

Antônio Lucas Dantas de Abrantes

09 de Novembro de 2023

Seazone Data Challenge

O desafio consiste em analisar os dados de “details” e de “rating” de anúncios no Booking, com o intuito de responder uma série de perguntas.

INÍCIO DESAFIO

Começo esse desafio criando um diretório no github, para os futuros commits, logo após no “vscode” faço o upload dos arquivos e criei um jupyter notebook para melhor análise dos dados. Além disso guardei os dados nas variáveis “df_Details” e “df_Ratings”.

Os dados em CSV não estavam sendo lidos direito, então, transformei para xlsx assim consegui trabalhar com os dados de forma mais fluida e precisa, sem erros.

ANÁLISE DOS DADOS

Início a análise importando a biblioteca pandas da linguagem Python. Faço a criação dos DataFrames e utilizo `df.head()` para visualizar melhor os dados na tabela e iniciar a limpeza. Removo todos os zeros na coluna 'number_of_ratings', pois esses zeros podem interferir nas análises futuras. Realizo a concatenação das tabelas para poder trabalhar com os dados conjuntos e exparto essa tabela para verificação no Excel. Utilizo a função `duplicated()` e `sum()` para verificar se há dados duplicados na coluna

'hotel_id', que representa o identificador do anúncio, e em seguida removo as duplicatas.

Ao analisar no Excel, identifico que existem IDs que não estão duplicados, mas precisam ser excluídos devido à presença das palavras 'excluir' ou 'teste' no nome. Faço a exclusão desses registros e replico no 'hotel_name'. Observo também que algumas cidades não estão preenchidas. Não encontrando outra solução, opto por abrir link por link e digitar as cidades, o que me permite responder à terceira questão com dados concretos, evitando especulações. Seguem as questões respondidas.

QUESTÕES

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

Hidrolândia: 0; Praia do Rosa: 0; Trancoso: 0; Santa Cruz Cabrália: 1; Taguatinga: 1; Águas Brancas: 1; Angra dos Reis: 2; Brasília: 2; Cabo Frio: 2; Garopaba: 2; Ilhéus: 2; Bertioga: 3; Porto Alegre: 3; Porto Belo: 3; Canela: 4; Petrópolis: 4; Itajaí: 5; Ubatuba: 5; Campos do Jordão: 7; Goiânia: 8; Bombinhas: 10; Gramado: 11; Balneário Camboriú: 20; Florianópolis: 207;

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

1. Florianópolis: 19447.64 | 2. Itapema: 2158.80 | 3. Porto Seguro: 2019.26 | 4. Balneário Camboriú: 2017.80 | 5. Bombinhas: 1628.00 | 6. Cabo Frio: 1343.00 | 7. Goiânia: 1107.00 | 8. Imbituba: 1050.00 | 9. Porto Belo: 876.00 | 10. Búzios: 874.00 | 11. Arraial d'Ajuda: 868.00 | 12. Ubatuba: 671.00 | 13. Canela: 591.00 | 14. Porto Alegre:

509.00 | 15. Garopaba: 506.00 | 16. Itajaí: 504.00 | 17. Gramado: 491.00 | 18. Campos do Jordão: 445.00 | 19. Trancoso: 384.00 | 20. Petrópolis: 267.00 | 21. Poços de Caldas: 207.00 | 22. Ilhéus: 194.74 | 23. Angra dos Reis: 160.00 | 24. Bertioga: 156.00 | 25. Pirenópolis: 130.00 | 26. Uru bici: 95.00 | 27. Brasília: 89.00 | 28. Maceió: 83.00 | 29. Santa Cruz Cabrália: 80.00 | 30. Meia Praia: 70.00 | 31. Águas Claras: 58.00 | 32. Caldas Novas: 50.00 | 33. Taguatinga: 50.00 | 34. Águas Brancas: 30.00 | 35. Hidrolândia: 0.00 | 36. Praia do Rosa: 0.00

