

Report for Lab-03 (Assignment):

Course Code: CSE303

Course Title: Statistics of Data Science

Section: 01

Submitted To:

Aminul Kader Bulbul

Lecturer

Department of Computer Science & Engineering

Submitted By:

Umme Mukaddisa


ID: 2022-3-60-317

To get access to the CSV file named **students_marks** provided in the classroom, this part has been implemented initially:


```
[10] import pandas as pd
      path="/content/drive/MyDrive/csv/students_marks.csv"
```

Question-01: (Read the Dataset from CSV file)

To read the dataset from CSV file, `pd.read_csv(path)` function has been used to get access to the data set & `df.head()` function is used to read the dataset.

 #Read the DataSet from CSV|

```
df=pd.read_csv(path)
print(f'The dataset is: ')
print(df.head(1001))
```

 The dataset is:

	Student_ID	Physics	Chemistry	Biology	Mathematics	English
0	STUD0001	78	73	85	98	65
1	STUD0002	91	67	54	45	44
2	STUD0003	68	70	48	80	74
3	STUD0004	54	47	54	78	67
4	STUD0005	82	78	88	43	68
..
995	STUD0996	43	76	75	72	42
996	STUD0997	40	64	42	43	67
997	STUD0998	88	87	48	64	45
998	STUD0999	79	40	47	59	83
999	STUD1000	71	92	72	73	83

[1000 rows x 6 columns]

Question-02: (Prompt the user to choose a subject (e.g., Physics))

For subject choosing `df.columns[1:]` has been used to show the subjects available in given csv file. User can write down their desired subject which they want to which they want to calculate the statistic for in this part.

```
#Prompt user to choose a subject

print("Aavailable subjects: ")
print(df.columns[1:])
subject=input("Enter the subject: ")

#Statistics for the subject chosen

if subject in df.columns:
    data=df[subject]
    mean=data.mean()
    median=data.median()
    mode=data.mode().tolist()

... Aavailable subjects:
Index(['Physics', 'Chemistry', 'Biology', 'Mathematics', 'English'], dtype='object')
Enter the subject: 
```

Question-03: (Calculate & display the following statistics for that subject-
mean, median, mode, variance & standard deviation)

To calculate the mean, median, mode, variance & standard deviation an “If”
“Else” condition has been implemented.

If the chosen subject in previous part is available in the data frame then if part will work:

```

#Prompt user to choose a subject

print("Aavailable subjects: ")
print(df.columns[1:])
subject=input("Enter the subject: ")

#Statistics for the subject chosen

if subject in df.columns:
    data=df[subject]
    mean=data.mean()
    median=data.median()
    mode=data.mode().tolist()

```

Available subjects:

Index(['Physics', 'Chemistry', 'Biology', 'Mathematics', 'English'], dtype='object')

Enter the subject:

```

▶ import numpy as np
import matplotlib.pyplot as plt
from collections import Counter

#Statistics for the subject chosen
if subject in df.columns:
    data=df[subject]
    mean_val=np.mean(data)
    median_val=np.median(data)
    mode_val=Counter(data).most_common(1)[0][0]
    var_val=np.var(data, ddof=1)
    std_val=np.std(data, ddof=1)

    print("Mean: ",mean_val)
    print("Mrdian: ",median_val)
    print("Mode: ", mode_val)
    print("Variance: ", var_val)
    print("Standard Deviation: ", std_val)

else:
    print(f'Subject is not the dataset')

```

```

↔ Mean: 70.2
Mrdian: 71.0
Mode: 97
Variance: 301.81781781781854
Standard Deviation: 17.372904702951043

```

But if the subject is not in the data frame then the **else** part will work & will show a subject not found message.

```
▶ #Prompt user to choose a subject

print("Aavailable subjects: ")
print(df.columns[1:])
subject=input("Enter the subject: ")

#Statistics for the subject chosen

if subject in df.columns:
    data=df[subject]
    mean=data.mean()
    median=data.median()
    mode=data.mode().tolist()
```

```
↗ Aavailable subjects:
Index(['Physics', 'Chemistry', 'Biology', 'Mathematics', 'English'], dtype='object')
Enter the subject: math
```

```
▶ import numpy as np
import matplotlib.pyplot as plt
from collections import Counter

#Statistics for the subject chosen
if subject in df.columns:
    data=df[subject]
    mean_val=np.mean(data)
    median_val=np.median(data)
    mode_val=Counter(data).most_common(1)[0][0]
    var_val=np.var(data, ddof=1)
    std_val=np.std(data, ddof=1)

    print("Mean: ",mean_val)
    print("Mrdian: ",median_val)
    print("Mode: ", mode_val)
    print("Variance: ", var_val)
    print("Standard Deviation: ", std_val)

else:
    print(f'Subject is not the dataset')
```

```
↗ Subject is not the dataset
```

