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.

Q1 – 25 percentile, Q3 – 75 percentile

Q1= numpy.percentile(delivery_times, 25)

Q3= numpy.percentile(delivery_times, 75)

Question 2:
 Find the Interquartile Range (IQR).

IQR = Q3 – Q1 (50% of data)

 Question 3:
 Detect Outliers using the IQR method.

Lower_bound= Q1 – 1.5*IQR
Upper_bound= Q3 + 1.5*IQR

If data is lower than Lower_bound value then it is Lower Outlier
If data is greater than Greater_bound value then it is Upper Outlier

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.

Mean = Total sum of scores/Total no.of elements in scores

Median = Check total no. of elements is odd or even. Odd= n+1/2, Even = center 2 values/2

Mode = Repeated values

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

In Mean value it is included with Outliers, Median value omit outlier so Median is Better than Mean.

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.

Customer Visit	Frequency
5	2
10	2
8	1
15	1
20	1
12	1
14	1
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?

Using Variance Inflation Factor(VIF) we can detect Multicollinearity.

VIF = 1 – Not Correlated,
VIF = 1 to 5 – Moderately Correlated
VIF = >5 – Highly Correlated

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?

We can do a t-test Unpaired test using Hypothesis Testing we can confirm the effectiveness.

Hypothesis Test: if ($p\_value <=0.05$) then it reject null hypothesis(Medicine works) else accept the null hypothesis(Medicine doesn't work).

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?

Lower_bound= Q1 – 1.5*IQR
Upper_bound= Q3 + 1.5*IQR

If data is lower than Lower_bound value then it is Lower Outlier
If data is greater than Greater_bound value then it is Upper Outlier

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?

Based on the mode we can finalize. In this 4 and 5 are repeated rating so rating of restaurant is hight.