3. Quais cidades têm mais avaliações?

Anitápolis: 78.0; Urubici: 1.0; Garopaba: 1.0; Brasília: 1.0; Balneário Camboriú: 2.0; Porto alegre: 2.0; Porto Alegre: 3.0; Ilhéus: 3.0; Cabo Frio: 3.0; Petrópolis: 8.0; Porto Belo: 8.0; Poços de Caldas: 4.0; Ubatuba: 4.0; Goiânia: 19.0; Gramado: 23.0; Balneário Camboriú: 54.0; Bombinhas: 127.0; Campos do Jordão: 135.0; Florianópolis: 3270.0;

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

Cidade com a maior média: Brasília - Média: 10.0

Cidade com a menor média: Urubici - Média: 4.0

a. Existe alguma explicação para isso? Conjecture.

A quantidade de anúncios em Florianópolis contribui para uma média mais baixa, contrastando com Brasília, que, apesar de ter poucos anúncios, ostenta notas altas. Dessa forma, Brasília conquistou a primeira posição devido às excelentes avaliações

de seus poucos anúncios. Por outro lado, em Urubici, apesar do reduzido número de anúncios, a média de avaliações não é positiva.

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

Certamente, todas as características de um imóvel influenciam em sua avaliação, incluindo a localização. Mesmo que o imóvel apresente excelentes comodidades, limpeza, entre outros, com avaliações positivas, a distância em relação ao centro ou a alguns pontos turísticos pode impactar negativamente em sua taxa de ocupação, possivelmente aquém do desejado.

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

Ao explorar a matriz de correlação entre as diversas categorias de avaliações, destaquei algumas relações interessantes entre a nota do anúncio e os recursos oferecidos nos imóveis. Uma observação notável foi a forte correlação positiva entre a categoria 'Comodidades' e a avaliação total ('Total'). Isso sugere que os hóspedes valorizam significativamente propriedades que oferecem boas comodidades, o que faz todo sentido. Além disso, identifiquei uma correlação positiva entre as avaliações de 'Limpeza', 'Conforto' e 'Custo-benefício' com a avaliação total. Isso indica que manter o local limpo, confortável e com uma boa relação custo-benefício pode influenciar positivamente as avaliações totais dos hóspedes. A análise também revelou alguma correlação com a categoria 'Localização', embora não tão forte quanto nas outras

áreas. Parece que, embora a localização seja relevante, outros aspectos podem ter um impacto mais expressivo nas avaliações.

Em resumo, a estratégia de investir em comodidades, manter a limpeza e o conforto, além de oferecer um bom custo-benefício, parece ser fundamental para conquistar avaliações mais positivas. Embora a localização seja importante, essa análise sugere que concentrar esforços nessas áreas específicas pode ser uma abordagem eficaz para aprimorar as avaliações dos imóveis.

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

Ao examinarmos a possível interação entre a nota do anúncio e sua localização, alguns insights se destacam: A localização emerge como um fator crucial na determinação da nota geral atribuída pelos hóspedes. Quando o local é bem posicionado, há uma tendência de receber avaliações mais elevadas. No entanto, é interessante observar que, apesar da importância da localização, outros aspectos específicos, como comodidades, limpeza, conforto e custo-benefício, não apresentam uma relação tão pronunciada com a localização. Outra descoberta intrigante é que uma localização excepcional pode, até certo ponto, compensar deficiências em outras áreas, resultando em uma nota geral mais positiva.

Em resumo, concluímos que a localização desempenha um papel-chave nas avaliações. Aprimorar a nota geral implica não apenas garantir uma localização conveniente, mas também dedicar atenção aos outros detalhes importantes que

impactam a experiência do hóspede. Essa abordagem equilibrada se revela essencial para aprimorar a satisfação do cliente e, por conseguinte, as avaliações recebidas.

