



47%

02

Análise visual

ATIVIDADES
2 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a

Transcrição

Nesta etapa, geraremos alguns gráficos para conseguirmos saber se de fato o **valor da gorjeta** é influenciado com o **pedido da sobremesa** ou não.

O primeiro será um **gráfico categórico** de "sim" ou "não" colocando todos os pontos dos clientes que pediram e dos que não pediram este item, em comparação com o valor da gorjeta em dois eixos x e y . Desta forma, poderemos visualizar melhor essa questão.

Para criar, chamaremos o Seaborn com `sns` seguido de `catplot()`. Dentro deste, passaremos o eixo `x='sobremesa'` e `y='gorjeta'`, finalizando com `data=gorjetas`.

```
sns.catplot(x='sobremesa', y='gorjeta', data=gorjetas)
```

COPIAR CÓDIGO



47%

ATIVIDADES
2 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

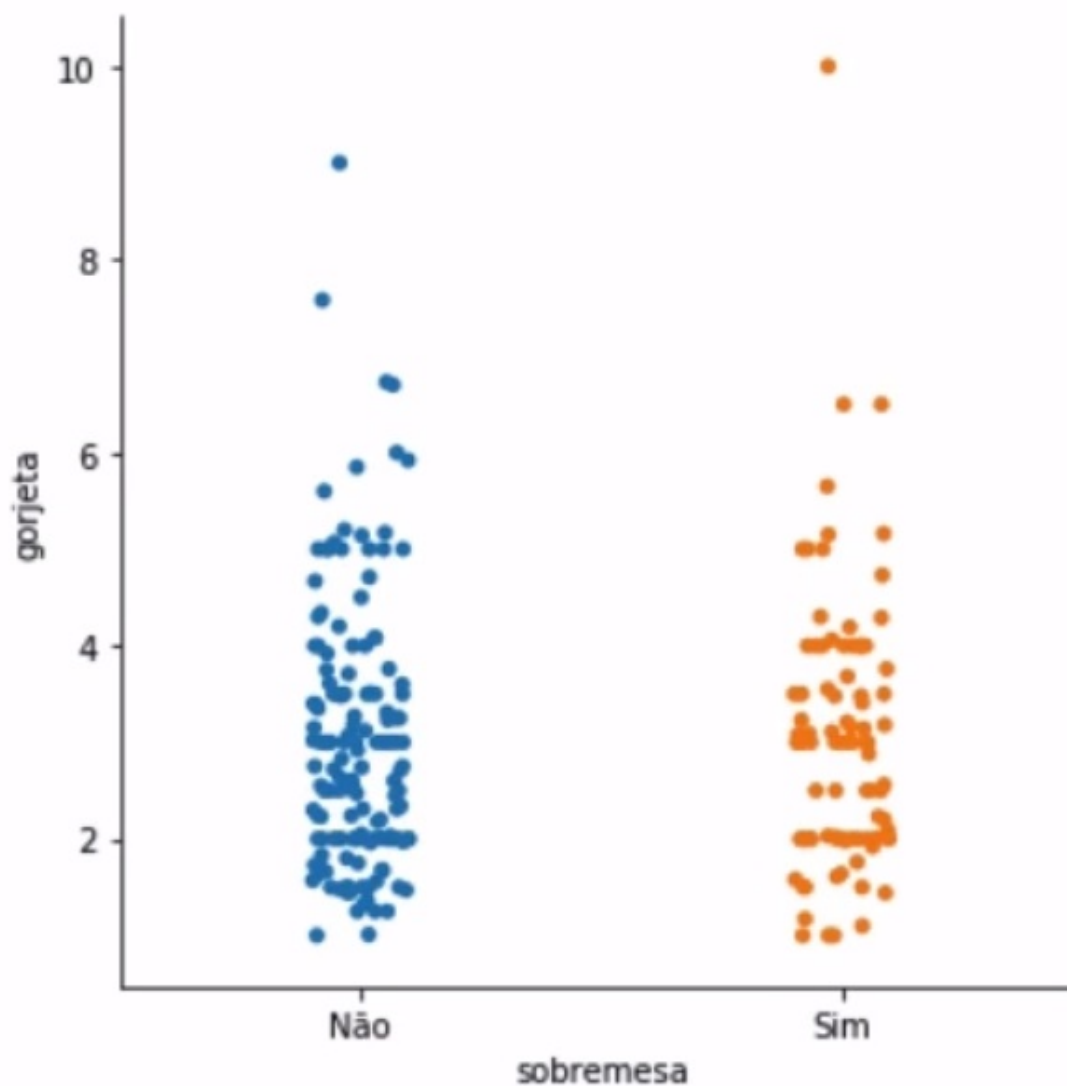
VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a



Executando com "Shift + Enter", teremos este primeiro gráfico categórico com o grupo de pessoas que não pediram sobremesa e os que pediram, em comparação ao valor da gorjeta em cada caso.

