

Patrick de Sousa Faria Gonçalves

20 de novembro de 2023

Seazone Data Challenge

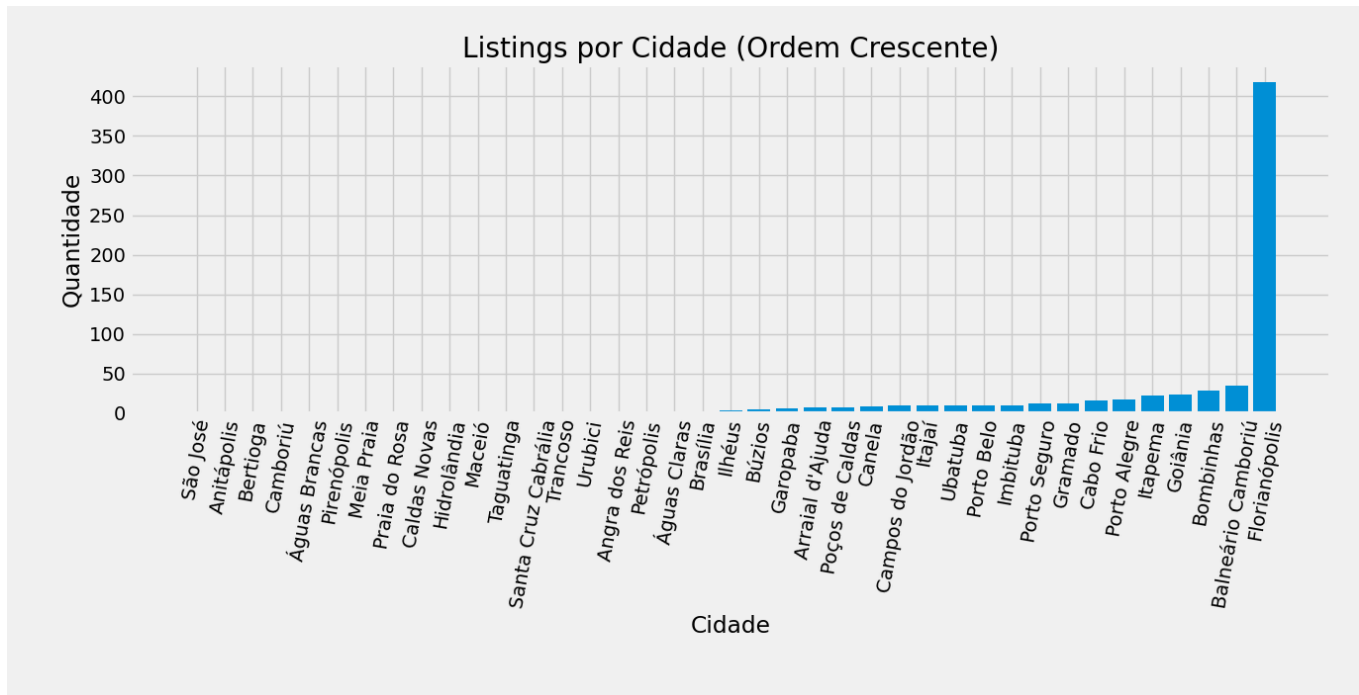
Desafio com o objetivo de analisar duas tabelas de dados ("desafio_details.csv" e "desafio_ratings.csv") e responder algumas perguntas

Início do desafio

Para iniciar o desafio, optei por utilizar Python como a linguagem principal para as análises. Posteriormente, converti as duas tabelas em DataFrames utilizando a biblioteca Pandas.

Inicialmente, eliminei algumas informações desnecessárias para aprimorar a efetividade das análises, como linhas duplicadas e nomes de hotéis que continham o termo "excluir" em seus títulos. Em seguida, procedeu à transformação dos DataFrames de volta para o formato CSV, nomeando os arquivos resultantes como "desafio_details_new.csv" e "desafio_ratings_new.csv".

