

DATA ANALYTICS

ANÁLISE EXPLORATÓRIA DE DADOS

AULA 05

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	4
HANDS ON	11
O QUE VOCÊ VIU NESTA AULA?	12
REFERÊNCIAS	13

EMSE

O QUE VEM POR AÍ?

Fala analytics expert, tudo certo?!

Chegamos em um ponto importante em nossa jornada, onde continuaremos nos aprofundando nas manipulações 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 nos plots nativos da própria biblioteca.

E, se por acaso, você não lembra direito o código exato, acesse-o aqui: <https://github.com/alura-tech/pos-datascience-analise-e-exploracao-de-dados/tree/aula4>

Nesta aula, você continuará entendendo melhor como realizar uma boa visualização gráfica. É muito importante que você sempre revise as aulas anteriores, faça os exercícios e bote em prática cada detalhe. Lembre-se que o hands on é fundamental e estamos com você em toda esta jornada, sempre à disposição no Discord para te ajudar com possíveis dúvidas!

Agora, vamos dominar esse maravilhoso mundo dos dados!

CONHEÇA SOBRE O ASSUNTO

Quando pensamos em qual é o melhor gráfico para representar o que quero dizer com meus dados, é comum termos dúvidas, já que existe uma infinidade deles.

Uma maneira interessante de aprender os tipos de gráficos e o que a própria função `plot()` pode fazer, é analisar a documentação oficial da função no endereço: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html>.

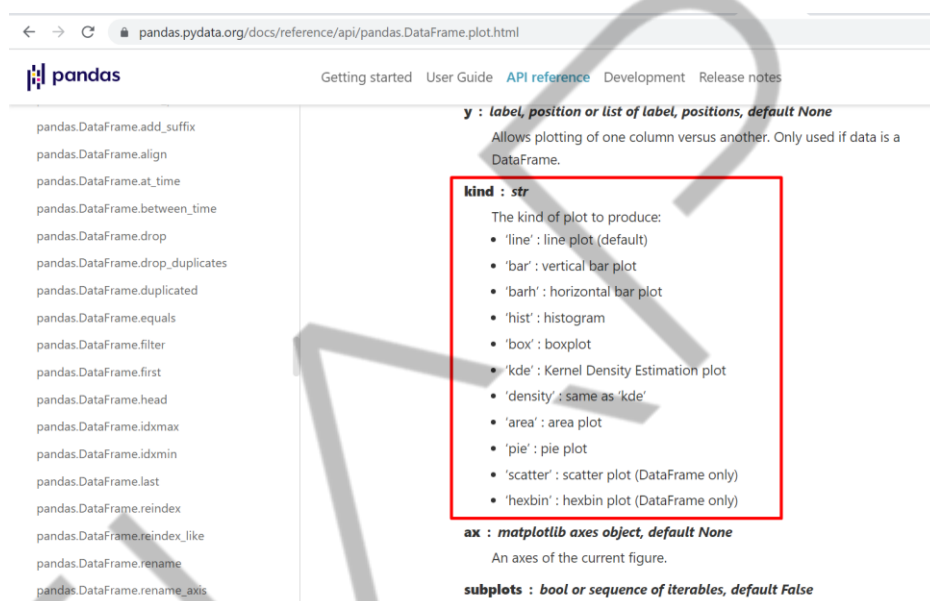


Figura 1 – Documentação `plot()` do pandas
Fonte: Elaborado pelo autor (2023)

Veja que no parâmetro `kind` você tem alguns tipos de gráficos e, para cada um deles, existe uma melhor escolha na hora de decidir qual utilizar. Por exemplo, se você tiver poucas categorias onde quer representar a porcentagem em relação ao todo, o ideal é o “pie”, mais conhecido como “gráfico de setor”. Agora, se você quer analisar uma série temporal, o ideal é que você utilize o “line”, que já é um parâmetro default, ou seja, não precisa passar esse parâmetro.

Além do tipo de gráfico, você tem uma série de parâmetros que ajudam a modificar o gráfico. Só tome cuidado, porque para certas modificações você precisará de bibliotecas gráficas, como o `matplotlib`.

Os gráficos mais comuns usados com a função `plot()` do Pandas incluem:

1. Line Plot (Gráfico de Linha): para exibir tendências em dados de séries temporais.
2. Bar Plot (Gráfico de Barras): para comparar valores categóricos.
3. Histogram (Histograma): para visualizar a distribuição de frequência de dados numéricos.
4. Pie Chart (Gráfico de Pizza): para representar a proporção de cada categoria em relação ao todo.
5. Scatter Plot (Gráfico de Dispersão): para visualizar relações entre dois conjuntos de dados numéricos.

