Date: 14/01/23

Assignment #01

Name: Sufyan Ahmad
Roll NO: 20198013
Subject: Data Mining
Section: A

Q# 03

Work Experience	Frequency	C.F	
1-2	100	100	
3-4	350	450	
5-6	600	1050	
7-8	400	1450	medain
9-10	300	1750	
11-12	400	2150	
13-14	15	2165	

- $n = 2165$
- $n/2 = \frac{2165}{2} = 1082.5$
- $l = 7$, • $C.F = 1050$
- $n/2 = 1082.5$, • $f = 400$
- width = Class limit = 2

Formula: $l + \frac{(n/2 - C.F) \times width}{f}$ (f = medain)

$$= 7 + \frac{(1082.5 - 1050)}{400} \times 2$$

$$= 7 + \left(\frac{32.5}{400}\right) \times 2$$

$$= 7 + (0.08125) \times 2$$

$$= 7 + 0.1625$$

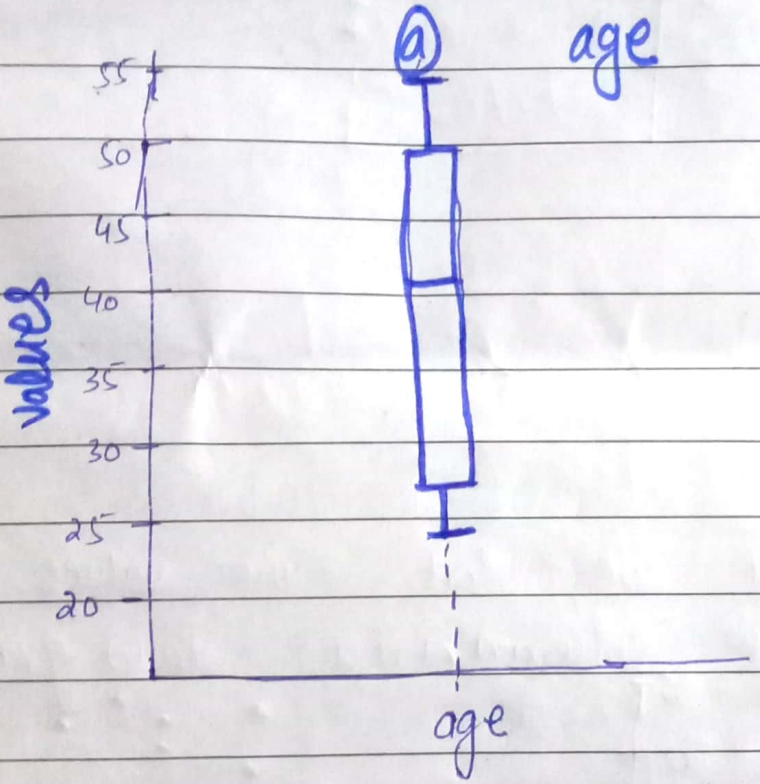
$$= 7.1625$$

Q#04

age	% fat
23	9.5
23	26.5
27	7.8
27	17.8
39	31.4
41	25.9
47	27.4
49	27.2
50	31.2
53	34.46
54	42.5
27	28.8

Date: 14/01/23

27	33.4
39	30.2
41	34.1
47	32.9
49	41.2
50	35.7



min = 23, max = 54, Q1 = 27

Q2 = 41, Q3 = 49, outliner = Not present.