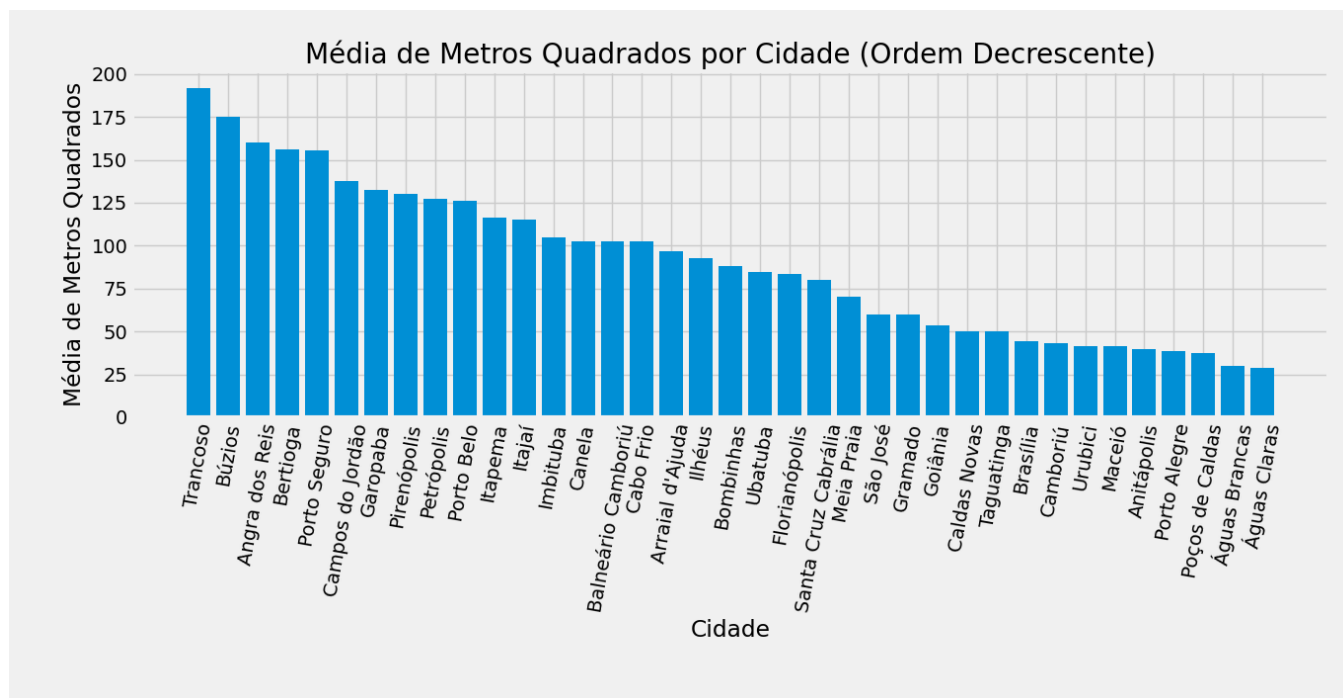
1. Ordene as cidades em ordem crescente de número de listings:



1. São José - 1
2. Anitápolis - 1
3. Bertioga - 1
4. Camboriú - 1
5. Águas Brancas - 1
6. Pirenópolis - 1
7. Meia Praia - 1
8. Praia do Rosa - 1
9. Caldas Novas - 1
10. Hidrolândia - 1
11. Maceió - 1
12. Taguatinga - 1
13. Santa Cruz Cabrália - 1
14. Trancoso - 2
15. Urubici - 2
16. Angra dos Reis - 2
17. Petrópolis - 2
18. Águas Claras - 2
19. Brasília - 3
20. Ilhéus - 4
21. Búzios - 5
22. Garopaba - 6
23. Arraial d'Ajuda - 8
24. Poços de Caldas - 8
25. Canela - 9
26. Campos do Jordão - 10
27. Itajaí - 10
28. Ubatuba - 10
29. Porto Belo - 10
30. Imbituba - 10
31. Porto Seguro - 12
32. Gramado - 13
33. Cabo Frio - 16
34. Porto Alegre - 18
35. Itapema - 22
36. Goiânia - 24
37. Bombinhas - 29
38. Balneário Camboriú - 35
39. Florianópolis - 418

OBS: Apenas as listagens que continham URLs foram utilizadas a partir do arquivo "desafio_ratings_new.csv". Para associar o nome das cidades a cada URL, empreguei a função pd.merge(), obtendo os dados da cidade do arquivo "desafio_details_new.csv" com base no hotel_id correspondente.

