



6%

ATIVIDADES  
3 de 7DISCORD  
ALURAFÓRUM DO  
CURSOVOLTAR  
PARA  
DASHBOARDMODO  
NOTURNO

15.1k xp

a

## Transcrição

Vamos iniciar nosso estudo a respeito de Time Series. Usaremos o Google Colab, que pode ser encontrado no link: <https://colab.research.google.com/notebooks/welcome.ipynb> (<http://>).

Também é possível usar o Anaconda para o desenvolvimento. Na atividade "Preparando o ambiente" será ensinado como desenvolver no Anaconda ou no laboratório do Google.

Para iniciar nosso estudo, precisaremos carregar os dados. Clicaremos na seta do lado esquerdo da tela do Google Colab, e será aberta uma aba. Então, clicaremos no botão "Files > Upload". O link para o download do material desse curso estará disponível também na atividade "Preparando o ambiente". Carregaremos esses arquivos. Eles serão: "alucar.csv", "alucel.csv", "cafelura.csv", "chocolura.csv", "dias\_final\_de\_semana.csv", "newsletter.csv" e "vendas.csv".

Selecionaremos os arquivos e clicaremos em "Open". Os dados serão carregados. Iniciaremos nossos estudos usando os dados da empresa Alucar. Poderemos fechar a aba à esquerda da tela.

Clicaremos no botão "Cell" na parte superior da tela, e criaremos uma célula. Para deixar nosso documento organizado, clicaremos no botão "Text" para que a célula seja do tipo "texto". Usando uma hashtag (#), digitaremos o título "Alucar - Analisando as vendas".

Para conseguirmos ler um arquivo `csv`, precisaremos importar o Pandas. Escreveremos, então:

```
import pandas as pd
```

Pela convenção, sempre chamaremos o Pandas de `pd` no código.

Leremos o arquivo, colocando a função para a leitura, o tipo do arquivo (`csv`) e chamando o arquivo (`alucar.csv`): `pd.read_csv('alucar.csv')`.

Se dermos um "Shift + Enter" e executarmos esse código agora, o programa mostrará todos os registros existentes nessa base de dados. Porém não sabemos qual é o tamanho desse arquivo e a quantidade desses registros. Por isso, usaremos o `.head` ao final, para que seja mostrado somente um "cabeçalho, ou seja, apenas os 5 primeiros registros: `pd.read_csv('alucar.csv').head()`

Eles serão abertos e terão duas colunas: uma referente a alguns meses, e outra com a quantidade de vendas relacionada a esses meses.

Copiaremos o trecho do código que contém todos os registros e armazenaremos isso numa variável, que chamaremos pelo próprio nome do nosso documento:

```
alucar = pd.read_csv('alucar.csv')
```

Agora, se dermos um `alucar.head()` visualizaremos o resultado daquelas 5 linhas das colunas de mês e vendas.

Precisaremos descobrir como tem sido o comportamento das vendas da empresa com o passar do tempo. Antes de criar um gráfico, faremos uma análise desses dados, pois ainda não sabemos quantas colunas, quantas linhas ou quanto tempo de dados temos.



6%

ATIVIDADES  
3 de 7

DISCORD  
ALURA

FÓRUM DO  
CURSO

VOLTAR  
PARA  
DASHBOARD

MODOS  
NOTURNO



15.1k xp

a

Vamos apagar o `alucar.head()` , que já visualizamos, e para descobrir a quantidade de linhas e colunas do arquivo, criaremos um `print`:

```
print('Quantidade de linhas e colunas', alucar.shape)
```

Quando rodarmos esse código, nos será devolvida a informação: "Quantidade de linhas e colunas: (24, 2)". E poderemos observar na visualização do "head" que haverá sempre a quantidade de vendas do último dia do mês para os 5 primeiros meses. Tendo 24 registros, serão 2 anos de dados.

