

Um primeiro approach aos dados: Análise Exploratória

agosto/2021



licenciado sob [Criative Comuns Atribuição 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Agenda do Dia

1 O que é?

2 Ferramentas

3 Análise

O que é ?



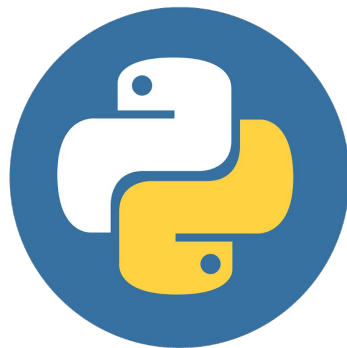


Entendo o Problema /
Contexto de Negócio

Escolhendo as Ferramentas



colab



 pandas

 matplotlib

NumPy 

 seaborn



Contexto desta Análise



Uma pessoa chef de cozinha gostaria de compreender o perfil de seu restaurante para realizar melhorias no atendimento.

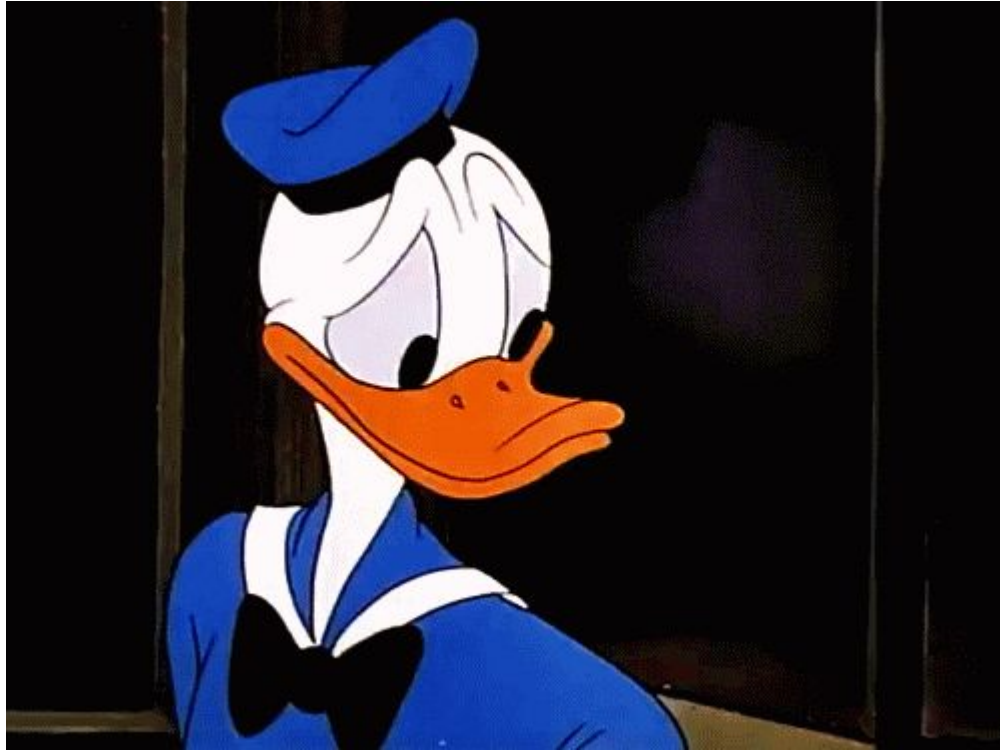
Para isso coletou dados de uma semana para analisar e ter seus insights

Nossos dados



Dataset Tips

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4



**Antes de seguir, pausa
para relembrar:**

Relembrando:



