# DATA SCIENCE ---2

1.FLIGHT DELAY ANALYSIS:

- **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.

**2. Employee Salary Analysis:**

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

**3.PRODUCT ANALYSIS:**

**Frequency Distribution:**

**Divide the sales data into intervals 5 or 10 units. Count the sales within each interval to understand how sales are distributed.**

**4. Student Exam Performance Analysis:**

**. Data Preprocessing:**

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

**Descriptive Statistics:**

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

**5. Clinical Trial for Diabetes Medication:**

**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).**