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

Ao analisar as notas dos imóveis, fica evidente que a localização desempenha um papel crucial. Imóveis bem localizados tendem a receber avaliações mais altas. Além disso, características específicas como comodidades, limpeza, conforto e custo-benefício também influenciam as notas, refletindo a variedade nas preferências dos hóspedes. Observamos que uma excelente localização pode compensar, em certa medida, avaliações menos favoráveis em outras categorias, resultando em notas gerais mais altas. A subjetividade das avaliações destaca a importância das preferências individuais dos hóspedes, onde o que é essencial para um hóspede pode não ter o mesmo peso para outro. Para melhorar as notas gerais, a estratégia pode envolver não apenas garantir uma localização ideal, mas também equilibrar outras comodidades e serviços para atender às expectativas variadas dos hóspedes. Em resumo, as notas não apenas refletem a importância da localização, mas também apontam para a necessidade de uma abordagem equilibrada para satisfazer as diversas expectativas dos hóspedes.

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

Anúncios com avaliações gerais consistentemente baixas podem ser considerados críticos. Isso pode indicar problemas significativos que afetam a satisfação dos hóspedes.

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

Se eu tivesse todos os dados preenchidos nas duas tabelas, eu poderia utilizar a biblioteca nltk (Natural Language Toolkit) para identificar as palavras mais frequentes nos anúncios, revelando o que os hóspedes geralmente buscam em um apartamento. Além disso, realizar uma análise das características mais comuns e bem avaliadas em diferentes cidades seria interessante para identificar padrões ou particularidades locais. Também consideraria uma análise de feedback negativo para ter uma visão mais clara das áreas que precisam de melhoria nos apartamentos. Essas abordagens proporcionariam insights valiosos para entender as preferências dos hóspedes e aprimorar a qualidade das acomodações oferecidas.

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

<https://lookerstudio.google.com/reporting/035c70e3-51f8-40eb-bf93-9acb1ca91ed9>

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

Caso eu tivesse acesso a dados adicionais, ampliaria minha análise. Incluiria informações demográficas dos hóspedes para entender melhor quem são os usuários que fazem avaliações. Além disso, consideraria dados climáticos, avaliações em outras plataformas, informações sobre preços e ocupação, dados econômicos locais, calendário de eventos e histórico de manutenção dos imóveis. Esses dados

proporcionariam insights mais abrangentes e ajudariam na identificação de padrões complexos, enriquecendo a compreensão do desempenho dos imóveis.

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

Para melhorar as notas dos imóveis, sugere-se investir em comodidades bem avaliadas, destacar pontos fortes nos anúncios, garantir limpeza e conforto, adaptar o marketing para destacar a localização, oferecer promoções, enviar pesquisas de satisfação, resolver problemas proativamente, ser transparente nas descrições, incentivar avaliações e estar aberto ao feedback constante. Essas ações visam aprimorar a experiência do hóspede e obter avaliações mais positivas.

CONCLUSÃO

Com esse desafio, tive a oportunidade de me divertir bastante, embora também tenha representado um desafio considerável. Como já fazia algum tempo desde minha última análise, precisei reacender algumas lembranças, mas foi recompensador perceber o quanto essa atividade pode ser envolvente e estimulante.

No geral, acredito que consegui atender às expectativas do desafio dentro de um prazo razoável. No entanto, enfrentei algumas dificuldades ao longo do processo. Uma delas foi minha falta de familiaridade ao criar um dashboard, o que demandou tempo extra para revisar e reaprender o procedimento.

Outro obstáculo significativo foi a limpeza dos dados, em particular, inserir as cidades manualmente. Reconheço que essa abordagem consumiu mais tempo do que eu

gostaria, mas, na ausência de uma solução mais eficiente, optei por esse método. Se examinar o código, notará que tentei resolver esse problema, mas as soluções não foram bem-sucedidas.

Essas experiências me proporcionaram valiosos aprendizados, e estou aberto a feedbacks construtivos para melhorar ainda mais minhas habilidades de análise de dados. Estou ansioso para futuros desafios e oportunidades de aprendizado.