Poderemos ver que o valor de gorjeta mais alto está no grupo que pediu a sobremesa. Para validar, iremos à tabela de descrição das ocorrências de `Sim` para o campo `sobremesa` e veremos que o valor máximo de gorjeta foi de \$10 conforme o campo de "max". Enquanto do grupo que não pediu, o valor máximo está em \$9, o que confirmará os dados do gráfico.



47%

ATIVIDADES
2 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a

Porém, ainda não conseguiremos identificar e afirmar que a sobremesa influencia diretamente no valor da gorjeta; é possível ver que há diferenças, mas a resposta não é exata.

Para fazermos esta confirmação, geraremos um segundo gráfico que nos permitirá comprovar essa questão por meio de **tonalidades** diferentes para cada grupo.

Chamaremos a classe do Seaborn com `sns` seguido de `relplot()`. Passaremos `x=valor_da_conta` e `y='gorjeta'` que irá mostrar um aumento conforme vimos anteriormente. Depois, adicionaremos um novo parâmetro `hue` do tipo `'sobremesa'` e finalizaremos com `data=gorjetas`.

```
sns.relplot(x='valor_da_conta', y='gorjeta', hue='sobremesa', data=gorjetas)
```

COPIAR CÓDIGO



47%

ATIVIDADES
2 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

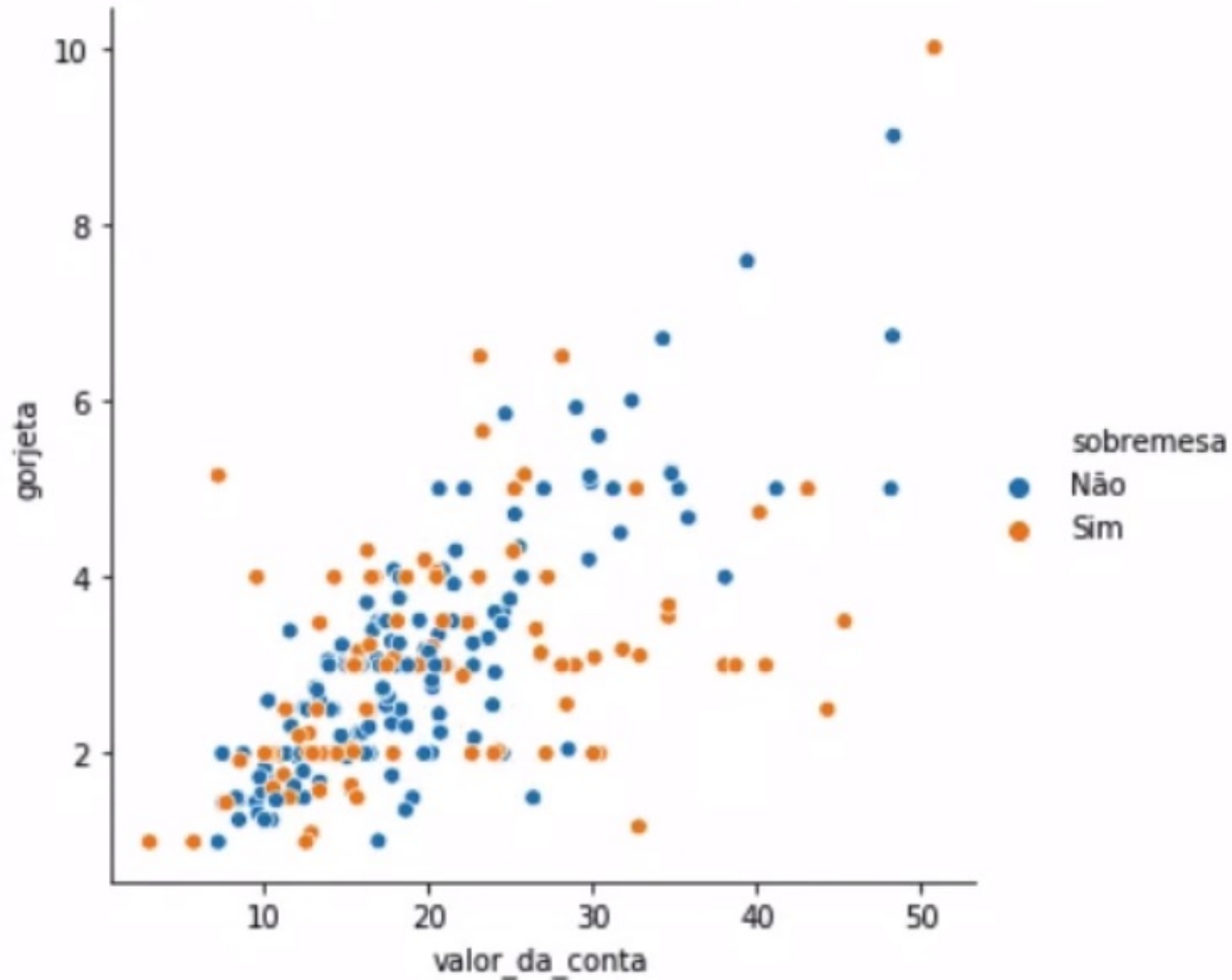
MODO
NOTURNO



14.3k xp

a

`<seaborn.axisgrid.FacetGrid at 0x7ff49c9f6128>`



Analisando este gráfico, poderemos visualizar que os pontos de "não" possuem uma distribuição mais linear do que os de "sim", os quais aparentam mais espaçamento sem grande linearidade.

Criaremos este mesmo gráfico em **duas colunas** para cada grupo. Copiaremos e colaremos a mesma sentença anterior, passando um quarto parâmetro `col` igual a `'sobremesa'`.



