

NEURON BootCamp 2020

Análise 1: Gorjetas

Equipe M

Equipe M

Antonio Danilo de Araújo Mota

Julia Bartolo

Makson Vinicio Ferreira de Sousa

Maria Paula Dias Rosa

Rogério Sampaio



Sumário

1. Apresentação

2. Objetivo

3. *Dataset*

3.1. Informações sobre o *dataset*

4. Alguns *Insights...*

4.1. Relação entre fumantes e gorjetas

4.2. Relação entre gênero e valor das gorjetas

4.3. Relação entre dias da semana e valor das gorjetas

4.4. Relação entre período do dia e valor das gorjetas

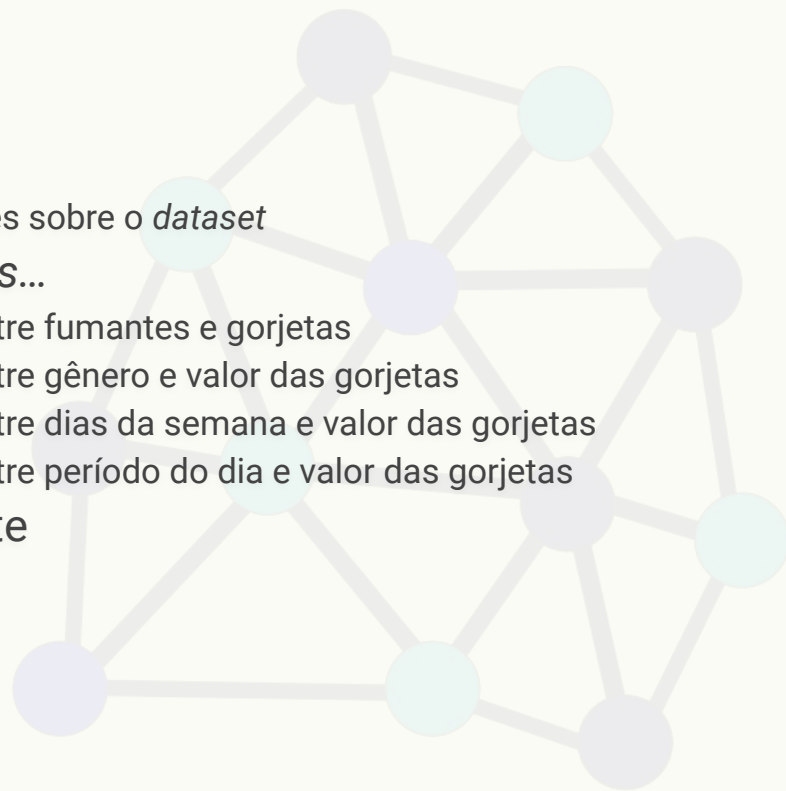
5. Perfil do Cliente

6. Promoção

7. Conclusão

Referência

Agradecimentos



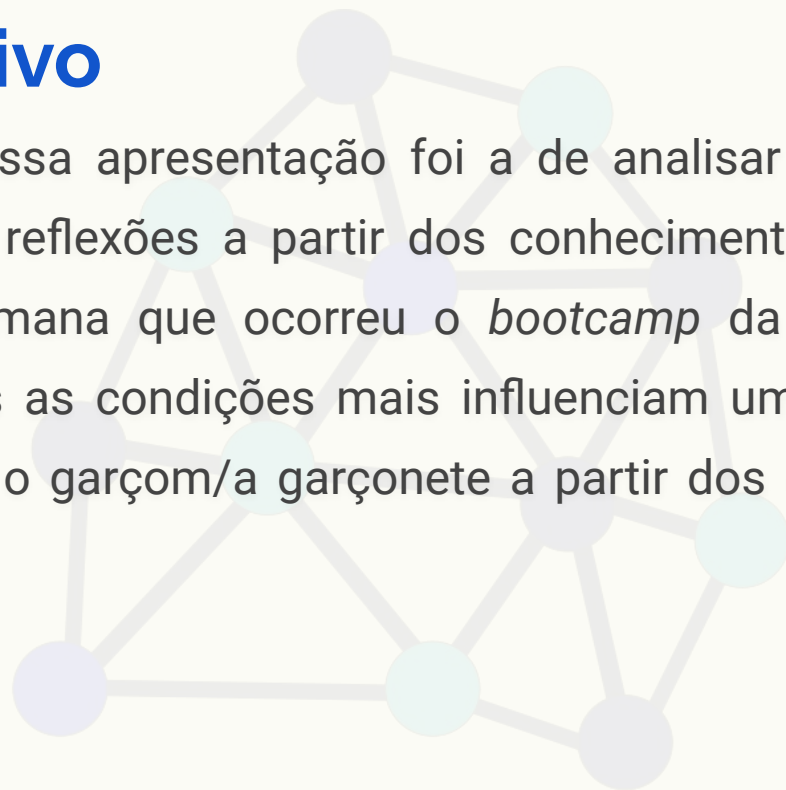
1. Apresentação

Durante os dias de 7 a 11 de agosto de 2020, participamos do bootcamp da Neuron, 100% on-line, onde tivemos diversas palestras com conteúdo intensivo sobre assuntos relacionados às áreas de Ciência de Dados e Inteligência Artificial.

Para o final de semana, dias 12 e 13 de agosto, tivemos a seguinte missão: avaliar três Datasets de acordo com o aprendido durante a semana. Esta apresentação conta com as análises que nós, do **Grupo M**, investigamos sobre o *dataset* **A Waiter's Tips** (Gorjetas de um garçom/uma garçonete), que será apresentado a seguir.

2. Objetivo

O objetivo dessa apresentação foi a de analisar o conteúdo de um *dataset*, criar reflexões a partir dos conhecimentos práticos obtidos durante a semana que ocorreu o *bootcamp* da *Neuron* e por fim, analisar quais as condições mais influenciam um cliente a dar mais gorjetas para o garçom/a garçonete a partir dos dados do dataset *A Waiter's Tips!*



3. Dataset

O *dataset* avaliado, *A Waiter's Tips*, traz sete informações: valor da conta (*total_bill*), valor da gorjeta (*tip*), gênero (*sex*), fumante (*smoker*), dia da semana (*day*), período do dia (*time*) e tamanho do grupo (*size_of_group*).

