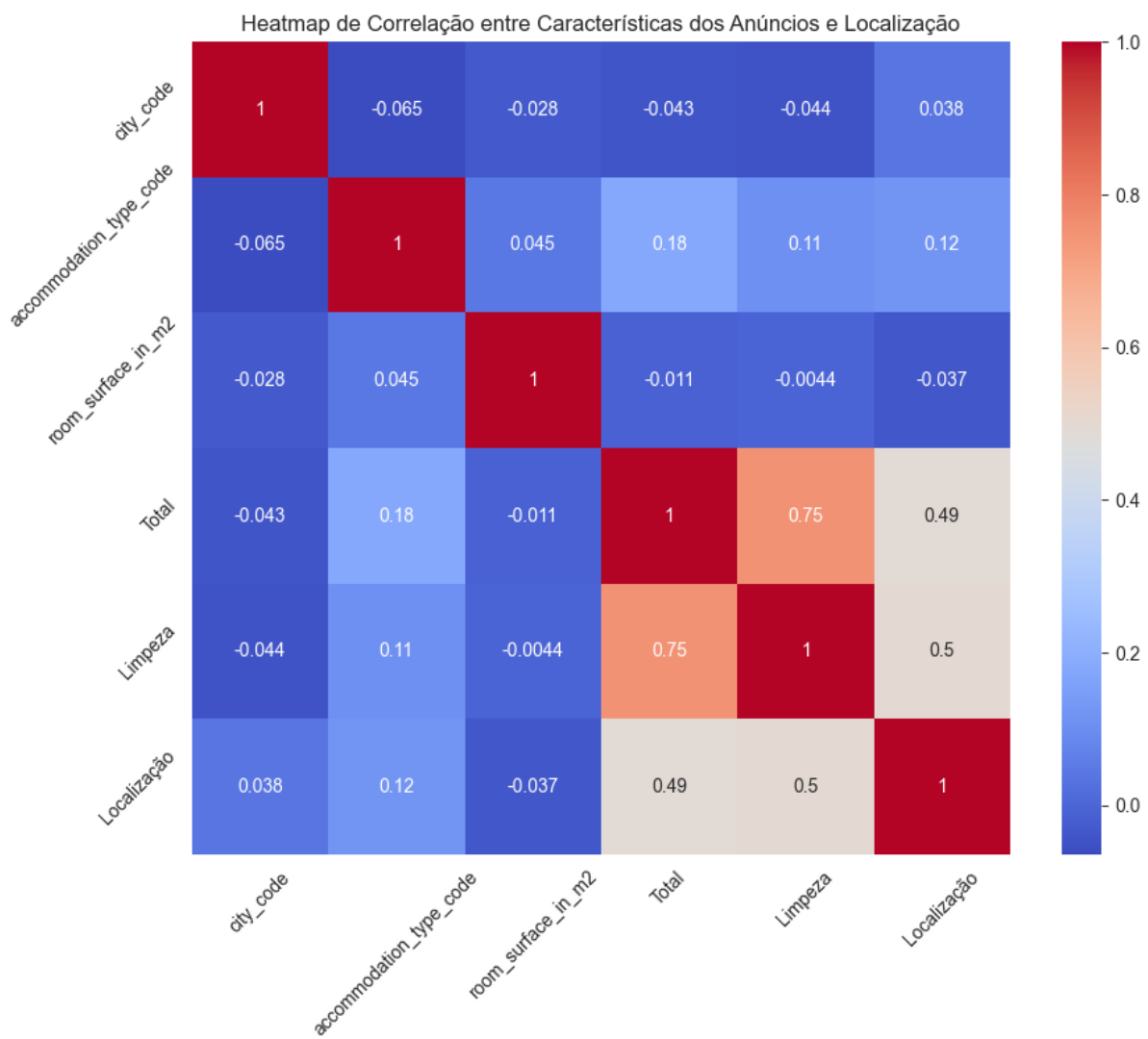


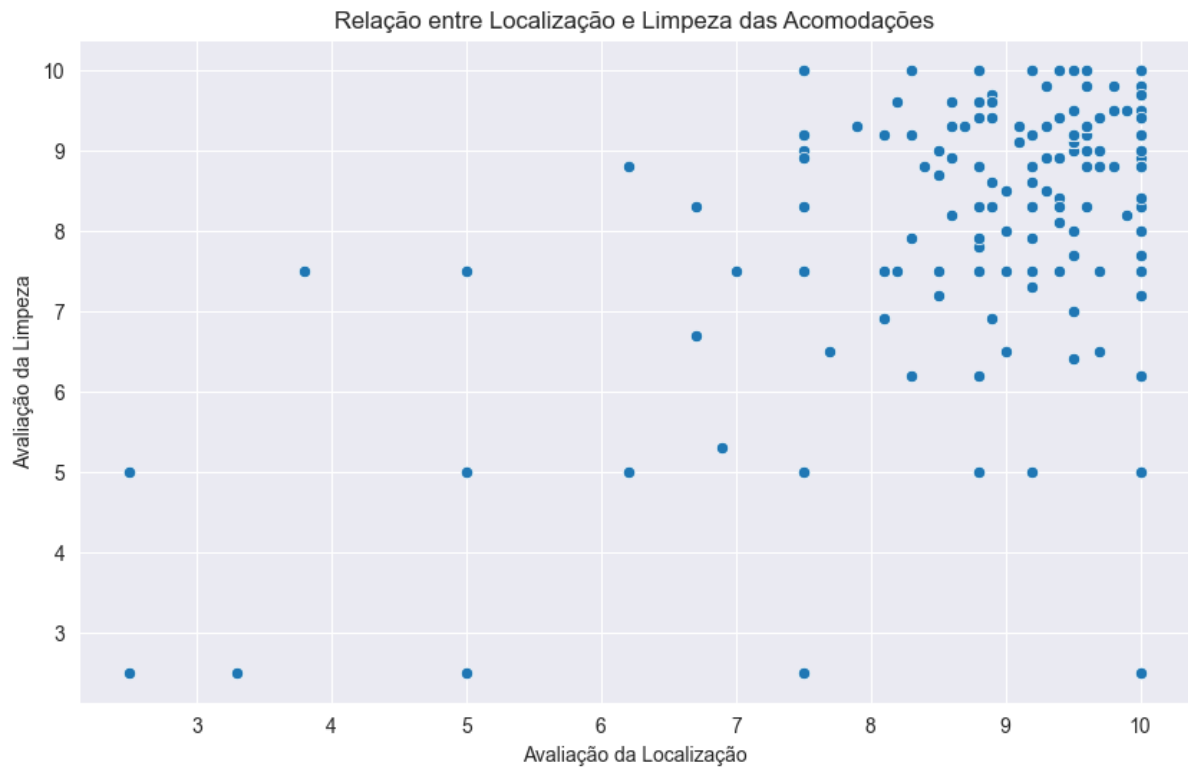
### Pergunta 5

**Existem correlações entre as características de um anúncio e sua localização?  
Quais? Explique.**

*Resposta:*

As análises indicaram correlações muito baixas entre a localização da cidade, o tipo de acomodação, a superfície do quarto e a nota total. Isso sugere que essas características não estão fortemente relacionadas com a localização geográfica. Correlação moderada entre Localização e Limpeza: A coluna Localização tem uma correlação positiva moderada com Limpeza. Isso sugere que as acomodações localizadas em determinadas áreas podem ser percebidas como mais limpas, o que pode estar relacionado a padrões de higiene específicos de determinadas regiões ou à percepção dos hóspedes sobre diferentes áreas.





