

Introdução ao PYTHON para Inteligência Artificial – PLP – IMT (2025.1)

2. Matplotlib

Matplotlib é uma biblioteca Python abrangente dedicada ao traçado gráfico em 2D para a criação de visualizações estáticas em Python. Ela foi projetada com a filosofia de permitir criar plotagens (gráficos) simples com poucos comandos. A ideia é que para ver um histograma de seus dados, por exemplo, você não precise dispendar muito esforço instanciando objetos, chamando métodos, definindo propriedades e assim por diante.

Para explorar a Matplotlib, construiremos exemplos de complexidade crescente que você deve experimentar de forma prática. Os experimentos vão utilizar os conhecimentos de bibliotecas como o NumPy e Pandas. NumPy é uma biblioteca para manipulação de vetores e matrizes formadas exclusivamente por valores numéricos.

Para conhecer e exercitar a biblioteca Matplotlib utilizaremos uma nova base de dados. Ela foi criada a partir de dados do ENADE de 2017, sendo um recorte das provas que envolveram os cursos de computação.

Em primeiro lugar carregaremos a base de dados com os comandos que já conhecemos e após criar o Dataframe usaremos os dados para plotar gráficos com a Matplotlib.

Como usaremos o Pandas para manipular a base de dados, temos que fazer o *import*:

```
import pandas as pd
```

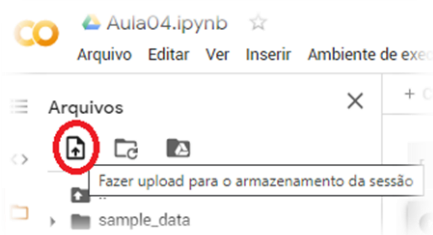
Em seguida carregamos a base. Você pode usar os comandos a seguir ou ainda, se tiver problemas com eles, usar as opções de arquivos no menu a esquerda do Colab:

```
from google.colab import files
uploaded = files.upload()
```

E selecione o dataset “ENADE2017.xlsx”. O arquivo referente a esse dataframe foi disponibilizado como material para essa aula.



Alternativamente...



Após carregar o arquivo crie o Dataframe “enade” que tem 1695 linhas e 21 colunas:

```
enade = pd.read_excel('ENADE2017.xlsx')
```

Use também o “shape” e o “head” para conhecer a base e navegue pelas primeiras linhas e as diversas colunas:

```
print(enade.shape)
```

```
(1695, 21)
```

```
enade.head()
```

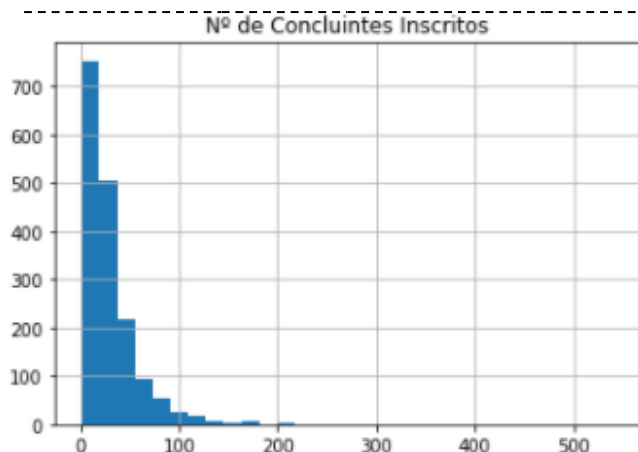
	Ano	Código da Área	Área de Avaliação	Código da IES	Nome da IES	Sigla da IES	Organização Acadêmica	Categoria Administrativa	Código do Curso	Modalidade de Ensino	Código do Município	Município do Curso	Sigla da UF	Nº de Concluintes Inscritos	Nº de Concluintes Participantes
0	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	550	UNIVERSIDADE ESTADUAL DA PARAIBA	UEPB	Universidade	Pública Estadual	1265453	Educação Presencial	2510808	Patos	PB	27	1
1	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	1066	INSTITUTO DE ENSINO E PESQUISA OBJETIVO	IEPO	Faculdade	Privada com fins lucrativos	58808	Educação Presencial	1721000	Palmas	TO	1	0
2	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	2676	FACULDADE LA SALLE	NaN	Faculdade	Privada sem fins lucrativos	73438	Educação Presencial	1302603	Manaus	AM	45	6
3	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	4656	FACULDADE ANHANGÜERA DE PIRACICABA	NaN	Faculdade	Privada com fins lucrativos	104836	Educação Presencial	3538709	Piracicaba	SP	2	1
4	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	575	UNIVERSIDADE FEDERAL DE MINAS GERAIS	UFMG	Universidade	Pública Federal	12946	Educação Presencial	3106200	Belo Horizonte	MG	79	72

Agora sim vamos importar a Matplotlib e começar a usá-la:

```
import matplotlib.pyplot as plt
```