A partir desses dados, pudemos avaliar algumas questões que serão discutidas nos slides a seguir.

Links úteis:

- [Dataset “A Waiter’s Tips” original \(Kaggle\);](#)
- [Notebook do Google Colab com as análises completas;](#)

3.1. Informações sobre o Dataset

A primeira avaliação que fizemos do *dataset* foi:

- Avaliar o tamanho do dataset: 244 linhas x 7 colunas;
- Descobrir se haviam valores duplicados: apenas 1;
- Descobrir se haviam valores nulos: não haviam;
- Valor somado das gorjetas: 731,58 dinheiros;

Assim, notamos que se tratava de um *dataset* pequeno mas nada que impedisse de se começar uma análise. 😊

4. Alguns *insights*...

...existe relação entre uma pessoa ser **fumante** e dar **gorjeta**?

...existe diferença entre os valores de gorjetas que **mulheres** e **homens** dão?

...existe relação entre **dia da semana** e valor da **gorjetas**? E entre os **período do dia** e **gorjetas**?

...qual o **perfil dos clientes** deste estabelecimento?

...quais seriam os **dias ideais** para se fazer uma **promoção** no estabelecimento avaliado?

Vejam os próximos slides! 😊

4.1. Relação entre fumantes e gorjetas

Uma das hipóteses que podem ser levantadas a partir do *dataset* é a de que se haveria alguma relação entre ser fumante e não fumante e dar gorjetas ou não.

Para essa análise *plotamos* o Gráfico 1 e notamos que, ser fumante ou não fumante, **não tem influência!**

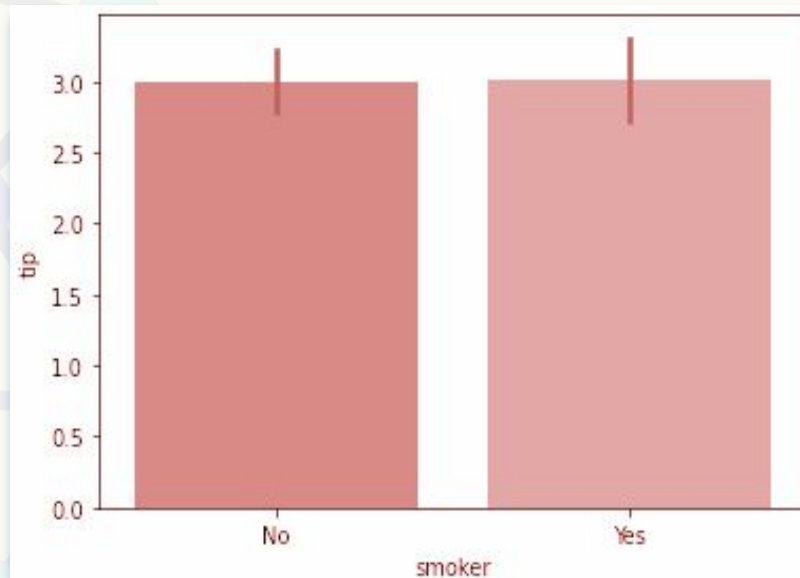


Gráfico 1 - smoker (fumante) x tip (gorjeta)

4.2. Relação entre gênero e valor das gorjetas

A próxima hipótese avaliada foi: será que existe relação entre o valor da gorjeta e o gênero do cliente que está dando a gorjeta?

