

DATA ANALYTICS
VISUALIZAÇÃO DE DADOS

AULA 04

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	3
HANDS ON	4
O QUE VOCÊ VIU NESTA AULA?	4
REFERÊNCIAS	6

EMSE

O QUE VEM POR AÍ?

Até aqui você teve uma noção de como carregar fontes de dados de maneiras distintas, aprendeu algumas manipulações base para sua jornada, também conheceu a biblioteca Seaborn do Python e aprendeu algumas funções que combinam Matplotlib e Seaborn.

Essa aula vai lhe trazer uma outra noção de modificações dentro de um Dataframe, manipulando datas e utilizando a função melt para melhorar nossas análises base.

A base de dados você encontra aqui <<https://github.com/alura-tech/pos-datascience-introducao-a-visualizacao/archive/refs/heads/dados.zip>>.

Agora vamos para cima desse maravilhoso mundo dos dados!

CONHEÇA SOBRE O ASSUNTO

Jovem padawan, um dos maiores desafios quando o assunto é data, é como tratar em formato de data efetivamente, porque muitas vezes o campo vem no formato de string, e não necessariamente adequado para manipular.

Por isso, o Python possui uma biblioteca maravilhosa chamada Datetime <<https://docs.python.org/3/library/datetime.html>> que nos ajuda a trabalhar com datas e tempos diversos. Não é a única biblioteca do Python, mas ela é a mais famosa e outras a utilizam como base.

Na primeira parte da videoaula, trabalhamos um jeito bacana de modificar nosso índice de datas, utilizando a função nativa map() e a biblioteca Datetime. Assista para aprender mais detalhes sobre esse assunto!

Se você observar na videoaula 1, o objeto que retorna naquela função para_dia() é algo do tipo Date (ano, mês, 1), sendo o último parâmetro um valor fixo, porque não nos interessa uma variação diária, e sim as variações mês a mês por anos.

Quando plotamos o gráfico em seguida, é interessante verificar se o formato de data padrão é bonito de ser apresentado.

Outro ponto bem interessante é a utilização da função `melt()`, que nos permite criar um novo Dataframe baseado em uma coluna para melhorar nossas análises.

O conceito fica bem claro na videoaula 2, então recomendamos que assista ao vídeo!

Veja que, no exemplo do vídeo, estávamos trabalhando com uma tabela onde cada coluna era uma unidade da federação diferente, no qual os valores numéricos estavam variando mês a mês. A função `melt()` nos ajudou a pivotar a tabela e deixar em uma visão clara nas variações mensais por UF.

HANDS ON

Agora chegou o momento de ver, na prática, como começar a importar os 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 da disciplina. A ideia é não se limitar apenas ao código explícito no hands on, então procure a documentação das bibliotecas, explore novas funcionalidades e muito mais! Lembrando que estamos disponíveis no Discord para te ajudar durante todo esse processo.

Confira a videoaula e faça as sugestões de desafios! O material completo da videoaula você encontra aqui <https://github.com/alura-tech/pos-datascience-introducao-a-visualizacao/archive/refs/heads/aula4.zip>.

O QUE VOCÊ VIU NESTA AULA?

Aprendemos a quais caminhos não seguir na hora de trabalhar com Datatime, como usar `Map` aplicando uma função no `DataFrame` e como usar o `Melt` no processamento dos dados.

Daqui em diante, é importante que você replique os conhecimentos adquiridos para fortalecer ainda mais suas bases e conhecimentos, já que uma pessoa boa em ciência de dados não é somente aquela que é uma enciclopédia humana, mas sim aquela 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 só nesta jornada! Te esperamos no Discord e nas lives com os nossos docentes, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

EMASP

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, wavy, translucent lines in shades of blue, yellow, and red. These lines flow from the left side towards the right, creating a sense of movement. Scattered throughout the composition are various geometric shapes: a thin vertical line, a circle containing the number '7', a small circle, an 'X' mark, a small yellow circle, and a hexagon in the bottom right corner.

POSTECH