### **Section 1.3 Homework: Excel Descriptive Statistics with Titanic Dataset**

#### **Objective: Use the** [**Titanic dataset**](https://www.kaggle.com/datasets/yasserh/titanic-dataset) **to practice descriptive statistics and Excel analysis techniques.**

### **Part 1: Data Cleaning**

1. **Remove Duplicates**: Check if duplicate rows using Excel’s **Remove Duplicates** feature.
2. **Handle Missing Data**:
   * Check for missing values in Age, Fare, or other columns.

### **Part 2: Descriptive Statistics**

1. **Basic Calculations**:
   * Calculate the **mean**, **median**, and **standard deviation** for Age and Fare using:
     + =AVERAGE(range)
     + =MEDIAN(range)
     + =STDEV.P(range).
2. **Survival Analysis**:
   * Calculate survival rate (=COUNTIF(Survived, 1)/COUNTA(Survived)).
   * Group survival rates by Sex and Pclass using **COUNTIF** or **PivotTables**.

### **Part 3: Data Visualization**

1. **Histogram**:
   * Create a histogram for Age to visualize the distribution of passenger ages.
2. **Bar Chart**:
   * Create a bar chart showing survival rates by Pclass.
3. **Pie Chart**:
   * Create a pie chart of passenger distribution by Sex.

### **Part 4: Advanced Analysis**

1. **Conditional Formatting**:
   * Highlight passengers who paid a fare higher than 75 using **Conditional Formatting**.
2. **Top 10 Analysis**:
   * Identify the 10 passengers who paid the highest Fare using Excel’s **SORT** function.

### **Reflection Questions**

1. What is the average and median age of passengers?
2. Which Pclass had the highest survival rate?
3. Are younger passengers more likely to survive? (Compare Age with Survived.)