Scenario 1: Delivery Time Analysis for an E-commerce Company

An e-commerce company tracks delivery times (in minutes) for 15 orders:

[25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95]

The company wants to analyze the delivery performance using percentiles and detect if there are any unusual delivery times.

Question 1:
Calculate Q1 and Q3.

Ans:

Q1(25th percentile)= 0.25(n+1)= 0.25(15+1)=4 i.e 4^{th} position value is 40, Q1=40

Q3(75th percentile)= 0.75(n+1)=0.75(16)=12 i.e 12^{th} position value is 80, Q3=80

Question 2:
Find the Inter quartile Range (IQR).

Ans: IQR = Q3-Q1= 80-40 = 40

Question 3:
Detect Outliers using the IQR method.

Ans:
Lower bound = Q1-1.5*IQR = 40 - 1.5*40 = -20 , there is no lower outlier present in the data
Upper bound = Q3+1.5*IQR = 80 + 1.5*40 = 140, there is no upper outlier present in the data

Conclusion: There is no outlier in this data.

Scenario 2: Student Score Analysis

A teacher is analyzing the mathematics scores of students in her class. The scores are:

[45, 50, 55, 60, 60, 62, 63, 65, 90, 95]

Question 1:
Calculate the mean, median, and mode of the scores.

Ans: Mean = 45 +50+55+60+60+62+63+65+90+95 / 10 = 64.5

Median= 60+62 / 2 =122/2 = 61

Mode = 60

Question 2:
Explain why the median might be a better representation than the mean in this case.

Ans: Mean does not include outlier but median include the outlier. Since outlier doesn't bother median value can provide better result.

Scenario 3: Grocery Store Customer Analysis

A grocery store manager tracks how many customers visit the store daily for a month:
`[5, 10, 8, 15, 20, 5, 12, 14, 10, 18]`

Question 1:
Create a frequency distribution table for this data

Ans:

[16]:

	No. of customer	Frequency
0	5	2
1	10	2
2	8	1
3	15	1
4	20	1
5	12	1
6	14	1
7	18	1

Scenario 4: Real Estate Model Analysis

A real estate model has three variables:

- House Size
- Number of Rooms
- Number of Bathrooms

Question 1:
How can you detect multicollinearity in this model?

Ans: Multicollinearity can be detected by VIF (variance inflation factor). Calculate the VIF, if VIF>5, consider as the highly correlated which impacts the model accuracy.

Scenario 5: Medicine Effectiveness Study

A company made a new medicine to lower blood pressure. They gave it to one group and gave a fake pill (placebo) to another group.

Question 1:
How can the company check if the new medicine works?

Ans: Using hypothesis function, we assign the group 1 as H0 who takes lower BP medicine and another group 2 as H1 who takes fake pills.

Run Ttest, based on the p value, we can decide whether it works or not. If the p <0.05, then medicine is working efficiently.

Scenario 6: Identifying Outliers in Sales Data

A company wants to find any unusual spikes in sales.

Question 1:
How can the company detect outliers in their sales data?

Ans:

Using Inter quartile calculation, we can able to find the outlier in the data.

Step1: Find Q3 and Q1 value. (formula: Q1=0.25(n+1) & Q3=0.75(n+1)
Step2: Then calculate IQR= Q3-Q1
Step3: calculate lower bound and upper bound.
Formulae: lower bound= Q1-1.5*IQR
Upper bound= Q3+1.5*IQR.
Step4:check if the lower value of data is greater than lower bound value, no outlier.Otherwise lower outlier is present.
If the highest value of the data is less than upper bound value, no outlier present. Otherwise upper outlier is present.

Scenario 7: Understanding Customer Satisfaction

A restaurant conducted a survey to rate customer satisfaction on a scale of 1 to 5:

[5, 4, 4, 5, 3, 4, 5, 2, 4, 3]

Question 1:

How can the restaurant summarize the overall satisfaction?

Ans:

We can use measure of central tendency concept, where we can mostly occurred rating using mode, average rating using mean or median.