47%

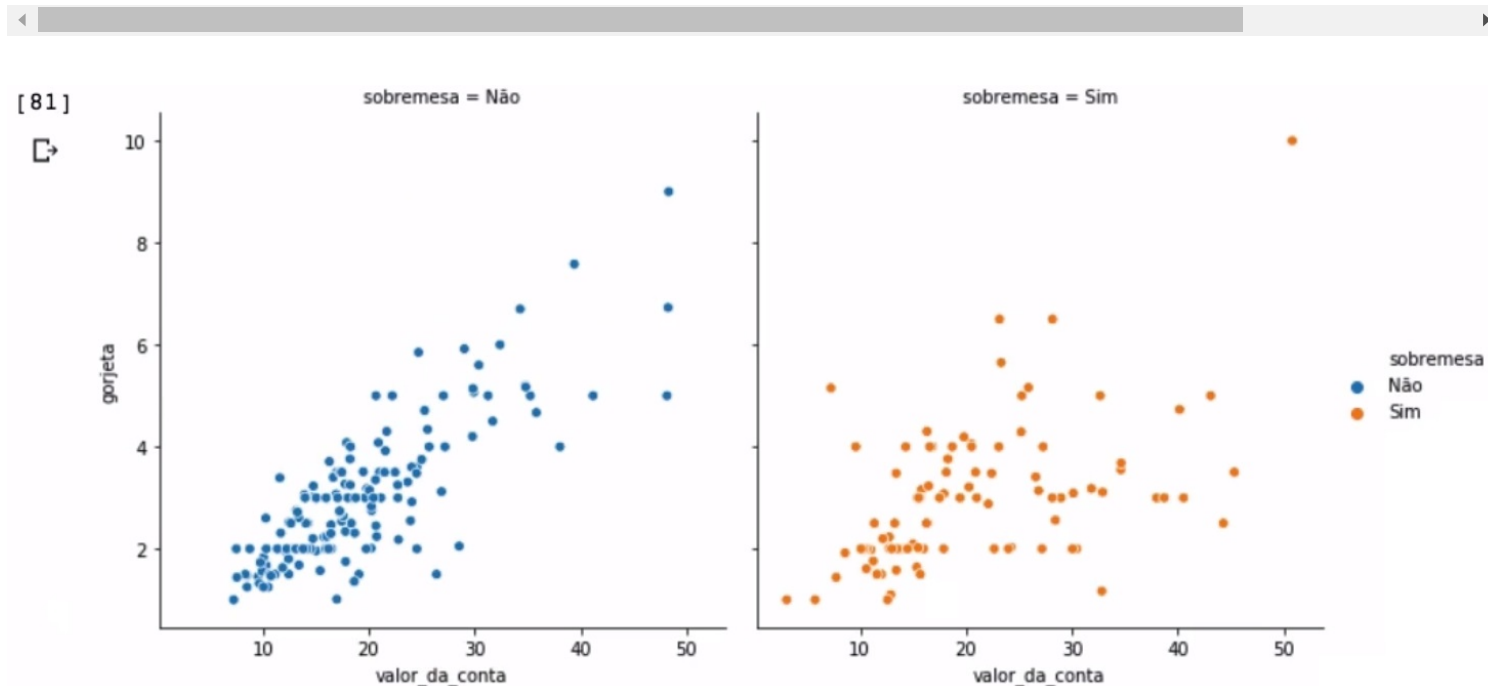
ATIVIDADES
2 de 6DISCORD
ALURAFÓRUM DO
CURSOVOLTAR
PARA
DASHBOARDMODOS
NOTURNO

14.3k xp

a

```
sns.relplot(x='valor_da_conta', y='gorjeta', hue='sobremesa', col='sobremesa', (
```

COPIAR CÓDIGO



No primeiro gráfico de tonalidades, tínhamos todos os *plots* na mesma categoria, e neste último temos cada *plot* em um gráfico diferente legendado.