Para essa análise *plotamos* o Gráfico 2 e notamos que, parece que **homens costumam dar um valor maior de gorjetas que mulheres**. Será essa conclusão tão simples?

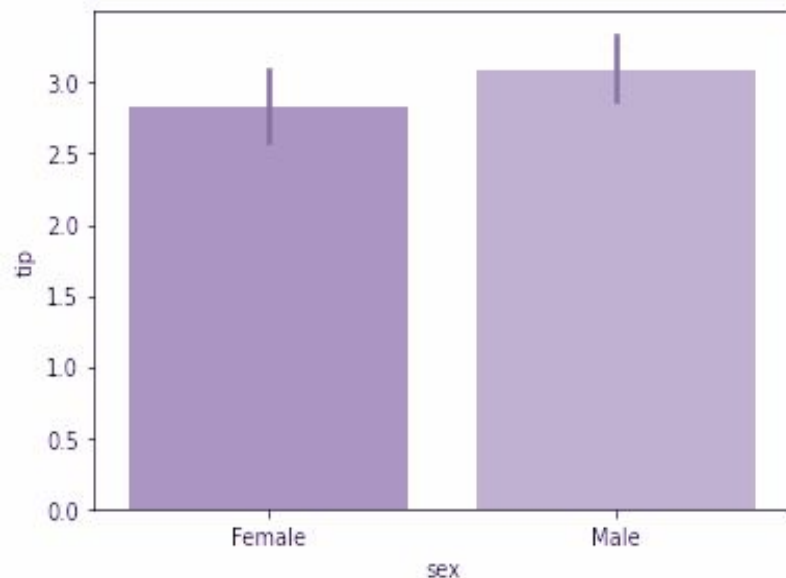


Gráfico 2 - sex (gênero) x tip (gorjeta)

4.2. Relação entre gênero e valor das gorjetas

Para complementar a análise do Gráfico 2, resolvemos avaliar também o perfil do cliente que frequenta o estabelecimento avaliado. Traçando o Gráfico 3, pudemos perceber que a quantidade de homens que frequenta o estabelecimento é quase o dobro da quantidade de mulheres.

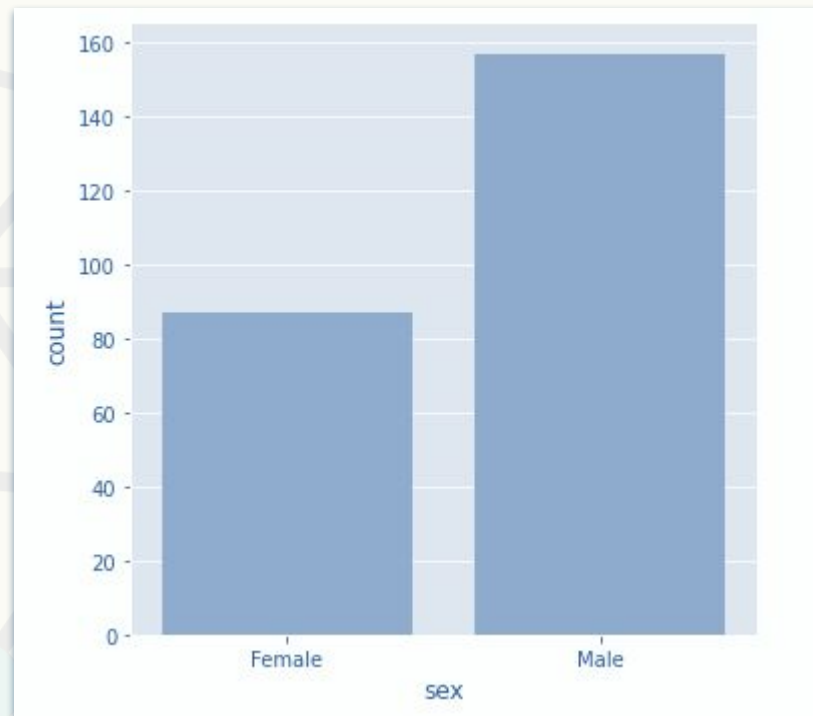


Gráfico 3 - sex(gênero) x count
(quantidade de pessoas)

4.2. Relação entre gênero e valor das gorjetas

Portanto, isso nos leva a crer que **gênero também não influencia no valor da gorjeta**, pois os dados acabam **enviesados**, já que a quantidade de homens que frequentam o estabelecimento é superior.

