

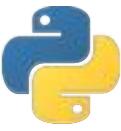
# Fuctura

## Módulo 3 - Introdução à Data Science e Machine Learning

---

Prof. Jéssica Andrade





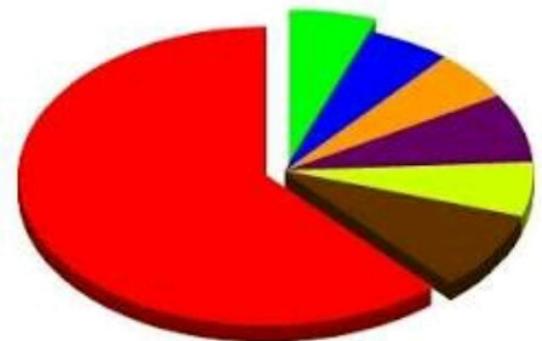
# matplotlib

cita

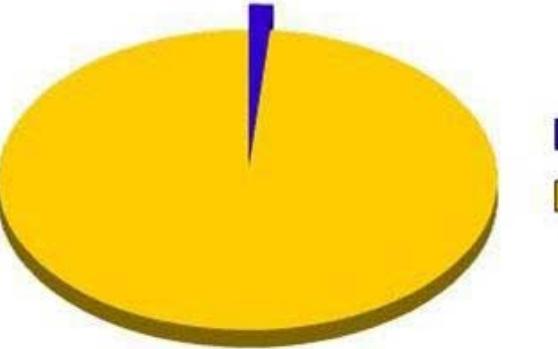


# Gráficos

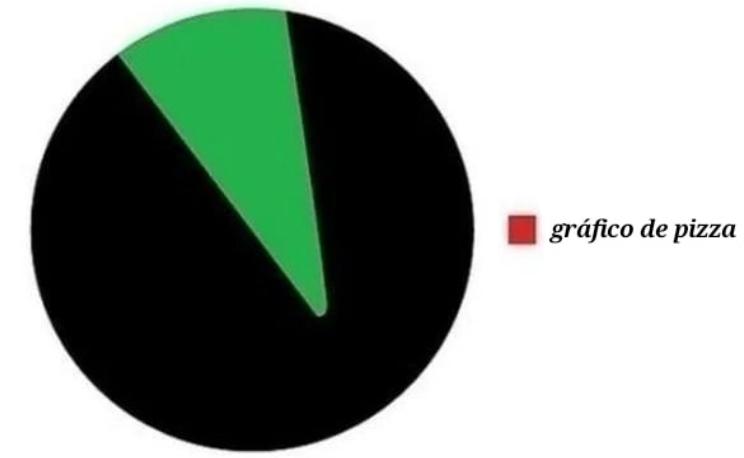
## Quando você olha para o seu celular



## O QUE VOCÊ FAZ QUANDO SE SENTE MAL



## Coisas que não entendo





# Objetivo

---

O objetivo dessa aula é **compreender o processo e fixar as etapas** de elaboração dos gráficos mais fundamentais do Matplotlib.



## O que é o Matplotlib?

**Matplotlib** é a biblioteca Python mais popular para **criação de gráficos** e **visualizações de dados** em geral.

Permite que você crie facilmente **gráficos**, **histogramas** e outras visualizações profissionais.

Ele possui suporte em todos os sistemas operacionais e também pode **exportar gráficos** para diversos formatos: **pdf, svg, jpeg, png, bmp, gif**, etc.





## O que é o Matplotlib?

No **Matplotlib**, o usuário tem total controle de **estilos de linhas, propriedades de fonte, atributos**, etc., através de uma interface **orientada a objetos** ou através de um **conjunto de funções** familiares aos usuários do MATLAB.

### **Dica:**

O objetivo final da visualização de dados, é **apresentar um resultado** (um insight). E você não precisa de um **gráfico poluído** para isso.

