Welcome everyone

⇒ Data-Science

⇒ Machine Learning } ⇒ Case Study

Case Studies ──┬→ EDA
               ├→ Classification
               ├→ Regression
               └→ Unsupervised

---
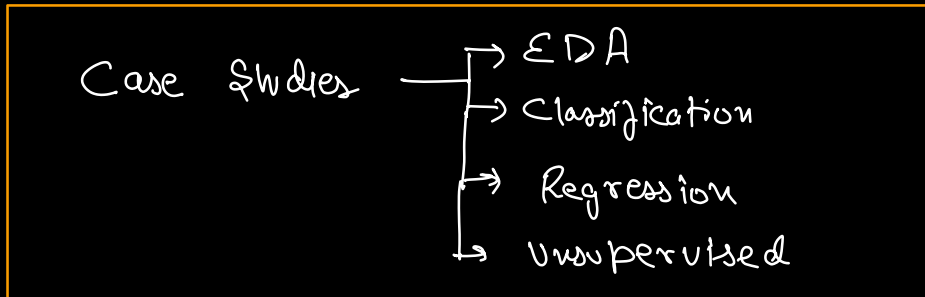
11 am – 1 pm ⇒ all codes
                notes ⇒ ☆
                reference material

|Classification Case Study|

|exploring|
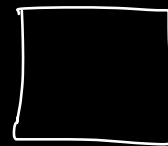
remembering

2 ways to take this case study Community Class

✶ 1. Enroll on Ineuron Platform ✶ ┐
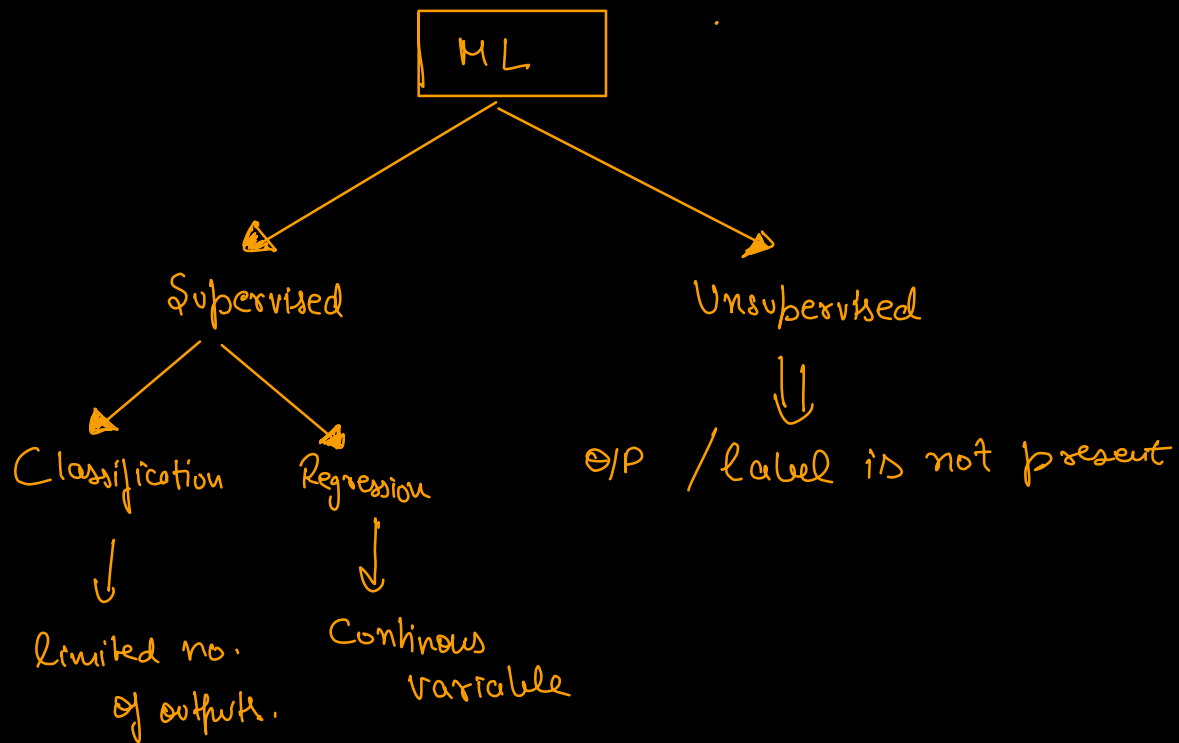   │
2. Via Youtube   │
      ⇓   ├→ Quizzes
    live   └→ Assignment
    classes
                              ┌──────────────┐
  11am – 1pm   │ Certificate │
                              └──────────────┘