Também poderemos analisar se há dados nulos, e quantos são eles. Para isso, escreveremos:

```
print ('Quantidade de dados nulos', alucar.isna().sum())
```

Dessa forma, será exibida o número de dados nulos que temos no mês (0) e o número de nulos nas vendas (0). Se quisermos que sejam exibidos os dados nulos no total, não importando se eles são para a coluna de mês ou de vendas, colocaremos mais um `.sum` :

```
print ('Quantidade de dados nulos', alucar.isna().sum().sum())
```

Então, será devolvido: "Quantidade de dados nulos: 0. Serão somados os 0 tanto dos dados nulos dos meses quanto das vendas.

Outra questão importante para nossos estudos será saber quais são os tipos de dados que teremos. Para descobri-los, vamos colocar: `alucar.dtypes` .

Pressionaremos "Shift + Enter" e será informado que o mês é do tipo *object*, e as vendas, do tipo *int64*. Como trabalharemos com Time Series, é importante que a coluna dos meses seja do tipo *datetime*. Então faremos a conversão do mês de um *pandas object* para *time*.



6%

ATIVIDADES  
3 de 7

DISCORD  
ALURA

FÓRUM DO  
CURSO

VOLTAR  
PARA  
DASHBOARD

MODOS  
NOTURNO



15.1k xp

a

Para essa modificação, vou chamar o mês, o pandas, e dizer para o mês que preciso convertê-lo para o tipo *datetime*:

```
alucar['mes'] = pd.to_datetime(alucar['mes'])
```

Depois da conversão conferiremos se o tipo foi alterado com: `alucar.dtypes` .

O mês que era tipo *object* agora estará como *datetime64*. Com essa prática, nossos estudos serão facilitados.

Nesse momento, instalaremos o `seaborn` para poder criar nosso primeiro gráfico e saber como as vendas estarão se comportando. Chamaremos o `pip install` com a exclamação na frente, e diremos que essa importação será na versão 0.9.0. Depois de trazer essa versão, diremos que desejamos importar o `seaborn` com o apelido `sns`, que se convencionou usar. Também usaremos o `matplotlib` . Para finalizar, importaremos o `pyplot` dessa biblioteca.

```
!pip install seaborn==0.9.0
import seaborn as sns
%matplotlib inline
from matplotlib import pyplot as plt
```

[COPIAR CÓDIGO](#)

Vamos rodar esse código e o `seaborn` será instalado. Podemos fazer uma verificação para ver qual versão dele teremos, já que o colaboratório vem com uma versão antiga.

Escreveremos, então: `print(sns.__version__)`. Pressionaremos "Enter" e será mostrado a versão correta, a 0.9.0. Caso fosse mostrada a 0.7.0, teríamos que, na barra superior do Colab, clicar em



6%

ATIVIDADES  
3 de 7

DISCORD  
ALURA

FÓRUM DO  
CURSO

VOLTAR  
PARA  
DASHBOARD

MODOS  
NOTURNO



15.1k xp

a

"Runtime > Restart and run all". Nossa máquina virtual, então, seria reiniciada para ficarmos com a versão certa do `seaborn`.



6%

ATIVIDADES  
3 de 7

DISCORD  
ALURA

FÓRUM DO  
CURSO

VOLTAR  
PARA  
DASHBOARD

MODO  
NOTURNO



15.1k xp

a

Faremos nosso primeiro plot a fim de descobrir se as vendas da Alucar aumentam ou diminuem com o passar dos meses. Escrevemos o plot e precisaremos passar dois parâmetros. O `x`, que será o mês, e o `y`, que serão as vendas. Passaremos, também, qual Dataset estaremos utilizando, Alucar.

```
sns.lineplot(x='mes', y='vendas', data=alucar)
```

Pressionaremos "Shift + Enter" e será plotado o gráfico das vendas.

Conforme o tempo passa, as vendas estarão aumentando. Esse resultado alegraria qualquer empresário devido ao resultado.

Entretanto, esse gráfico parecerá pobre por não ter um título e as labels dos eixos `x` e `y` estarem muito pequenas. Poderemos melhorar a aparência dele para apresentá-lo num relatório. Faremos isso na próxima aula.