Exp - 3

DWM

Aim : To solve problems in data exploration & data preprocessing.

Obj : To enable students to effectively identify sources of data & process it for data mining.

Q. 1) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are

→ Data : 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

(a) what is the mean of the data? what is the median?

→ $\text{mean} = \frac{\sum f_i x_i}{\sum f_i}$

$$= \frac{13 + 15 + 16 + 16 + 19 + 20 + 20 + 21 + 22 + 22 + 25 + 25 + 25 + 25 + 30 + 33 + 33 + 35 + 35 + 35 + 35 + 36 + 40 + 45 + 46 + 52 + 70}{27}$$

$$= 30.962$$

median = mid value is 25.

(b) what is the mode of the data? Comment on the data modality?

→ mode = highest occurence is 25 & 35 is 4 times.

mode is 25 or 35 & modality is bimodal.

(c) what is the mid range of data?

→ midrange = $\frac{\text{min value} + \text{max value}}{2}$

$= \frac{13 + 70}{2}$

$= \frac{83}{2}$

$= 41.5$

(d) Can you find the first quartile & the third quartile of the data?

→ $Q1 = 20$

$Q2 = 25$

$Q3 = 35$

(e) Give the five-number summary of data?

→ i) Minimum - the smallest value in the dataset. 13

ii) first q Q1 - the value below which 25% of the data falls. 20

iii) median q Q2 - the middle value of the dataset i.e. 25

iv) third q Q3 - the value below which 75% of the data falls. 35

v) Maximum - the largest value in the dataset 70

(f) Show a boxplot of the data.

→