                                 recorded

Classification Case Study

```
                    ┌──────────┐
                    │   ML     │
                    └──────────┘
                   ╱            ╲
                  ╱              ╲
            Supervised        Unsupervised
           ╱          ╲             ⇓
          ╱            ╲       O/P / label is not present
    Classification   Regression
         ↓              ↓
    limited no.     Continous
    of outputs.     varriable
```

Classification Problem

1. Binary Classification

Y	M	P	
N	F	NP	} 2 classes

2. Multiclass Classification

M	S	One	
F	V	Two	} multiple
O	Ve	Three	classes

$$\Downarrow$$

1 vs other

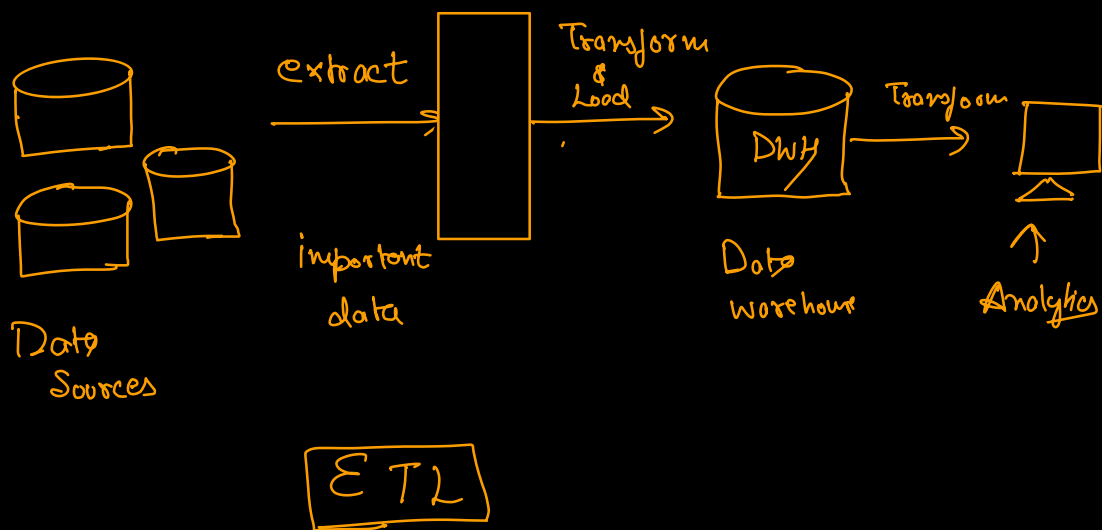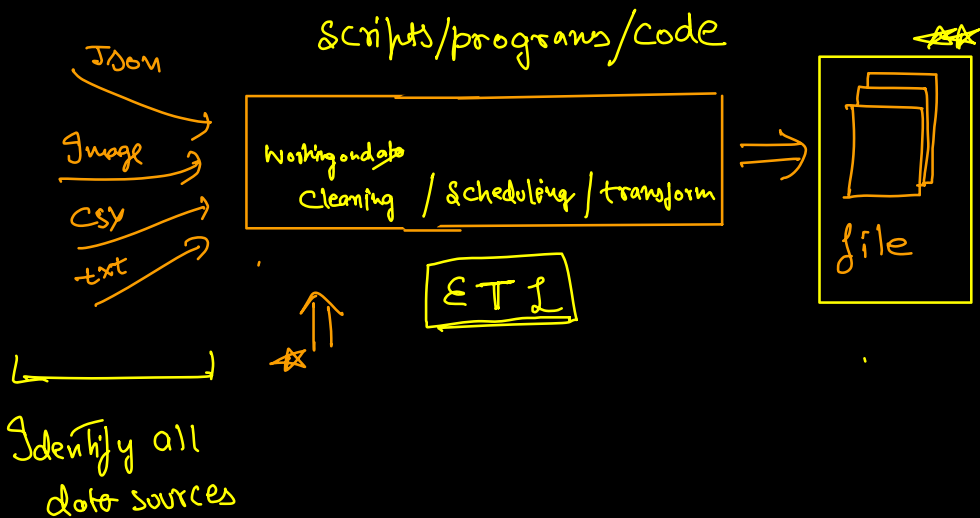|	
M	1
M	1
F	0
H	0

