

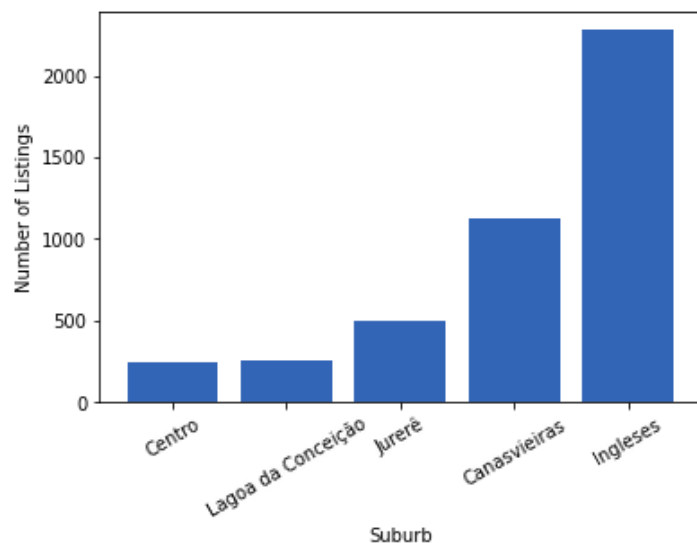
Desafio de processo seletivo para vaga de estágio em Data Science - Seazone

Rômulo de Oliveira Carneiro - <https://github.com/carneiroRomulo/SeazoneChallenge>

1. Ordene os bairros em ordem crescente de número de listings

```
# Ordena os dataset em ordem ascendente de acordo com o número de listings em cada suburb
df3_details = df3_details.groupby('suburb', as_index=False).airbnb_listing_id.count()
df3_details.sort_values(['airbnb_listing_id'], inplace=True)
df3_details
```

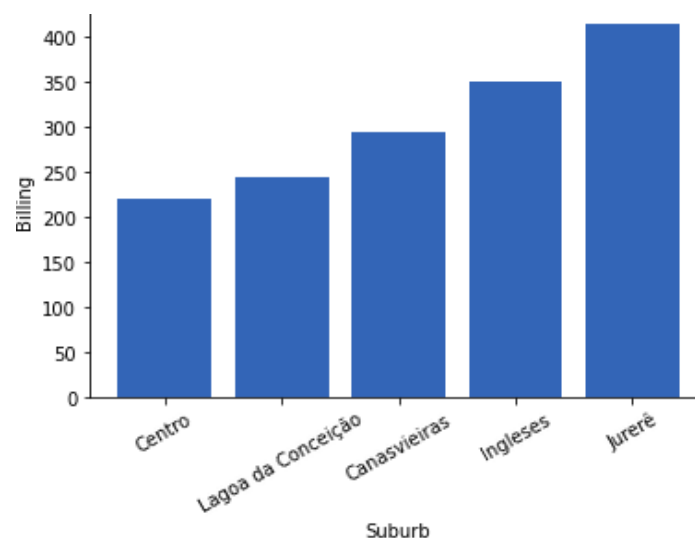
	suburb	airbnb_listing_id
1	Centro	238
4	Lagoa da Conceição	258
3	Jurerê	498
0	Canasvieiras	1125
2	Ingleses	2281



2. Ordene os bairros em ordem crescente de faturamento médio dos listings

```
# Calcula a média de faturamento dos anúncios por bairro
df2_merged = df_merged.groupby('suburb', as_index=False).billing.mean().round(2)
df2_merged.sort_values(['billing'], inplace=True)
df2_merged
```

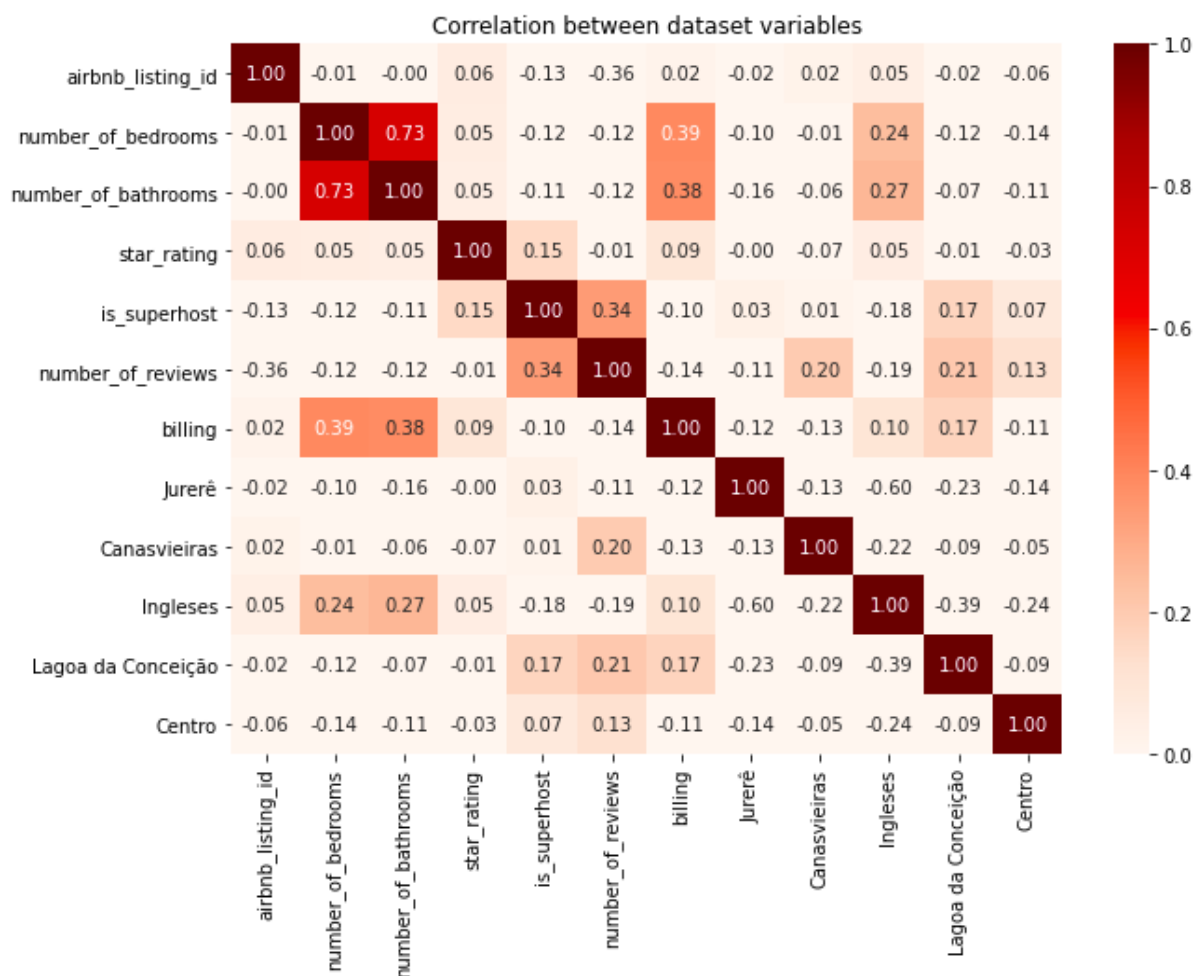
	suburb	billing
1	Centro	219.70
4	Lagoa da Conceição	243.83
0	Canasvieiras	294.30
2	Ingleses	350.27
3	Jurerê	414.28