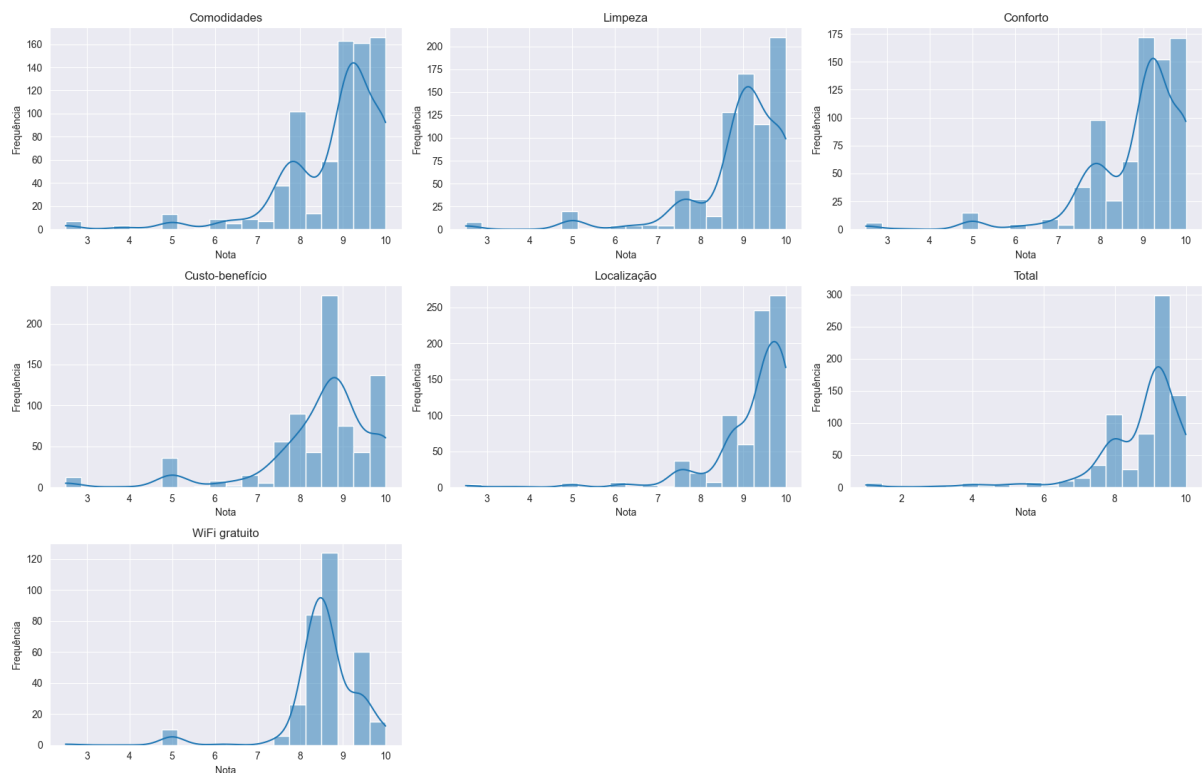
### Pergunta 6

**Existem relações entre a nota do anúncio e os recursos disponíveis no imóvel?  
Quais? Explique.**

*Resposta:*

Algumas comodidades parecem ter um efeito positivo nas avaliações dos hóspedes, mas essa influência varia e não é consistente em todas as categorias de avaliação. É interessante realizar uma análise mais aprofundada.





## Pergunta 7

**Existe alguma relação entre a nota recebida e a localização? Quais? Explique.**

*Resposta:*

Existe uma correlação moderada, indicando que: embora a localização seja um fator relevante na determinação da nota geral de uma acomodação, ela não é o único fator decisivo. Outros aspectos, como as comodidades, a limpeza e o conforto, também desempenham papéis importantes na avaliação geral.