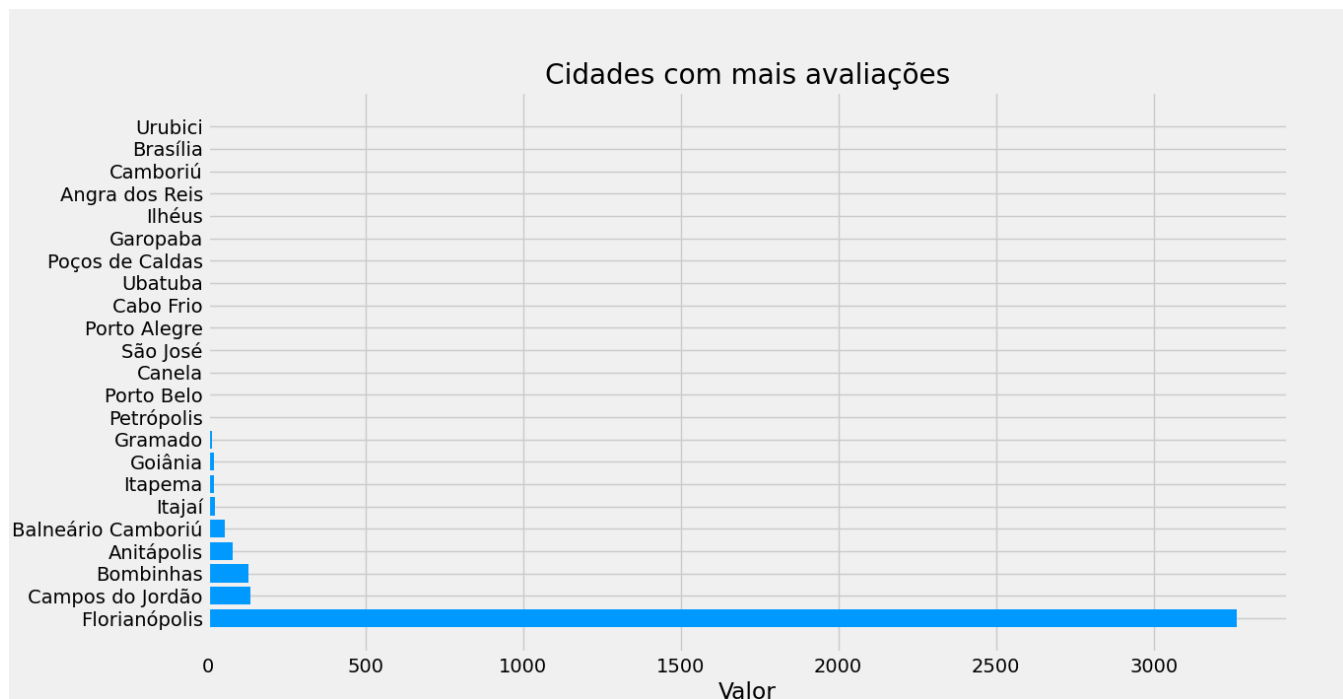
2. Ordene as cidades em ordem decrescente de metros quadrados:



1. Trancoso - 192.0
2. Búzios - 174.8
3. Angra dos Reis - 160.0
4. Bertioga - 156.0
5. Porto Seguro - 155.33
6. Campos do Jordão - 137.5
7. Garopaba - 132.67
8. Pirenópolis - 130.0
9. Petrópolis - 127.33
10. Porto Belo - 126.1
11. Itapema - 116.08
12. Itajaí - 115.31
13. Imbituba - 105.0
14. Canela - 102.67
15. Balneário Camboriú - 102.33
16. Cabo Frio - 102.31
17. Arraial d'Ajuda - 96.44
18. Ilhéus - 92.44
19. Bombinhas - 88.07
20. Ubatuba - 84.55
21. Florianópolis - 83.41
22. Santa Cruz Cabrália - 80.0
23. Meia Praia - 70.0
24. São José - 60.0
25. Gramado - 59.61
26. Goiânia - 53.44
27. Caldas Novas - 50.0
28. Taguatinga - 50.0
29. Brasília - 44.5
30. Camboriú - 43.0
31. Urubici - 41.67
32. Maceió - 41.5
33. Anitópolis - 40.0
34. Porto Alegre - 38.59
35. Poços de Caldas - 37.13
36. Águas Brancas - 30.0
37. Águas Claras - 29.0

OBS: Hidrolândia e Praia da Rosa foram eliminadas por não possuir informações.

