include library

```
import pandas as pd
import numpy as np
```

How We Can Create Series

```
temp=[4,5,6,78,2]
ser=pd.Series(temp)
print(f'The Table is {ser}')
print(f'The Data Type of The Df is {type(ser)}')
```

How We Can Create DataFrame

```
data={
    "Name":["Malek","Ali","Ahmed"],
    "Age":[23,25,28]
}
df=pd.DataFrame(data)
print(f'The Table is {df}')
print(f'The Data Type of The Df is {type(df)}')
# df.iloc[0]
# df.iloc[0,1]
```

```
The Table is     Name  Age
0  Malek   23
1    Ali   25
2  Ahmed   28
The Data Type of The Df is <class 'pandas.core.frame.DataFrame'>
```

Read Data From JSON File

```
myJsonFile=pd.read_json('output.json')
myJsonFile.head()
```

```
    Name  Age      City
0   John   25  New York
1  Alice   30    London
2    Bob   35     Paris
```

Read Data From XML File

```
myXMLFile=pd.read_xml('output.xml')
myJsonFile.head()
```

```
    Name  Age      City
0   John   25  New York
1  Alice   30    London
2    Bob   35     Paris
```

Read Data From Text File

```
myTxtFile = pd.read_fwf('data.txt')
print(myTxtFile)
```

```
    John  25  170
0  Alice  28  165
1    Bob  30  180
```

include file

Read Data From CSV File

```
dataset=pd.read_csv('students.csv')
print(dataset)
```

```
          sex race_ethnicity parental_level_of_education         lunch
\
0      female        group B           bachelor's degree      standard

1      female        group C                some college      standard

2      female            NaN             master's degree      standard

3        male        group A          associate's degree  free/reduced

4        male        group C                         NaN      standard

...       ...            ...                         ...           ...

20999  female        group E             master's degree      standard

21000    male        group C                 high school  free/reduced

21001  female        group C                 high school  free/reduced

21002  female        group D                some college      standard
```

```
21003  female        group D                some college  free/reduced


      test_preparation_course  math digree  reading digree  writing
score
0                        none           72            72.0
74
1                   completed           69            90.0
88
2                        none           90            95.0
93
3                        none           47            57.0
44
4                        none           76            78.0
75
...                       ...          ...             ...
...
20999               completed           58            99.0
95
21000                    none           62            55.0
55
21001               completed           49            71.0
65
21002               completed           68            48.0
77
21003                    none           77            89.0
86

[21004 rows x 8 columns]
```

Know Your Data

Get The Shape

```
dataset.shape
```

```
(21004, 8)
```

Get The Head Of The Data Set

```
dataset.head()
```

```
      sex race_ethnicity parental_level_of_education         lunch  \
0  female        group B           bachelor's degree      standard
1  female        group C                some college      standard
2  female            NaN             master's degree      standard
3    male        group A          associate's degree  free/reduced
```

```
4    male        group C                         NaN      standard

  test_preparation_course  math digree  reading digree  writing score

0                    none           72            72.0             74

1               completed           69            90.0             88

2                    none           90            95.0             93

3                    none           47            57.0             44

4                    none           76            78.0             75
```

```
dataset.head(20)
```

```
       sex race_ethnicity parental_level_of_education         lunch  \
0   female        group B           bachelor's degree      standard
1   female        group C                some college      standard
2   female            NaN             master's degree      standard
3     male        group A          associate's degree  free/reduced
4     male        group C                         NaN      standard
5   female        group B          associate's degree      standard
6   female        group B                some college      standard
7     male        group B                some college  free/reduced
8     male        group D                 high school  free/reduced
9   female        group B                 high school  free/reduced
10    male        group C          associate's degree      standard
11    male        group D          associate's degree      standard
12  female        group B                 high school      standard
13    male        group A                some college      standard
14  female        group A             master's degree      standard
15  female        group C           some high school      standard
16    male        group C                 high school      standard
17  female        group B           some high school  free/reduced
18    male        group C             master's degree  free/reduced
19  female        group C          associate's degree  free/reduced

   test_preparation_course  math digree  reading digree  writing score

0                     none           72            72.0             74

1                completed           69            90.0             88

2                     none           90            95.0             93

3                     none           47            57.0             44

4                     none           76            78.0             75

5                     none           71            83.0             78
```

```
6                completed           88            95.0            92
7                     none           40            43.0            39
8                completed           64            64.0            67
9                     none           38            60.0            50
10                    none           58            54.0            52
11                    none           40            52.0            43
12                    none           65            81.0            73
13               completed           78            72.0            70
14                     NaN           50            53.0            58
15                    none           69            75.0            78
16                    none           88             NaN            86
17                    none           18            32.0            28
18               completed           46            42.0            46
19                    none           54            58.0            61
```