## Pergunta 8

**O que você pode inferir sobre as notas dos imóveis?**

*Resposta:*

Estatísticas descritivas:

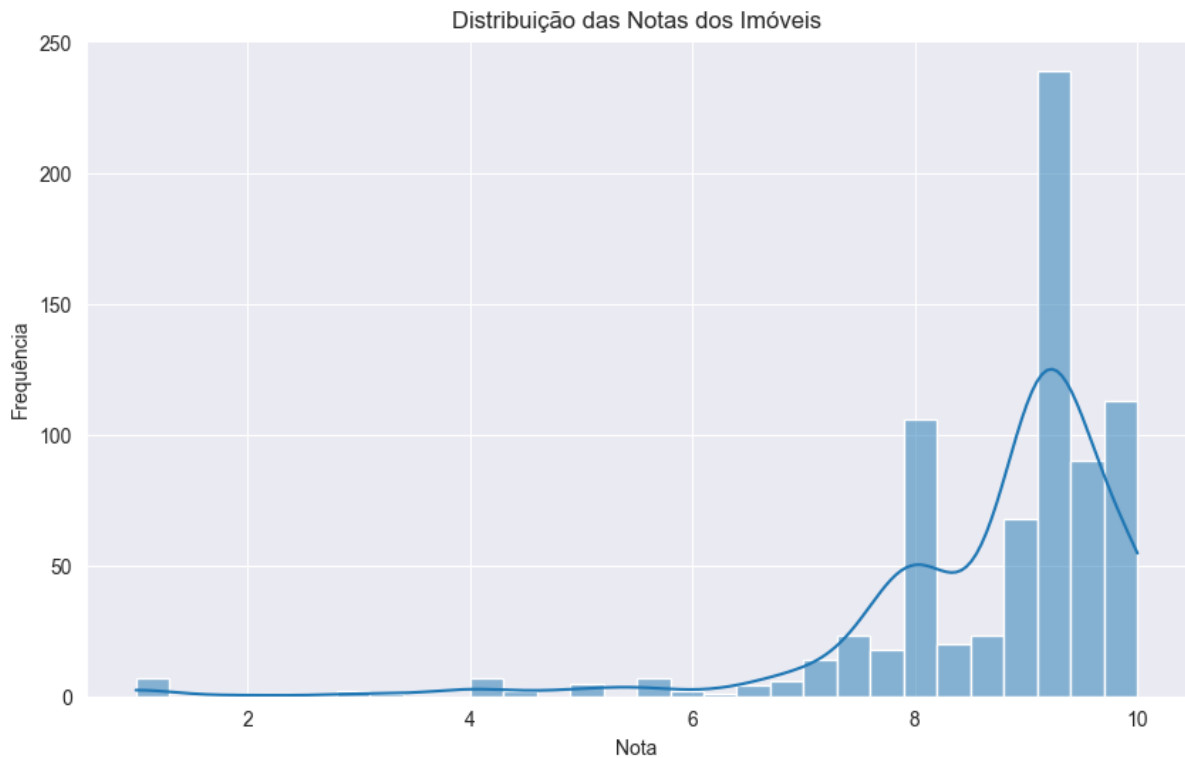
- **Quantidade de Avaliações:** Temos um total de 759 avaliações em nosso conjunto de dados.
- **Média das Notas:** A média das notas é de aproximadamente 8.71, o que sugere que, em geral, as avaliações tendem a ser positivas.
- **Variação das Notas:** O desvio padrão é cerca de 1.36, indicando uma moderada variação nas notas.
- **Mínimo e Máximo das Notas:** As notas variam de 1 a 10, o que está de acordo com um sistema de avaliação de 1 a 10.
- **Quartis:** 25% das notas são iguais ou inferiores a 8.0, 50% (a mediana) são iguais ou inferiores a 9.1, e 75% são iguais ou inferiores a 9.4. Isso indica que a maioria das avaliações se concentra no espectro mais elevado.