.1° fat

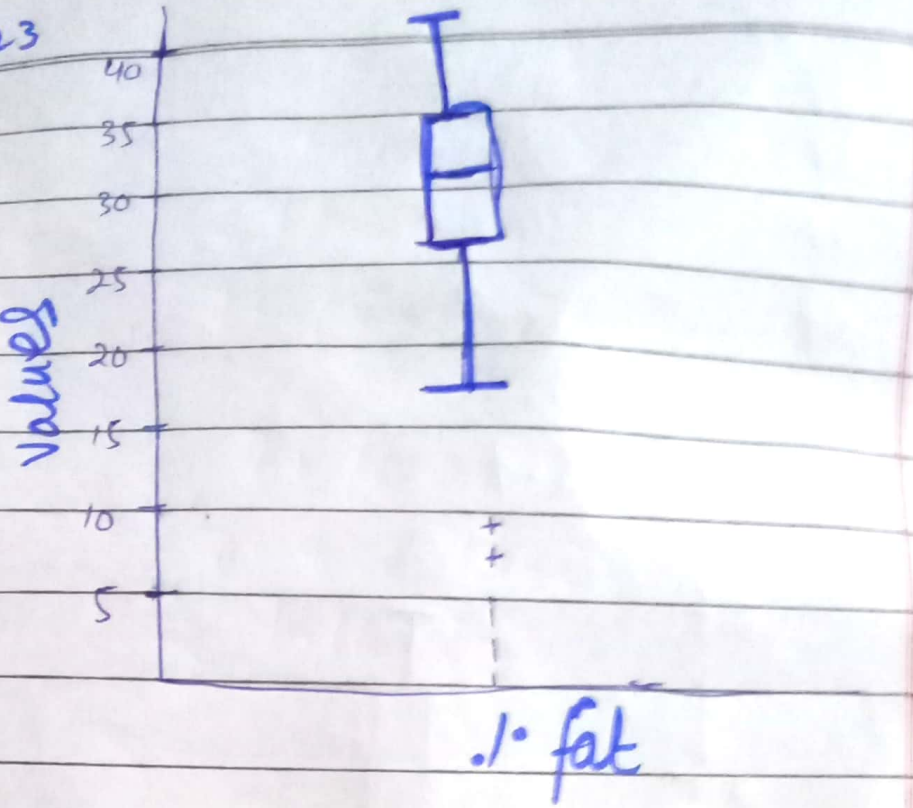
min = 7.8, Max = 42.5, Q1 = 26.5

Q2 = 30.7, Q3 = 34.1

.Outliers = 9.5, 7.8

Date: 14/01/23

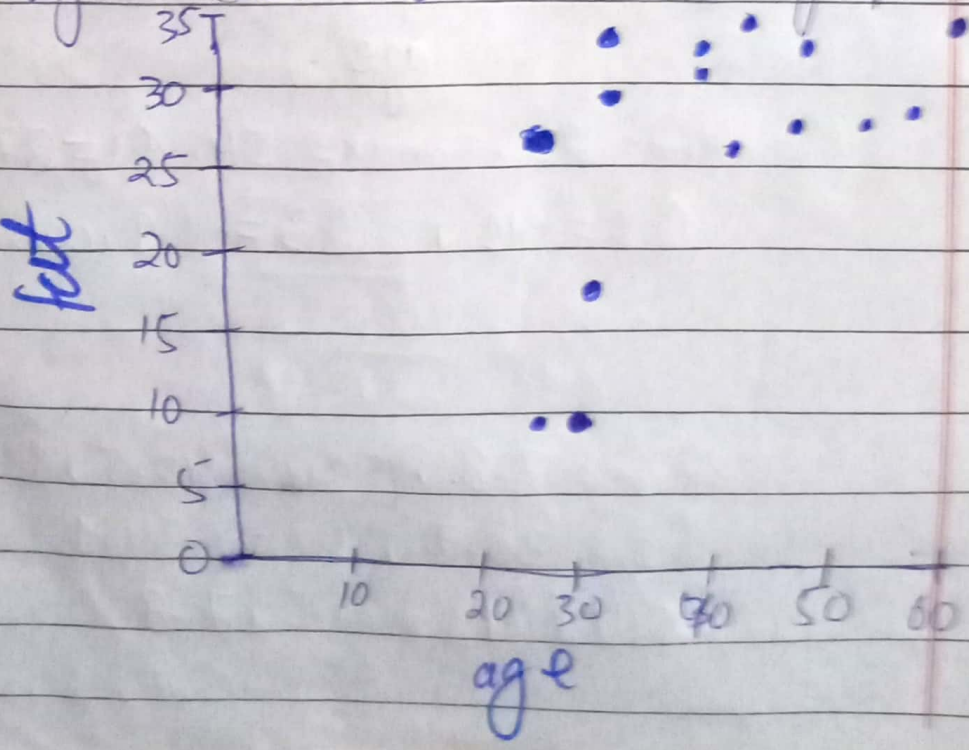
(a)



(b)

