Variables

- Quantitetive
  - Descrite qua. vari. (0-9)
  - continuos qua. vari. (Decimal)
- qualitetive
  - Nominal qual. vari (yes/no) (T/F) (P/F)
  - ordinal queli. vari (industry - intern, junior, senior, Assd. mang. manaj.

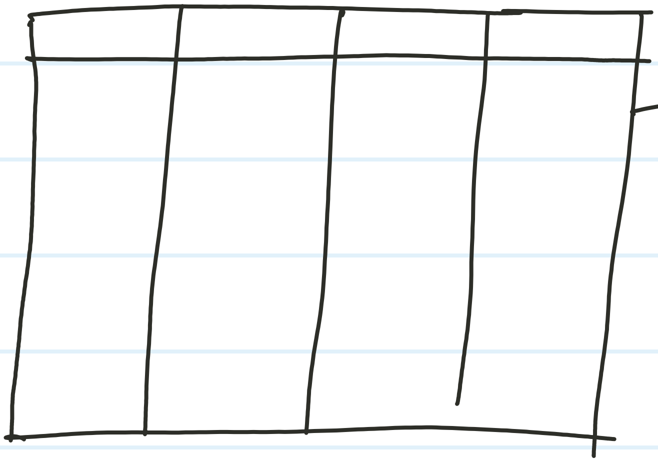
Salary	Age	No. of Account	Designation	employee Y/N
↑ C.q.v.	↑ C.q.v.	↑ D.q.v.	↑ O.q.v.	↑ N.q.v.

Data

information → image, voice, pdf, word, video etc.

Type of Data

1. structure
2. Semi structure
3. unstructure



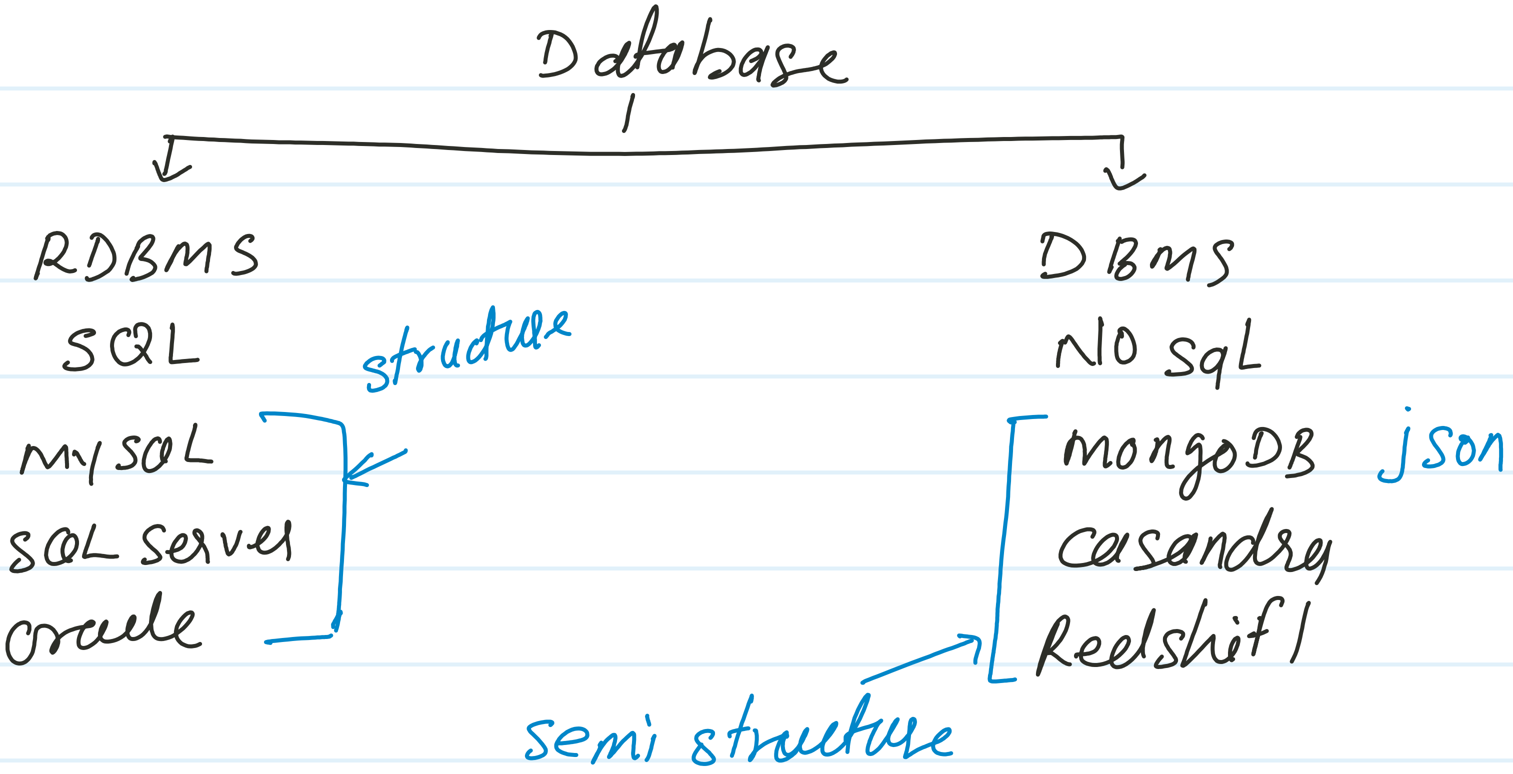
Tabular Data

90%

Semi structure - json, parque, aBq, xml etc.

