Date: SEPT 24, 2023

ASSIGNMENT # 01

NAME : ISHA MALIK

CLASS : BSE - 7A

REG NO : FA20-BSE-013

COURSE : INTRODUCTION TO DATA SCIENCE

SUBMITTED TO : MA'AM MAHA RASHEED

Q1: You are tasked with ... frequency of 10 as necessary.

DATASET : AGES OF INDIVIDUALS

40, 23, 31, 27, 19, 45, 36, 20, 29, 22, 43, 38, 25, 18, 47, 27, 35, 30, 42, 33, 26, 21, 43, 28, 32, 44, 19, 41, 37, 18, 49, 34, 23, 22, 46, 18, 20, 50, 39, 27, 19, 31, 27, 22, 45, 38, 23, 18, 47, 36.

* **PREPROCESSING:**

First of all, we will preprocess this data set by sorting this data into ascending order and then dividing into equal depth bins i.e, frequency 10.

→ SORTED DATA :

18, 18, 18, 18, 19, 19, 19, 20, 20, 21, 22, 22, 22, 23, 23, 23, 23, 25, 26, 27, 27, 27, 28, 29, 30, 31, 31, 32, 33, 34, 35, 36, 36, 37, 38, 38, 39, 40, 41, 42, 43, 44, 45, 45, 46, 47, 47, 49, 50.

Now, we will apply bin smoothing techniques on this sorted dataset.

Date:

* Equi-Depth Bins:

Bin 1: 18, 18, 18, 18, 19, 19, 19, 20, 20, 21
Bin 2: 22, 22, 22, 23, 23, 23, 23, 25, 26, 27
Bin 3: 27, 27, 28, 29, 30, 31, 31, 32, 33, 34
Bin 4: 35, 36, 36, 37, 38, 38, 39, 40, 41, 42
Bin 5: 43, 43, 44, 45, 45, 46, 47, 47, 49, 50

* Smoothing By Bin Mean.

→ MEANS:

Bin 1 : 19
Bin 2 : 23.6 ≈ 24
Bin 3 : 30.2 ≈ 30
Bin 4 : 38.2 ≈ 38
Bin 5 : 45.9 ≈ 46

→ Bin 1:

19, 19, 19, 19, 19, 19, 19, 19, 19, 19

→ Bin 2:

24, 24, 24, 24, 24, 24, 24, 24, 24, 24

→ Bin 3:

30, 30, 30, 30, 30, 30, 30, 30, 30, 30

→ Bin 4:

38, 38, 38, 38, 38, 38, 38, 38, 38, 38

Date:

Bin 5:

46, 46, 46, 46, 46, 46, 46, 46, 46, 46

* Smoothing By Boundary.

Boundaries:

Bin 1: 18 - 21
Bin 2: 22 - 27
Bin 3: 27 - 34
Bin 4: 35 - 42
Bin 5: 43 - 50

Bin 1:

18, 18, 18, 18, 18, 18, 18, 21, 21, 21

Bin 2:

22, 22, 22, 22, 22, 22, 22, 27, 27, 27

Bin 3:

27, 27, 27, 27, 27, 34, 34, 34, 34, 34

Bin 4:

35, 35, 35, 35, 35, 35, 42, 42, 42, 42

Bin 5:

43, 43, 43, 43, 43, 43, 50, 50, 50, 50

This is the representation of data after applying bin smoothing techniques.