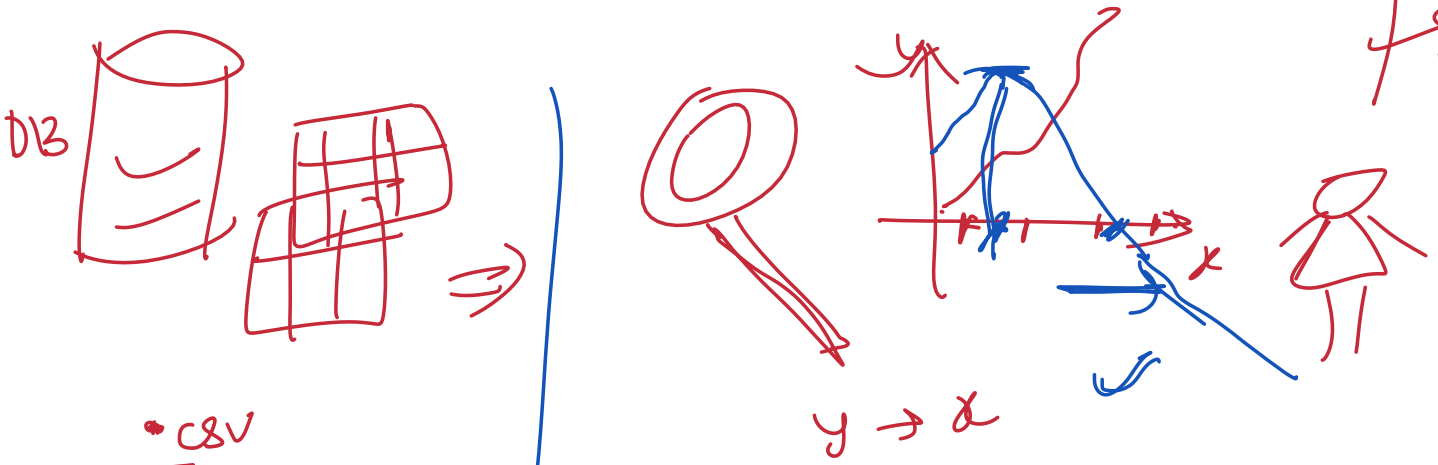
28-Feb-25

Friday, February 28, 2025 6:22 PM

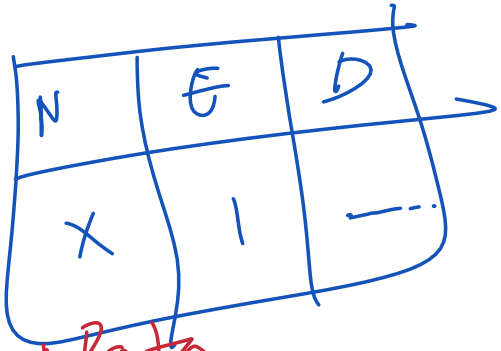
Data Analysis with python?

