① Discuss the privacy issues in data mining with examples.

Ans) Privacy Issues in data mining

1. minimal protection Setup:

Most of the time data is protected by security measures like anti-Viruses, usernames, passwords, etc. which doesn't protect the data in long term.

2. Access controls:

Access controls verifies the identity of the person ~~buy~~ trying to access data. Single layer access control is not a very secured option.

3. Non-Verified Data updation:

Many times data is collected throughly and updated without the ~~Verified~~ - e Verification of Source.

4. Security - Architect Evaluation:

To save money and time, a lot of Organizations skip the process of audit of the Security architect, which makes easier to hack the data collection.

5. Filtering and Validating External Sources:

Wherever an unauthorised device is able to connect to the security system, it gives an entry point for vulnerabilities.

Ex:- People bring office work home and access the official data via their personal devices, which can create a loophole.

(2) for a certain dataset, the values of the attribute Age is given as follows.

25, 30, 15, 16, 33, 35, 70, 52, 13, 25, 33, 25, 40, 36, 35, 19, 25, 16, 20, 13, 22, 45, 21, 35, 20.

(a) find the mean, median, ~~~~ mode and mid-range of the data.

$$\text{mean} = \frac{1}{25} \sum_{i=1}^{25} x_i$$

$$\Rightarrow \bar{x} = \frac{1}{25}\left[2\times13+15+2\times16+19+2\times20+21+22+4\times25+30+2\times33+3\times35+36+40+45+52+70\right]$$

$$\Rightarrow \bar{x} = \frac{1}{25} \times 719$$

$$\Rightarrow \bar{x} = 28.76.$$

$\therefore n \rightarrow$ odd

$$\text{median} = x_{\left[\frac{n+1}{2}\right]} = x_{\left[\frac{25+1}{2}\right]} = x_{13}$$

$\therefore$ median = 25. //

age	no. of students / people
13	2
15	1
16	2
19	1
20	2
21	1
22	1
25	4
30	1
33	2
35	3
36	1
40	1
45	1
52	1
70	1

→ $\therefore$ highest no. of people are of age 25.

$\therefore$ Mode = 25. //

$$\text{Mid range} = \frac{\min + \max}{2}$$

$$\Rightarrow \text{midrange} = \frac{13+70}{2} = \frac{83}{2}$$

$$\therefore \text{midrange} = 41.5$$

(b) Calculate the Variance and Standard deviation.

$$n = 25$$
$$\Sigma x = 719.$$
$$\bar{x} = 28.76.$$

$$\text{Variance } (\sigma^2) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\Rightarrow \sigma^2 = \frac{1}{24} [(13-28.76)^2 \times 2 + (15-28.76)^2 +$$
$$(16-28.76)^2 \times 2 + (19-28.76)^2 +$$
$$(20-28.76)^2 \times 2 + (21-28.76)^2 +$$
$$(22-28.76)^2 + (25-28.76)^2 \times 4 + (30-28.76)^2 +$$
$$(33-28.76)^2 \times 2 + (35-28.76)^2 \times 3 +$$
$$(36-28.76)^2 + (40-28.76)^2 + (45-28.76)^2$$
$$+ (52-28.76)^2 + (70-28.76)^2]$$

$$\Rightarrow \sigma^2 = \frac{1}{24}\Big[2\times 248.3776 + 189.3376 + 162.8176\times 2 + 95.2576 + 2\times 76.7376 + 60.2176 + 45.6976 + 4\times 14.1376 + 1.5376 + 2\times 17.9776 + 3\times 38.9376 + 52.4176 + 126.3376 + 263.7376 + 540.0976 + 1,700.7376\Big]$$

$$\Rightarrow \sigma^2 = \frac{1}{24}\times 4260.56$$

$$\Rightarrow \sigma^2 = 177.5233.$$

Standard deviation$(\sigma) = \sqrt{\sigma^2}$

$$\Rightarrow \sigma = \sqrt{177.5233}$$

$$\Rightarrow \sigma = 13.32.$$

(c) Give the five-number summary of the data.

