Data Science using Python

Industry Expert : Arun MV. (informatica)

Instructor : Suraj Hasija (Mastercard, Scaler)

.Challenges

→ Industry gap , workarounds , real-time projects.
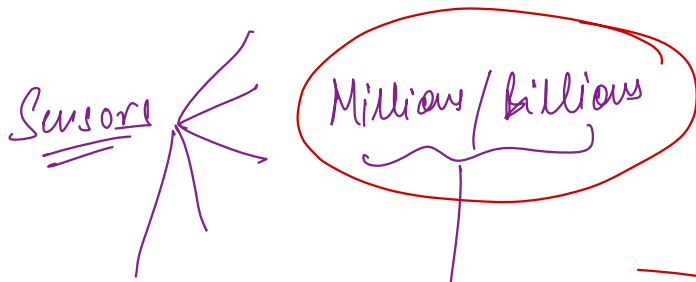
How to Overcome?

Kaggle , groups —.

Agenda

1. What is Data Science?
2. Technologies in Data Science
3. Python — Basics / getting started
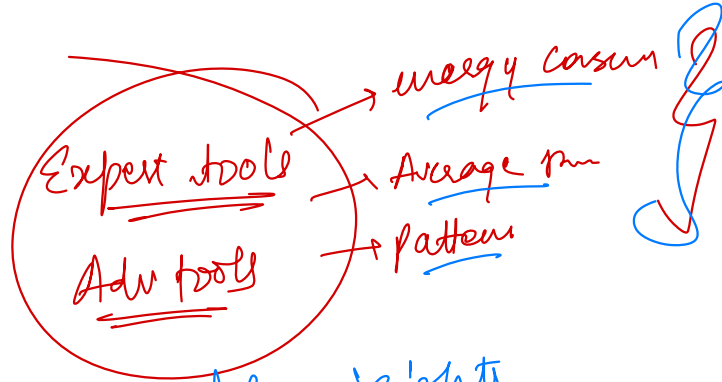4. Analysis using Python for Data Science.

Case study

↳ Takeaway / living

2 hr

Sensors

Million/billions

Small X

Expert tools
Adv tools

→ energy consum
→ Average m
→ Pattern

Turning raw data into beautiful insights

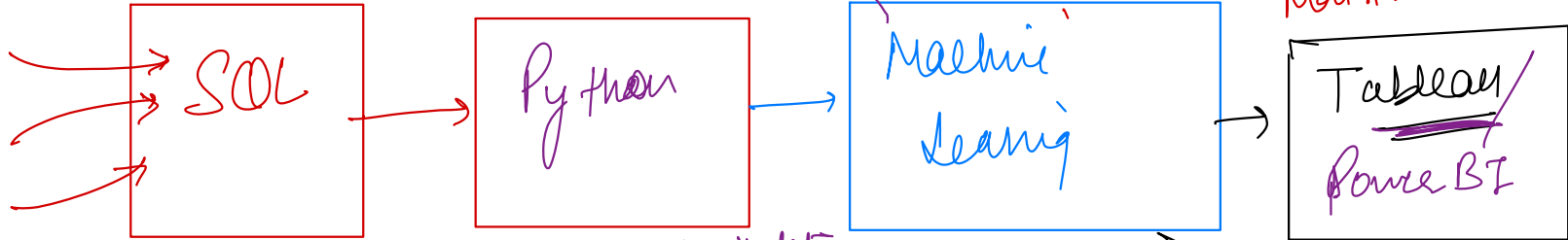Data Analytics + Machine Learning (giving power to a machine to predict)

Data Science

Technologies → To become a good data Science.

3-4 easily

MLOps , CV, NLP

Statistics,
Mathematics — Differentiation
Matrix

| SQL | → | Python | → | Machine Learning | → | Tableau / Power BI |

AWS, GCP

extraction & basic aggregation

Data Manipulation, Sanitisation (Cleaning)

Challenge

① Missing data / values

↳ fix it
    ↳ (Python)

Data Cleaning using Python

① Drop the entire

② fill it with Zero

Classes	Percentages
1	70
2	60
3	65
4	70
5	Mean ← — one of the ways.
6	67
7	69
8	75
9	68

③ fill it with mean

(2) Some outlier / extremity / exceptional

Class A

Students	Age	
A	12	→
B	13	→
C	12	→
D	11	→
E	12	→
F	13	
G	75	←
H	12	
I	13	

Avg

Mathematical

Basic way

Std dev

overall Avg → 13

After cleaning → Analysis

↳ finding patterns in the data
& giving insights along
with recommendations

Eg

(2 G) → Data Analyst

(Task) → give reasons, for selling of ACs in Delhi.
factors.

