

Importando os dados



5%

ATIVIDADES
3 de 8DISCORD
ALURAFÓRUM DO
CURSOVOLTAR
PARA
DASHBOARDMODO
NOTURNO

13.9k xp

a

Transcrição

Dando início à nossa **análise exploratória de dados**, acessaremos a ferramenta **Colaboratory** do **Google** conforme orientado na [atividade anterior de "Preparando o ambiente"](https://cursos.alura.com.br/course/data-visualization-com-seaborn/task/47270) (<https://cursos.alura.com.br/course/data-visualization-com-seaborn/task/47270>).

Clicaremos em "*File > New Python 3 notebook*" para registrar a nova análise no *notebook* e renomearemos para "EDA-Gorjetas.ipynb".

Nos basearemos em gorjetas e sabemos aplicamos 10% do valor total da conta no Brasil, enquanto em demais países os clientes têm costume de dar caixinha livremente, sem uma porcentagem específica. Com isso, analisaremos diversas influências sobre este valor conforme citamos anteriormente.

Para criar nossas análises, geraremos alguns gráficos a partir do fornecimento de nossa biblioteca. Os dados ali contidos foram tratados pela equipe da Plataforma Alura para fins didáticos e estão disponíveis na mesma atividade "Preparando o ambiente".

Clicaremos na *setinha* lateral para carregarmos os dados.



5%

ATIVIDADES
3 de 8

DISCORD
ALURA

FÓRUM DO
CURSO

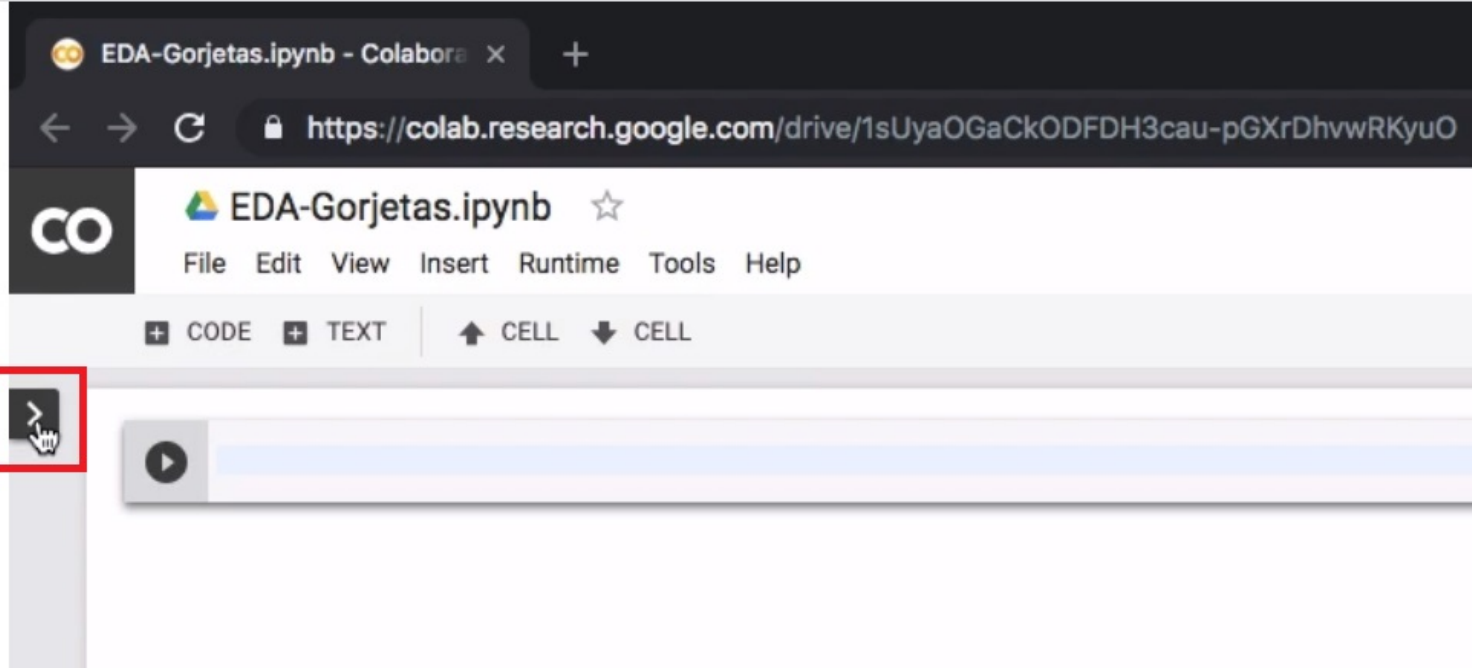
VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



13.9k xp

a



Clicando em "*File*" nas opções desta aba lateral, aguardaremos a conexão para podermos clicar em "*upload*" e selecionar o material `tips.csv` baixado na atividade deste curso. Em seguida, abriremos com "*Open*" para receber uma caixa de diálogo que deve ser concluída com "*OK*". Feito isso, o arquivo aparece na lista desta mesma aba lateral.

Para conseguirmos importar e utilizar estes dados, precisamos do **Pandas**; Antes, adicionaremos um texto para nossa organização clicando em "+ TEXT" e depois em "CELL" com a indicação para deixar a célula no topo. Para padronizarmos e sabermos todas as nossas ações com clareza, escreveremos "# Importante base de dados" e finalizaremos com "Shift + Enter". Desta forma, temos um `h1` para importarmos nossa base de dados.

Na primeira célula de edição, adicionamos o comando `import pandas as pd` e executamos a importação do Pandas. Agora, visualizaremos como este dados estão distribuídos na biblioteca.

Na célula seguinte, chamaremos a biblioteca `pd` com `read` para o sistema ler o arquivo `csv()` , recebendo como parâmetro o nome `'tips.csv'` .



5%

```
pd.read_csv('tips.csv')
```

COPIAR CÓDIGO

Executando com "Shift + Enter", veremos que há 244 linhas com índice numérico e 6 colunas: `"total_bill"`, `"tip"`, `"dessert"`, `"day"`, `"time"` e `"size"`. Em um primeiro momento, entenderemos do que se tratam cada uma.

1. `total_bill`: valor total da conta paga pelo cliente
2. `tip`: valor da gorjeta em si
3. `dessert`: se a pessoa pediu sobremesa ou não
4. `day`: dia da semana
5. `time`: horário de almoço ou horário de jantar
6. `size`: quantidade de clientes na mesa

O primeiro item possui um valor total de conta de \$16,99 dólares cuja gorjeta soma \$1,01. O segundo item possui um total da conta menor e caixinha maior. Com isso, já poderemos ver que nossa análise de dados não é tão simples.

Para que não precisemos carregar a leitura do arquivo todas as vezes, salvaremos o mesmo comando em uma variável `dados` .

```
dados = pd.read_csv('tips.csv')
```

COPIAR CÓDIGO

ATIVIDADES
3 de 8

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODOS
NOTURNO



13.9k xp

a

Desta forma, bastará apenas escrever `dados` na célula para carregar. Como não queremos visualizar a lista completa, escreveremos com a extensão `dados.head()` para ver apenas os cinco primeiros itens.



5%

ATIVIDADES
3 de 8

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



13.9k xp

a