Get The Data The Tail of The DataSet

```
dataset.tail()
```

```
          sex race_ethnicity parental_level_of_education         lunch
\
20999  female        group E             master's degree      standard
21000    male        group C                 high school  free/reduced
21001  female        group C                 high school  free/reduced
21002  female        group D               some college      standard
21003  female        group D               some college  free/reduced


      test_preparation_course  math digree  reading digree  writing
score
20999               completed           58            99.0
95
21000                    none           62            55.0
```

```
55
21001                completed            49             71.0
65
21002                completed            68             48.0
77
21003                     none            77             89.0
86

dataset.tail(20)
```

Get The Columns Name

```
dataset.columns

Index(['sex', 'race_ethnicity', 'parental_level_of_education',
'lunch',
       'test_preparation_course', 'math digree', 'reading digree',
       'writing score'],
      dtype='object')
```

Get The DataSet Information

```
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21004 entries, 0 to 21003
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   sex                          21002 non-null  object
 1   race_ethnicity               20962 non-null  object
 2   parental_level_of_education  20962 non-null  object
 3   lunch                        21004 non-null  object
 4   test_preparation_course      20983 non-null  object
 5   math digree                  21004 non-null  int64
 6   reading digree               20983 non-null  float64
 7   writing score                21004 non-null  int64
dtypes: float64(1), int64(2), object(5)
memory usage: 1.3+ MB
```

Get The DataSet Describe

```
dataset.describe()

        math digree  reading digree  writing score
count  21004.000000    20983.000000   21004.000000
mean      66.105408       69.151027      68.068511
std       15.169625       14.598066      15.190079
min        0.000000       17.000000      10.000000
```

```
25%       57.000000       59.000000       58.000000
50%       66.000000       70.000000       69.000000
75%       77.000000       79.000000       79.000000
max      100.000000      100.000000      100.000000
```

Select Columns From Table With DataType

```
dataset.select_dtypes('int64').columns

Index(['math digree', 'writing score'], dtype='object')
```

Get All unique Value From The Column

```
dataset.sex.unique()

array(['female', 'male', 'انثی','ذكر', nan], dtype=object)

dataset.sex.value_counts()

sex
female    10875
male      10119
6          انثی
2          ذكر
Name: count, dtype: int64

dataset.lunch.unique()

array(['standard', 'free/reduced'], dtype=object)

dataset.lunch.value_counts()

lunch
standard        13549
free/reduced     7455
Name: count, dtype: int64
```

ReName The Column Names

```
dataset.rename(columns={'math digree':'math_digree'}, inplace=True)
dataset.rename(columns={'reading digree':'reading_digree'},
inplace=True)
dataset.rename(columns={'writing score':'writing_score'},
inplace=True)

dataset

          sex race_ethnicity parental_level_of_education         lunch
\
```

```
0      female        group B           bachelor's degree      standard
1      female        group C                some college      standard
2      female            NaN             master's degree      standard
3        male        group A          associate's degree  free/reduced
4        male        group C                         NaN      standard
...       ...            ...                         ...           ...
20999  female        group E             master's degree      standard
21000    male        group C                high school  free/reduced
21001  female        group C                high school  free/reduced
21002  female        group D                some college      standard
21003  female        group D                some college  free/reduced

      test_preparation_course  math_digree  reading_digree
writing_score
0                        none           72            72.0
74
1                   completed           69            90.0
88
2                        none           90            95.0
93
3                        none           47            57.0
44
4                        none           76            78.0
75
...                       ...          ...             ...
...
20999               completed           58            99.0
95
21000                    none           62            55.0
55
21001               completed           49            71.0
65
21002               completed           68            48.0
77
21003                    none           77            89.0
86

[21004 rows x 8 columns]
```

Data Cleaning

Missing Vlaues

Get The Number of Null Value

```
dataset.isnull().sum()
```

```
sex                           2
race_ethnicity               42
parental_level_of_education  42
lunch                         0
test_preparation_course      21
math digree                   0
reading digree               21
writing score                 0
dtype: int64
```

