

# ASSIGNMENT#1

**Possible Points: 100**

**Deadline: 19<sup>th</sup> February 2025**

**All are required to do the assignments on an individual basis.**

## **Question 1 (40 Points)**

You are provided with a sales report containing information about different types of clothes. Use the pandas for data manipulation and matplotlib for data visualization. Answer the questions below:

1. Load the data from the csv file provided. (3 Points)
2. Get statistical summary of column stock, check for missing values in each column (5 Points)
3. Visualize stock availability per color using a bar chart. (10 Points)
4. Add a column “Status” where “Low” if stock is less than 5, “Medium” between 5 and 10 and “High” if greater than 10. (7 Points)
5. Generate a pie chart to represent the distribution of status column. Retrieve all the rows where stock is “medium”. Store it into a new dataset. Display top 5 rows. (10 Points)
6. Save the new dataset into a csv file. (5 Points)

## **Question 2 (30 Points)**

You are provided with a sample data containing information about marks obtained by the students. Answer the questions below:

Name	Math	Science	English	History
Alice	85	90	88	92
Bob	78	82	80	76
Charlie	92	95	96	94
David	70	75	72	68

1. Store the data as list of dictionaries, the name should be used as keys and the marks should be stored in a list as value for each name. (7 Points)

For example:

```
students = [
    {"name": "Alice", "scores": [85, 90, 88, 92]},
    {"name": "Bob", "scores": [78, 82, 80, 76]},
    {"name": "Charlie", "scores": [92, 95, 96, 94]},
    {"name": "David", "scores": [70, 75, 72, 68]},
    {"name": "Eve", "scores": [88, 91, 85, 90]}
]
```

2. Extract the test score from each dictionary and store it into a NumPy array. Compute the average score of each student. (10 Points)
3. Calculate average score of each student. Calculate maximum and minimum score for each student. (5 points)
4. Get average score of each subject. Determine which subject has the highest standard deviation (5 Points)
5. Calculate median score of each subject and median score of each student (3 points)

### **Question 3 (30 Points)**

Generate a dataset which contains monthly temperatures of a city. Analyze the data and answer the data below:

1. Generate a NumPy array containing temperature for 12 months. For each month add 10 entries. The random values should be between 10 degrees Celsius and 40 degrees Celsius (5 Points)

For example:

```
[[28.36271364 16.37657556 19.93765878 12.98773517 36.56135741 24.85204956
 28.41713682 39.0745361 30.15350851 14.33829076]
 [34.10790409 25.13308573 12.45907367 39.24393822 30.75928293 32.23960188
 26.70615383 15.19262271 28.8488068 31.17765459]
 [35.37903434 14.22262288 22.56104959 32.05813553 28.06484619 33.12937494
 25.16855632 16.23958689 11.25392817 36.43137411]
 ...
 [27.40824715 12.74668498 39.22858271 14.31994631 16.27666979 24.42888195
 23.74682265 12.3024893 21.62306291 31.90789329]]
```

2. Get average temperature for each month. Which month has the highest and lowest temperature? Plot a line chart for average monthly temperature. (10 Points)
3. Get mean, mode, median values for each month (5 Points)
4. Split the data into three time periods, Winter/Spring (Jan-Apr), Summer (May-August) and Fall (September to December). Get the average temperature of these three periods. Plot a bar chart for comparison. (10 Points)