unstructure

image, voice
video, word



Database

RDBMS
SQL
mySQL
SQL server
oracle
structure

DBMS
NO Sql
mongoDB json
casandra
Redshift
semi structure

Missing value

Features

ID	Name	salary	Age
1	Rahul	10K	19
2	vinay	20k	=
3	Nitesh	40K	22
4	Sashank	45k	28
5	Ravi	30K	30
6	Pankaj	50K	=

mean (Avg)

median

mode

19 22 28 30

$$\text{Media} = \frac{22+28}{2} = 25$$

mode = [1, 2, 3, 3, 2, 1, 1]

⇒ 1

Age = [10, 20, 20, 30, 40, – , – , –25, 30, –, –, –, 90, 100]

⇒ [10, 20, 20, 25, 30, 30, 40, 90, 100] ⇒ 30 / 40.5

$$\frac{10+20+15+90}{4} = 33.75$$

$$\frac{10+20+15+30}{4} = 18.75$$

Result = [yes, no $\underset{\underset{Y}{\uparrow}}{\_\_}$ yes, no $\underset{\underset{Y}{\uparrow}}{\_\_}$ yes, yes]

mode = yes

[18,20,30,37,28,22,35,33,21,27,29,31,38,37,28
89, 1, 2]

outlies / anomalies

$Q1 = 25\%$

$Q_3 = 75\%$

$IQR = Q_3 - Q_1$

upper fence = $Q_3 + 1.5\, IQR$

lower fence = $Q_1 - 1.5\, IQR$

$Q_1 = 21.45$

$Q_3 = 34.5$

$IQR = 34.5 - 21.45$

$= 13.05$

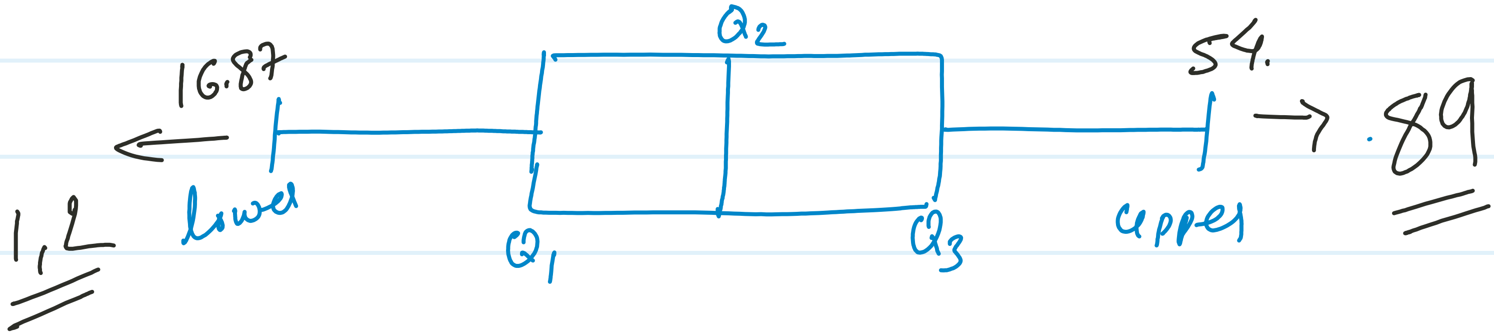
upper fence $= 34.5 + (1.5 \times 13.05)$

$= 54.07$

lower fence $= 21.45 - (1.5 \times 13.05)$

$= 16.87$

Box and whisker plot



median $= 28.5$

Encoding in data

ID	Grade
1	A
2	C
3	B
4	A
5	D
6	C
7	B

Encode
0
2
1
0
3
2
1

labelencoder

one-hot encoding

Grade	A	B	C	D
A	1	0	0	0
B	0	1	0	0
C	0	0	1	0
D	0	0	0	1
C	0	0	1	0
B	0	1	0	0
D	0	0	0	1

Scaling

$Y = mx + c$

X_1	X_2	X_3	Y
1	100	1009	
3	150	1500	
4	180	1800	
2	350	1650	
5	440	1250	

x = Features | Ind. feat.

Y = Target | Dep. feat.

Age	Eye sight
20	increasing.
30	incr.
40	inc.
50	Dec.
60	De

0 - 1

$\frac{1}{2000}$

$\frac{4}{2000} = 0.002 \Rightarrow$

$\frac{150}{2000} = 0.079$

$\boxed{0 - 1}$

$\frac{1800}{2000} = 0.9$

$X\uparrow - Y\downarrow$

$X\uparrow + Y\uparrow$

correlation

standar scaler