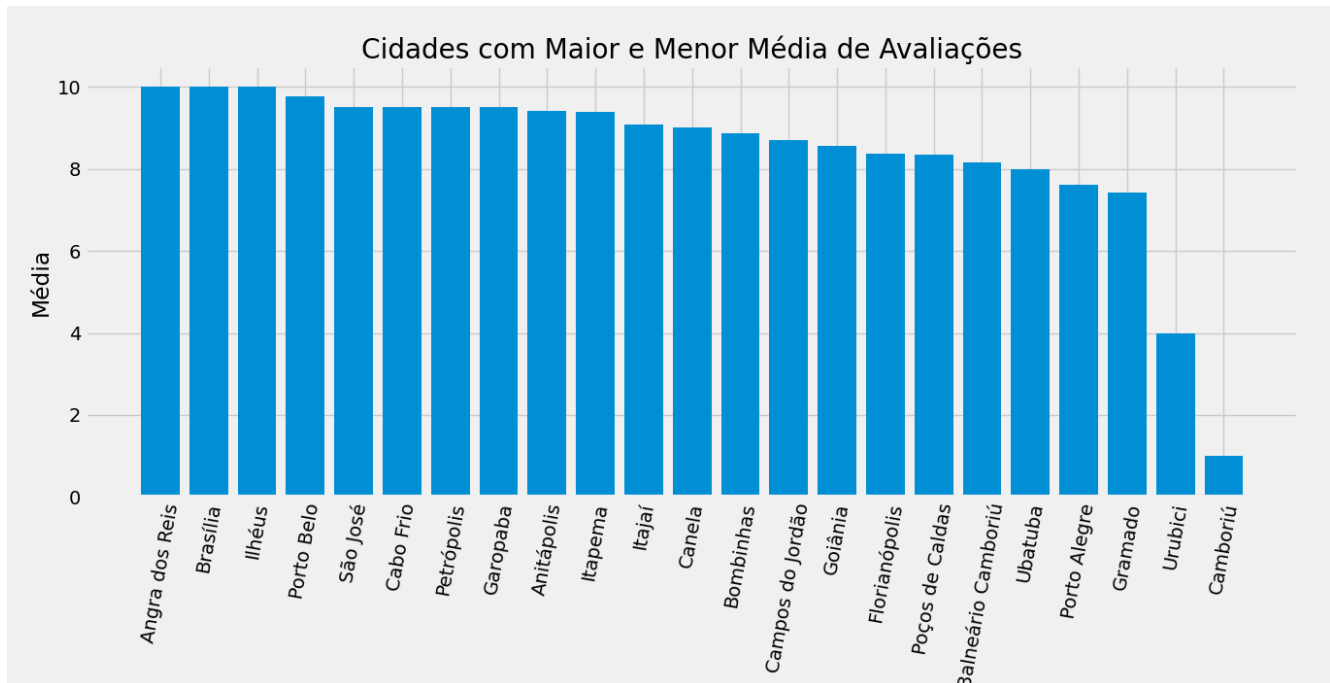
3. Quais cidades têm mais avaliações?



1. Florianópolis - 3262
2. Campos do Jordão - 135
3. Bombinhas - 128
4. Anitápolis - 78
5. Balneário Camboriú - 54
6. Itajaí - 23
7. Itapema - 19
8. Goiânia - 19
9. Gramado - 15
10. Petrópolis - 8
11. Porto Belo - 8

12. Canela - 6
13. São José - 6
14. Porto Alegre - 5
15. Cabo Frio - 4
16. Ubatuba - 4
17. Poços de Caldas - 3
18. Garopaba - 3
19. Ilhéus - 3
20. Angra dos Reis - 1
21. Camboriú - 1
22. Brasília - 1
23. Urubici - 1

4. Quais cidades têm a maior média de avaliações? E a menor média?



Cidade com a maior média: Brasília - Angra dos Reis - Ilhéus - Média: 10.

Cidade com a menor média: Camboriú - Média: 1.

a. Existe alguma explicação para isso? Conjecture.

A prevalência de anúncios em Florianópolis influencia uma média inferior, em contraste com Brasília, Angra dos Reis e Ilhéus, que, apesar de contar com poucos anúncios, destacam-se por avaliações excepcionais. Nesse contexto, Brasília, Angra dos Reis e Ilhéus alcançaram a primeira posição devido à qualidade superior de seus escassos anúncios. Por outro lado, em Camboriú, apesar do número reduzido de anúncios, a média de avaliações não se revela positiva.

