

SHABARI PRAKASH SV 21C085

Import Pandas and Read/Import Multiple File Format

```
In [ ]: import pandas as pd
```

READ CSV FILE

```
In [ ]: df=pd.read_csv('samplecsv.csv')
```

```
In [ ]: print(df.head())
```

```
<bound method NDFrame.head of      Age  Gender AcademicPerformance TakingNoteInClass
DepressionStatus \
0      23    Male           Average             No      Sometimes
1      23    Male      Excellent      Sometimes             Yes
2      24    Male           Average             No      Sometimes
3      20  Female           Good             Yes      Sometimes
4      24  Female           Average            Yes             Yes
..    ...    ...             ...             ...             ...
94     21    Male           Good             Yes      Sometimes
95     21  Female           Good             Yes             Yes
96     25    Male           Good             Yes      Sometimes
97     21    Male           Good             Yes             No
98     22  Female           Average            Yes             No

      FaceChallengesToCompleteAcademicTask LikePresentation SleepPerDayHours \
0                                     Yes             Yes             12
1                                     No             Yes             8
2                               Sometimes             No             8
3                                     Yes             No             5
4                                     Yes             Yes             5
..                                ...             ...             ...
94                                     Yes             Yes             6
95                                     No             Yes             6
96                                     No             Yes             4
97                               Sometimes             No            10
98                                     No             Yes             7

      NumberOfFriend LikeNewThings
0                NaN           Yes
1              80.0           Yes
2              10.0           Yes
3              15.0           Yes
4               2.0           Yes
..              ...             ...
94               4.0           Yes
95               6.0           Yes
96               9.0            No
97              23.0           Yes
98               3.0           Yes
```

[99 rows x 10 columns]>

## READ JSON FILE

```
In [ ]: jsondataset=pd.read_json('./samplejson.json')
```

```
In [ ]: print(jsondataset.head)
```

```
<bound method NDFrame.head of
DepressionStatus \
0    23    Male    Average    No    Sometimes
1    23    Male    Excellent    Sometimes    Yes
2    24    Male    Average    No    Sometimes
3    20    Female    Good    Yes    Sometimes
4    24    Female    Average    Yes    Yes
..    ...    ...    ...    ...    ...
94   21    Male    Good    Yes    Sometimes
95   21    Female    Good    Yes    Yes
96   25    Male    Good    Yes    Sometimes
97   21    Male    Good    Yes    No
98   22    Female    Average    Yes    No
```

```
FaceChallengesToCompleteAcademicTask LikePresentation SleepPerDayHours \
0    Yes    Yes    12
1    No    Yes    8
2    Sometimes    No    8
3    Yes    No    5
4    Yes    Yes    5
..    ...    ...    ...
94    Yes    Yes    6
95    No    Yes    6
96    No    Yes    4
97    Sometimes    No    10
98    No    Yes    7
```

```
NumberOfFriend LikeNewThings
0    Yes
1    80    Yes
2    10    Yes
3    15    Yes
4    2    Yes
..    ...    ...
94    4    Yes
95    6    Yes
96    9    No
97    23    Yes
98    3    Yes
```

```
[99 rows x 10 columns]>
```

## READ EXCEL FILE

```
In [ ]: xlsdata=pd.read_excel('./sampleexcel.xlsx')
```

```
In [ ]: print(xlsdata.head)
```

```

<bound method NDFrame.head of
DepressionStatus \
0      23      Male      Average      No      Sometimes
1      23      Male      Excellent      Sometimes      Yes
2      24      Male      Average      No      Sometimes
3      20      Female      Good      Yes      Sometimes
4      24      Female      Average      Yes      Yes
..      ...      ...      ...      ...      ...
94     21      Male      Good      Yes      Sometimes
95     21      Female      Good      Yes      Yes
96     25      Male      Good      Yes      Sometimes
97     21      Male      Good      Yes      No
98     22      Female      Average      Yes      No