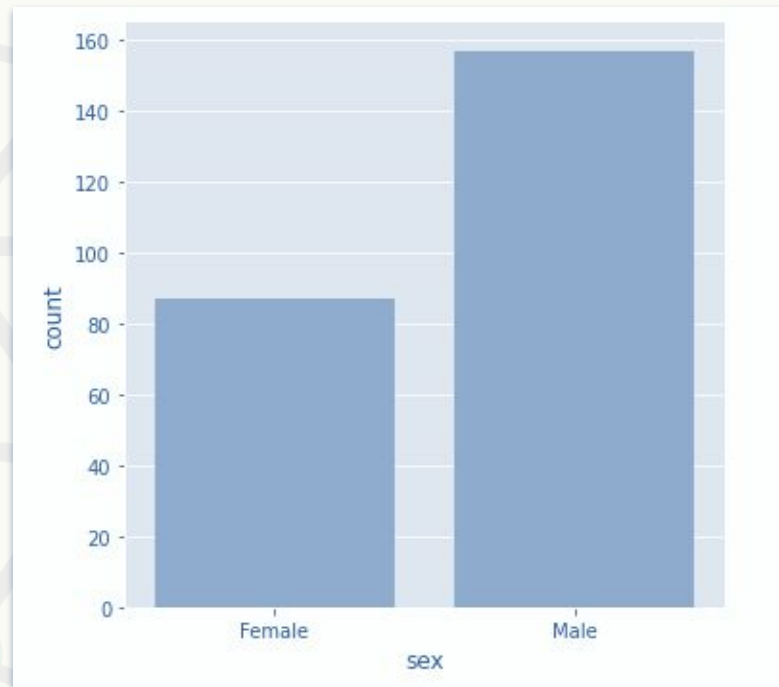


Gráfico 3 - sex (gênero) x count (quantidade de pessoas)

4.3. Relação entre dias da semana e valor das gorjetas

Pelo **Gráfico 4**, nota-se que os dias em que são dados maiores valores de gorjetas são aos domingos e aos sábados, respectivamente.

Então, para os garçons e garçonetes deste estabelecimento, **trabalhar aos sábados e domingos é mais lucrativos!!**

