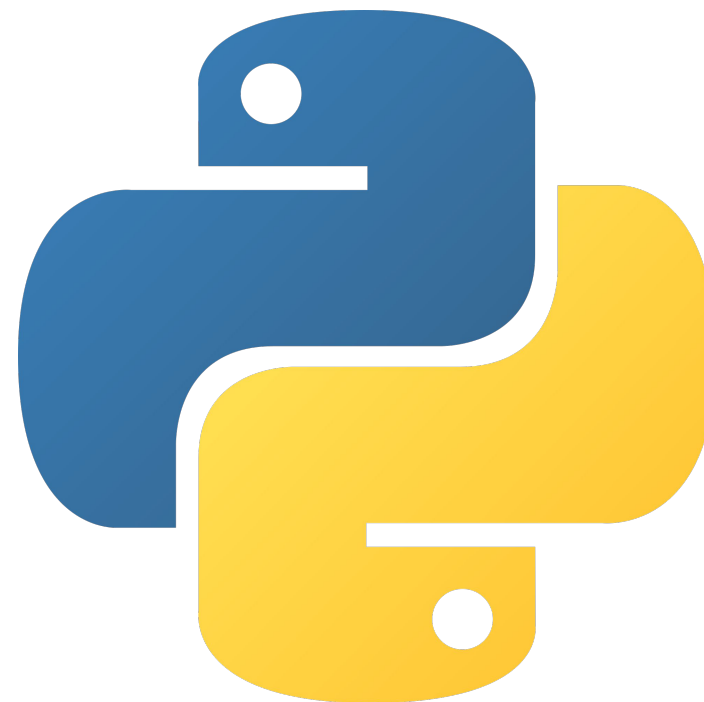


# Fuctura

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

---

Prof. Jéssica Andrade



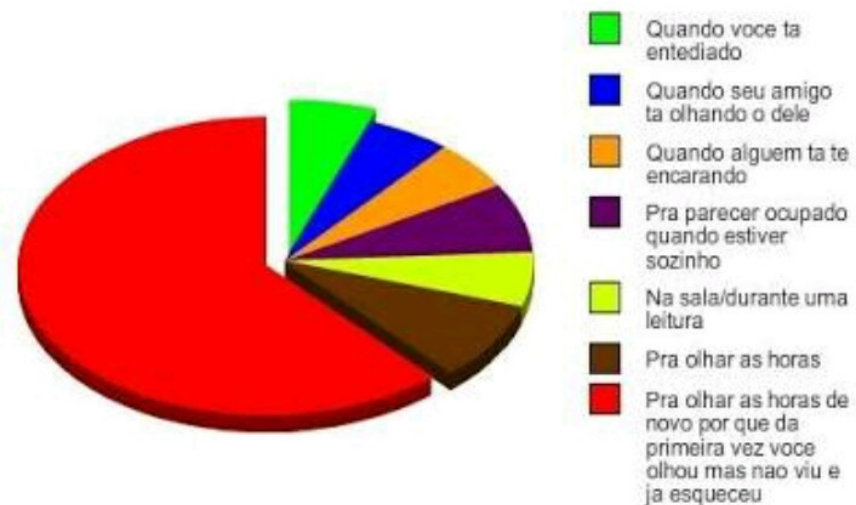


**matplotlib**

# Gráficos



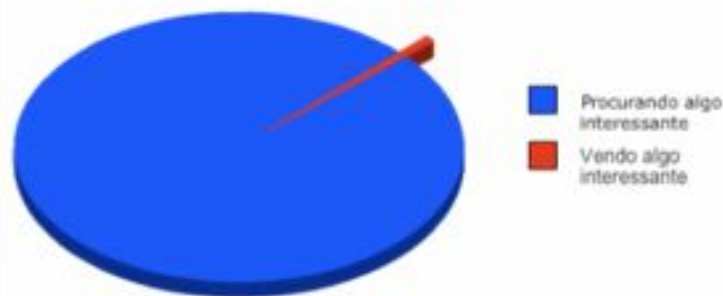
## Quando voce olha para o seu celular



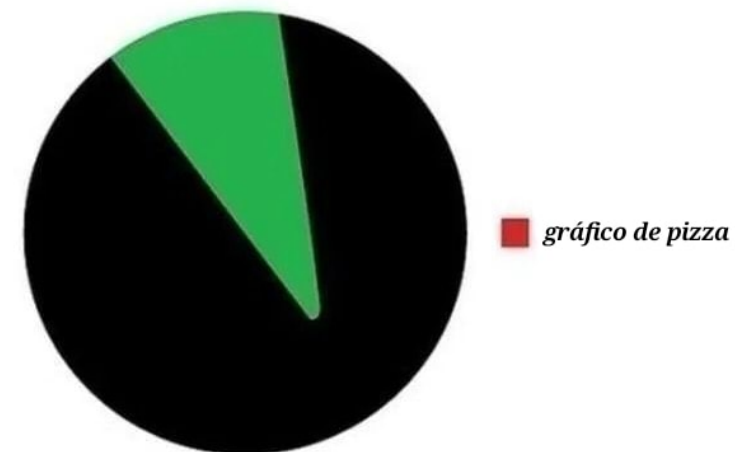
## O QUE VOCÊ FAZ QUANDO SE SENTE MAL



## Tempo gasto na internet



## 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.

*cit~*

*cit~*



## 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:**

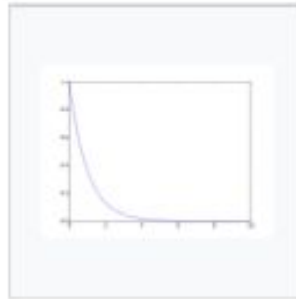
