

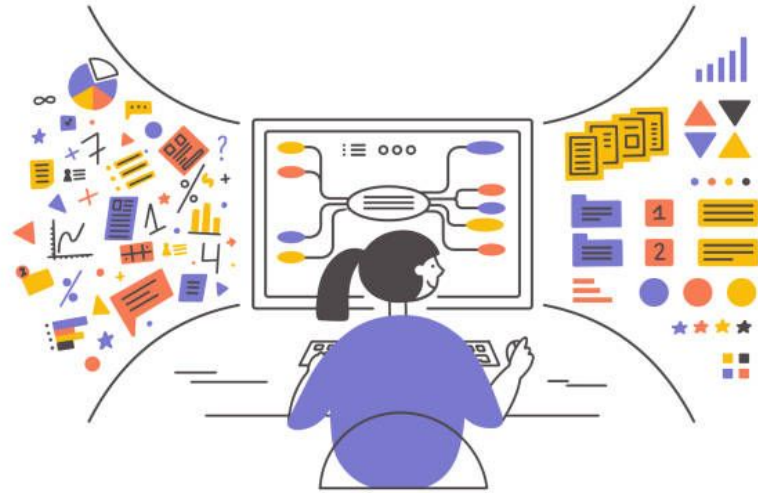


# Análise Exploratória de Dados (AED)



# Objetivos

- Resumir as características de um dataset com métodos visuais;
- Identificação de padrões;
- Detecção de valores discrepantes;
- Suposições e Hipóteses;



# Estatística X AED

## ESTATÍSTICA:

- Coleta e interpretação de dados;
- Utiliza de modelos simples e verificação na consistência de dados com a predição desses modelos;
- Foca em quantificar a incerteza;
- Tem por maior objetivo identificar causas de algo;