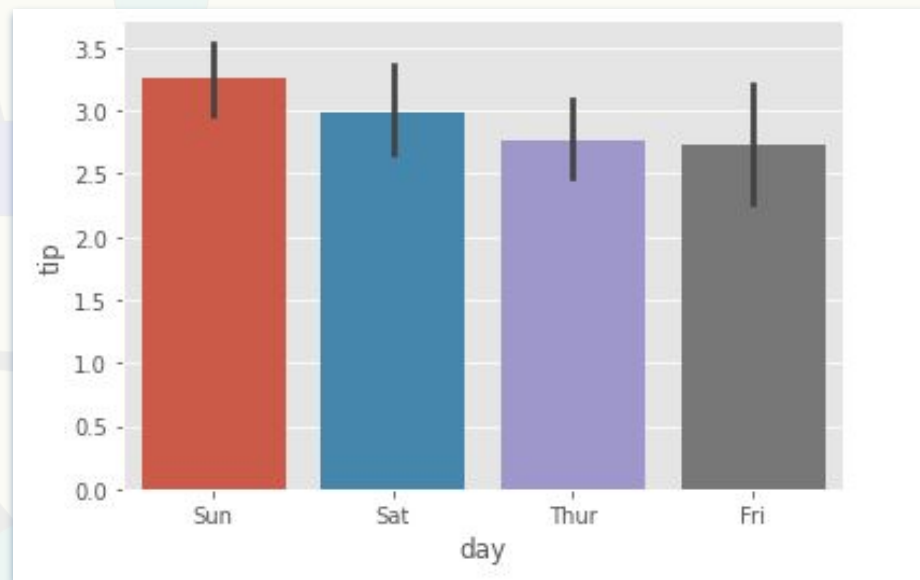
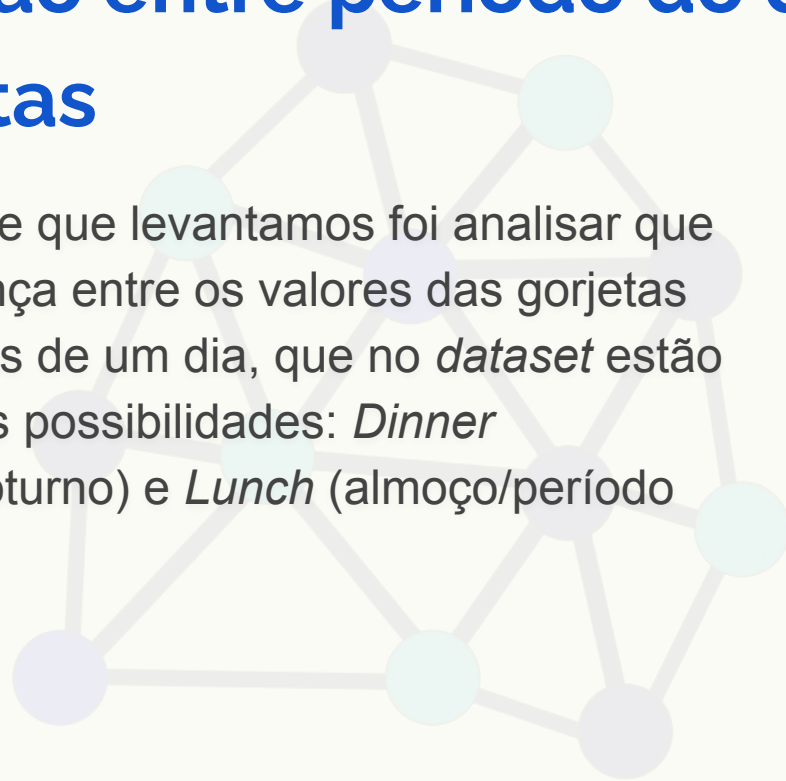


Gráfico 4 - day (dia da semana) x tip (valor da gorjeta)

4.4. Relação entre período do dia e valor das gorjetas

A terceira hipótese que levantamos foi analisar que se haveria diferença entre os valores das gorjetas dadas no períodos de um dia, que no *dataset* estão divididos em duas possibilidades: *Dinner* (jantar/período noturno) e *Lunch* (almoço/período diurno).



4.4. Relação entre período do dia e valor das gorjetas

Pode-se avaliar, pelo **Gráfico 5**, que o **horário onde os valores das gorjetas são mais altos, é o período da noite.**

Ou seja, para os garçons e garçonetes deste estabelecimento, **trabalhar no período noturno é mais lucrativo!!** 😊

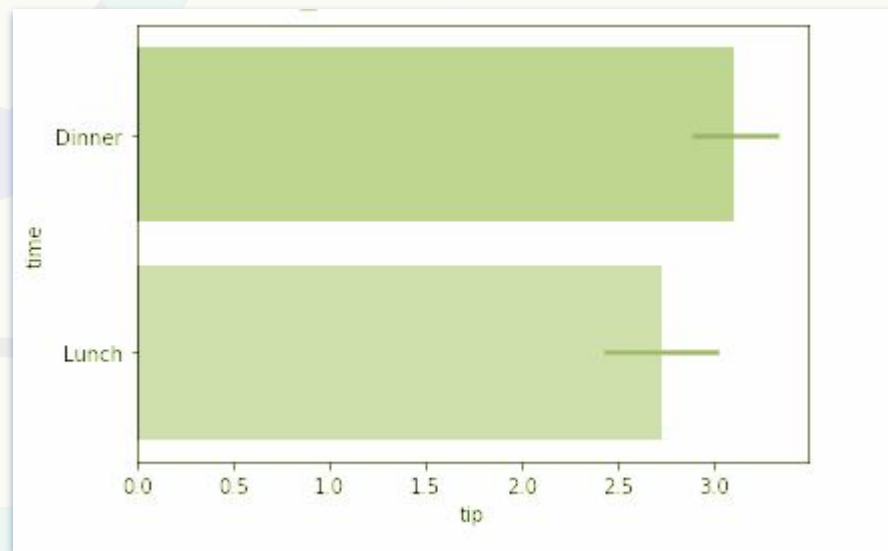
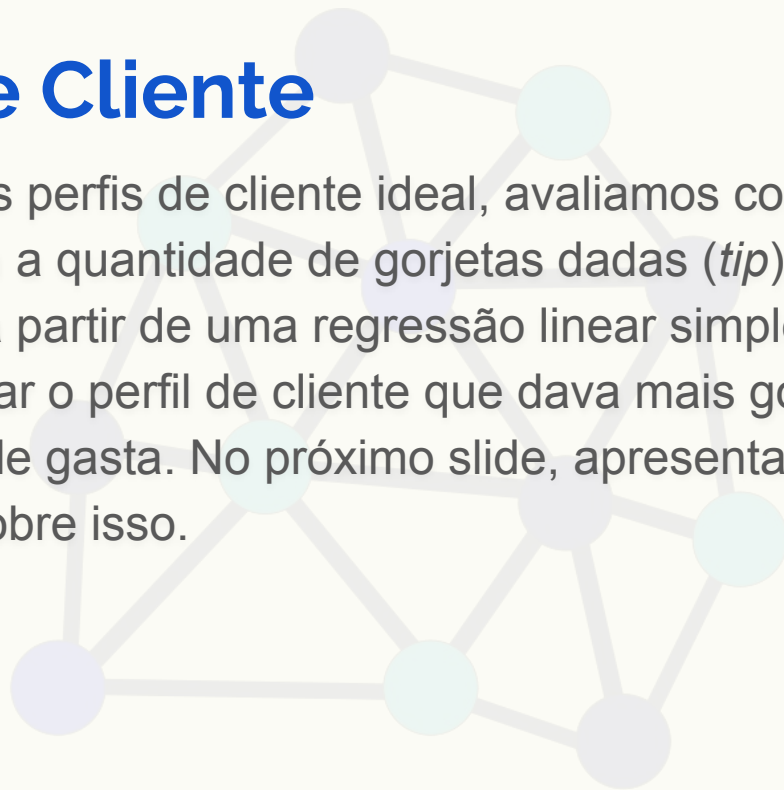