Soln

13, 13, 15, 16, 16, 19, 20, 20, 21, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 52, 70.

Minimum = 13.

Quartile 1 = $\left(\frac{N+1}{4}\right)$th term

= (6·5)th term

i.e. between 19 & 20 [6th - 7th]

= $\frac{19+20}{2}$ = 19·5.

Median = 25.

Quartile 3 = $\frac{(N+1)3}{4}$ th term

= 19·5th term

i.e. between 35 & 35 [19th - 20th]

= $\frac{35+35}{2}$ = 35//

Maximum = 70.

d) Create a box plot of the data.

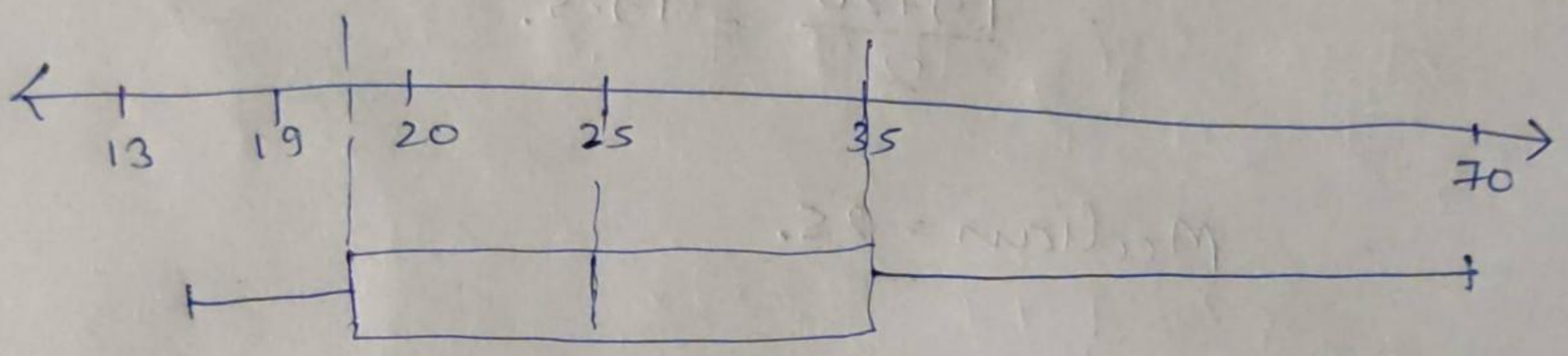
Soln) Minimum = 13.