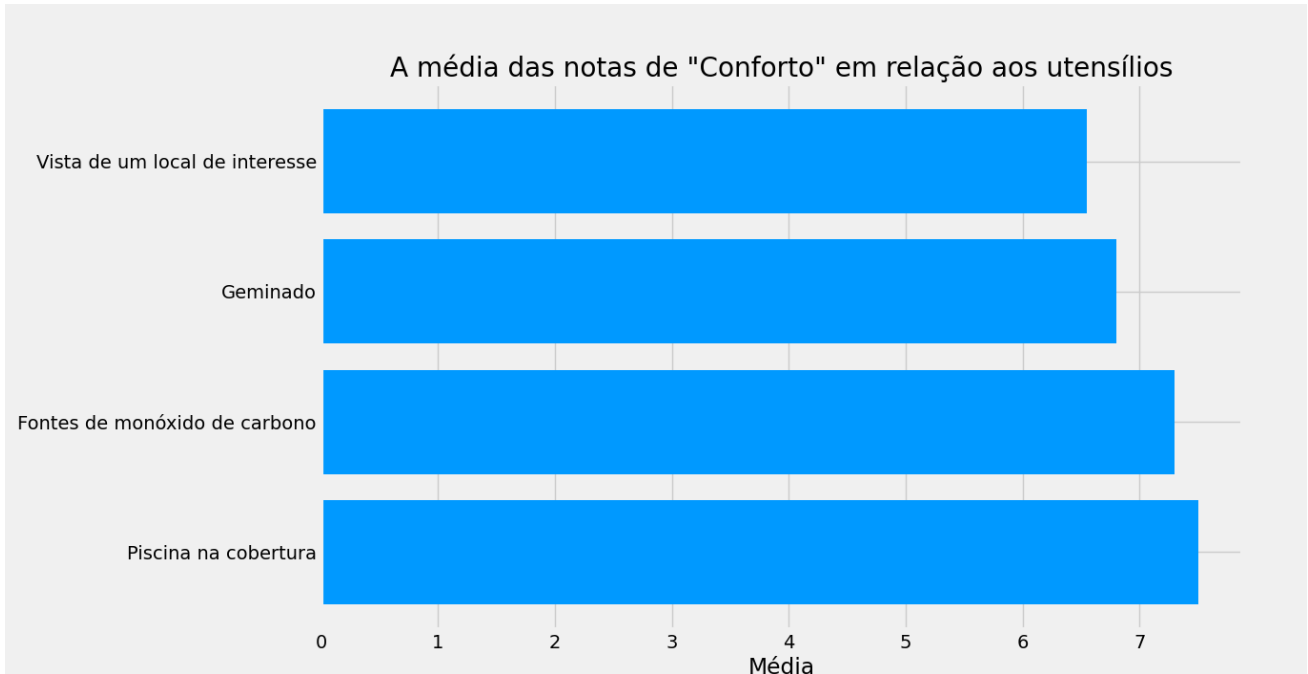
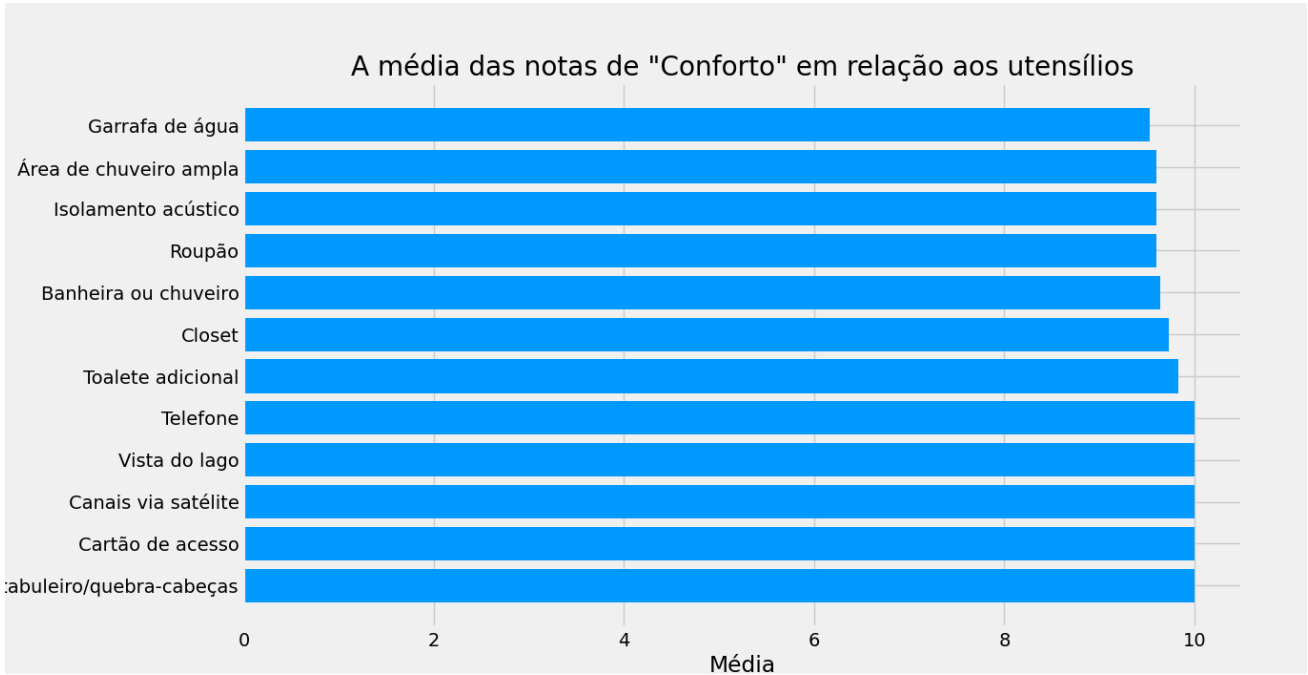
5. Existem correlações entre as características de um anúncio e a sua localização?

a. Quais? Explique.

Sim, todas as características de um imóvel exercem influência em sua avaliação, e a localização é uma peça chave nesse quebra-cabeça. Mesmo que o imóvel apresente excelentes comodidades, limpeza e outros atrativos, obtendo avaliações positivas, a proximidade em relação ao centro ou a pontos turísticos específicos pode impactar negativamente em sua taxa de ocupação. É possível que, apesar das qualidades do imóvel, a localização menos estratégica resulte em uma ocupação abaixo do desejado,

demonstrando a importância de considerar não apenas as características internas, mas também a acessibilidade e conveniência proporcionadas pela localização do imóvel.