Com isso, poderemos ver melhor a questão de linearidade de cada caso. Se removermos o parâmetro `hue`, os gráficos serão feitos de uma única cor. É por meio dos parâmetros que realizamos as **edições visuais** também.

```
sns.relplot(x='valor_da_conta', y='gorjeta', col='sobremesa', data=gorjetas)
```

COPIAR CÓDIGO



47%

ATIVIDADES
2 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

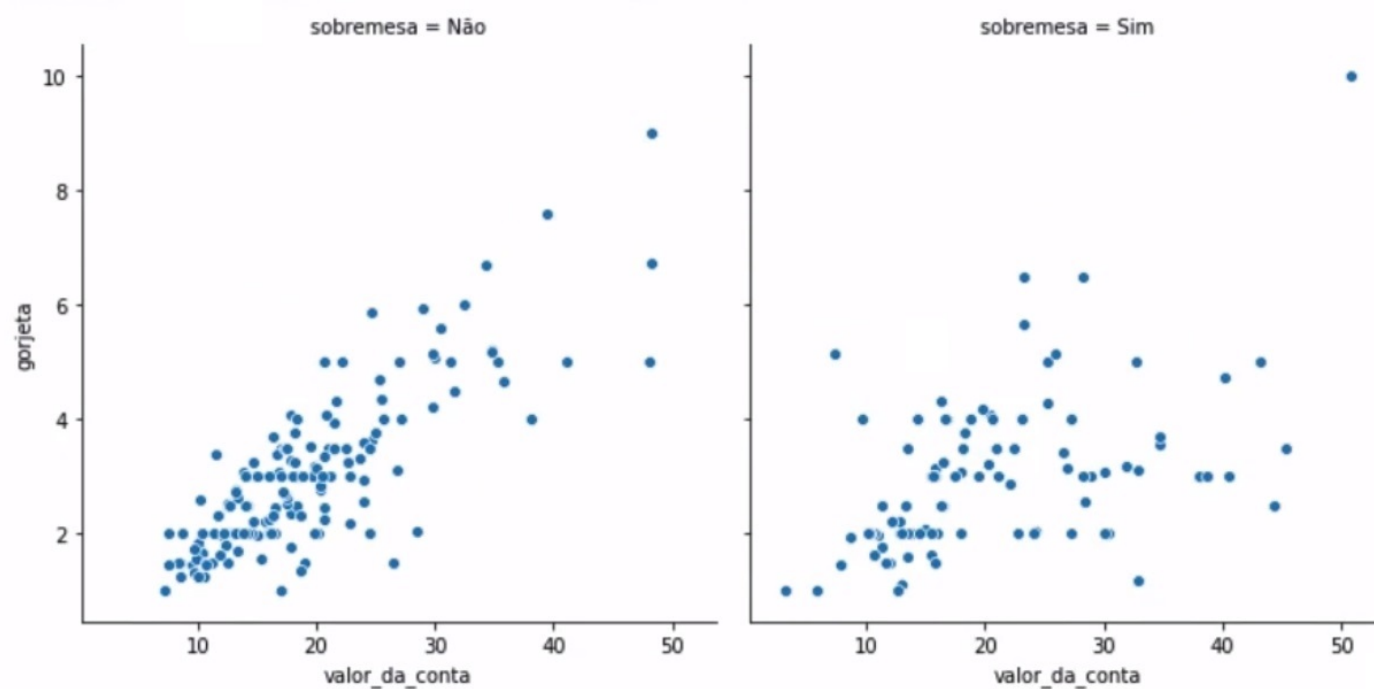
MODOS
NOTURNO



14.3k xp

a

[82]



