## Scenario 1:Flight Delay Analysis

**Question:**
An airline tracks flight delays (in minutes) for 20 flights. How do you analyze the flight delays to calculate percentiles, detect outliers, and evaluate the overall distribution?

**Answer:**

- **Percentiles Calculation:**
  Calculate the 10th, 25th, 50th (Median), 75th, and 90th percentiles to understand how flight delays are distributed at different levels.

- **IQR Calculation:**
  Compute **IQR = Q3 - Q1**. Detect outliers using the formula:

    - Outliers are values outside Q1−1.5×IQRQ1 - 1.5 \times IQRQ1−1.5×IQR or Q3+1.5×IQRQ3 + 1.5 \times IQRQ3+1.5×IQR.

- **Distribution:**
  Use a **box plot** to visualize the spread, median, quartiles, and potential outliers. A **histogram** helps assess the distribution shape, skewness, and frequency of delays.

## Scenario 2: Employee Salary Analysis

**Question:**
A company wants to analyze the salary distribution of its employees to understand the central tendency and determine if the data is skewed. How should this be done?

**Answer:**

- **Central Tendency Calculation:**
  Calculate **Mean**, **Median**, and **Mode** to understand the central value of the salaries.

- **Skewness Check:**

- If **Mean > Median**, the data is **Right Skewed** (higher salaries skew the mean).
- If **Mean < Median**, the data is **Left Skewed** (lower salaries pull the mean down).
- If **Mean ≈ Median**, the data is **Symmetrical**.

- **Best Representation:**
  Use the **Median** when there are outliers, as it better represents the central tendency in the presence of extreme values.

### Scenario 3: Product Sales Analysis

**Question:**
A retail store records product sales over 15 days. How do you create a frequency distribution table and visualize the sales data?

**Answer:**

- **Frequency Distribution:**
  Divide the sales data into intervals (e.g., 5 or 10 units). Count the sales within each interval to understand how sales are distributed.

- **Visualization:**

  - **Histogram**: Displays the frequency distribution across sales intervals.
  - **Bar Plot**: Shows trends in sales, helping to visualize changes over time.

## Scenario 4: Student Exam Performance Analysis

### Question:

### A school wants to analyze the exam performance of students across three subjects: Mathematics, Science, and English. How can Data Science concepts be applied to understand their performance?

### Answer:

**1. Data Preprocessing:**

- **Handle missing values by imputing with the mean or median.**
- **Convert categorical values (if any) using label encoding or one-hot encoding.**

**2. Descriptive Statistics:**

- **Calculate mean, median, mode, and standard deviation for each subject to understand the central tendency and dispersion.**

**3. Visualization:**

- **Use box plots to detect outliers.**
- **Create histograms to visualize the distribution of scores.**
- **Plot scatter plots to check relationships between subjects.**

**4. Correlation Analysis:**

- **Compute correlation coefficients to see how subject scores relate to each other.**
- **Use a heatmap to visualize correlations.**

## Scenario 5: Clinical Trial for Diabetes Medication

**Question:**
A pharmaceutical company conducted a clinical trial with two groups: one receiving medication and the other a placebo. How do you perform a hypothesis test to determine the effectiveness of the medication?

**Answer:**

- **Hypothesis Test:**
    - $H_0$ **(Null Hypothesis):** No difference between medication and placebo.
    - $H_1$ **(Alternative Hypothesis):** Medication lowers blood sugar more than the placebo.
- **T-Test:**
    - If $p \leq 0.05$, reject $H_0$ (indicating medication is effective).
    - If $p > 0.05$, fail to reject $H_0$ (no significant difference).