Quantitativos - quantifica ou mede

Qualitativos - Característica ou qualidade

dados/atributos/variáveis

O que temos

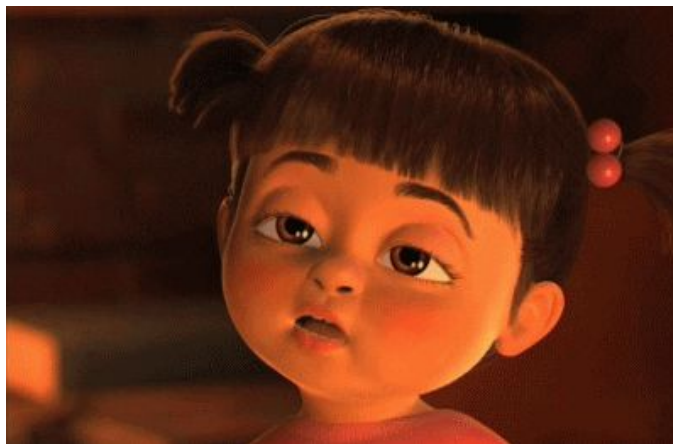
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   total_bill  244 non-null    float64
1   tip         244 non-null    float64
2   sex         244 non-null    category
3   smoker      244 non-null    category
4   day         244 non-null    category
5   time       244 non-null    category
6   size        244 non-null    int64