Se quisermos traçar uma linha que permite uma visualização mais clara da situação, utilizaremos o `lplot()` já citado, passando os eixos `x` e `y` como `'valor_da_conta'` e `'gorjeta'` respectivamente, separando os dois grupos com `col` em tonalidades `hue` e finalizando com `data=gorjetas` .

```
sns.lplot(x='valor_da_conta', y='gorjeta', col='sobremesa', hue='sobremesa', d:
```

COPIAR CÓDIGO





47%

ATIVIDADES
2 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

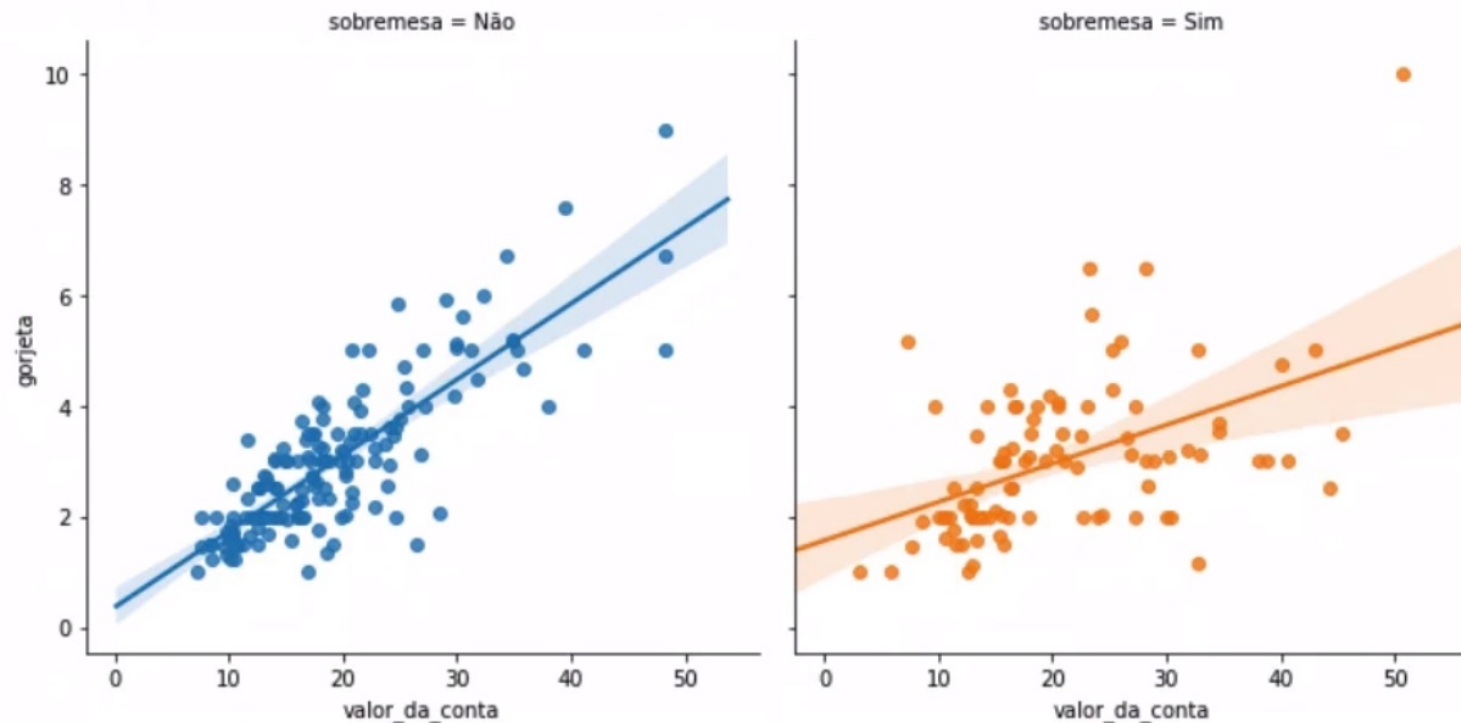
VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a



Com isso, poderemos ver que existe uma diferença de fato; visualmente, identificaremos que aqueles que pediram sobremesa possuem a distribuição diferente em comparação aos que não pediram.