3. Existem correlações entre as características de um anúncio e seu faturamento?
- a. Quais? Explique .

A correlação entre as features foi obtida através do coeficiente de relação de Pearson, com a ajuda da biblioteca Seaborn para representar o resultado em um gráfico de heatmap.

Através do gráfico de heatmap é perceptível que há sim uma correlação entre algumas das features e o faturamento. Pode-se observar que: 'number_of_bedrooms', 'number_of_bathrooms' tem uma forte correlação entre si e com o faturamento, e a feature 'star_rating', junto dos bairros: Ingleses e Lagoa da Conceição apesar de possuírem uma correlação mais fraca com o faturamento, ela também é considerável. Pode-se absorver também dessa análise que ao alugar através de um anúncio no airbnb, as pessoas costumam procurar por um número de quartos e banheiros equivalente, o que pode estar relacionado com um maior aluguel de suítes e geralmente localizados no bairro Ingleses já que este possui forte correlação com esses cômodos



4. Qual a antecedência média das reservas?
- a. Esse número é maior ou menor para finais de semana?

A antecedência média das reservas foi de 8.19 dias, e esse número foi maior para finais de semanas.

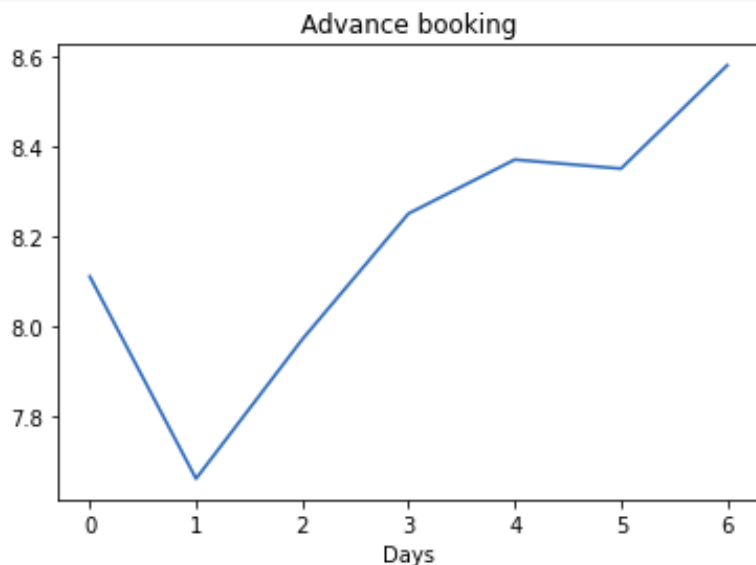
```
# Removendo outliers
df5_price = df4_price[df4_price.advance_booking < 36]
```

```
# Média de advance_booking
df5_price['advance_booking'].mean()
```

8.193997537043657

```
# Média de reservas realizadas por dia da semana
df6_price = df5_price.groupby('days', as_index=False).advance_booking.mean().round(2)
df6_price = df6_price.sort_values(['days'])
df6_price
```

	days	advance_booking
0	0	8.11
1	1	7.66
2	2	7.97
3	3	8.25
4	4	8.37
5	5	8.35
6	6	8.58



● Feedback

Pessoalmente achei a proposta bastante desafiadora, isso porque meu conhecimento nas ferramentas utilizadas precisavam melhorar muito, então tive que correr atrás e pesquisar bastante para que no final tudo se encaixasse. Entretanto, não vejo isso como algo a ser revisto e nem alteraria nada, penso que a proposta foi bem clara e me trouxe um ganho de conhecimento enorme, foi uma experiência muito gratificante independente do resultado.