Steps in Supervised ML algorithm

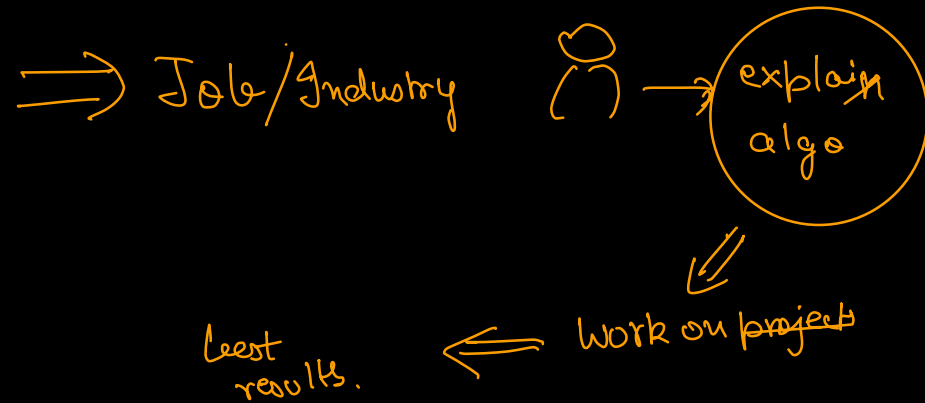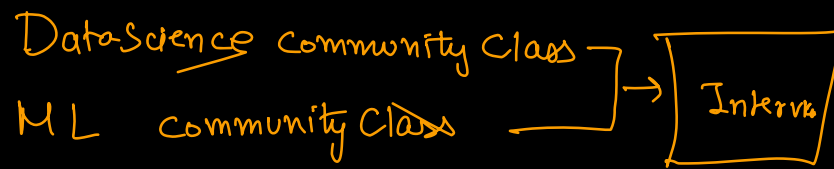0. **Understanding Business Problem**

1. **Obtaining/Collecting your data**

CSV
Json
Avro
Images
txt

→ finite
→ infinite

internal ←
external ←

2. Collate your all data at a single place & in a manner that it can be worked upon

Scripts/programs/code

JSON
Image
CSV
txt

Working on data
Cleaning / scheduling / transform

ETL

file

Identify all data sources

Data Sources

extract

important data

Transform & Load

DWH

Transform

Data warehouse

Analytics

ETL

DataScience community Class ⌐→ Interv
ML community Class ⌐

⟹ Job/Industry 🧑→ (explain algo)

least results. ⟸ Work on project

3. Cleaning & EDA on your data

→ clean the data

→ Statistical analysis

→ [Missing value, outliers , norm/ stand]

→ Feature engineering
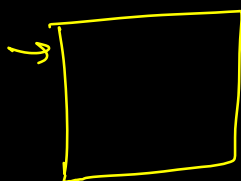
→ Univariate / Bivariate / Multivariate

X   y

$x_i \Rightarrow$ independent

$y \Rightarrow$ dependent

petal
length
↓↑
petal
width.