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

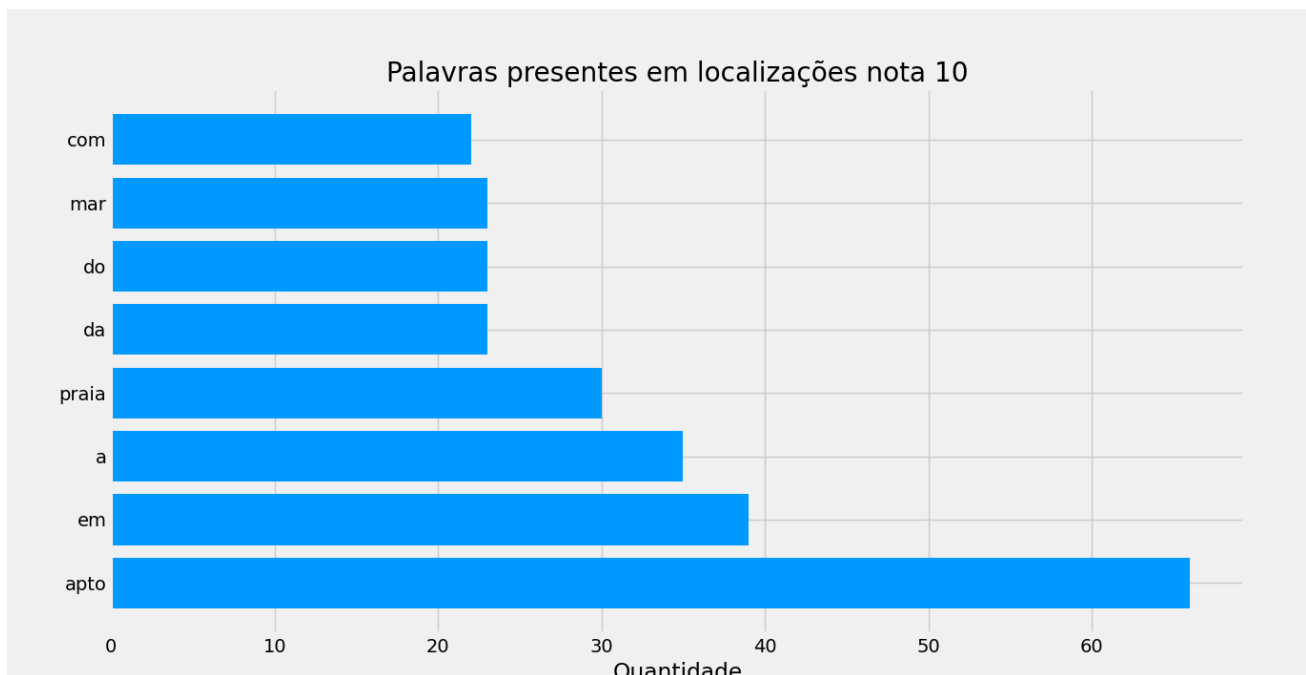


Observando atentamente, torna-se evidente que a interação entre os utensílios presentes em uma residência está diretamente ligada à qualidade do conforto oferecido durante a hospedagem. Mesmo o utensílio com a avaliação mais baixa ainda mantém uma nota superior à metade, indicando que, ao aprimorar e investir em mais utensílios, é possível

elevar ainda mais a experiência dos hóspedes. Essa melhoria não apenas contribui para atrair um número maior de visitantes, mas também se mostra crucial, uma vez que uma localização privilegiada, um bom wifi gratuito, limpeza exemplar e um excelente custo-benefício perdem seu valor quando confrontados com um nível de conforto insatisfatório.

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

Lugares situados próximo à praia ou em áreas turísticas ganham destaque pela conveniência de acesso. Ao analisar as palavras mais frequentes nos nomes dos anúncios em relação às avaliações de nota 10, identifiquei padrões que levaram aos seguintes resultados:



Como pode ser observado, as palavras 'mar' e 'praia' destacam-se entre as mais frequentes nos anúncios, ressaltando a importância de especificar a localização. Ao considerar a nota geral, a média dessas duas palavras foi a seguinte:

	hotel_name	Localização
447	praia	9.078947
368	mar	9.098039

Ressaltando ainda mais a importância de uma hospedagem com uma boa localização.

8. O que você pode inferir sobre as notas dos imóveis?

As avaliações dos imóveis refletem a importância da localização, mas também revelam a necessidade de uma abordagem equilibrada para atender às expectativas dos hóspedes.