Drop Rows With Missing Values

```
# define a dictionary with sample data which includes some missing
values
data = {
    'A': [1, 2, 3, None, 5],
    'B': [None, 2, 3, 4, 5],
    'C': [1, 2, None, None, 5]
}

df = pd.DataFrame(data)
print("Original Data:\n",df)
print()

# use dropna() to remove rows with any missing values
df_cleaned = df.dropna()

print("Cleaned Data:\n",df_cleaned)
```

```
Original Data:
     A    B    C
0  1.0  NaN  1.0
1  2.0  2.0  2.0
2  3.0  3.0  NaN
3  NaN  4.0  NaN
4  5.0  5.0  5.0

Cleaned Data:
     A    B    C
```

```
1  2.0  2.0  2.0
4  5.0  5.0  5.0

myFirestCopy=dataset.copy()
myFirestCopy

          sex race_ethnicity parental_level_of_education         lunch
\
0      female        group B           bachelor's degree      standard

1      female        group C                some college      standard

2      female            NaN             master's degree      standard

3        male        group A          associate's degree  free/reduced

4        male        group C                         NaN      standard

...       ...            ...                         ...           ...

20999  female        group E             master's degree      standard

21000    male        group C                 high school  free/reduced

21001  female        group C                 high school  free/reduced

21002  female        group D                some college      standard

21003  female        group D                some college  free/reduced


      test_preparation_course  math digree  reading digree  writing
score
0                        none           72            72.0
74
1                   completed           69            90.0
88
2                        none           90            95.0
93
3                        none           47            57.0
44
4                        none           76            78.0
75
...                       ...          ...             ...
...
20999               completed           58            99.0
95
21000                    none           62            55.0
55
21001               completed           49            71.0
65
```

```
21002                completed           68             48.0
77
21003                     none           77             89.0
86

[21004 rows x 8 columns]
```

```
df_cleaned = myFirestCopy.dropna()
print("Cleaned Data:\n",df_cleaned)
```

```
Cleaned Data:
           sex race_ethnicity parental_level_of_education
lunch  \
0      female        group B           bachelor's degree      standard

1      female        group C                some college      standard

3        male        group A           associate's degree  free/reduced

5      female        group B           associate's degree      standard

6      female        group B                some college      standard

...       ...            ...                         ...           ...

20999  female        group E             master's degree      standard

21000    male        group C                high school  free/reduced

21001  female        group C                high school  free/reduced

21002  female        group D                some college      standard

21003  female        group D                some college  free/reduced


      test_preparation_course  math digree  reading digree  writing
score
0                        none           72             72.0
74
1                   completed           69             90.0
88
3                        none           47             57.0
44
5                        none           71             83.0
78
6                   completed           88             95.0
92
...                       ...          ...              ...
...
20999               completed           58             99.0
```

```
95
21000                      none              62              55.0
55
21001                 completed              49              71.0
65
21002                 completed              68              48.0
77
21003                      none              77              89.0
86

[20897 rows x 8 columns]
```

```
df_cleaned.isnull().sum()
```

```
sex                            0
race_ethnicity                 0
parental_level_of_education    0
lunch                          0
test_preparation_course        0
math digree                    0
reading digree                 0
writing score                  0
dtype: int64
```

