

## DS - Scenario Set Question 1

### 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.

#### Question 2:

Find the Interquartile Range (IQR).

#### Question 3:

Detect Outliers using the IQR method.

#### Answer:

#### Method:

1. Calculate **Q1** = 25th percentile of the data
  2. Calculate **Q3** = 75th percentile of the data
  3. Find **IQR** =  $Q3 - Q1$
  4. Compute **Lower Bound** =  $Q1 - 1.5 \times IQR$
  5. Compute **Upper Bound** =  $Q3 + 1.5 \times IQR$
  6. If any value  $<$  Lower Bound  $\rightarrow$  **Lower Outlier**
  7. If any value  $>$  Upper Bound  $\rightarrow$  **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.

#### Question 2:

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

**Answer:**

**Question 1: Method**

1. Calculate **Mean** = (Sum of all scores) ÷ (Number of scores)
2. Calculate **Median** = Middle value (or average of two middle values) after arranging data in order
3. Calculate **Mode** = Value that appears most frequently

**Question 2: Method**

1. Compare mean and median values
  2. If extreme (very high or low) scores exist, they can distort the mean
  3. Median is less affected by outliers
  4. Therefore, **median** gives a better representation of the central performance when data has outliers
- 

**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.

**Answer:**

**Question 1: Method — Frequency Distribution Table**

1. **List all unique values** in the data in ascending order
  2. **Count the number of times** each value occurs
  3. **Create a table** with two columns:
    - **Unique Value (Number of Customers)**
    - **Frequency (Number of Days it Occurred)**
  4. Optionally, calculate **relative frequency** = (Frequency ÷ Total Days) × 100
-

## 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?

### Answer:

- **Correlation Matrix:**

- Compute pairwise correlations between independent variables (House Size, Number of Rooms, Number of Bathrooms).
- High correlation (close to +1 or -1) indicates possible multicollinearity.

- **Variance Inflation Factor (VIF):**

- Calculate VIF for each independent variable.
  - $VIF > 5$  (or 10) indicates significant multicollinearity.
- 

## 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?

### Answer :

#### Question 1: Method - How to Check if the Medicine Works

1. **Compare Groups:** Give the medicine to one group and a placebo to another.
2. **Measure Blood Pressure:** Record before and after treatment.
3. **Set Hypotheses:**
  - **Null Hypothesis ( $H_0$ ):** The medicine has **no effect** (mean change in blood pressure is the same as placebo).

- **Alternative Hypothesis ( $H_1$ ):** The medicine **reduces blood pressure** (mean change is greater than placebo).
  - 4. **Statistical Test:** Use a **t-test** to compare the two groups.
  - 5. **Check p-value:**
    - If **p-value** < **0.05**, reject  $H_0 \rightarrow$  medicine works.
    - If **p-value**  $\geq$  **0.05**, fail to reject  $H_0 \rightarrow$  no significant effect.
  - 6. **Optional:** Visualize results with charts (bar or boxplot) for clarity.
- 

## 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?

Answer:

### Question 1: Method — How to Detect Outliers

1. **Use a Boxplot** to visually spot unusual sales.
  2. **IQR Method:**
    - Find Q1 (25th percentile) and Q3 (75th percentile)
    - Calculate  $IQR = Q3 - Q1$
    - Lower Bound =  $Q1 - 1.5 \times IQR$ , Upper Bound =  $Q3 + 1.5 \times IQR$
    - Sales outside this range  $\rightarrow$  outliers
  3. **Z-Score Method:**
    - Find how far each sale is from the mean in terms of standard deviation
    - Usually,  $|z| > 3 \rightarrow$  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?

Answer:

### How to Summarize Overall Satisfaction

1. **Calculate Average (Mean):** Shows the general satisfaction level.
2. **Find Median:** Middle value when ratings are ordered; less affected by extremes.
3. **Find Mode:** Most common rating to see what most customers feel.
4. **Frequency Distribution:** Count how many customers gave each rating.
5. **Optional Visualization:** Use a **bar chart** or **pie chart** to show rating distribution clearly.