- Numericals.

Q1] sales: (2300, 495, 675, 543, 454, 7877, 5434, 345, 2342, 654, 547, 545)
Normalize data using min-max scaling

$X_N = \frac{X - X_{min}}{X_{max} - X_{min}}$ eg: $X_N = \frac{2300 - 345}{7877 - 345} = 0.2596$

// Do this for all vals.

Ans: [0.2596, ...]

Q2] Data given. [5, 10, 13 ..... 85]
Draw histograms of bin size 5 and 8.
Explain effect of bin size on hist.

⇒ bin size = 5

5 - 9
10 - 14
15 - 19
20 - 24
⋮

bin size = 8

5 - 12
13 - 20
21 - 28
⋮

(Draw hist. for both)

Smaller bin → finer details visible, might be diff. to interpret if too many bins.

Larger → smooth represent^n. Extreme vals might get merged.

Q3] Data = [4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34]
No. of bins = 3

i] Partition into equal freq. bins.

∴ data size = 12, ∴ 12/3 = 4 → bin size.

bin 1 = [4, 8, 9, 15]
bin 2 = [21, 21, 24, 25]
bin 3 = [26, 28, 29, 34]

i] Smoothing by bin means.

bin 1 mean = $\frac{4+8+9+15}{4} = \frac{36}{4} = 9$

Bin 2 mean = $\frac{21+21+24+25}{4} = 22.75 = 23$

bin 3 mean = $\frac{26+28+29+34}{4} = 29.25 \approx 29$

smoothed data

[9, 9, 9, 9, 23, 23, 23, 23, 29, 29, 29, 29]

ii] smooming by bin boundaries. Replace each val with closest boundary.

Bin 1 : 4, 8, 9, 15 → 4, 4, 4, 15

Bin 2 : 21, 21, 24, 25 → 21, 21, 25, 25

Bin 3 : 26, 28, 29, 34 → 26, 26, 26, 34

Eg: Nearest boundary for 9 is 4 as 9 - 4 = 5 ✓

15 - 9 = 6 ✗

Ans: [4, 4, 4, 15, 21, 21, 25, 25, 26, 26, 26, 34]

Q4] z-score normalisatn / standarization.

$z = \frac{x - \mu}{\sigma}$ → mean

→ S.D.

Eg: 10, 20, 30, 40, 50

$\mu = 30$, $\sigma = 14.14$

$z_{10} = \frac{10-30}{14.14} = -1.41$, $z_{20} = \frac{20-30}{14.14} = -0.71$

$z_{30} = 0$, $z_{40} = \frac{40-30}{14.14} = 0.71$, $z_{50} = 1.41$

Z-score normalised.

[-1.41, -0.71, 0, 0.71, 1.41]

mean = 0, $\sigma = 1$

Q5] Age: [5, 10, 13, 15, 16, 16, 20 ..... 85]

i] Mean = $\frac{\sum x}{n}$ = 30.68

ii] Median = 15th val = 25

iii] $\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$ = 16.88

iv] Q_1 : Median of values of first half excluding means

= 20

v] Q_3 : 35 (Median of upper half)

IQR = $Q_3 - Q_1$ = 15

vi] Low outliers = $Q_1 - 1.5 \times IQR$

= 20 - 1.5 × 15 = -2.5

High outlier = $Q_3 + 1.5 \times IQR$

= 35 + 22.5 = 57.5

∴ (70, 85) ⟶ outliers