**Acredite, visualização de dados é quase uma arte!**

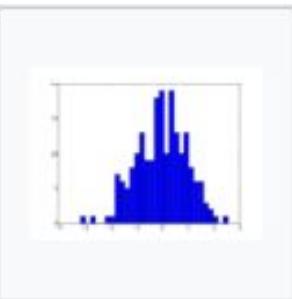




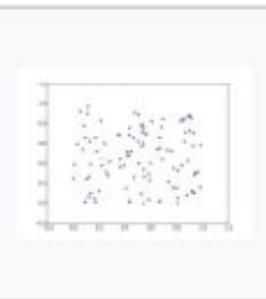
# Exemplos



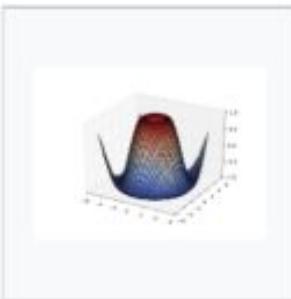
Line plot



Histogram



Scatter plot



3D plot

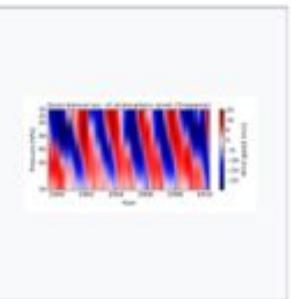
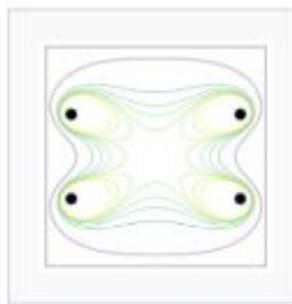
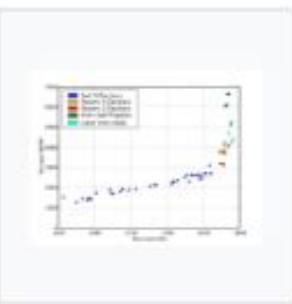


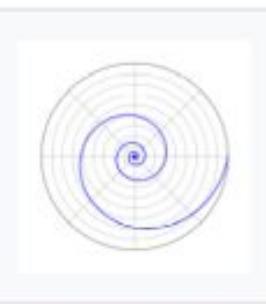
Image plot



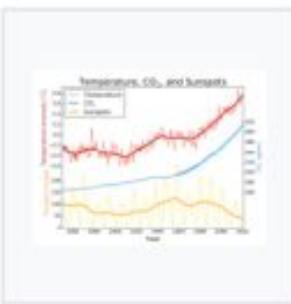
Contour plot



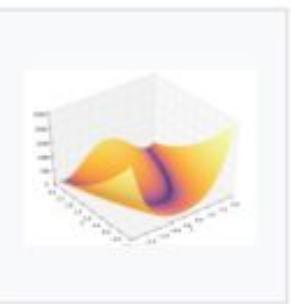
Scatter plot



Polar plot



Line plot



3-D plot

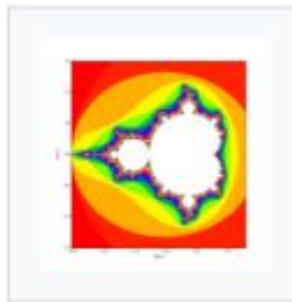


Image plot



# Análise de dados

---

Matplotlib, Pandas e Numpy juntos, constituem as principais ferramentas do que chamamos:

**PyData Stack**



# Análise de dados

---

O processo de análise de dados é um conjunto de etapas que tem como objetivo transformar dados brutos em informações úteis e relevantes para tomada de decisões.



## As principais etapas do processo de análise de dados são:

---

- **Definição do problema:** é importante ter clareza sobre o objetivo da análise de dados e sobre as questões que precisam ser respondidas.
- **Coleta de dados:** a coleta de dados pode ser feita de diversas formas, como por meio de pesquisas, questionários, dados de sistemas, entre outros.



## As principais etapas do processo de análise de dados são:

---

- **Limpeza de dados**: é comum que os dados coletados estejam incompletos, inconsistentes ou com erros. Nessa etapa, é preciso identificar e corrigir esses problemas para garantir a qualidade dos dados.
- **Análise exploratória**: nessa etapa, os dados são visualizados e resumidos por meio de gráficos, tabelas e estatísticas descritivas. O objetivo é identificar padrões e tendências nos dados, bem como possíveis relações entre as variáveis.



## As principais etapas do processo de análise de dados são:

---

- **Modelagem**: é a etapa em que são desenvolvidos modelos estatísticos ou algoritmos de aprendizado de máquina para prever ou explicar um fenômeno. É importante escolher o modelo mais adequado para o problema em questão.
- **Interpretação dos resultados**: os resultados da análise devem ser interpretados para que possam ser comunicados de forma clara e compreensível aos tomadores de decisão. É importante destacar as principais conclusões e recomendações para a empresa ou organização.



