

SEAZONE CODE CHALLENGE DATA SCIENCE

Geovanio Claudino Cavalcante Filho

E-mail: geo.cavalcante@hotmail.com

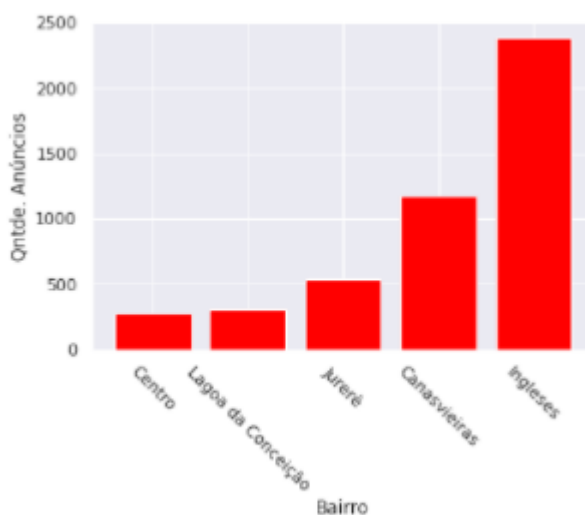
Fone: (84) 994633273

No desafio proposto consiste em analisar os dados, fornecidos a partir de 2 arquivos do tipo csv, de ocupação e preço de anúncios no Airbnb, a fim de responder uma série de perguntas:

Questão 1 - Ordene os bairros em ordem crescente de número de listings (anúncios):

A partir dos dados fornecidos, foi apenas necessário fazer um agrupamento pelo bairro e colocar os dados gerados em uma ordem crescente. Logo, temos um ranking do bairro com a menor quantidade de anúncio até o maior. O resultado pode ser visto, tanto na tabela gerada, como no gráfico.

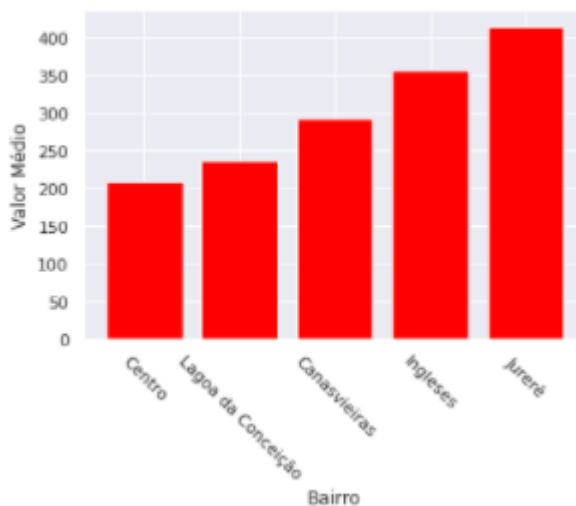
	suburb	airbnb_listing_id
1	Centro	278
4	Lagoa da Conceição	309
3	Jurerê	539
0	Canasvieiras	1177
2	Inglese	2388



Questão 2 - Ordene os bairros em ordem crescente de faturamento médio dos listings:

Já para essa questão, foi necessário fazer um merge (junção) das duas tabelas para analisar os preços médio de cada dos anúncios que já foi alugado (faturamento médio).

	suburb	price_string
1	Centro	206.894422
4	Lagoa da Conceição	234.387587
0	Canasvieiras	292.236096
2	Inglese	355.183420
3	Jurerê	414.080865



É possível de se observar que a ordem dos bairros se mante a mesma da questão 1, tendo em vista Jurerê tendo a maior quantidade de anúncios e o Centro com a menor quantidade. Caso o número de anúncios se mantenha com a mesma proporção e os valores de aluguel não sejam discrepantes, um dos outros, tal ordem já era esperado. Contudo se faz necessário mais análises para validar tão afirmação.