O objetivo final da visualização de dados, é **apresentar um resultado** (um insight). E você não precisa de um **gráfico poluído** para isso.

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

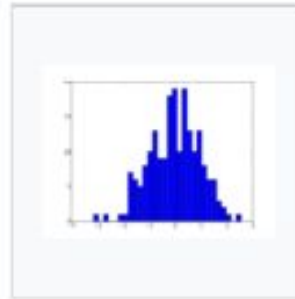
*cit~*

*cit~*

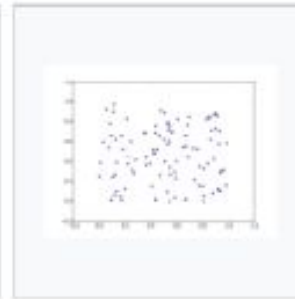
# Exemplos



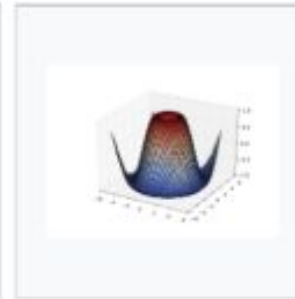
Line plot



Histogram



Scatter plot



3D plot

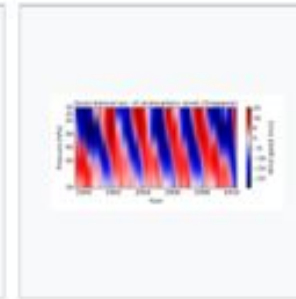
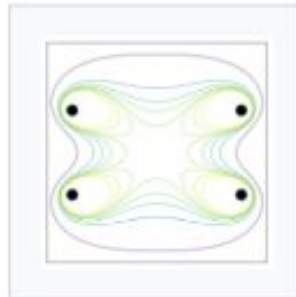
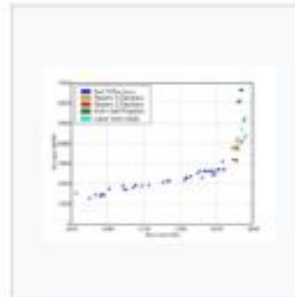