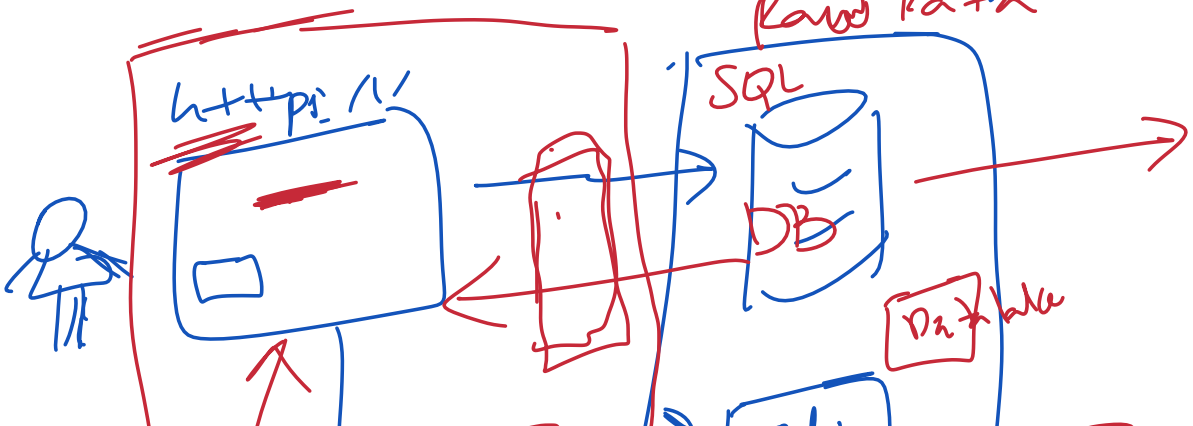


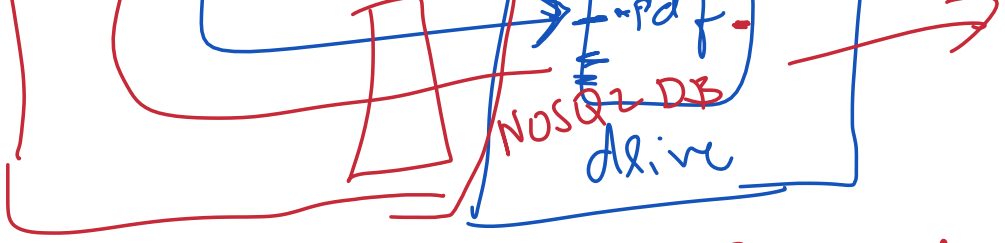
Basic python
SQL
~~SVC~~

DB, files, Data lakes..

Raw Data







Messy

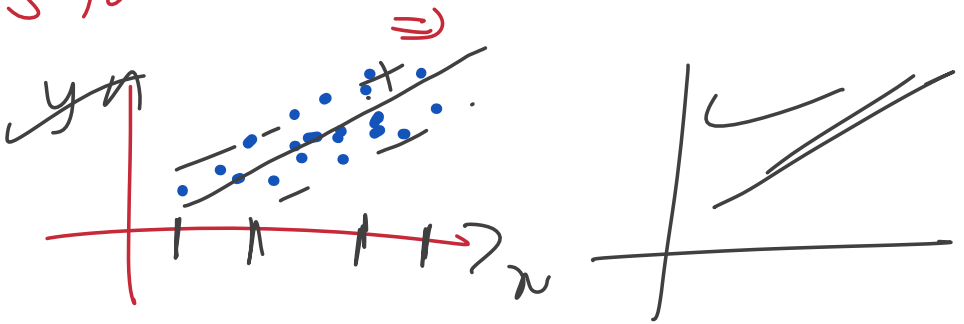
✓ Null Values

Data Engineers

⇒ ✓ Duplicates

Data Store

⇒

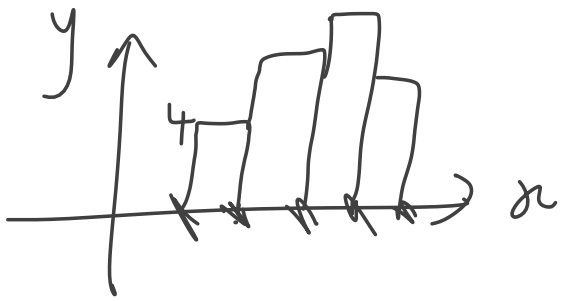


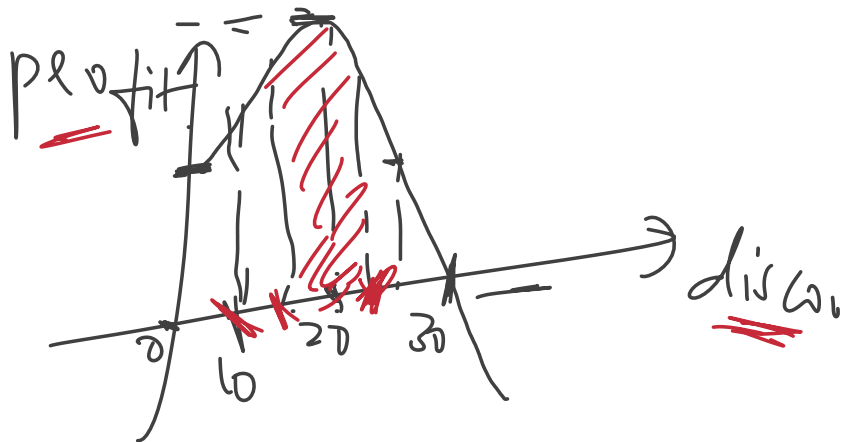
output input

$y \rightarrow x$??

inputs

$y \rightarrow x_1, x_2, \ldots$??





