Python Data Structures and Libraries

Task 1 – Tuple, List, Set, Dictionary, Pandas and NumPy Practice

Muhammad Haseeb

Task Objectives

- **Tuple Practice**
  - Create a tuple for the first 5 students containing their StudentID and Name.
- **List Practice**
  - Store the marks (Math, Science, English) of top 5 students in a list.
  - Calculate and print the average score using this list.
- **Set Practice**
  - Extract all unique genders from the dataset and store them in a set.
- **Dictionary Practice**
  - Pick any one student and create a dictionary with keys: `"Name"`, `"Gender"`, and `"Marks"` (another dictionary for their subjects and scores).
- **Pandas Practice**
  - Load the dataset using Pandas.
  - Display the first 10 rows.
  - Check missing values.
  - Add a new column called `"Total"` which sums all subject scores.
  - Find and print details of the top 5 students with the highest total marks.
- **NumPy Practice**
  - Convert all numeric scores into a NumPy array.
  - Calculate and print the mean, max, and standard deviation for each subject.
  - Use filtering to display all students with `"Math"` $>=$ 85 and `"Science"` $>=$ 90.
  - Count how many male and female students scored above 240 total marks.

1 Importing Libraries and Reading the Dataset

```
import pandas as pd
import numpy as np

df = pd.read_csv("Dataset 1.csv")
print("First 10 rows:")
print(df.head(10))
```

Output

```
First 10 rows:
   StudentID       Name Gender  Math  Science  English
0          1      Megan      M    94       97       50
1          2  Katherine      F    53       53       89
2          3     Robert      F    59       69       71
3          4   Jonathan      M    86       73       56
4          5    William      M    74       74       62
5          6    Richard      M    51       88       89
6          7    Kristen      F    73       96       74
7          8      Kevin      M    67       87       75
8          9     Thomas      F    63       58       59
9         10     Brandy      M    70       66       55
```

2 Top 5 Students by Total Marks

```
df[’Total’] = df[[’Math’, ’Science’, ’English’]].sum(axis=1)
top_5 = df.sort_values(by=’Total’, ascending=False).head(5)
print(top_5)
```

Output

```
    StudentID      Name Gender  Math  Science  English  Total
95         96   Sabrina      M    90       86       98    274
36         37  Jennifer      F    99       86       84    269
78         79    Sandra      M    78       97       90    265
69         70     Kelly      F    88       97       79    264
94         95     Wendy      M    98       85       77    260
```

3 Tuple Practice

```
student_tuples = tuple(zip(df[’StudentID’][:5], df[’Name’][:5]))
print(student_tuples)
```

Output

```
((1, ’Megan’), (2, ’Katherine’), (3, ’Robert’), (4, ’Jonathan’), (5, ’William’))
```

4 List Practice

```
marks_list = top_5[[’Math’, ’Science’, ’English’]].values.tolist()
all_scores = [score for student in marks_list for score in student]
average_score = sum(all_scores) / len(all_scores)
print("Average score of top 5 students:", round(average_score, 2))
```

Output

```
Average score of top 5 students: 88.8
```

5 Set Practice

```
genders = set(df[’Gender’].dropna())
print(genders)
```

Output

```
{’M’, ’F’}
```

6 Dictionary Practice

```
student = df.iloc[0]
student_dict = {
    "Name": student[’Name’],
    "Gender": student[’Gender’],
    "Marks": {
        "Math": student[’Math’],
        "Science": student[’Science’],
        "English": student[’English’]
    }
}
print(student_dict)
```

Output

```
{’Name’: ’Megan’, ’Gender’: ’M’, ’Marks’: {’Math’: 94, ’Science’: 97, ’English’: 50}}
```

7 NumPy Practice

```
scores = df[[’Math’, ’Science’, ’English’]].to_numpy()
print("Mean per subject:", np.mean(scores, axis=0))
print("Max per subject:", np.max(scores, axis=0))
print("Std Dev per subject:", np.std(scores, axis=0))
```

Output

```
Mean per subject: [73.73 75.89 72.7 ]
Max per subject: [99 98 98]
Std Dev per subject: [14.62043433 14.96789564 14.17779955]
```

8 Filtering Students with High Scores

```
filtered_df = df[(df[’Math’] > 85) & (df[’Science’] > 90)]
print(filtered_df)
```

Output

```
    StudentID   Name Gender  Math  Science  English  Total
0           1  Megan      M    94       97       50    241
69         70  Kelly      F    88       97       79    264
```

9 Gender Count of High Scorers

```
high_scorers = df[df[’Total’] > 240]
gender_counts = high_scorers[’Gender’].value_counts()
print(gender_counts)
```

Output

```
Gender
F    14
M    13
Name: count, dtype: int64
```