E para ficar claro onde queremos chegar, nada melhor do que verificar alguns exemplos de código em Python que gerem esses gráficos:

Line Plot:

```
import pandas as pd
import matplotlib.pyplot as plt

data = {'Year': [2010, 2011, 2012, 2013, 2014],
        'Sales': [1000, 1100, 1050, 1075, 1150]}
df = pd.DataFrame(data)

df.plot(x='Year', y='Sales', kind='line')
plt.show()
```

Resultado:



Figura 2 – Gráfico de linha
Fonte: Elaborado pelo autor (2023)

Bar Plot:

```
import pandas as pd
import matplotlib.pyplot as plt

data = {'City': ['New York', 'London', 'Paris', 'Berlin'],
        'Sales': [1000, 800, 900, 700]}
df = pd.DataFrame(data)

df.plot(x='City', y='Sales', kind='bar')
plt.show()
```

Resultado:

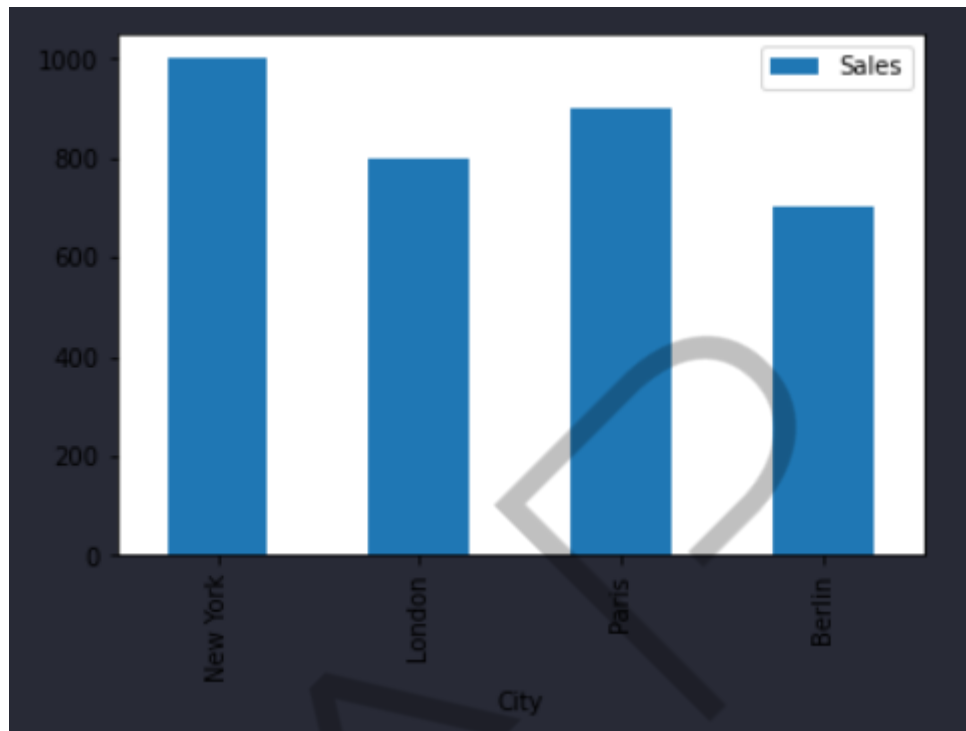


Figura 3 – Gráfico de Barras
Fonte: Elaborado pelo autor (2023)

Histogram:

```
import pandas as pd
import matplotlib.pyplot as plt

data = [10, 20, 15, 30, 25, 40, 35, 50]
df = pd.DataFrame(data, columns=['Data'])

df.plot(kind='hist', bins=15)
plt.show()
```

Resultado:

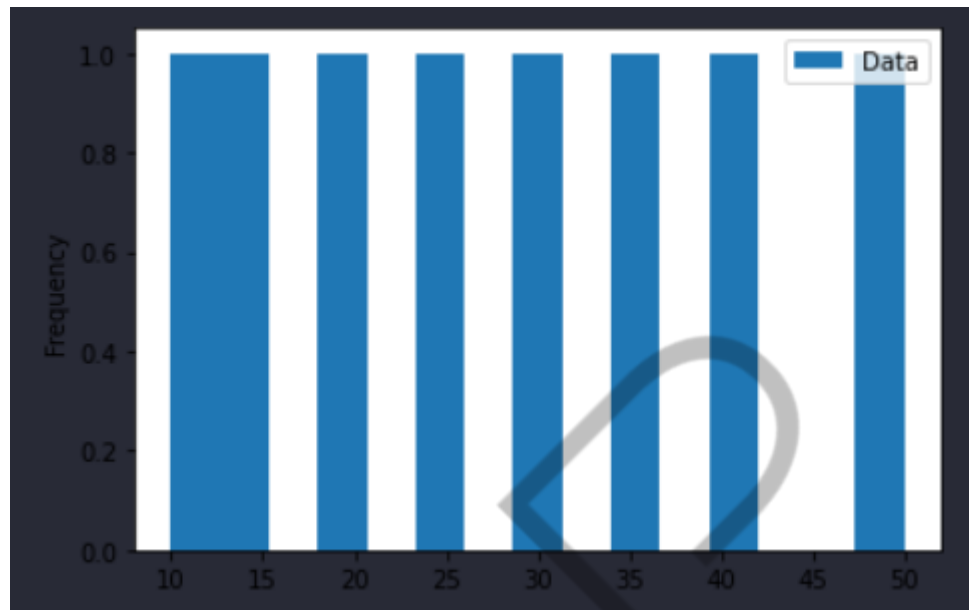


Figura 4 – Histograma
 Fonte: Elaborado pelo autor (2023)

