Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2025-26

Class:	**TE**	**Semester:**	**V**
Course Code:	**CSC504**	**Course Name:**	**Data Warehousing and Mining**

Name of Student:	**Shravani Sandeep Raut**
Roll No. :	**51**
Experiment No.:	**03**
Title of the Experiment:	**Tutorial on a)Data exploration b)Data Preprocessing**
Date of Performance:	
Date of Submission:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Performance	5	
Understanding	5	
Journal work and timely submission	10	
Total	20	

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Meet Expect Below Expectations (BE)
Performance	4-5	2-3	1
Understanding	4-5	2-3	1
Journal work and timely submission	8-10	5-8	1-4

Checked by
Name of Faculty : Ms. Neha Raut
Signature :
Date:

Q1. 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

Mean = $\frac{\Sigma x_i}{n}$

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

$= \frac{809}{27}$

$= 29.96$

Median = 25

Mode = 25 and 35

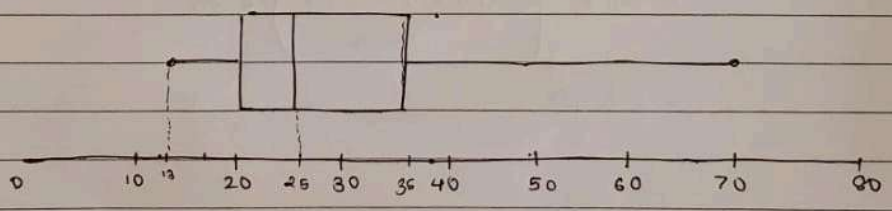
It is bimodal respectively.

Midrange = $\frac{\text{high}+\text{low}}{2} = \frac{70+13}{2} = 41.5$

Q1 = 20

Q2 = Median = 25

Q3 = 35



Q2.

Age	Frequency
1-5	200
6-15	450
16-20	300
21-50	1500
51-80	700
81-110	44

Total no. of observations (N) = 200 + 450 + 300 + 1500 + 700 + 44
= 3194

Median class = $\frac{N}{2} = \frac{3194}{2} = 1597$

Cummulative freq.

Interval	Frequency	CF
1-5	200	200
6-15	450	650
16-20	300	950
21-50	1500	2450

$$\text{Median} = L + \left(\frac{N/2 - F}{f}\right) \times W$$

$$= 21 + \left(\frac{(3194/2) - 950}{1500}\right) \times 30$$

$$= 33.94$$

Q3. Euclidean distance between points :

Formula. $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

Distances -

P_1 & P_2 : $\sqrt{(2-0)^2 + (0-2)^2} = 2.83$

P_1 & P_3 : $\sqrt{(3-0)^2 + (0-2)^2} = 3.16$

P_1 & P_4 : $\sqrt{(5-0)^2 + (1-2)^2} = 5.10$

P_2 & P_3 : $\sqrt{(3-2)^2 + (1-0)^2} = 1.41$

P_2 & P_4 : $\sqrt{(5-2)^2 + (1-0)^2} = 3.16$

P_3 & P_4 : $\sqrt{(5-3)^2 + (1-1)^2} = 2$

Q4 Given :

Min = 12000 , Max = 98000 , value = 73600

Formula .

$$\text{Normalized} = \frac{X - \min}{\max - \min} = \frac{73600 - 12000}{98000 - 12000}$$

$$= 0.716$$

Q5. Data = 2, 10, 18, 18, 19, 20, 22, 25, 28

Partition into bins of 3 :

Bin 1 : 2, 10, 18

Bin 2 : 18, 19, 20

Bin 3 : 22, 25, 28

Smoothing by Bin median :

Bin 1 Median = 10 → [10, 10, 10]

Bin 2 Median = 19 → [19, 19, 19]

Bin 3 Median = 25 → [25, 25, 25]

Submitted data = [10, 10, 10, 19, 19, 19, 25, 25, 25]

Bin Boundaries:

- Bin 1 boundaries = [2, 18]
  10 → nearest boundary = 18
  [2, 18, 18]

- Bin 2 boundaries = [18, 20]
  19 → nearest boundary = 20
  [18, 20, 20]

- Bin 3 boundary = [22, 28]
  25 → nearest boundary → 28
  [22, 28, 28]

Smoothed by bin boundaries [2, 18, 18, 18, 20, 20, 22, 28, 28].