Agora, precisaremos constatar se os dois casos realmente influenciam no valor da gorjeta. Isso só será possível através de um **cálculo estatístico**.

Antes, geraremos mais um gráfico; estamos sempre utilizando um valor da conta e outro de gorjeta, e sabemos que quando um aumenta o outro também aumenta, ainda que sem uma proporção exata. Copiaremos a última sentença escrita e colaremos na célula seguinte, substituindo 'gorjeta' por 'porcentagem' no eixo y .

```
sns.lmplot(x='valor_da_conta', y='porcentagem', col='sobremesa', hue='sobremesa')
```



47%

COPIAR CÓDIGO

ATIVIDADES
2 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

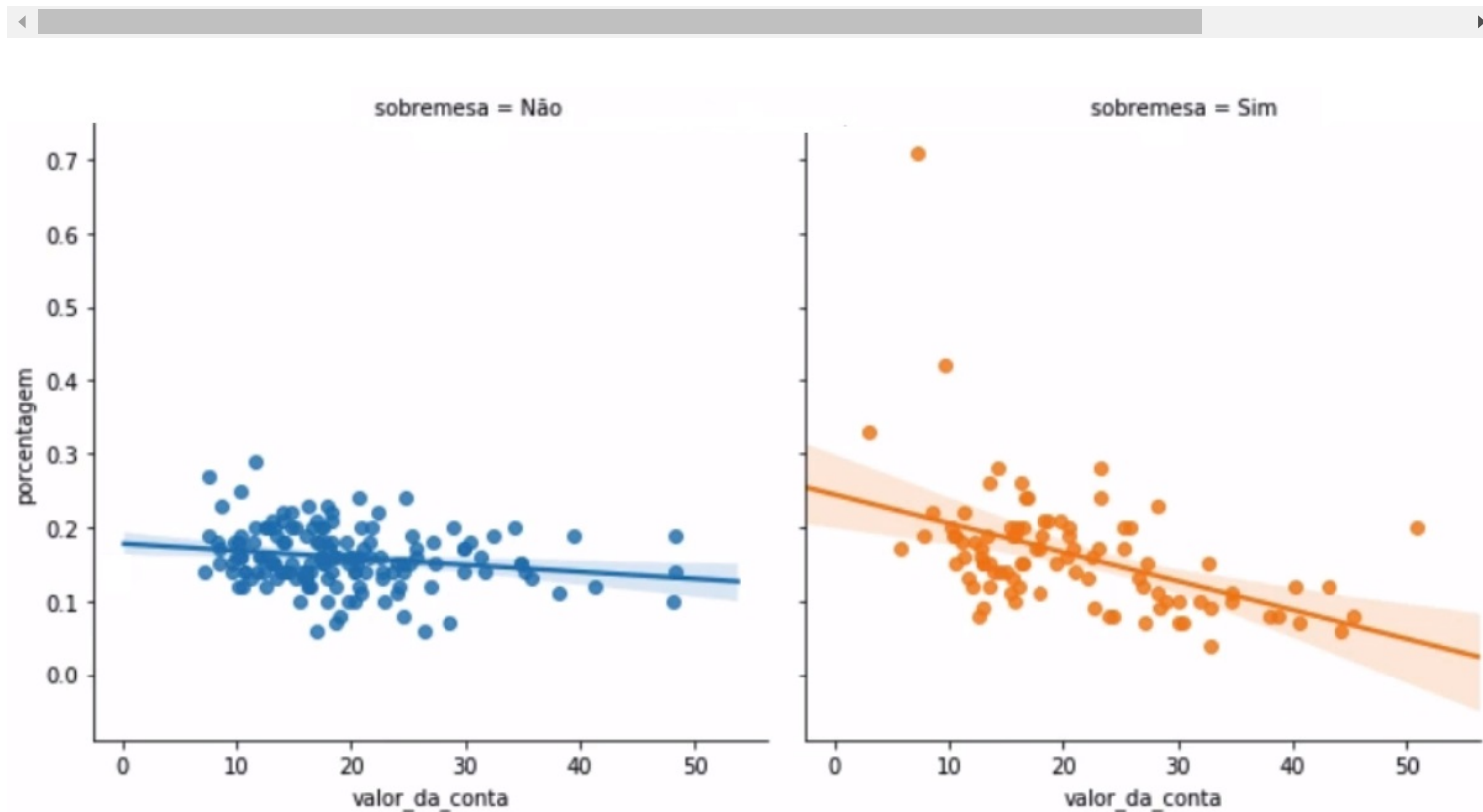
VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a



Desta forma, conseguiremos visualizar que o grupo que não pediu sobremesa possuirá uma linha mais horizontal, enquanto os que pediram apresentará maior inclinação da linha.