#### Histograma das Notas:

O histograma nos mostra como as notas dos imóveis estão distribuídas. Observamos uma concentração de notas elevadas, próximas a 9 e 10, o que sugere que a maioria dos imóveis recebe avaliações positivas. A distribuição é ligeiramente inclinada para a esquerda, com menos imóveis recebendo notas mais baixas.

#### Conclusão:

Com base nas informações apresentadas, podemos concluir que, de modo geral, os imóveis têm sido bem avaliados pelos hóspedes. A predominância de notas mais altas indica uma satisfação geral com as acomodações oferecidas. A mediana acima de 9 sugere que a maioria dos hóspedes está bastante satisfeita com os imóveis.



### Pergunta 9

Quais são os anúncios que parecem críticos? Explique.

Resposta:

- **Áreas com Problemas:** Os anúncios críticos tendem a ter problemas principalmente nas áreas de limpeza e conforto.
- **Localização Bem Avaliada:** É interessante observar que a localização é geralmente bem avaliada, mesmo nos anúncios com notas gerais mais baixas.
- **Tipo de Imóvel:** A maioria desses anúncios críticos são apartamentos, especialmente em cidades como Florianópolis e Goiânia.

Implicações:

Essas informações podem ser úteis para os proprietários desses imóveis focarem em melhorias específicas, principalmente em limpeza e conforto, para elevar suas avaliações gerais. Melhorias nessas áreas podem ajudar a melhorar a satisfação dos hóspedes e, conseqüentemente, suas avaliações e classificações.

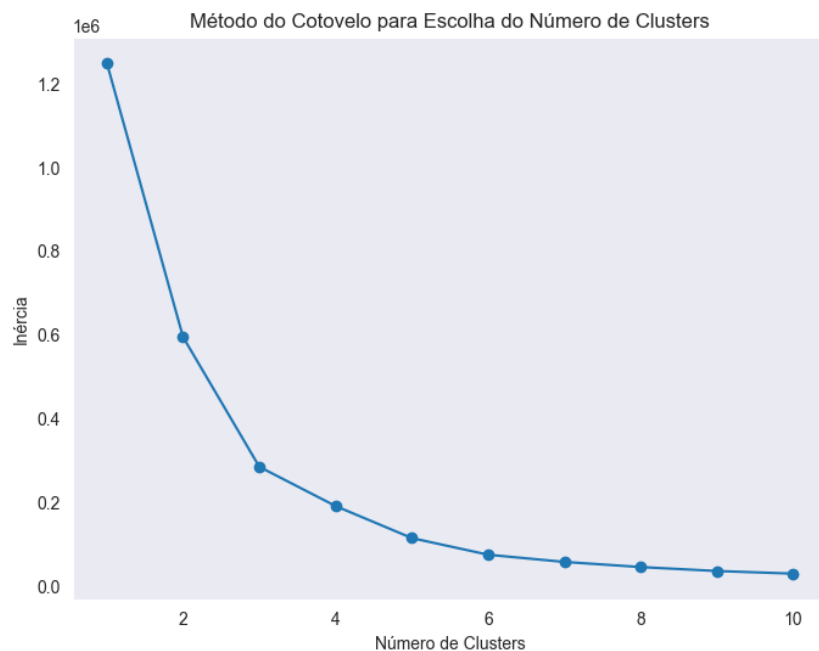
## Pergunta 10

**Quais outras análises você faria desses dados? Use sua criatividade.**

*Resposta:*

Neste cluster, encontramos imóveis com quartos muito espaçosos. Uma característica notável é o tamanho generoso dessas acomodações, o que pode torná-las mais caras em comparação com outros imóveis.