Gráfico 5 - tip (valor da gorjeta) x time (período do dia)

5. Perfil de Cliente

Para descrever os perfis de cliente ideal, avaliamos como os dados relacionam a quantidade de gorjetas dadas (*tip*) e o total da conta (*total_bill*) a partir de uma regressão linear simples para assim, tentar traçar o perfil de cliente que dava mais gorjeta e a quantidade que ele gasta. No próximo slide, apresentaremos algumas ideias sobre isso.



5. Perfil de Cliente

A partir da regressão linear simples, os valores obtidos foram:

- Erro médio quadrático (MSE): 0,75
- Coeficiente de determinação R^2 : 0,43

Podemos perceber que os maiores valores de gorjetas estão concentrados em contas que tenham ficado entre 10 e 30 dinheiros.

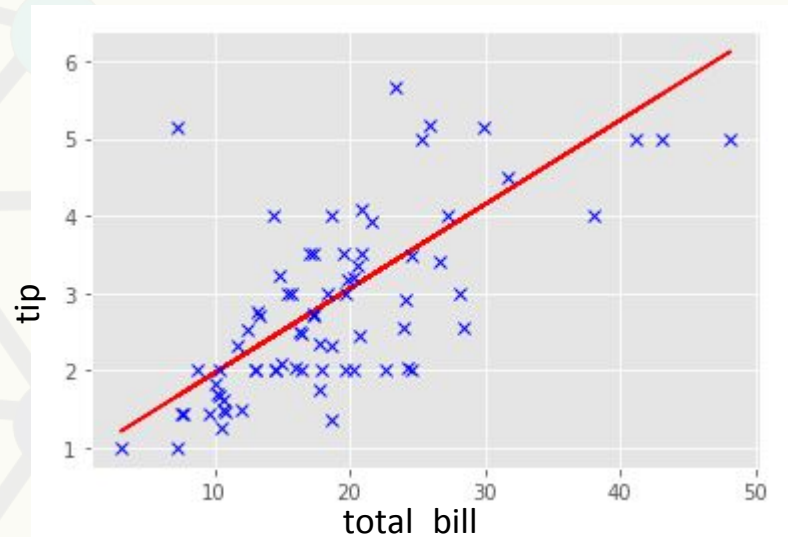


Gráfico 6 - total bill (valor total da conta) x tip (valor das gorjetas)

5. Perfil de Cliente

A partir da regressão linear múltipla, o valor para o coeficiente de determinação foi de $R^2 = 0,36$.

Pode-se notar que os perfis de cliente que mais dão gorjeta portanto são os que a conta ficam entre 10 e 30 dinheiros, em grupos que tenham entre 2 e 4 pessoas.

