Data toolkit

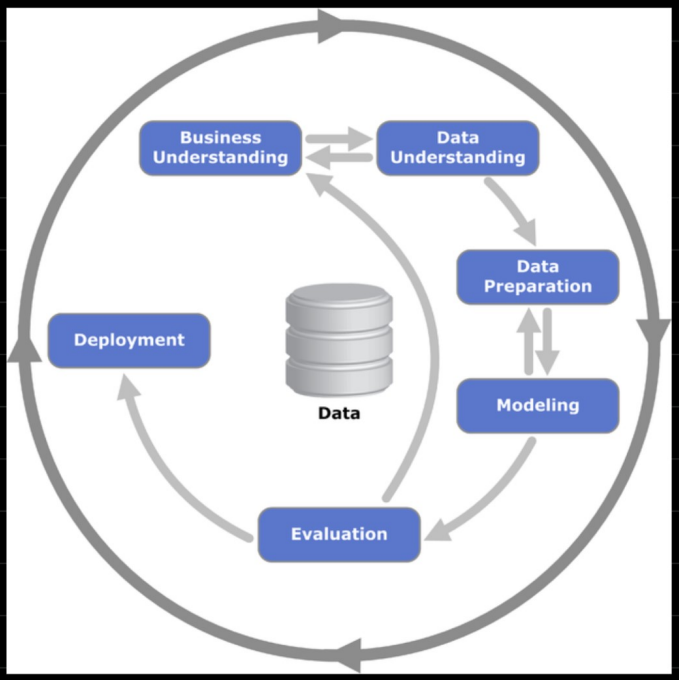
- → Numpy
- → Pandas
- → Matplotlib
- → Seaborn
- → Plotly
- → Bokeh.

* Data science process

* CRISP-DM Framework ⇒ Cross-industry process for data mining.

⇓

getting insights from data.



① Business Understanding.

Problem statement :-

* To Predict price of House.
  - ① Area of house
  - ② No of rooms.

② Data Understanding ✓

* How area of house decides the price of house
  - Area of House ↑ Price of Houses ↑
  - No. of rooms ↑ Price ↑
  - No of rooms ↑ Area ↑

→ Exploratory data Analysis (EDA)

③ Data preparation ✓

④ Modelling

⑤ Evaluation

⑥ Deployment

* library → A library is a collection of pre-written code, that is used to perform common task.

basicCalculations { addition, subtraction, multiplication } ↓ modules

{ Linear Algebra } (LA) ↓ module

{ Calculus } (C) ↓ module

→ All of these modules kept together is called as package

Analogy

→ Multiple package together is called as library

Library (multiple books) ⇒ Packages

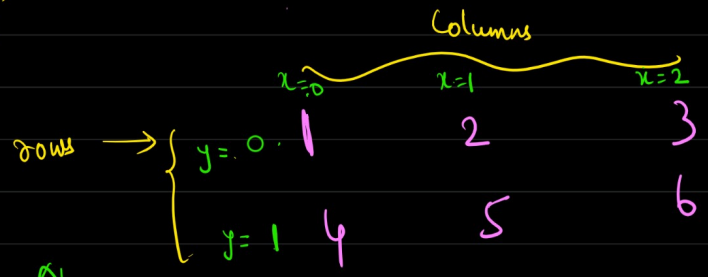
* Library Contains Python modules and packages.

Lessons/chapters ⇒ modules.

Paragraphs ⇒ functions.

* arr = [1, 2, 3, 4, 5] = only rows

* arr = [[1, 2, 3],
[4, 5, 6]]



np.fromfunction(lambda i, j : i == j (3,3))

i = row
j = colm

i ↓ / j →	0	1	2
0	0 == 0 True	False	False
1	False	1 == 1 True	False
2	False	False	2 == 2 True

np.from(lambda i,j : i*j ,(3,3)

i \ j	0	1	2
0	0x0 0	0x1 0	0x2 0
1	1x0 0	1x1 1	1x2 2
2	2x0 0	2x1 2	2x2 4