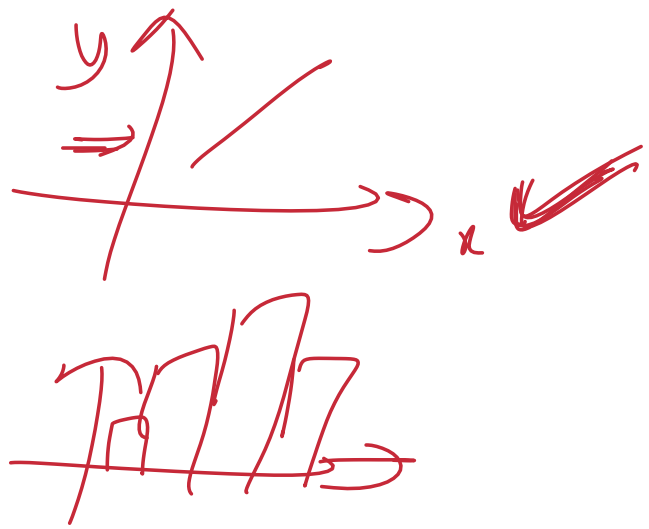
Business Requirement

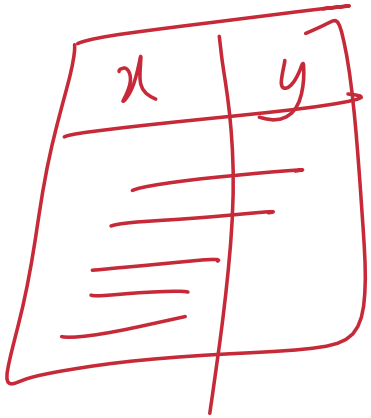
Data Source ⟹ Data Collection (Raw Data) ⟹ Data Cleaning ⟹ Data Visualization ⟹

** ⟹ (python) ⇒ null values

⇒ tableau ⇒ Duplicates

⇒ PowerBI





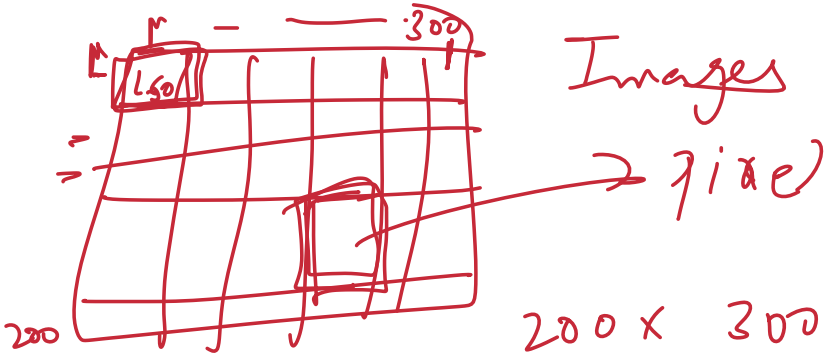
1000

→ PyPI

Visualization → matplotlib, seaborn

Anaconda
or Repo

Data Analysis → Pandas [numpy]



Images → size

200 x 300

array = [[1 2] [3 4]] → (2 x 2)