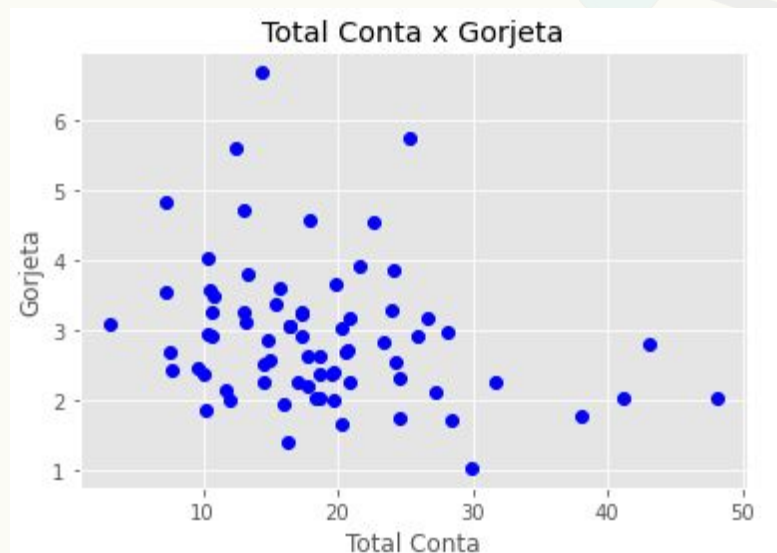


Gráfico 7 - Total conta x Gorjeta

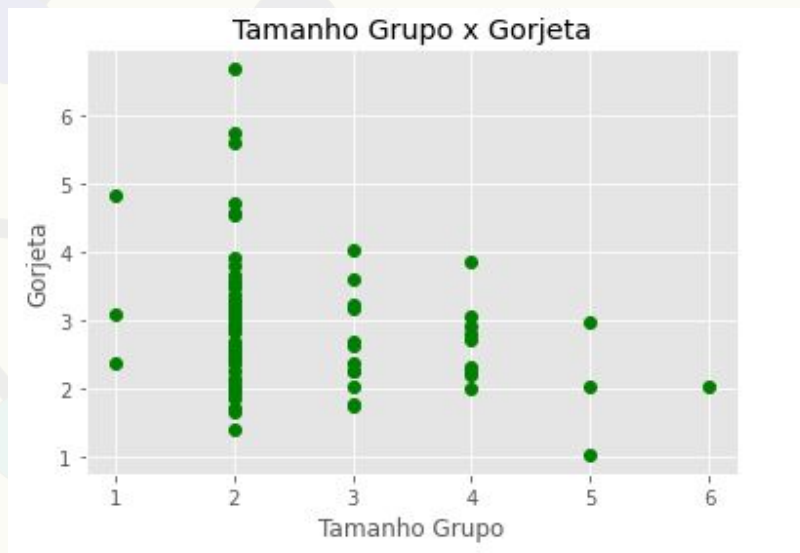
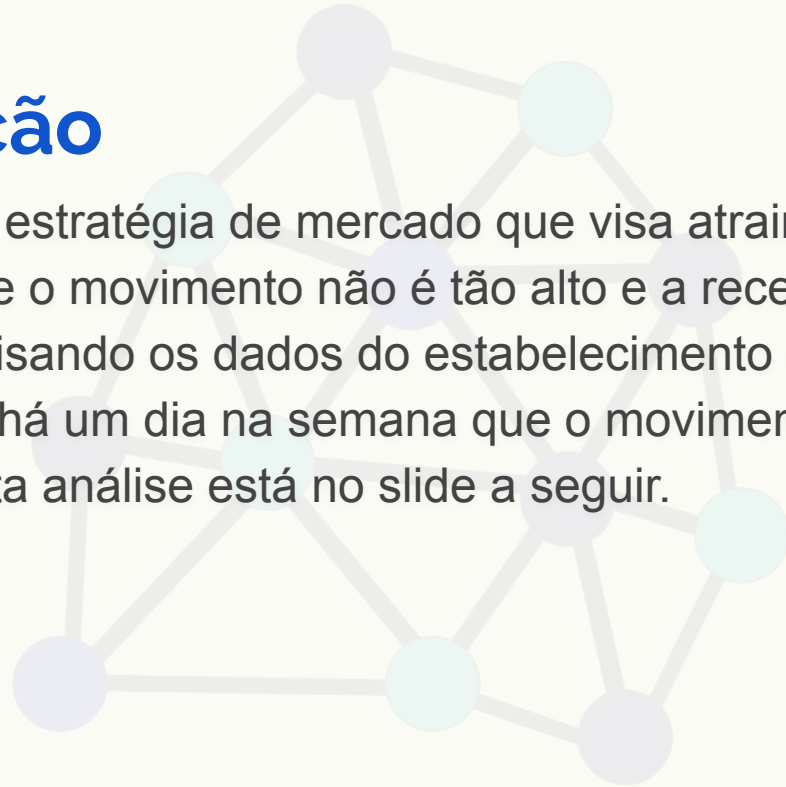