dtypes: category(4), float64(2), int64(1)
memory usage: 7.3 KB
```

Amostra com 244 observações

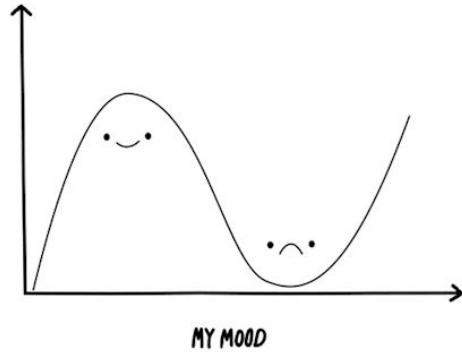
Em Resumo:

	total_bill	tip	size
count	244.000000	244.000000	244.000000
mean	19.785943	2.998279	2.569672
std	8.902412	1.383638	0.951100
min	3.070000	1.000000	1.000000
25%	13.347500	2.000000	2.000000
50%	17.795000	2.900000	2.000000
75%	24.127500	3.562500	3.000000
max	50.810000	10.000000	6.000000

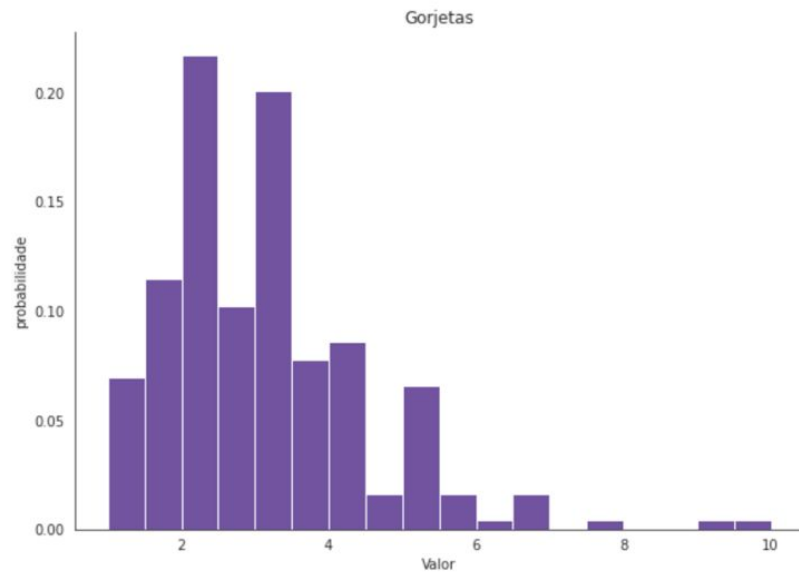