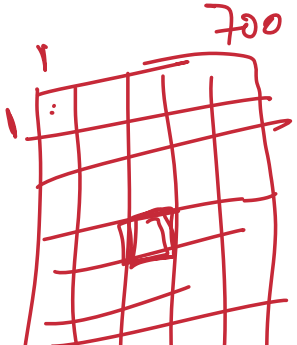
0 - 255 → B/W

R - 0 - 255 - 255

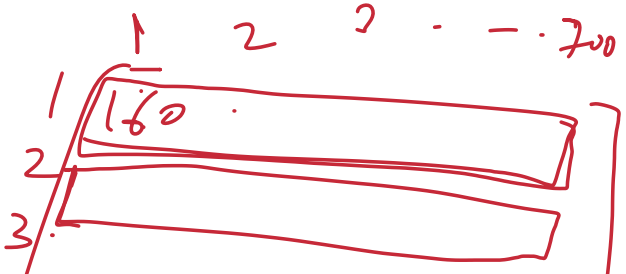
G - 0 - 255 - 0

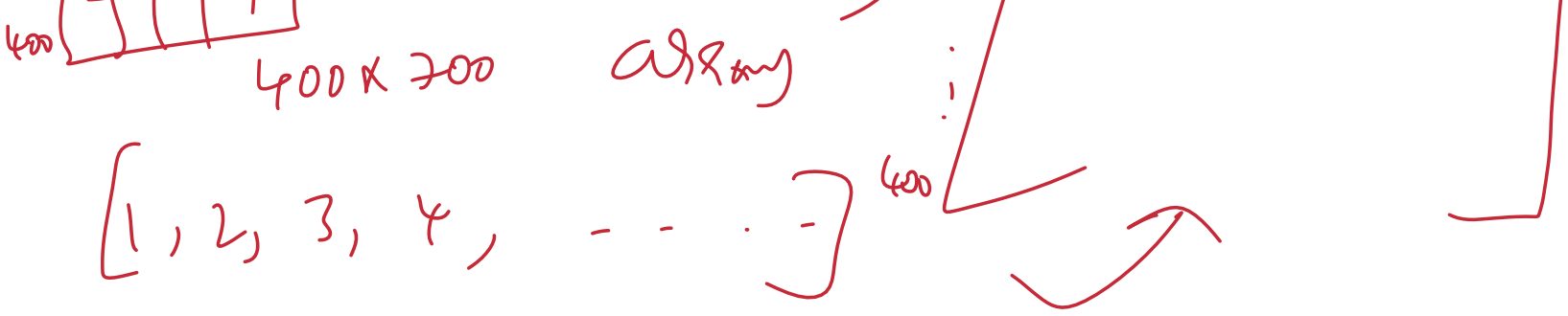
B - 0 - 255 - 0

255	0
255	0
255	0

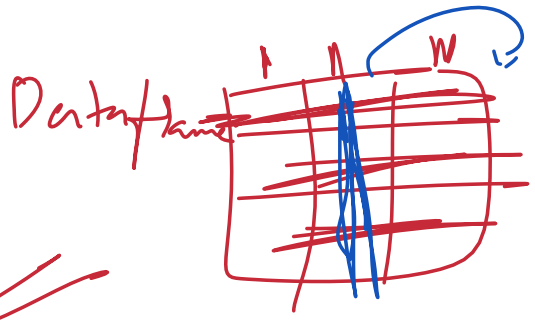


→ B/W →





a = 1
b = 2
c = 3



Basic python →

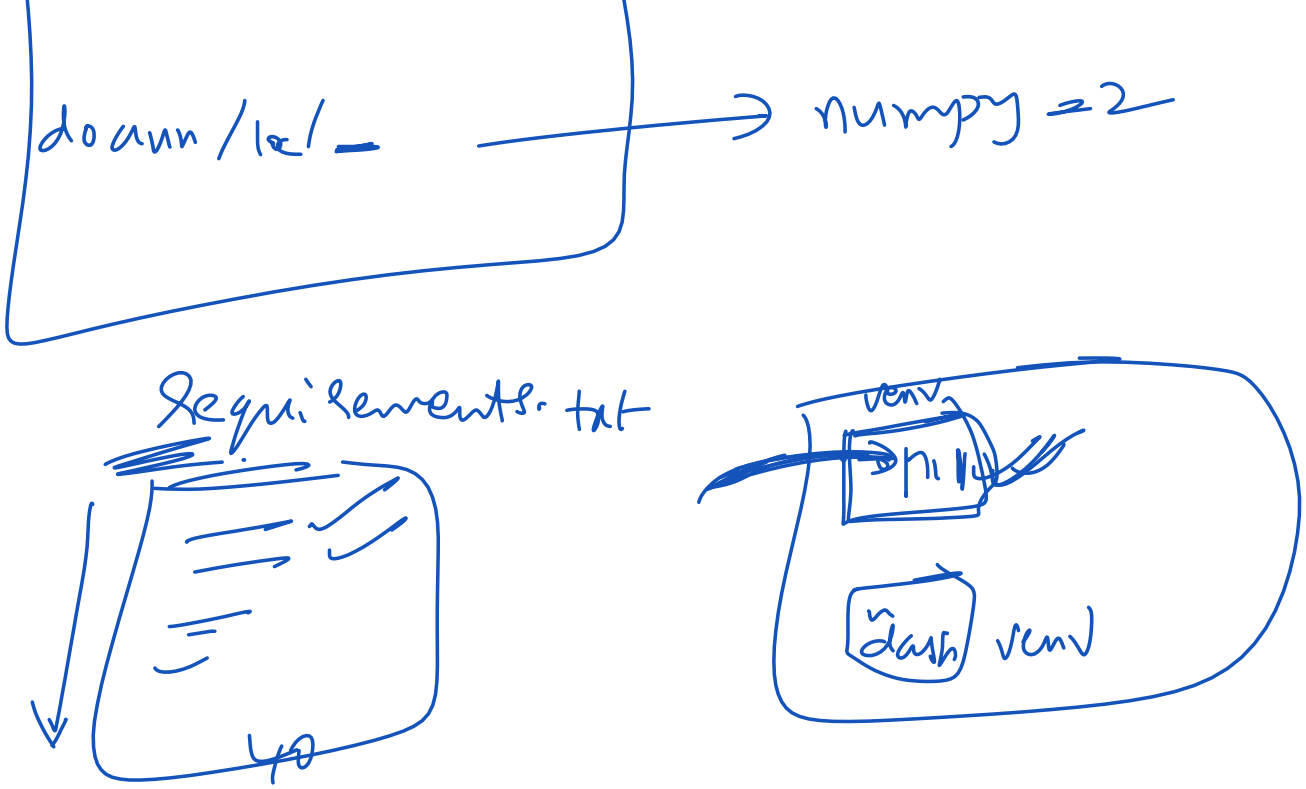
Python packages

Data Analysis → Numpy *
Pandas ***

Matplotlib
seaborn

documents/09_dv/DA

numpy22


docum/ic/_
numpy=2
requirements.txt
venv
pip
dash
venv