Visualmente, parece que existe uma diferença na gorjeta dependendo se a sobremesa é pedida ou não. Da mesma maneira que temos feito, adicionaremos um comentário textual entre `**` com a análise dos gráficos.

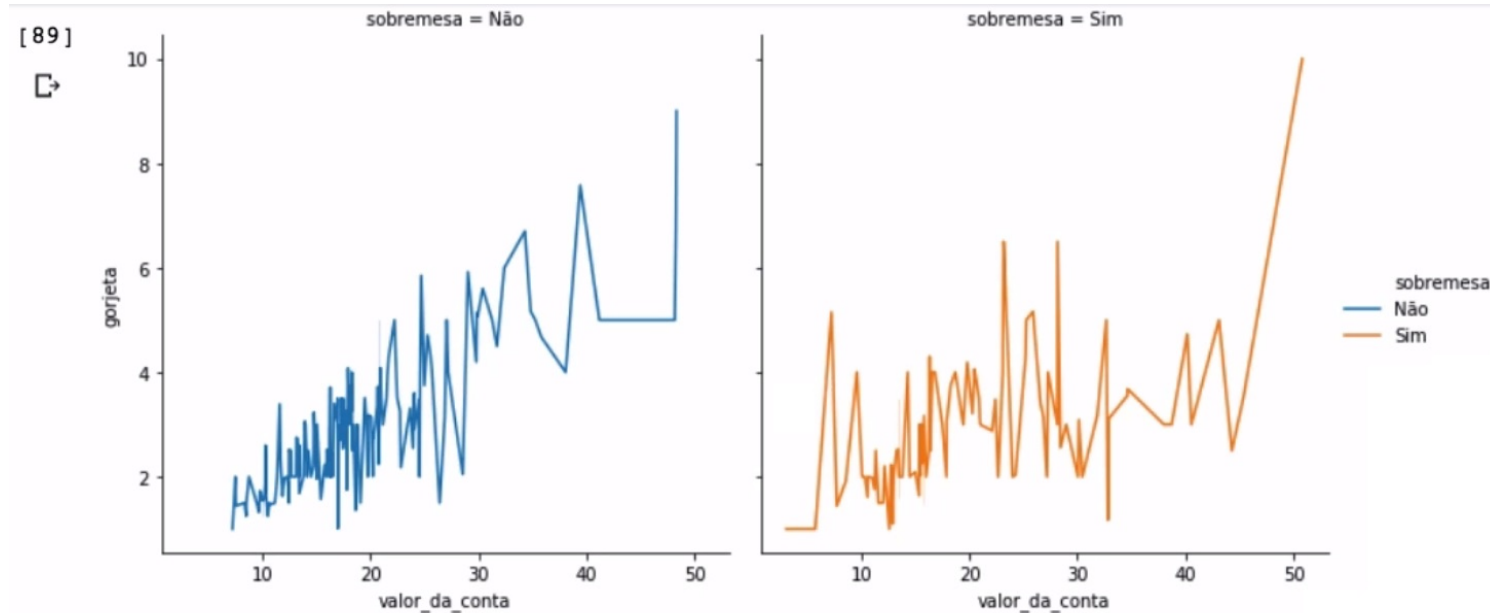
Em todos os gráficos, utilizamos pontos e linhas. Mas em `relplot()` há um parâmetro capaz de transformar **totalmente em linhas** com uso, chamado `kind` do tipo `'line'`.



47%

```
sns.relplot(x='valor_da_conta', y='gorjeta', col='sobremesa', hue='sobremesa', l
```

COPIAR CÓDIGO



Nossa conclusão será que, de fato, existe uma diferença no valor da gorjeta de acordo com o pedido ou não de sobremesa, mas que é bastante sutil. Para termos certeza da influência de um sobre o outro, faremos um **cálculo estatístico** a seguir.



14.3k xp

a