Fill Missing Values With Const Value

```
# define a dictionary with sample data which includes some missing
values
data = {
    'A': [1, 2, 3, None, 5],
    'B': [None, 2, 3, 4, 5],
    'C': [1, 2, None, None, 5]
}
df = pd.DataFrame(data)
print("Original Data:\n", df)
# filling NaN values with 0
df.fillna(0, inplace=True)
print("\nData after filling NaN with 0:\n", df)
```

```
Original Data:
      A    B    C
0  1.0  NaN  1.0
1  2.0  2.0  2.0
2  3.0  3.0  NaN
3  NaN  4.0  NaN
4  5.0  5.0  5.0

Data after filling NaN with 0:
      A    B    C
0  1.0  0.0  1.0
```

```
1  2.0  2.0  2.0
2  3.0  3.0  0.0
3  0.0  4.0  0.0
4  5.0  5.0  5.0
```

```
dataset.isnull().sum()
```

```
sex                           2
race_ethnicity               42
parental_level_of_education  42
lunch                         0
test_preparation_course      21
math digree                   0
reading digree               21
writing score                 0
dtype: int64
```

```
mis=dataset[dataset['reading digree'].isna()]
mis
```

```
       sex race_ethnicity parental_level_of_education     lunch  \
16     male        group C                 high school  standard
1016   male        group C                 high school  standard
2016   male        group C                 high school  standard
3016   male        group C                 high school  standard
4016   male        group C                 high school  standard
5016   male        group C                 high school  standard
6016   male        group C                 high school  standard
7016   male        group C                 high school  standard
8016   male        group C                 high school  standard
9016   male        group C                 high school  standard
10016  male        group C                 high school  standard
11016  male        group C                 high school  standard
12016  male        group C                 high school  standard
13016  male        group C                 high school  standard
14016  male        group C                 high school  standard
15016  male        group C                 high school  standard
16016  male        group C                 high school  standard
17016  male        group C                 high school  standard
18016  male        group C                 high school  standard
19018  male        group C                 high school  standard
20018  male        group C                 high school  standard

      test_preparation_course  math digree  reading digree  writing
score
16                       none           88             NaN
86
1016                     none           88             NaN
86
2016                     none           88             NaN
```

```
86
3016                        none            88              NaN
86
4016                        none            88              NaN
86
5016                        none            88              NaN
86
6016                        none            88              NaN
86
7016                        none            88              NaN
86
8016                        none            88              NaN
86
9016                        none            88              NaN
86
10016                       none            88              NaN
86
11016                       none            88              NaN
86
12016                       none            88              NaN
86
13016                       none            88              NaN
86
14016                       none            88              NaN
86
15016                       none            88              NaN
86
16016                       none            88              NaN
86
17016                       none            88              NaN
86
18016                       none            88              NaN
86
19018                       none            88              NaN
86
20018                       none            88              NaN
86
```

```
mySecondCopy=dataset.copy()
mySecondCopy.loc[:, "reading digree"] = mySecondCopy["reading
digree"].fillna(20.0)
mySecondCopy.isnull().sum()
```

```
sex                           2
race_ethnicity               42
parental_level_of_education  42
lunch                         0
test_preparation_course      21
math digree                   0
reading digree                0
```

```
writing score                   0
dtype: int64
```

```
mis=mySecondCopy[mySecondCopy['reading digree']==20.0]
mis
```

```
        sex race_ethnicity parental_level_of_education     lunch  \
16     male        group C                 high school  standard
1016   male        group C                 high school  standard
2016   male        group C                 high school  standard
3016   male        group C                 high school  standard
4016   male        group C                 high school  standard
5016   male        group C                 high school  standard
6016   male        group C                 high school  standard
7016   male        group C                 high school  standard
8016   male        group C                 high school  standard
9016   male        group C                 high school  standard
10016  male        group C                 high school  standard
11016  male        group C                 high school  standard
12016  male        group C                 high school  standard
13016  male        group C                 high school  standard
14016  male        group C                 high school  standard
15016  male        group C                 high school  standard
16016  male        group C                 high school  standard
17016  male        group C                 high school  standard
18016  male        group C                 high school  standard
19018  male        group C                 high school  standard
20018  male        group C                 high school  standard

      test_preparation_course  math digree  reading digree  writing
score
16                       none           88            20.0
86
1016                     none           88            20.0
86
2016                     none           88            20.0
86
3016                     none           88            20.0
86
4016                     none           88            20.0
86
5016                     none           88            20.0
86
6016                     none           88            20.0
86
7016                     none           88            20.0
86
8016                     none           88            20.0
86
9016                     none           88            20.0
```

```
86
10016                    none           88            20.0
86
11016                    none           88            20.0
86
12016                    none           88            20.0
86
13016                    none           88            20.0
86
14016                    none           88            20.0
86
15016                    none           88            20.0
86
16016                    none           88            20.0
86
17016                    none           88            20.0
86
18016                    none           88            20.0
86
19018                    none           88            20.0
86
20018                    none           88            20.0
86
```

```
mySecondCopy.iloc[16,1]='group A'
mySecondCopy.head(20)
```

```
       sex race_ethnicity parental_level_of_education         lunch  \
0   female        group B           bachelor's degree      standard
1   female        group C                some college      standard
2   female            NaN             master's degree      standard
3     male        group A          associate's degree  free/reduced
4     male        group C                         NaN      standard
5   female        group B          associate's degree      standard
6   female        group B                some college      standard
7     male        group B                some college  free/reduced
8     male        group D                 high school  free/reduced
9   female        group B                 high school  free/reduced
10    male        group C          associate's degree      standard
11    male        group D          associate's degree      standard
12  female        group B                 high school      standard
13    male        group A                some college      standard
14  female        group A             master's degree      standard
15  female        group C           some high school      standard
16    male        group A                 high school      standard
17  female        group B           some high school  free/reduced
18    male        group C             master's degree  free/reduced
19  female        group C          associate's degree  free/reduced

   test_preparation_course  math digree  reading digree  writing score
```

```
0                    none          72          72.0          74
1               completed          69          90.0          88
2                    none          90          95.0          93
3                    none          47          57.0          44
4                    none          76          78.0          75
5                    none          71          83.0          78
6               completed          88          95.0          92
7                    none          40          43.0          39
8               completed          64          64.0          67
9                    none          38          60.0          50
10                   none          58          54.0          52
11                   none          40          52.0          43
12                   none          65          81.0          73
13              completed          78          72.0          70
14                    NaN          50          53.0          58
15                   none          69          75.0          78
16                   none          88          20.0          86
17                   none          18          32.0          28
18              completed          46          42.0          46
19                   none          54          58.0          61
```