Q1 = 19.5

Q2(Median) = 25

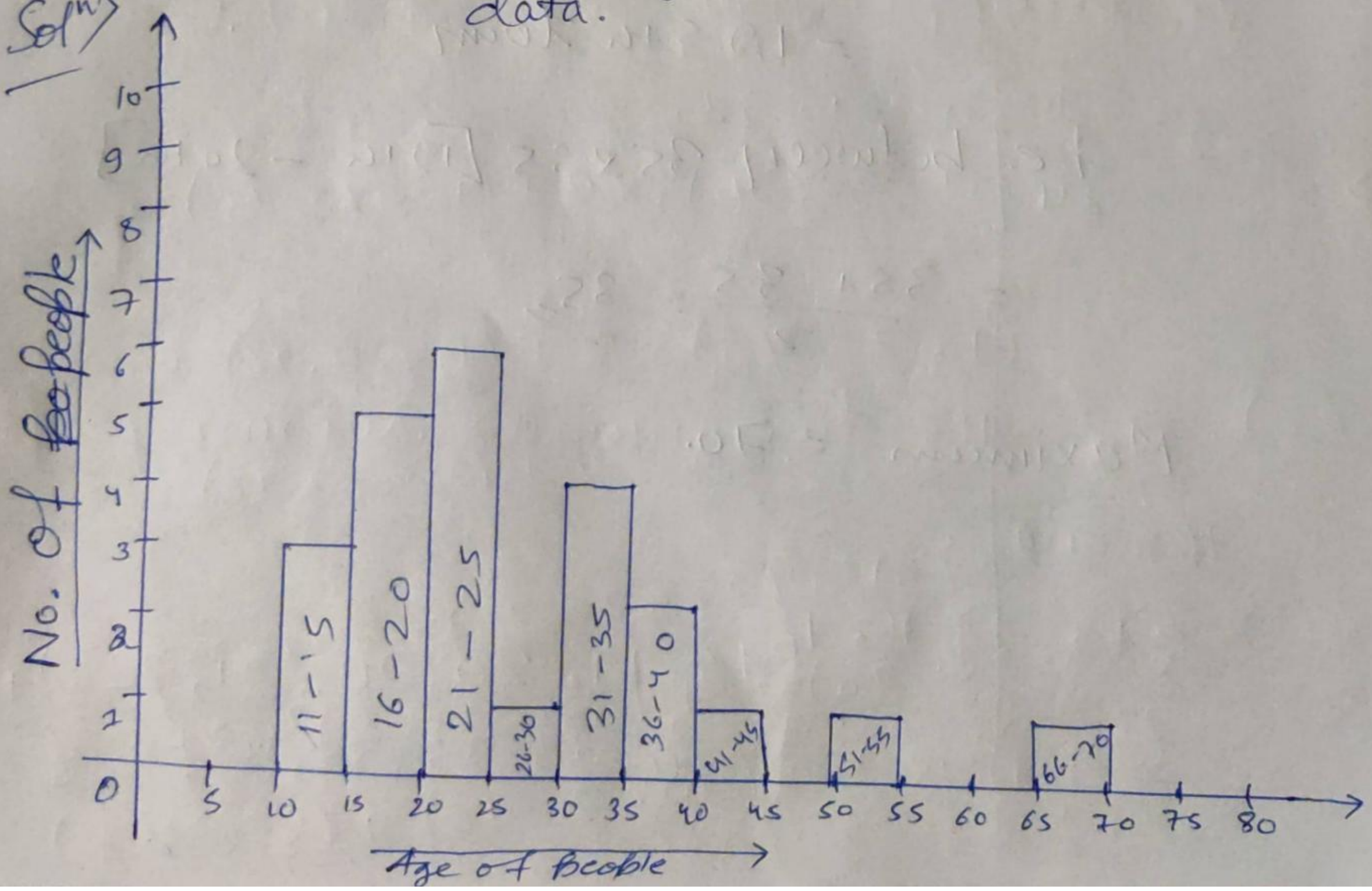
Q3 = 35.

Maximum = 70.



e) Draw a histogram to represent the data.

Soln)



(3) Create a dissimilarity matrix for the given data:-

Item	Colour	Price	Size
1	Blue	500	Small
2	Green	100	Large
3	Red	300	Small
4	Green	600	Medium

$\left[d = \frac{P - M}{P}\right]$

Sol^{n} $P \rightarrow$ no. of attributes $= 03$

$$d(2,1) = \frac{3-0}{3} = 1$$

$$d(3,1) = \frac{3-1}{3} = 0.67$$

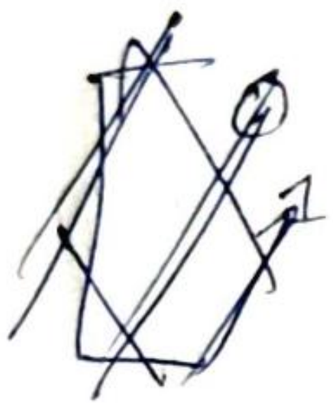
$$d(3,2) = \frac{3-0}{3} = 1$$

$$d(4,1) = \frac{3-0}{3} = 1$$

$$d(4,2) = \frac{3-1}{3} = 0.67$$

$$d(4,3) = \frac{3-0}{3} = 1.$$

$$\begin{bmatrix} d(1,1) & & & \\ d(2,1) & d(2,2) & & \\ d(3,1) & d(3,2) & d(3,3) & \\ d(4,1) & d(4,2) & d(4,3) & d(4,4) \end{bmatrix}$$

Ex 

$$= \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 0.67 & 1 & 0 & \\ 1 & 0.67 & 1 & 0 \end{bmatrix}$$

Ans