Embora propriedades bem localizadas geralmente recebem notas mais altas, características como comodidades, limpeza e custo-benefício desempenham papéis cruciais. Uma localização excelente pode, em parte, compensar avaliações menos favoráveis em outras categorias. Além disso, a presença de WiFi gratuito emerge como um fator significativo nas avaliações, indicando a valorização da conectividade pelos hóspedes. A análise detalhada também revela variações nas avaliações entre diferentes tipos de acomodações e a influência do número de avaliações na confiabilidade dos resultados. Essas nuances destacam a importância de uma gestão holística e personalizada para garantir avaliações positivas e a satisfação duradoura dos clientes.

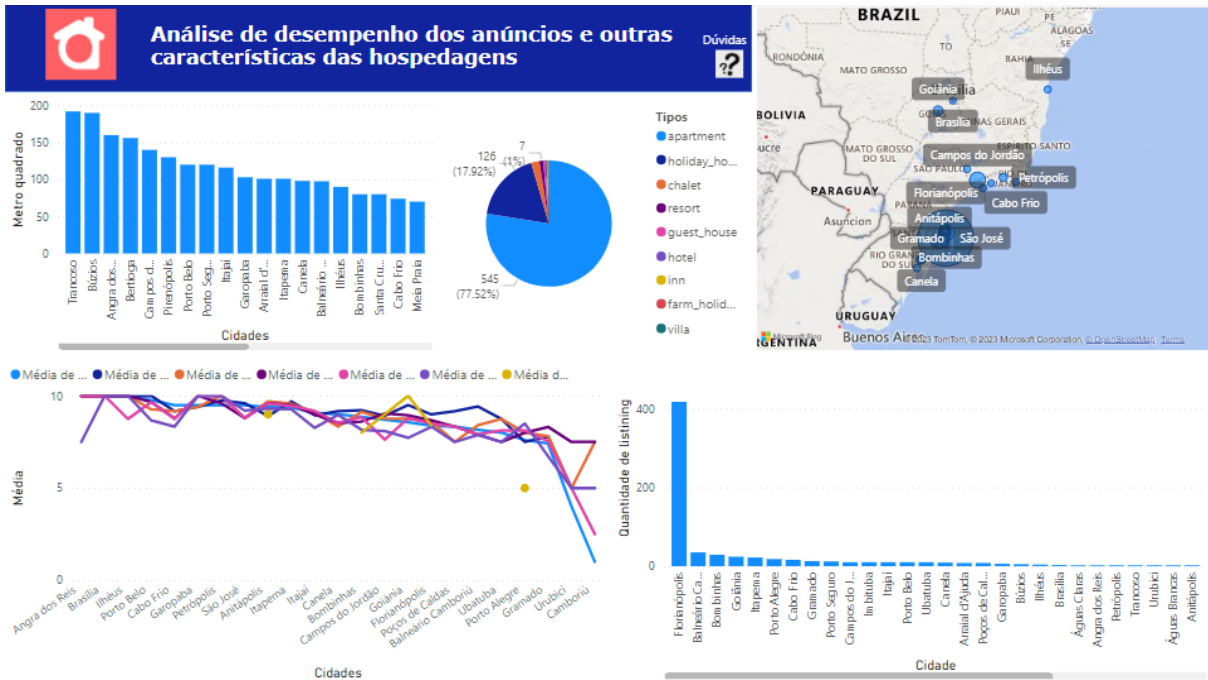
9. Quais são os anúncios que te parecem críticos? a. Explique.