## As principais etapas do processo de análise de dados são:

---

- **Comunicação dos resultados**: é a etapa final em que os resultados são apresentados aos stakeholders envolvidos no processo. É importante escolher uma forma adequada de apresentação dos resultados, como relatórios, apresentações ou dashboards



# Se concentrem no processo...

The image displays a comprehensive grid of logos for various open-source databases and data processing platforms, categorized into several groups:

- Data Analysis & Platforms**: Includes HPCC Systems, Storm, Dremel, Apache Drill, IKANOW, Hadoop, MapReduce, Apache SAMOA, Hortonworks, and others.
- Databases / Data warehousing**: Includes bigdata, INFOBRIGHT, Cassandra, 4Store, Infinispan, riak, H2, Oracle Berkeley DB, Firebird, MariaDB, Drizzle, SQLite, RethinkDB, HyperSQL, and monetdb.
- In-Memory Computing**: Includes GlobalsDB, hazelcast, TERRACOTTA, NMemory, and GORA.
- ERP BI Solutions**: Includes talend, Jaspersoft, Palo, jedox, pentaho, openbi.org, KEEL, rapidminer, mahout, togaware, SPMF, and BIRT.
- Business Intelligence**: Includes openbi.org, open intelligence, rapidminer, mahout, WEKA, KEEL, togaware, SPMF, and BIRT.
- Data Mining**: Includes orange, KNIME, rapidminer, mahout, WEKA, KEEL, togaware, SPMF, and BIRT.
- Big Data search**: Includes lucene, Apache Solr, elasticsearch, and northgate.
- Multivalue database**: Includes Rocket, U2, REVELATION, QM, iBASE INTERNATIONAL, and ScarletDME.
- Programming**: Includes R and julia.
- Data aggregation**: Includes ogoop.
- KeyValue**: Includes AEROSPIKE, leveldb, redis, Chordless, Tokyo Cabinet, memcached, SCALIEN, Project Voldemort, hamsterdb, RAPTORDB, FairCom, HyperDex, IQLECT, openLDAP, ioremap.net, and Scalaris.
- Document Store**: Includes mongoDB, COUCHBASE, Tokutek, JasDB, SchemafreeDB, RaptorDB, EJDB, pdjon, CouchDB, and Raven DB.
- Graph databases**: Includes Gephi, Gremlin, Franz Inc, InfiniteGraph, HYPERGRAPHDB, Neo4j, GraphBuilder, FlockDB, GraphBase, BrightstarDB, and Sparksee.
- Operational**: Includes VOLTDB, Apache Kafka, ThinkUp, and Corona.
- Object databases**: Includes db4objects, ZOPE, mobject, Magma, PERSEVERE, EyeDB, Picolisp, NEOPPOD, siaqodb, and STARMOUNT.
- Multimodel**: Includes ArangoDB, existdb, BASE, Qizx, alchemydatabase, sedna, LIQUIBASE, and Galaxy.
- XML Databases**: Includes ArangoDB, existdb, BASE, Qizx, alchemydatabase, sedna, LIQUIBASE, and Galaxy.
- Multidimensional**: Includes FIS, SciDB, rasdaman, and raster data manager.
- Grid Solutions**: Includes GIGASPACE.

cita



# Abordagem para criação de gráficos

---

- Sempre priorizem códigos com boa apresentação estética e comprehensíveis, priorize a criação de variáveis e/ou data frames para os gráficos.

A black and white portrait of Blaise Pascal is on the right. To his left is a white rectangular box containing a quote in Portuguese. At the bottom, there is a dark footer bar with the name 'Blaise Pascal' and the word 'PENSADOR' next to a small icon.

Se eu tivesse mais tempo, teria escrito uma carta mais curta.

Blaise Pascal

PENSADOR



# Abordagem para criação de gráficos

---

- Lembre-se: **tenha um roteiro!!!**

A maioria das pessoas que têm uma curva de aprendizado mais lenta estão sem foco, copiando códigos enormes que poderiam ser mais simples. Comece aprendendo o estritamente necessário, e vá aprofundando conforme **necessidade, isso é estratégia!**



# Criação de gráficos

---

- Entendendo a construção básica:

```
dataframe.plot()  
plt.title('SEU TÍTULO')  
plt.xlabel('NOME DO EIXO X')  
plt.ylabel('NOME DO EIXO Y')  
plt.show()
```



# Links

---

- Cores disponíveis na Matplotlib:

[https://matplotlib.org/stable/gallery/color/named\\_colors.html](https://matplotlib.org/stable/gallery/color/named_colors.html)

- Mais sobre loc e iloc:

[https://github.com/CamilaSDantas/loc\\_iloc/blob/master/loc\\_iloc\\_utilizacao.ipynb](https://github.com/CamilaSDantas/loc_iloc/blob/master/loc_iloc_utilizacao.ipynb)