```

```

FaceChallengesToCompleteAcademicTask LikePresentation SleepPerDayHours \
0      Yes      Yes      12
1      No      Yes      8
2      Sometimes      No      8
3      Yes      No      5
4      Yes      Yes      5
..      ...      ...      ...
94     Yes      Yes      6
95     No      Yes      6
96     No      Yes      4
97     Sometimes      No      10
98     No      Yes      7

```

```

NumberOfFriend LikeNewThings
0      NaN      Yes
1      80.0      Yes
2      10.0      Yes
3      15.0      Yes
4      2.0      Yes
..      ...      ...
94     4.0      Yes
95     6.0      Yes
96     9.0      No
97     23.0      Yes
98     3.0      Yes

```

[99 rows x 10 columns]>

CONVERT TO PANDAS DATAFRAME

```

In [ ]: df=pd.DataFrame(df)
df

```

Out[ ]:

	Age	Gender	AcademicPerformance	TakingNoteInClass	DepressionStatus	FaceChallan
0	23	Male	Average	No	Sometimes	
1	23	Male	Excellent	Sometimes	Yes	
2	24	Male	Average	No	Sometimes	
3	20	Female	Good	Yes	Sometimes	
4	24	Female	Average	Yes	Yes	
...	...	...	...	...	...	...
94	21	Male	Good	Yes	Sometimes	
95	21	Female	Good	Yes	Yes	
96	25	Male	Good	Yes	Sometimes	
97	21	Male	Good	Yes	No	
98	22	Female	Average	Yes	No	

99 rows × 10 columns



CHECK FOR NULL VALUES

```
In [ ]: df.isnull()
```

Out[ ]:

	Age	Gender	AcademicPerformance	TakingNoteInClass	DepressionStatus	FaceChallan
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	False	False	False	False	
...	...	...	...	...	...	...
94	False	False	False	False	False	
95	False	False	False	False	False	
96	False	False	False	False	False	
97	False	False	False	False	False	
98	False	False	False	False	False	

99 rows × 10 columns



## SHOW NULL VALUES IN EACH ATTRIBUTE

```
In [ ]: df.isnull().sum()
```

```
Out[ ]: Age                                0
        Gender                             0
        AcademicPerformance                 0
        TakingNoteInClass                   0
        DepressionStatus                    0
        FaceChallangesToCompleteAcademicTask 0
        LikePresentation                    0
        SleepPerDayHours                    0
        NumberOfFriend                       4
        LikeNewThings                        0
        dtype: int64
```

## FOR FILLING THE NULL VALUES , FIND MEAN

```
In [ ]: nooffmean=df['NumberOfFriend'].mean()
        nooffmean
```

```
Out[ ]: 16.189473684210526
```

## FILL NULL VALUES WITH MEAN VALUE

```
In [ ]: df.fillna({'NumberOfFriend':nooffmean},inplace=True)
```

## NOW CHECK FOR NULL VALUES

```
In [ ]: df.isnull().sum()
```

```
Out[ ]: Age                                0
        Gender                             0
        AcademicPerformance                 0
        TakingNoteInClass                   0
        DepressionStatus                    0
        FaceChallangesToCompleteAcademicTask 0
        LikePresentation                    0
        SleepPerDayHours                    0
        NumberOfFriend                       0
        LikeNewThings                        0
        dtype: int64
```

## EXPORT THE PROCESSED DATAFRAME TO OTHER FORMATS - CSV,JSON,EXCEL,HDF

```
In [ ]: df.to_csv('./ProcessedDataset/ExportCSV.csv',index=True)
```

```
In [ ]: df.to_json('./ProcessedDataset/ExportJSON.json')
```

```
In [ ]: df.to_excel('./ProcessedDataset/ExportXSL.xlsx')
```

```
In [ ]: df.to_hdf('./ProcessedDataset/ExportHDF.h5',key='df',mode='a')
```