Logo de início faremos nosso primeiro histograma a partir da coluna “Nº de Concluintes Inscritos”. Diferentemente de um gráfico de barras comum, cada coluna de um histograma representa um intervalo de valores e não um valor específico. Note que indicamos ao “hist” qual coluna de dados usar com o parâmetro “column=”. O parâmetro “bins” define o número de bins (barras de intervalo) no histograma:

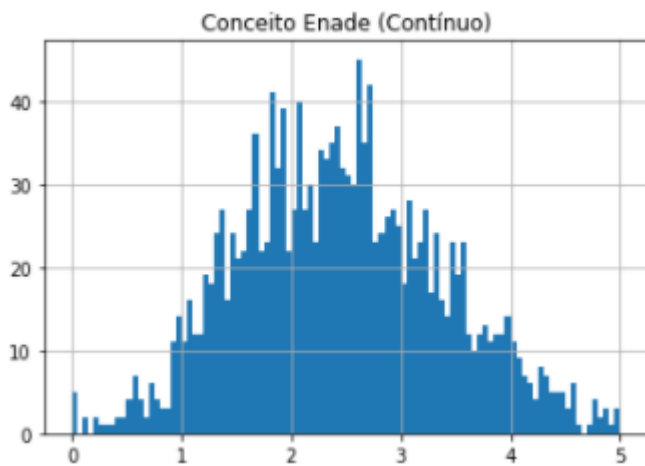
```
enade.hist(column='Nº de Concluintes Inscritos',bins = 30)  
plt.show()
```



Percebemos pelo histograma que a grande maioria dos cursos tem de 0 a 100 inscritos e a maioria dos cursos tem poucos inscritos.

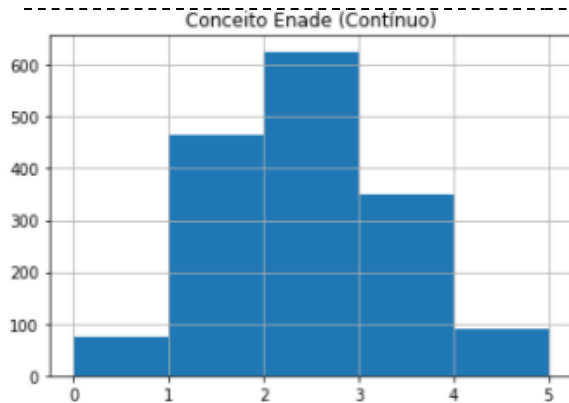
Se fizermos o mesmo para a coluna “Conceito Enade (Contínuo)” teremos:

```
enade.hist(column='Conceito Enade (Contínuo)',bins = 100)  
plt.show()
```



Ao alterar o parâmetro “bins” para “bins = 5”, teremos:

```
enade.hist(column='Conceito Enade (Contínuo)',bins = 5)  
plt.show()
```



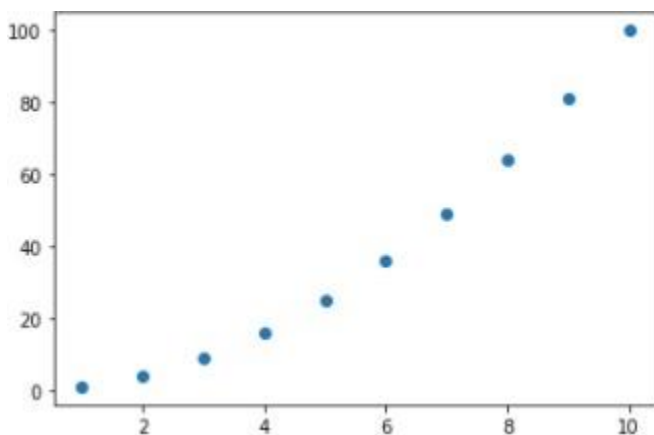
Veremos agora como construir gráficos com valores para x e y representando a relação entre duas variáveis. Para isso inicialmente formamos dois conjuntos de valores considerando a função $y=x^2$:

```
x = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
y = []
for i in x:
    y.append(i**2)
print(y)
```

```
[1, 4, 9, 16, 25, 36, 49, 64, 81, 100]
```

Gráficos que representam a relação entre funções ($y=x^2$) e variáveis (x e y) são chamados de gráficos de dispersão, ou “scatter”:

```
plt.scatter(x, y)
plt.show()
```



Veja agora um exemplo de manipulação de dados e apresentação de gráficos de dispersão para encontrar tendências. Em primeiro lugar selecionamos as escolas:

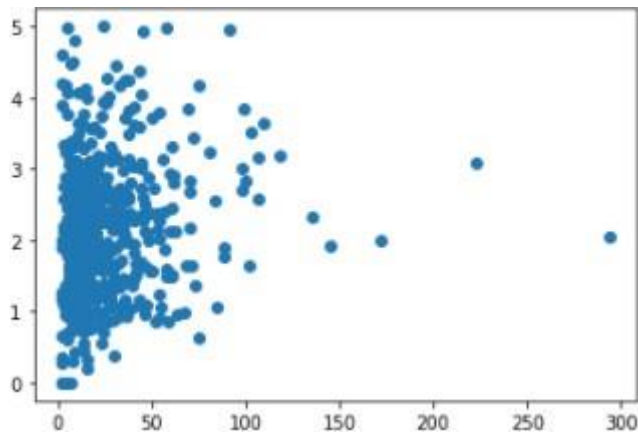
```
escolassp = enade.loc[enade['Sigla da UF']=='SP']
escolassp.head()
```

	Ano	Código da Área	Área de Avaliação	Código da IES	Nome da IES	Sigla da IES	Organização Acadêmica	Categoria Administrativa	Código do Curso	Modalidade de Ensino	Código do Município	Município do Curso	Sigla da UF	Nº de Concluintes Insritos	Nº de Concluintes Participantes	Nota Bruta - FG
3	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	4656	FACULDADE ANHANGUERA DE PIRACICABA	NaN	Faculdade	Privada com fins lucrativos	104836	Educação Presencial	3538709	Piracicaba	SP	2	1	49.400000
5	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	54	UNIVERSIDADE ESTADUAL DE CAMPINAS	UNICAMP	Universidade	Pública Estadual	2717	Educação Presencial	3509502	Campinas	SP	48	45	64.480000
13	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	591	UNIVERSIDADE FEDERAL DE SÃO PAULO	UNIFESP	Universidade	Pública Federal	110528	Educação Presencial	3549904	São José dos Campos	SP	45	38	64.542105
14	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	7	UNIVERSIDADE FEDERAL DE SÃO CARLOS	UFSCAR	Universidade	Pública Federal	112680	Educação Presencial	3552205	Sorocaba	SP	44	43	61.920930
28	2017	4004	CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)	4925	FUNDAÇÃO UNIVERSIDADE FEDERAL DO ABC	UFABC	Universidade	Pública Federal	1102342	Educação Presencial	3547809	Santo André	SP	93	69	64.811594

Será que existe uma relação entre as colunas “Número de Concluintes Participantes” e “Nota Padronizada”?

```
npartic = escolassp['Nº de Concluintes Participantes']  
nota = escolassp['Nota Padronizada - CE']
```

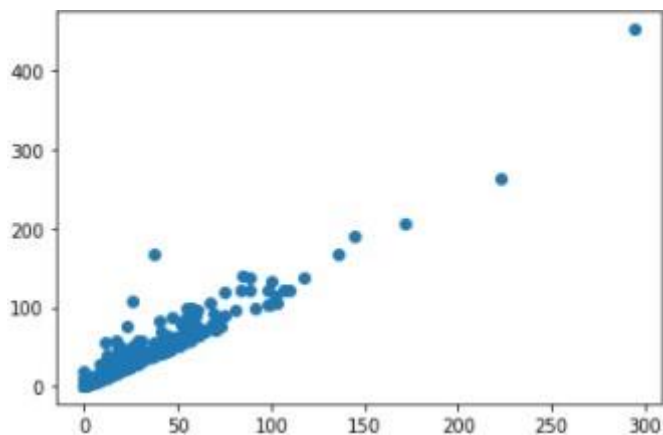
```
plt.scatter(npartic, nota)  
plt.show()
```



A coluna “Nota Padronizada” representa a nota de qualidade da instituição, podendo variar de 0 até 5. Podemos observar que indetificar que não necessariamente uma nota maior implica em uma maior quantidade de participantes pois as instituições com a maior quantidade de participantes possuem um valor de nota entre 1,5 e 3,5.

Por fim vamos apresentar a relação entre número de concluintes participantes e o número de concluintes inscritos.

```
ninscritos = escolassp['Nº de Concluintes Inscritos']  
plt.scatter(npartic, ninscritos)  
plt.show()
```



Neste caso parece existir uma relação entre x e y: Quanto maior o número de inscritos, maior o número de participantes (alunos que foram de fato realizar a prova e não apenas se inscreveram).

Veja agora alguns comandos para personalizar os gráficos criados:

- Colocar nome no eixo x e y com “xlabel” e “ylabel”

```
plt.ylabel('Inscritos')
```

```
plt.title('Participantes')
```

- Indicação dos limites dos eixos apresentados no gráfico com “xlim” e “ylim”:

```
plt.xlim(0, 400)
```

```
plt.ylim(0, 500)
```

- Grade de apresentação com “grid”:

```
plt.grid(True)
```

Veja o exemplo de uso no último gráfico apresentado:

```
plt.ylabel('Concluintes')
plt.title('Participantes')
plt.xlim(0, 400)
plt.ylim(0, 500)
plt.grid(True)
plt.scatter(nconcluintes, nparticipantes)
plt.show()
```

