Name: Abdelrahman Ahmed Hassan Mohamed **Sec : 3**

Exercise: You are given the following dataset representing students' scores in a Statistics exam:

Data = [45, 48, 52, 54, 56, 58, 60, 62, 65, 68, 70, 72, 75, 78, 80]

Discretize the data into 3 bins using the following two methods:

- Equal Width Binning
- Equal Frequency Binning **For each binning method:**

- **Sort the values in each bin in ascending order.**
- **Apply data smoothing using the following techniques:**
  - Smoothing by Mean
  - Smoothing by Median
  - Smoothing by Boundaries

Solution

Discretizing:

1. **Equal Width Binning: Range** = 80 – 45 = 35 ➔ 35/ 3 = 11.67
   Bin 1: {45, 56.67} ➔ [45, 48, 52, 54, 56]
   Bin 2: {56.67, 68.34} ➔ [58, 60, 62, 65, 68]
   Bin 3: {68.34, 80} ➔ [70, 72, 75, 78, 80]
2. **Equal Frequency Binning:** 15/3 = 5 ➔ **so five values per bin**
   Bin 1: [45, 48, 52, 54, 56]
   Bin 2: [58, 60, 62, 65, 68]
   Bin 3: [70, 72, 75, 78, 80]

And as it seems both ways resulted the same bins
So the smoothing will be the same
(no need to do it twice I mean)

Smoothing:

1. **By Mean:**
   - **Bin 1: (45 + 48 + 52 + 54 + 56) / 5 = 51 ➔ Bin 1: [51, 51, 51, 51, 51]**
   - **Bin 2: (58 + 60 + 62 + 65 + 68) / 5 = 62.6 ➔ Bin 2: [62.6, 62.6, 62.6, 62.6, 62.6]**
   - **Bin 3: (70 + 72 + 75 + 78 + 80) / 5 = 75 ➔ Bin 3: [75, 75, 75, 75, 75]**
2. **By Median:**
   - **Bin 1 median = 52 ➔ Bin 1: [52, 52, 52, 52, 52]**
   - **Bin 2 median = 62 ➔ Bin 1: [62, 62, 62, 62, 62]**
   - **Bin 3 median = 75 ➔ Bin 1: [75, 75, 75, 75, 75]**
3. **By Boundaries:**
   - **Bin 1: min = 45, max = 56 ➔ Bin 1: [45, 45, 56, 56, 56]**
   - **Bin 2: min = 58, max = 68 ➔ Bin 1: [58, 58, 58, 68, 68]**
   - **Bin 3: min = 70 , max = 80 ➔ Bin 1: [70, 70, 80, 80, 80]**