Species
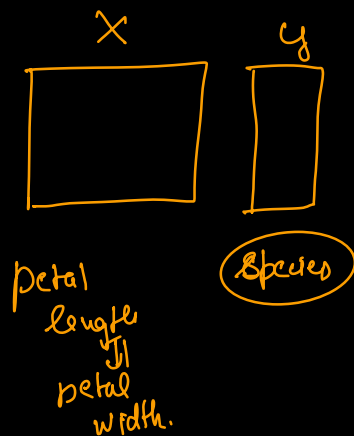
$x_r, x_j \Rightarrow$ highly dep.

Bivariate

** Don't force your observation/experience/knowledge on the data while doing EDA

→ single $\Rightarrow 3/y$
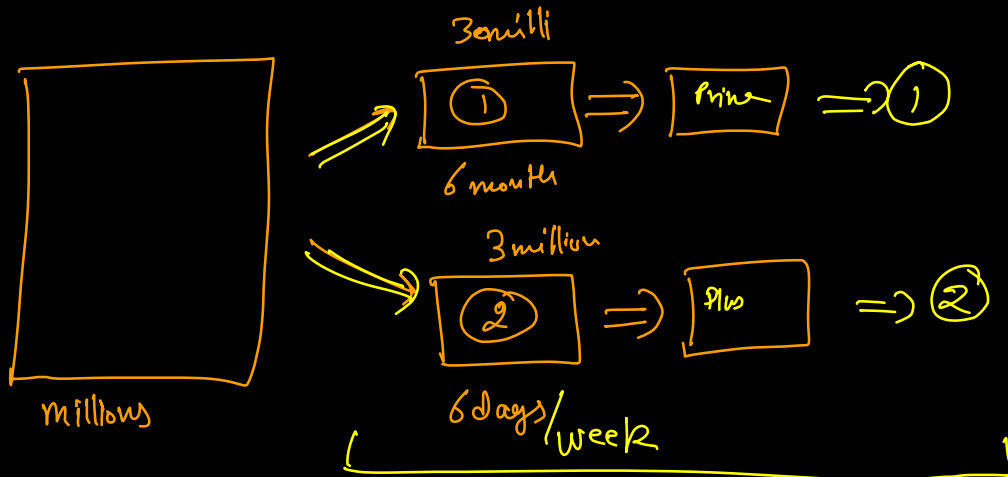
single $\Rightarrow 5$

verif

Bussiness knowledge   Domain knowledge

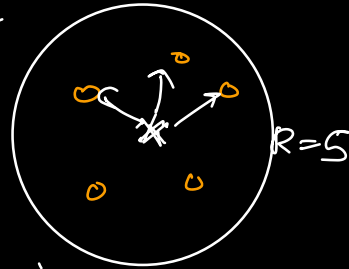Swiggy / Zomato   Amazon / flipkort

Super   Gold   Prime / Plus

3 milli

```
┌──────────┐        ┌─────────┐      ┌──────┐
│          │   ──→  │   ①     │ ⟹   │ Prime│ ⟹ ①
│          │        └─────────┘      └──────┘
│          │          6 month
│          │
│          │         3 million
│          │   ──→  ┌─────────┐      ┌──────┐
│          │        │   ②     │ ⟹   │ Plus │ ⟹ ②
└──────────┘        └─────────┘      └──────┘
  millions              6 days/week
```

Analytics
on our ⟹ 2nd set
date   of customers.

Missing Value analysis

1. Mean / Median $\Rightarrow$ numerical
2. Mode $\Rightarrow$ categorical

KNN $\Rightarrow$

R = 5

mean / median of values

mode of category