Pie chart (pizza):

```
import pandas as pd
import matplotlib.pyplot as plt

data = {'Country': ['USA', 'UK', 'France', 'Germany'],
        'Sales': [1000, 800, 900, 700]}
df = pd.DataFrame(data)

df.plot(kind='pie', y='Sales', labels=df['Country'])
plt.axis('equal')
plt.show()
```

Resultado:

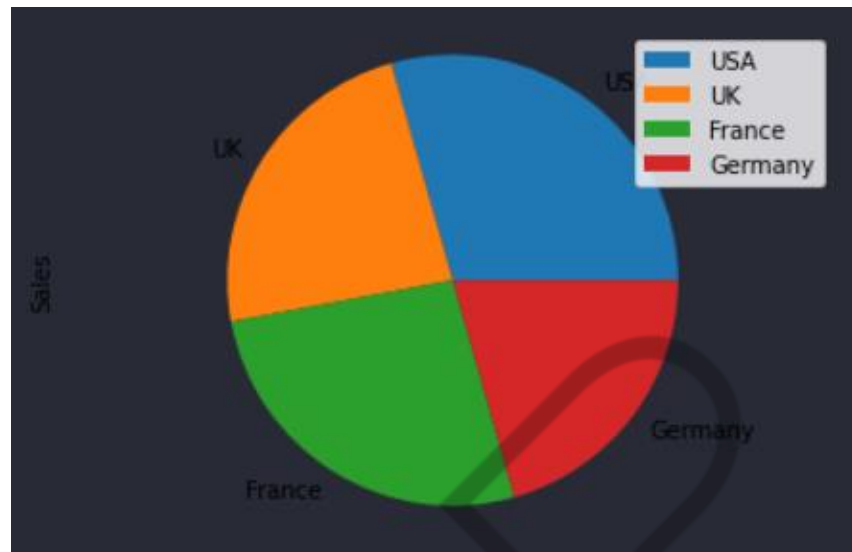


Figura 5 – Gráfico de Pizza
Fonte: Elaborado pelo autor (2023)

Scatter Plot:

```
import pandas as pd
import matplotlib.pyplot as plt

data = {'X': [10, 20, 30, 40, 50],
        'Y': [15, 25, 35, 45, 55]}
df = pd.DataFrame(data)

df.plot(x='X', y='Y', kind='scatter')
plt.show()
```

Resultado:



Figura 6 – Scatter Plot
Fonte: Elaborado pelo Autor (2023)

HANDS ON

Agora, chegou o momento de ver, na prática, como começar a importar nossos dados e trabalhar com eles via programação. A ideia é não se limitar apenas ao código apresentado no hands on, então, procure a documentação das bibliotecas, explore novas funcionalidades e muito mais!

A seguir, temos o exemplo de um vídeo que se aproxima do que foi feito. É muito importante que você assista às aulas para compreender totalmente o assunto!

Lembre-se de que, apesar dos exemplos de gráficos no decorrer deste material, existem várias combinações possíveis e que geram a melhor visualização a partir do que precisamos, e indicam que você precisa se adaptar às mudanças necessárias.

Vamos para cima!

O QUE VOCÊ VIU NESTA AULA?

Você viu como os dados represados podem impactar os dados, quais as dificuldades de interpretar e realizar comparações em um gráfico de pizza e como utilizar uma tabela para transmitir informações.

Material completo: notebook_completo <https://github.com/alura-tech/pos-datascience-analise-e-exploracao-de-dados/blob/aula5/Produ%C3%A7%C3%A3o_Hospitalar-51.ipynb>.

Daqui para a frente, é importante que você replique os conhecimentos adquiridos para fortalecer mais suas bases e conhecimentos, já que uma pessoa ótima 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 sozinho ou sozinha nesta jornada! Te esperamos no Discord e nas lives com os nossos especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

REFERÊNCIAS

GOOGLE COLAB. <<https://colab.research.google.com/>>. Acesso em: 02 dez. 2022

DOCUMENTAÇÃO PANDAS. <<https://pandas.pydata.org/>>. Acesso em: 02 dez. 2022

DATASUS. <<https://datasus.saude.gov.br/>>. Acesso em: 02 dez. 2022

TABNET. <<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>>. Acesso em: 02 dez. 2022

PALAVRAS-CHAVE

Python. Pandas. Dataframe.

EMENDAS

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 motion and depth. Scattered throughout the composition are various geometric shapes: a circle with the number '7' inside, a circle with an 'X' inside, a circle with the number '0' inside, and a hexagon in the bottom right corner. The overall aesthetic is futuristic and technological.

POSTECH