Gráfico 8 - Tamanho Grupo x Gorjeta

6. Promoção

Promoção é uma estratégia de mercado que visa atrair clientes e gerar lucro nos dias que o movimento não é tão alto e a receita acaba ficando prejudicada. Analisando os dados do estabelecimento estudado, percebemos que há um dia na semana que o movimento de pessoas é muito baixo, e esta análise está no slide a seguir.



6. Promoção

Avaliando o Gráfico 6, podemos perceber que os dias que o menor movimentos de pessoas ocorre às sextas (movimento muito baixo) e quintas (movimento normal).

Sendo assim, **o melhor dia para o estabelecimento criar uma promoção é durante à sexta-feira!**

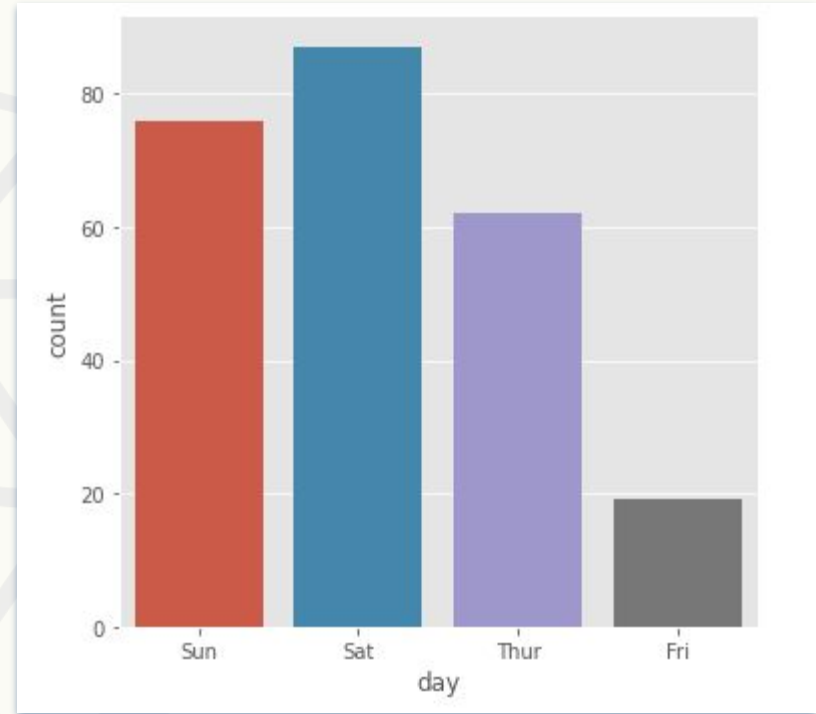


Gráfico 6 - day (dia da semana) x count (quantidade de pessoas)

7. Conclusão

Os *insights* mais interessantes obtidos através do dataset analisado, foram:

- **Não faz diferença ser fumante ou não fumante** ao dar gorjetas;
- Percebemos que **homens dão gorjetas maiores**, mas que esses dados estão **enviesados**, já que a **maior parte do público do estabelecimento é masculino**, portanto, não é seguro tirar uma conclusão sobre;
- Os **dias da semana mais lucrativos são**, nesta ordem: **domingo**, sábado, quinta e sexta;
- A refeição que traz **maiores quantidades de gorjeta** é a janta, ou seja, **período noturno** é mais lucrativo;
- Os clientes que gastam entre **10 e 20 dinheiros**, em **grupos de 2 a 4 pessoas**, são os clientes que dão mais gorjetas;
- O melhor dia para **promoção** é **sexta-feira**, no período **diurno(almoço)**, que é o momento com menos movimento neste estabelecimento.

Referências

1. [Notebook de análises Grupo M \(Google Colab\)](#);
2. [Regressão na Prática - Nicholas Richers \(Github\)](#);
3. [Drive Dados Bootcamp \(Neuron\)\(dataset e guias\)](#);
4. Bibliotecas Python: [Pandas](#), [NumPy](#), [Matplotlib](#), [seaborn](#);



Agradecemos sua atenção!

Equipe M