

DATA ANALYTICS

VISUALIZAÇÃO DE DADOS

AULA 06

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	3
HANDS ON	5
O QUE VOCÊ VIU NESTA AULA?	5
REFERÊNCIAS	6

EMASP

O QUE VEM POR AÍ?

Fala, analytics expert, tudo certo?!

Chegamos em mais um ponto interessante da nossa jornada, onde continuaremos a exploração de uma melhor visualização de dados.

Até o momento trabalhamos com muitos pontos envolvendo a análise de Dataframes de maneira preliminar com o Pandas, passando por algumas funções e também nos plots nativos da própria biblioteca, passando pelos pontos iniciais do Seaborn, manipulando datas e etc.

E, se por um acaso você não lembra direito o código exato, acesse o notebook_completo em: <https://github.com/alura-tech/pos-datascience-introducao-a-visualizacao/archive/refs/heads/aula5.zip> e corra para o abraço!

Nesta aula você continuará entendendo melhor algumas possibilidades de melhorar seus gráficos com o Seaborn, e com isso, fecharemos este conteúdo de visualização de dados com chave de ouro.

Agora vamos para cima desse maravilhoso mundo dos dados!

CONHEÇA SOBRE O ASSUNTO

Sabe quando queremos analisar nossos dados agrupando-os por categoria, média, soma, contagem e por aí vai?

É nisto que consiste a função `groupby()` do Pandas!

Mas assim como em um banco de dados, sempre que formos agrupar os nossos dados por uma coluna, precisamos, posteriormente, declarar mais uma função que nos forneça o tipo de agregação que estamos nos referindo.

Por exemplo, se quisermos agrupar pela soma, podemos utilizar um comando que está contido na nossa videoaula:

```
gastos_por_ano = mensal_do_meu_estado.groupby("ano").sum()
```

Essa linha nos dá um agrupamento do nosso Dataframe pela coluna “ano”, e veja que a função `sum()` vem logo em seguida. Isso significa que tudo o que for valor “numérico” deste Dataframe será somado e agrupado pela coluna “ano”, só que você verá que isso pode ser um problema. Esse Dataframe possui uma coluna chamada “mês” que é numérica e, assim que agrupamos, ele acaba somando a coluna também.

Se isso não for um problema, é possível dropar essa coluna. Mas se você não quiser que isso aconteça, talvez seja melhor pensar de uma outra forma.

Para entender melhor, assista a nossa aula, onde falamos sobre esse assunto e mostramos algumas visualizações bem bacanas com o Seaborn, mudando paleta de cores e tipos de gráfico.

HANDS ON

Agora, chegou o momento de ver, na prática, como começar a importar nossos dados e trabalhar com eles via programação. O ambiente utilizado é o Google Colab, e as bases de dados que foram disponibilizadas no início do material. A ideia é não se limitar apenas ao código explícito no hands on, então é sempre bom procurar a documentação das bibliotecas, explorar novas funcionalidades e muito mais! Nas videoaulas temos mais exemplos do que vimos nesta aula.

O QUE VOCÊ VIU NESTA AULA?

- Como usar o Groupby;
- Como alterar a paleta de cores do seu gráfico com Seaborn.
- Como usar a paleta de cores para ajudar na interpretação dos gráficos.
- O que é e como usar o catplot para comparar categorias.

Material completo: [notebook completo <https://github.com/alura-tech/pos-datascience-introducao-a-visualizacao/archive/refs/heads/aula6.zip>](https://github.com/alura-tech/pos-datascience-introducao-a-visualizacao/archive/refs/heads/aula6.zip).

Daqui em diante, é importante que você replique os conhecimentos adquiridos para fortalecer ainda mais suas bases e conhecimentos, já que um bom ou uma boa cientista de dados não é somente aquele(a) que é uma enciclopédia humana, mas sim aquele(a) que sabe ler um problema e atuar com eficácia.

IMPORTANTE: não esqueça de praticar com o desafio da disciplina, para que assim você possa aprimorar os seus conhecimentos!

Você não está sozinho(a) nesta jornada! Te esperamos no Discord e nas lives com os especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

REFERÊNCIAS

DOCUMENTAÇÃO PANDAS. <<https://pandas.pydata.org/>>. Acesso em: 09 fev. 2023.

GOOGLE COLAB. <<https://colab.research.google.com/>>. Acesso em: 09 fev. 2023.

EMENDAS

PALAVRAS-CHAVE

Python. Pandas. Dataframe.

EMAP

The background is a dark blue field filled with numerous small, light blue dots, resembling a starry sky. Overlaid on this are several large, flowing, wavy lines in shades of teal, blue, and yellow. These lines create a sense of movement and depth. Scattered throughout the composition are various geometric shapes: a thin vertical line, a circle containing the number '7', a small circle, a cross, a small circle, and a hexagon.

POSTECH