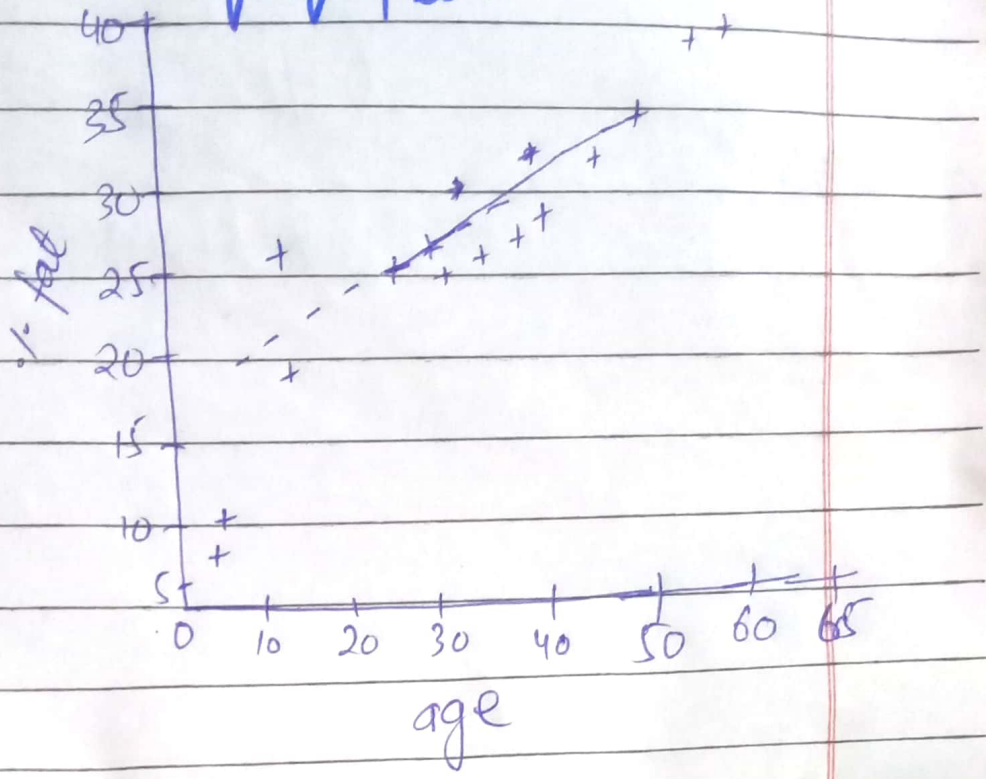
Scatter plot

- first we sort the values in age column and draw graph



Date: 14/01/23

q-q plot



Q#01

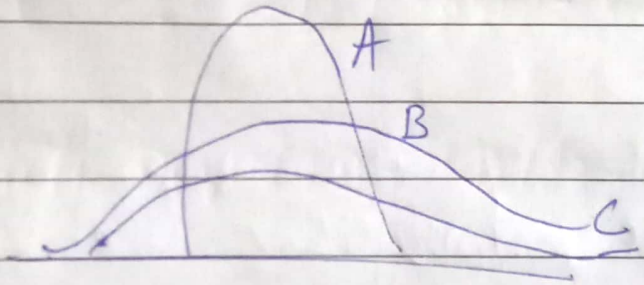
	Data warehouse	Database
purpose	Analysis	Reporting
Database	OLAP(online Analytical processing)	OLTP(online transactional processing)
Type of collection	Subject-oriented	Application-oriented
Query	Complex analytical queries	Simple transaction queries

Q#02

Solution

Dispersion:-

Dispersion is the state of getting dispersed or spread. Statistical dispersion means the the extent to which numerical data is likely to vary about an average value



How measure Dispersion of Data:-

Range:- It is simply the difference between the maximum value and the minimum value given in a data set Example 1, 3, 5, 6, 7

Range = 7-1=6

Variance:-

Deduct the mean from each data in the set, square each of them and add each square and finally divide them by the total no of values in the dat set.

Standard Deviation:- The square root of the variance is know as the standard deviation.

Quartiles and Quartile Deviation:- The quartiles are values that divide a list of numbers into quarters. The quartile deviation is half of the distance between the third and the first quartile.

Date:

Mean and Mean Deviation:-

The average of number is know as the mean and the arithmetic mean of the absolute deviations of the observations from a measure of central tendency is know as the mean deviation (also called absolute deviation).

- Range = C.D (Xmax - Xmin // Xmax + Xmin)
- Quartile Deviation = C.D (Q3 - Q1) // (Q3 + Q1)
- Standard Deviation (S.D) = C.D = S.D / Mean
- Mean Deviation = C.D = Mean deviation / Averag

Mean:- The most common and effective numevic measure of the cente of a set of data is the mean

Date:

$$L = \sum_{i=1}^{N}$$

Median:-

Middle value of data.

- Median = $L1 + \left(\frac{N/2 - (\Sigma freq)}{freq_{median}}\right) \times width$
- Sort the element array first pick the middle value.

Mode:- Most Repeated value of the element array.