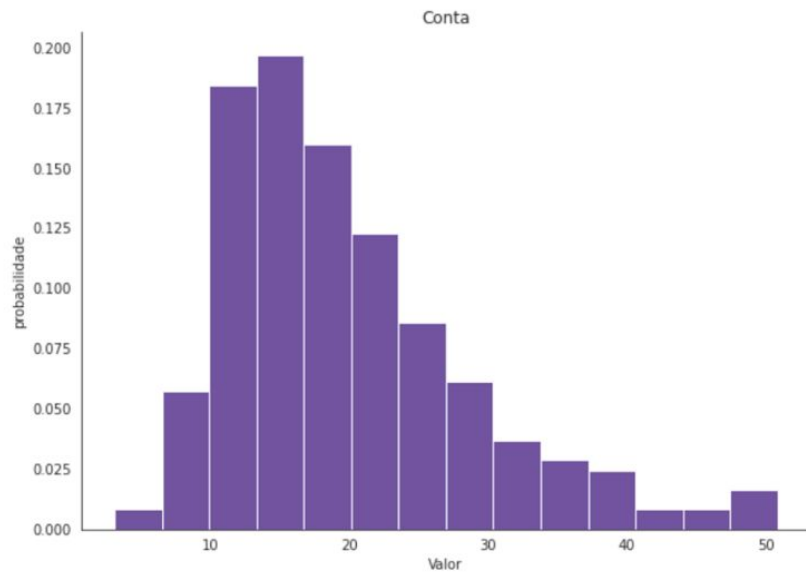
Dá para ser mais divertido?



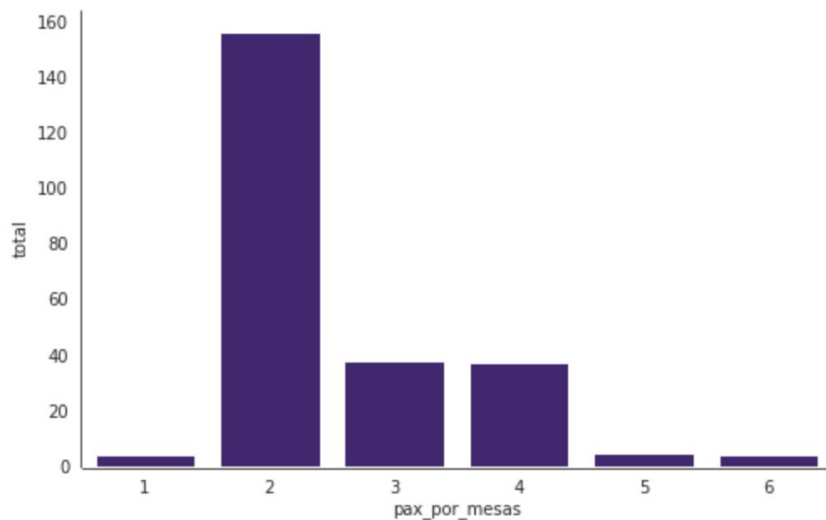
Sim!!!!!! Use gráficos!



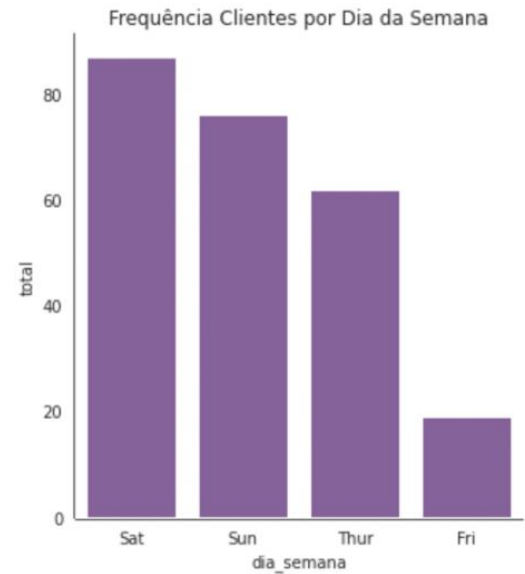
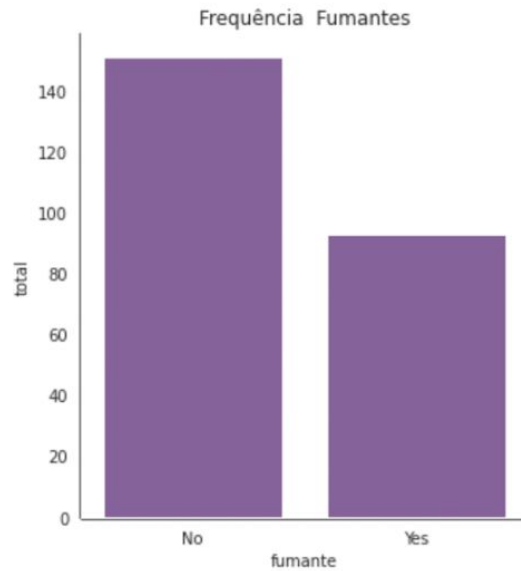
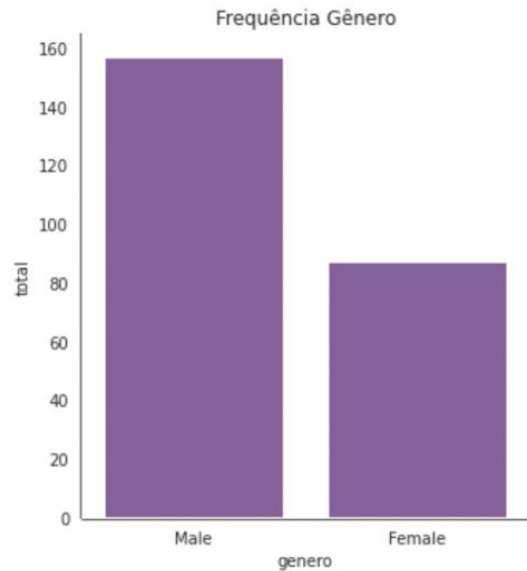
Gráficos:



Mais gráficos:



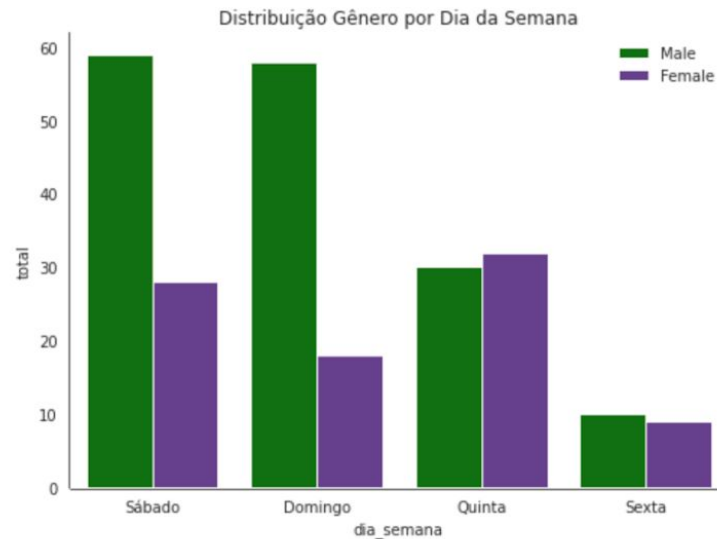
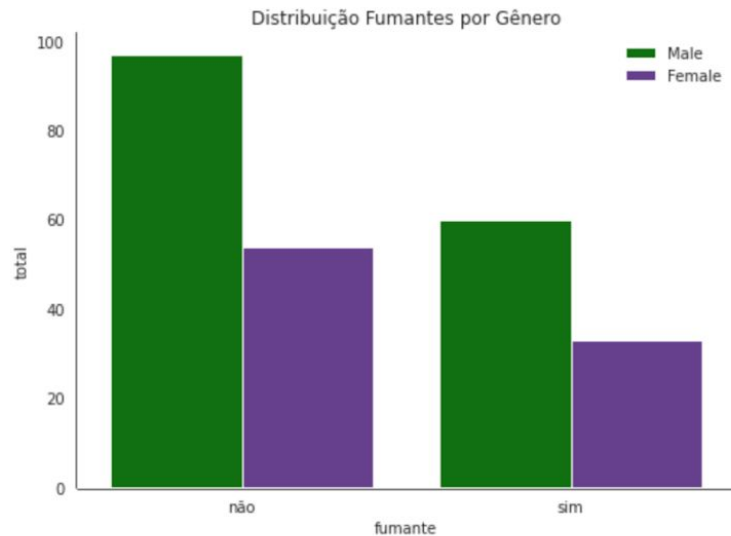
Por fim, mais gráficos:



E o que os dados nos
contam até agora?



1 é bom, mas 2 pode ser melhor