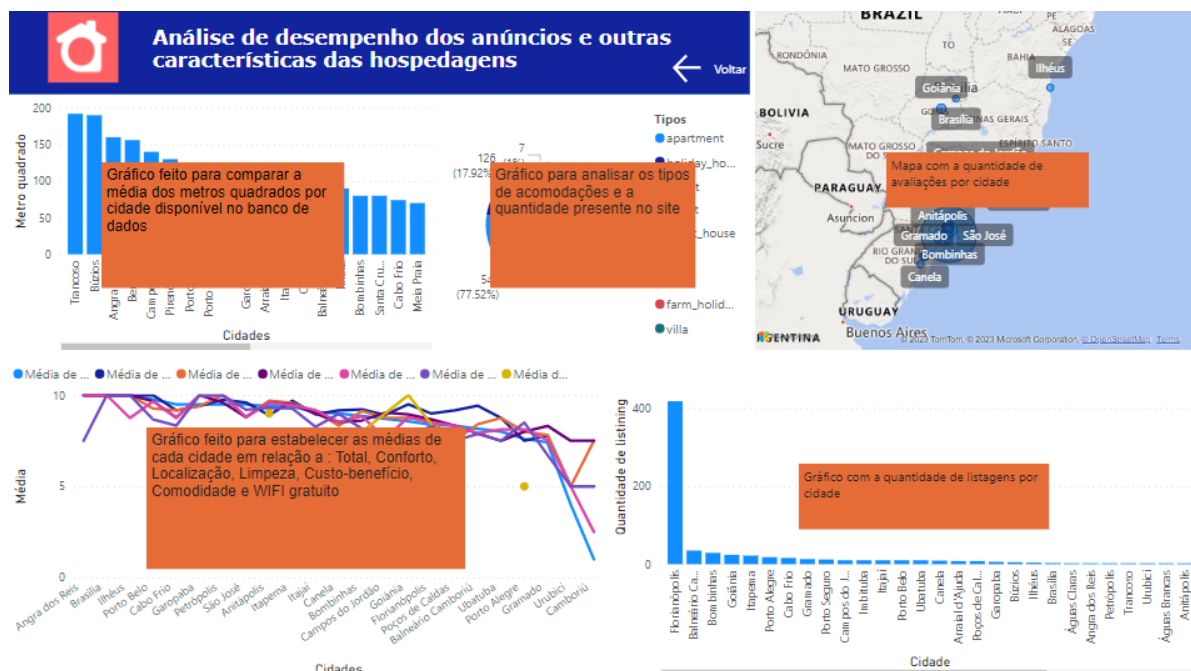
Anúncios com avaliações persistentemente baixas são considerados críticos, sinalizando a presença de questões substanciais que têm um impacto significativo e negativo na experiência e satisfação dos hóspedes. Essas avaliações indicam não apenas a existência de problemas, mas também a necessidade urgente de intervenção e melhoria.

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

Uma análise pertinente envolveria a comparação entre 'room_surface_in_m2' e 'comodidades', a fim de investigar se o tamanho de uma hospedagem tem impacto em suas avaliações. Além disso, considerar a relação entre 'accommodation_type' e 'number_of_ratings', somando a quantidade de avaliações por tipo, proporciona insights sobre tendências preferenciais entre apartamentos e casas, ajudando a entender as preferências dos hóspedes em relação ao número de avaliações recebidas.

11. Como você projetaria um dashboard para mostrar essas informações?





Power BI não permite compartilhar link, devido a conta ser gratuita.

O arquivo do Power BI está em : <https://github.com/P4TRIIICK/Desafio-Seazone>

12. Quais outras informações/dados você relacionaria com essas bases, caso tivesse acesso?

- Estação do Ano, para explorar se as estações do ano afetam as avaliações, considerando fatores climáticos ou eventos sazonais que possam influenciar a experiência dos hóspedes.
- Preços médios, analisando se os hóspedes estão procurando hospedagens por localização ou preço ou até mesmo um meio termo entre os dois
- Comentários dos hóspedes, visando encontrar os pontos que prejudicam a avaliação de um local e consecutivamente melhorar esse ponto.
- Tipo de Viagem, com o objetivo diferenciar as avaliações com base no tipo de viagem (negócios, lazer, familiar) para entender as expectativas específicas de cada categoria de hóspede.

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

Investir em utensílios aprimora significativamente as notas de conforto, proporcionando aos hóspedes uma experiência mais acolhedora. Oferecer uma conexão WiFi rápida e confiável é uma estratégia eficaz para atender às expectativas dos hóspedes, garantindo conectividade sem interrupções. Uma limpeza eficiente não apenas torna o local mais agradável, mas também cria a possibilidade de fidelizar o cliente, incentivando seu retorno. Destacar pontos turísticos próximos ao local é uma prática que enriquece a experiência do hóspede, proporcionando informações valiosas sobre as atrações ao redor. Além disso, a

implementação de promoções especiais não apenas atrai mais pessoas para o local, mas também cria oportunidades para experiências diferenciadas, incentivando o interesse e a satisfação dos visitantes.

Feedback sobre o desafio

O desafio me cativou profundamente, pois não se tratava de uma tarefa fácil, acessível a qualquer um, mas também não era excessivamente difícil. Consegui completá-lo antes do prazo estimado, o que reforçou minha confiança, e destacou minhas habilidades na área de análise de dados, utilizando Python e a biblioteca Pandas para uma análise abrangente. É justo admitir que enfrentei alguns obstáculos durante a resolução de certas questões, especialmente na questão 6, onde inicialmente tentei utilizar as funções `"pd.get_dummies()"` e `".corr()"`, mas acabei gerando uma série de colunas nos dados. Diante desse desafio, optei por uma abordagem diferente que se mostrou mais eficaz. Expresso minha gratidão à Seazone por proporcionar essa experiência enriquecedora, que ampliou significativamente meu conhecimento. Estou ansioso para enfrentar novos desafios semelhantes, pois acredito que cada um deles contribui para meu crescimento profissional e pessoal.