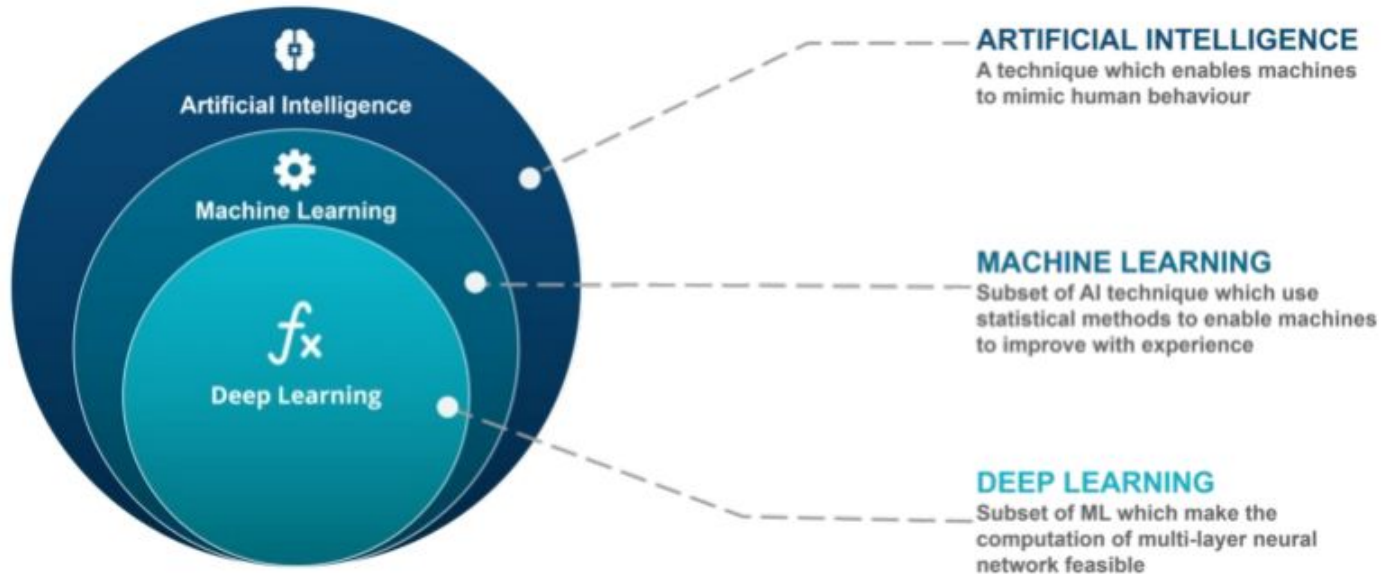


## AED:

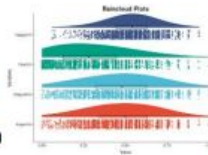
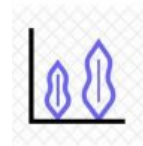
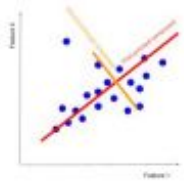
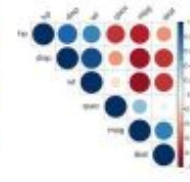
- Campo multidisciplinar;
- Uso de métodos, processo e sistemas científicos na extração de informações dos dados;
- Foca em encontrar o melhor modelo que faça a predição dos dados;
- Utiliza de conceitos da estatística;



# Relação com IA e Machine Learning

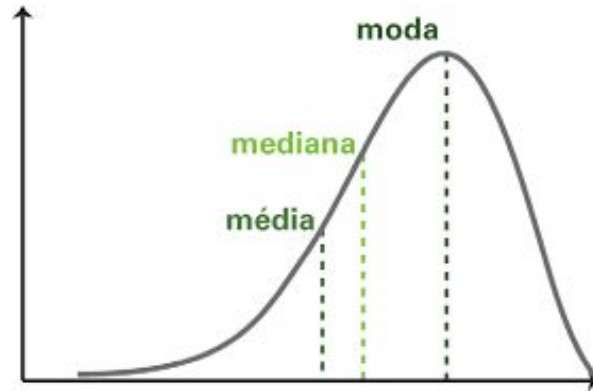


# Como Funciona?



# Conceitos Utilizados

- Tipos de variáveis (quantitativas ou qualitativas)
- Médias simples e ponderada;
- Moda e mediana;
- Variância e desvio padrão;
- Quartis (0.25, 0.5, 0.75);
- Intervalos de Confiança;



# Principais Ferramentas



- Numpy;
- Matplotlib;
- Pandas;
- Seaborn;



- RODBC;
- DPLYR;
- CARET;
- GGLOT2;

# Exemplo em Python

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

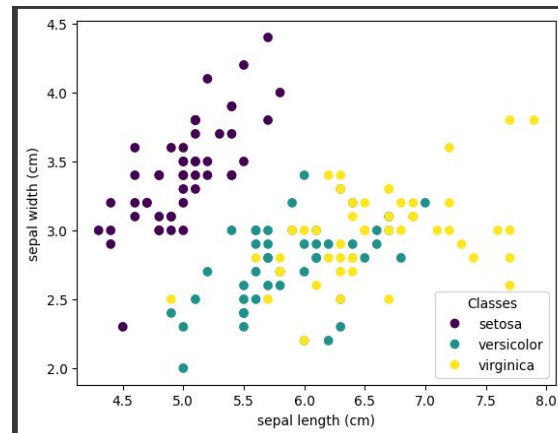
Imagem do Data Frame criado pelo Pandas.

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

Principais características do Data Frame, para cada atributo

```
RangeIndex: 150 entries, 0 to 149
Data columns (total 4 columns):
#   Column              Non-Null Count  Dtype
---  -
0   sepal length (cm)    150 non-null    float64
1   sepal width (cm)     150 non-null    float64
2   petal length (cm)    150 non-null    float64
3   petal width (cm)     150 non-null    float64
dtypes: float64(4)
```

Características de cada atributo.



Visualização 2D da relação de tamanho entre largura e altura das sépalas.