↑ Temperature   → Type of City   → Purchasing Power
↑ Per capita income   → Humidity   → high selling of
                                           Ice-cream.

Correlation | Causation

high selling of ice-cream | Temperature

	ACs	Ice-cream
December	↓	✓ ↗
Temp ↓		

Temp ↑   AC ↗   (Ice - cream) ↗

↓

Practical

Steps 1: Login with gmail on google colab

2) give worksheet a name

2) Try runing 1-2 cells.

Case Study : (Store) → Retail
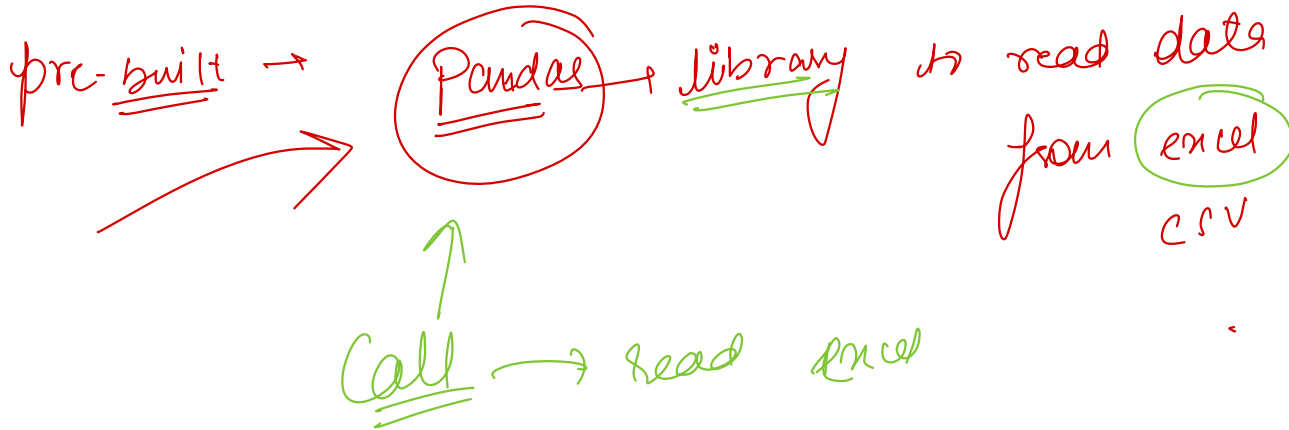
{ Cust id   Order id   Date   City State   Product, Sales
                                                Profit ....

① Factors that affects the Sales of e-commerce
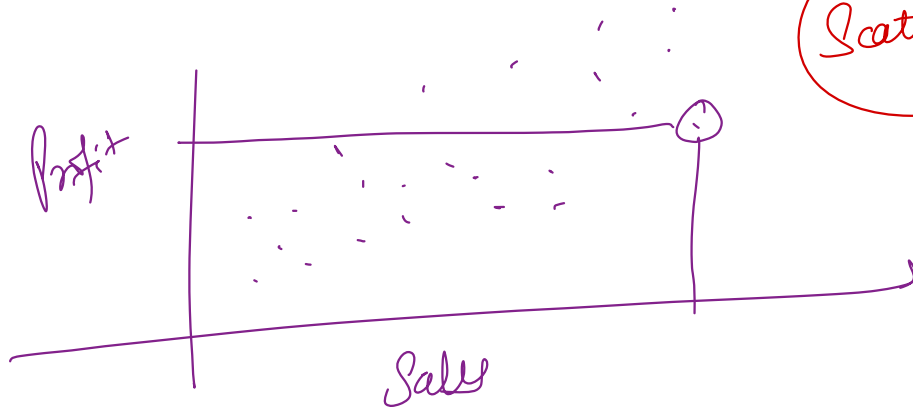   Store.

To read the (file)

x = 10
y = 20

z = x + y
print(z) → (30) }

pre-built → (Pandas) → library to read data
from (excel)
CSV

Call → read excel

1) State wise description of sales

2) Sub-category wise description of profit.

Seaborn for charts