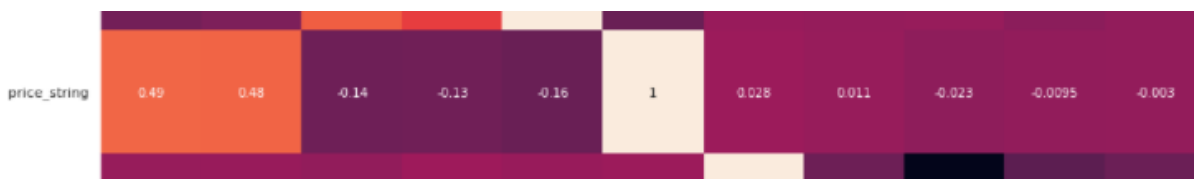
Questão 3 - Existem correlações entre as características de um anúncio e seu faturamento?

Sim, existe correlação entre dados de faturamento, contudo essa correlação é de correlação **média**. Foi atribuído uma divisão de Correlação (C), sendo:

- $C \geq 0,6$ como alta
- $0,4 \geq C > 0,6$ sendo média
- $0 > C > 0,4$ sendo baixa.

a. Quais? Explique

Analisando com base na média de faturamento de cada um dos anúncios, é possível de se analisar a correlação a seguir. Onde a quantidade de quartos apresentou uma correção de **0,49** e quantidade de banheiros de **0,48**.



Questão 4 - Qual a antecedência média das reservas?

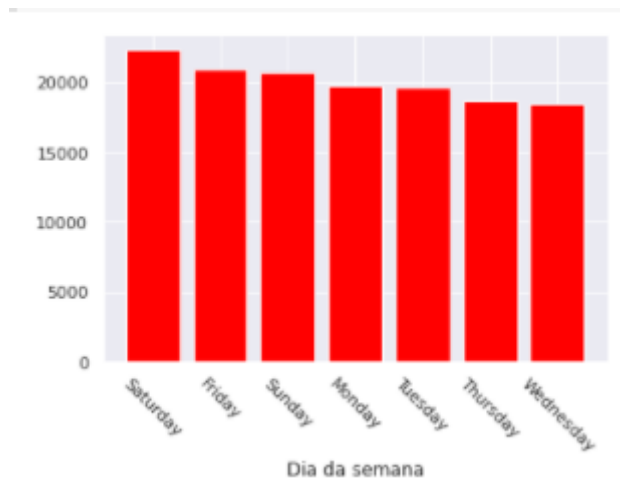
Tomando como base a quantidade total de registros dos anúncios fornecidos por uma das tabelas e analisada a partir desses registros aqueles que foram alugados, foi possível de analisar uma antecedência média de:

```
media = df_alugado.shape[0] / df_priceav.shape[0]
print(media*100, "%")

39.633307006656885 %
```

a. Esse número é maior ou menor para finais de semana?

A partir disso é necessário analisar qual dia da semana está relacionado a cada uma das datas e assim foi possível de ver que, as 3 maiores quantidade de anúncios alugados estão nos sábados, sextas-feiras e nos domingos, respectivamente, possível observar com o gráfico a seguir.



Validando com números, a partir do sábado e do domingo em relação aos demais dias da semana se tem:

```
media = df_alugado_fds.shape[0] / df_fds.shape[0]  
print(media*100, "%")
```

```
41.91928530416419 %
```

Validando que, sim. O número antecedência média nos fins de semana é maior.

FeedBack

Gosto muito de fazer desafios assim, pois aprendo bastando. Acho que só tem alguns pontos a melhorar, como na descrição dos dados não batem exatamente com as colunas que estão os arquivos, um exemplo disso é a coluna available foi trocada pela occupied.

Onde o entendimento da nova coluna foi compreendido e trabalhado da seguinte forma, sendo:

“available”: Booleano de ocupação. 1 significa livre e 0 ocupado.

“occupied”: Booleano de ocupação. 1 significa ocupado e 0 livre.

No geral foi muito gratificante fazer esse desafio, espero fazer parte da equipe.

Grato.