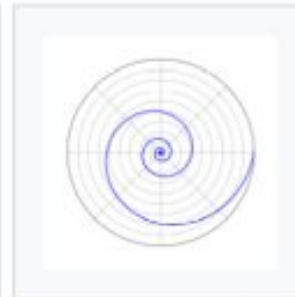
Image plot



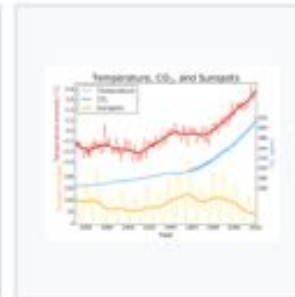
Contour plot



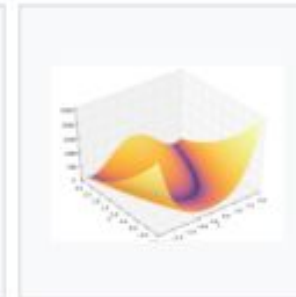
Scatter plot



Polar plot



Line plot



3-D plot

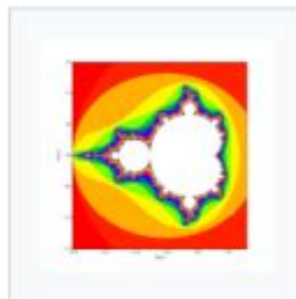


Image plot



# 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



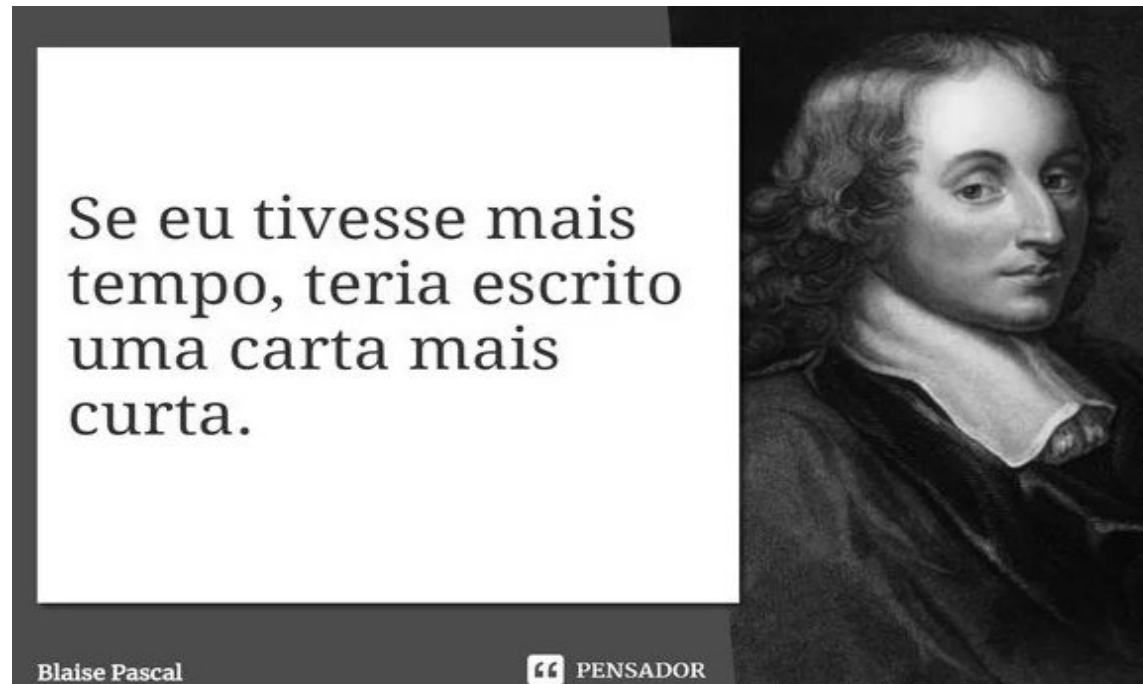
# Se concentrem no processo...

<b>Data Analysis &amp; Platforms</b> HPCC Systems Storm Dremel Hadoop Hive Spark SAMOA APACHE DRILL Hortonworks IKANOW   BRILLIANT DECISIONS	<b>Databases / Data warehousing</b> bigdata INFOBRIGHT Cassandra 4store H2 InfiniDB riak Infinispan HYPERTABLE MariaDB Drizzle SQLite RethinkDB Firebird ORACLE BERKELEY DB HyperSQL monetdb GlobalsDB	<b>In-Memory Computing</b> GridGain hazelcast TERRACOTTA NMemory GORA
<b>ERP BI Solutions</b> talend spagobi Jaspersoft Palo jedox pentaho BIRT	<b>Business Intelligence</b> openi.org open intelligence Data Mining orange KNIME rapidminer mahout WEKA KEEL togaware SPMF	<b>Big Data search</b> Lucene Apache Solr elasticsearch.
<b>Key/Value</b> AEROSPIKE leveldb redis Chordless Tokyo Cabinet SCALIEN Project Voldemort hamsterdb RAPTORDB FairCom STS DB HyperDex OpenLDAP iorem.net Scalaris	<b>Document Store</b> mongoDB Couchbase Raven DB CLUSTERPOINT Tokutek RaptorDB EJDB djondb JasDB SchemafreeDB sisodb Apache CouchDB relax	<b>Graph databases</b> Gephi Gremlin GraphBuilder FRANZ INC Sparksee InfiniteGraph INFO GRID HYPERGROPH-DB Neo4j FlockDB GraphBase BrightstarDB
<b>Object databases</b> db4objects Zope mobject Magma Picolisp NEOPOD siaoqodb RAMER D PERSEVERE EyeDB STARCOUNTIE Sterling NDatabase C# Lightweight Object Database	<b>Operational</b> VOLTDB Social Apache Kafka ThinkUp Corona	<b>Multidimensional</b> FIS SciDB rasdaman raster data manager
<b>Multimodel</b> ArangoDB alchemydatabase	<b>XML Databases</b> eXistdb BASE Qizx sedna LIQUIBASE	<b>Grid Solutions</b> GIGASPACE Galaxy

# Abordagem para criação de gráficos

---

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







# 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)