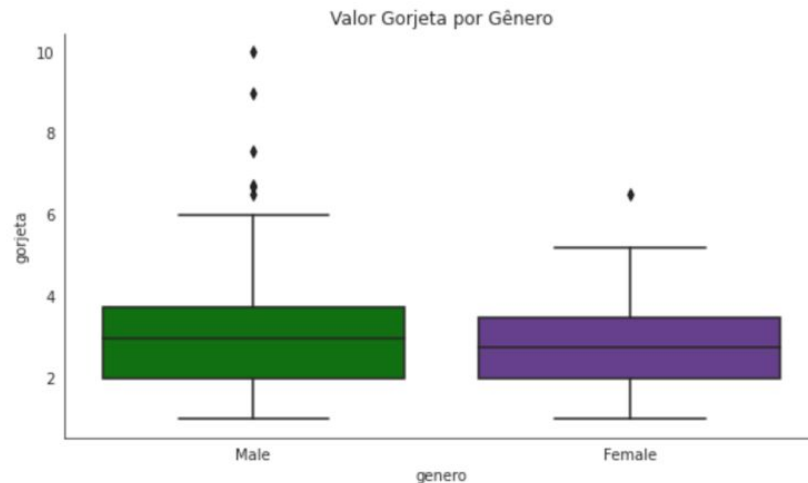
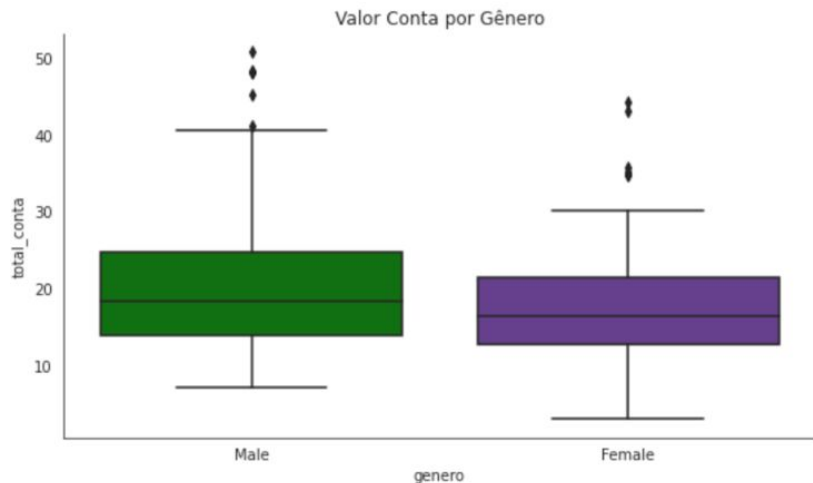


1 é bom, mas 2 pode ser melhor

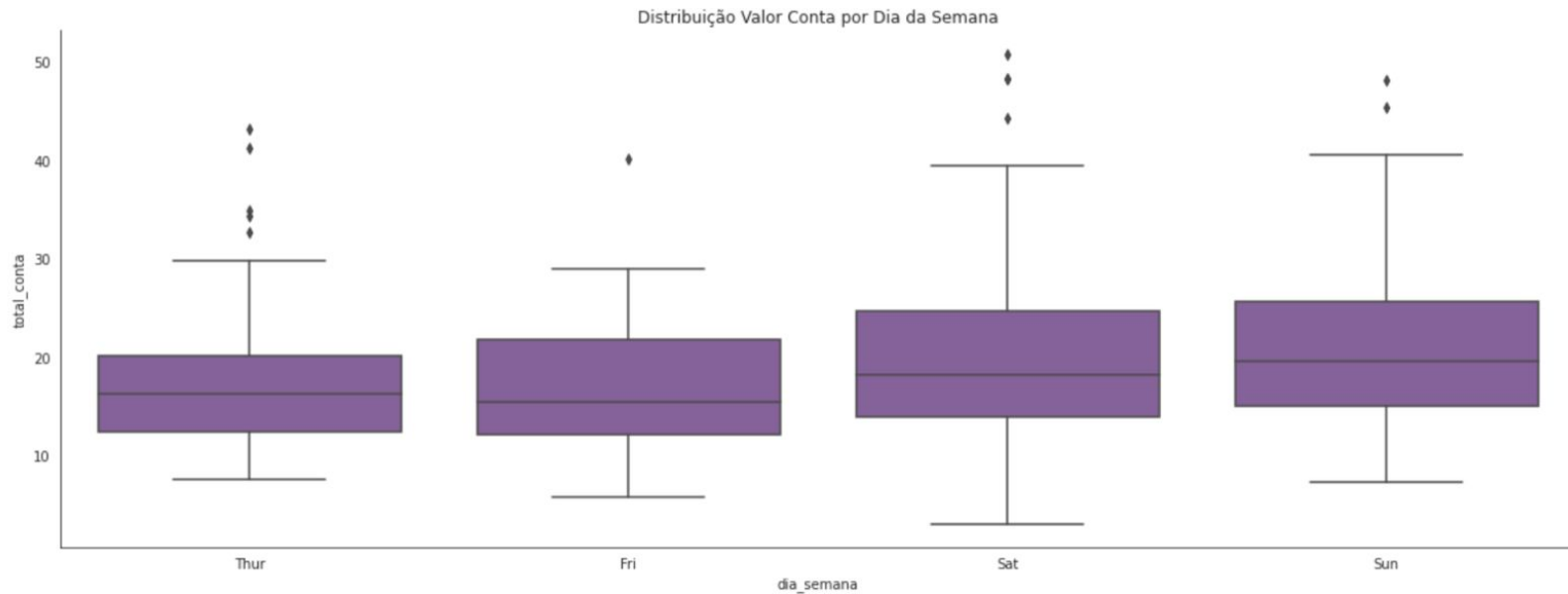
genero porcentagem

0 Male 0.643443

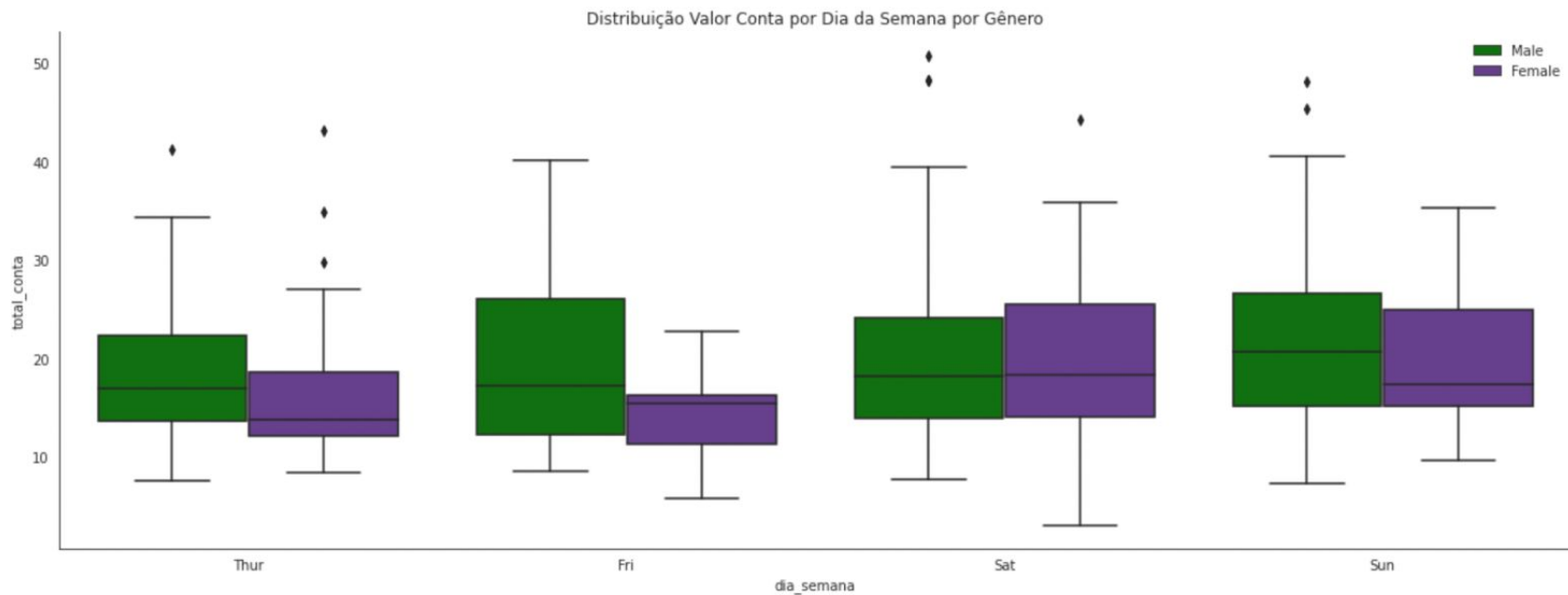
1 Female 0.356557



Vários, pode ser legal



Mas cuidado,



a amostra pode não ser significativa

dia_semana	genero	
Thur	Female	32
	Male	30
Fri	Male	10
	Female	9
Sat	Male	59
	Female	28
Sun	Male	58
	Female	18

dia_semana	genero	
Thur	Female	0.516129
	Male	0.483871
Fri	Male	0.526316
	Female	0.473684
Sat	Male	0.678161
	Female	0.321839
Sun	Male	0.763158
	Female	0.236842

Male	0.526316
Female	0.473684

Por fim, hipóteses

as mesas das
pessoas
consumidoras do
sexo masculino
gastam mais do
que as mesas das
pessoas
consumidoras do
sexo feminino

as mesas das
pessoas
consumidoras
consomem mais,
em média, aos
finais de semana.

as mesas das
pessoas
consumidoras do
sexo feminino dão
mais gorjetas do
que as mesas das
pessoas
consumidoras do
sexo masculino