No entanto, as avaliações deste grupo são relativamente mais baixas. Isso sugere que, apesar do espaço amplo oferecido, os hóspedes podem não estar tão satisfeitos em comparação com os imóveis de outros clusters.





## Pergunta 11

**Como você projetaria um dashboard para mostrar essas informações?**

*Resposta:*

Distribuição de Notas por Cidade

- Um gráfico de barras interativo mostraria a distribuição das notas de avaliação para cada cidade. Poderia selecionar uma cidade específica e ver como as avaliações são distribuídas nessa localização.

Distribuição de Notas por Tipo de Acomodação

- Um gráfico de barras empilhadas permitiria que os dados fossem explorados como as notas variam entre diferentes tipos de acomodação, como apartamentos, casas, hotéis, etc.

Distribuição de Notas por Tamanho do Imóvel

- Um gráfico de histograma interativo exibiria a distribuição das notas com base no tamanho do imóvel em metros quadrados.

## Análise Temporal

- Um gráfico de linha interativo representaria a tendência das avaliações ao longo do tempo. Poderiam selecionar um período específico e analisar as mudanças nas notas.

## Mapa de Calor de Avaliações

- Um mapa interativo de calor mostraria visualmente as áreas mais populares e as menos populares com base nas avaliações dos hóspedes. As cores mais quentes indicariam áreas com avaliações mais altas.

## Feedback dos Hóspedes

- Um gráfico de dispersão interativo permitiria que os usuários explorassem o feedback dos hóspedes em relação a diferentes aspectos, como limpeza, conforto, serviço, etc., para identificar áreas de força e fraqueza.

A inclusão desses gráficos interativos tornaria o dashboard uma ferramenta poderosa para a análise de dados de avaliações, oferecendo insights valiosos de maneira visual e intuitiva.

## Pergunta 12

**Quais outras informações/dados você relacionaria com essas bases caso tivesse acesso?**

*Resposta:*

- **Dados Temporais:** A inclusão das datas das avaliações nos permitiria examinar tendências sazonais e mudanças de longo prazo na qualidade dos imóveis.
- **Informações de Preço:** Analisar os preços em relação às avaliações ajudaria a entender a percepção de valor dos hóspedes e a relação preço-qualidade.
- **Dados Demográficos dos Hóspedes:** Compreender quem são nossos hóspedes e suas preferências nos ajudaria a personalizar serviços e aprimorar a experiência do cliente.

- **Taxas de Ocupação e Histórico de Manutenção:** Esses dados nos permitiriam correlacionar a demanda e as condições dos imóveis com as avaliações, identificando como diferentes fatores influenciam a satisfação do hóspede.

### **Pergunta 13**

**(Extra) Com base nesses dados e nos anúncios fornecidos, como você melhoraria as notas?**

*Resposta:*

1. **Identificação de Áreas para Melhoria:** Primeiramente, identificaria áreas específicas para melhorias com base nas avaliações recebidas. Focaria em aspectos como comodidades, serviço ao cliente e limpeza, que são fundamentais para a satisfação dos hóspedes.
2. **Implementação de Melhorias Direcionadas:** Implementaria melhorias direcionadas nessas áreas identificadas. Trabalharia para aprimorar a qualidade das comodidades oferecidas, treinar a equipe de atendimento ao cliente e elevar os padrões de limpeza e manutenção.
3. **Destaque dos Pontos Fortes:** Utilizaria os anúncios para destacar os pontos fortes de cada hotel. Criaria uma imagem positiva realçando as melhorias feitas e as características exclusivas que tornam cada hotel especial.
4. **Resposta Construtiva aos Feedbacks:** Enfatizaria a importância de responder aos feedbacks dos clientes de forma construtiva e proativa. Isso demonstraria nosso compromisso com a excelência e o aprimoramento contínuo.

### **Feedback do processo**

No geral, achei a experiência bastante interessante e desafiadora. Enfrentei algumas questões que me fizeram pensar bastante, o que foi ótimo para meu crescimento e aprendizado.

O que realmente me empolgou foi a liberdade que tive para abordar as tarefas. Isso me permitiu explorar diferentes caminhos e encontrar soluções criativas para os problemas apresentados.