Replace Missing Values With Mean, Median and Mode

```
data = {
    'A': [1, 2, np.nan, 4, 5],
    'B': [np.nan, 2, 3, 4, 5],
    'C': [1, 2, 3, np.nan, 5],
    'D': [1, 2, 3, 4, 5]
}
df = pd.DataFrame(data)
df_copy=df.copy()
```

```
print(f'The Data Before \n{df}')
print('----------------------------------')
# replace missing values with mean
# df_copy['A']=df['A'].fillna(value=df['A'].mean() )

# # replace missing values with median
# df_copy['B']=df['B'].fillna(value=df['B'].median() )

# # replace missing values with mode
# df_copy['C']=df['C'].fillna(value=df['C'].mode()[0] )

# print(df_copy)
```

```
The Data Before
     A    B    C  D
0  1.0  NaN  1.0  1
1  2.0  2.0  2.0  2
2  NaN  3.0  3.0  3
3  4.0  4.0  NaN  4
4  5.0  5.0  5.0  5
----------------------------------
```

```
dataset.describe()
```

```
        math digree  reading digree  writing score
count  21004.000000    20983.000000   21004.000000
mean      66.105408       69.151027      68.068511
std       15.169625       14.598066      15.190079
min        0.000000       17.000000      10.000000
25%       57.000000       59.000000      58.000000
50%       66.000000       70.000000      69.000000
75%       77.000000       79.000000      79.000000
max      100.000000      100.000000     100.000000
```

```
mis=dataset[dataset["reading digree"].isna()]
mis
```

```
        sex race_ethnicity parental_level_of_education     lunch  \
16     male        group C                 high school  standard
1016   male        group C                 high school  standard
2016   male        group C                 high school  standard
3016   male        group C                 high school  standard
4016   male        group C                 high school  standard
5016   male        group C                 high school  standard
6016   male        group C                 high school  standard
7016   male        group C                 high school  standard
8016   male        group C                 high school  standard
9016   male        group C                 high school  standard
10016  male        group C                 high school  standard
11016  male        group C                 high school  standard
12016  male        group C                 high school  standard
```

```
13016  male        group C                high school  standard
14016  male        group C                high school  standard
15016  male        group C                high school  standard
16016  male        group C                high school  standard
17016  male        group C                high school  standard
18016  male        group C                high school  standard
19018  male        group C                high school  standard
20018  male        group C                high school  standard

      test_preparation_course  math digree  reading digree  writing
score
16                       none           88             NaN
86
1016                     none           88             NaN
86
2016                     none           88             NaN
86
3016                     none           88             NaN
86
4016                     none           88             NaN
86
5016                     none           88             NaN
86
6016                     none           88             NaN
86
7016                     none           88             NaN
86
8016                     none           88             NaN
86
9016                     none           88             NaN
86
10016                    none           88             NaN
86
11016                    none           88             NaN
86
12016                    none           88             NaN
86
13016                    none           88             NaN
86
14016                    none           88             NaN
86
15016                    none           88             NaN
86
16016                    none           88             NaN
86
17016                    none           88             NaN
86
18016                    none           88             NaN
86
```

```
19018                    none           88              NaN
86
20018                    none           88              NaN
86

mis=dataset.copy()
mis["reading digree"]=dataset["reading
digree"].fillna(value=dataset["reading digree"].min())
mis.head(20)

       sex race_ethnicity parental_level_of_education         lunch  \
0   female        group B           bachelor's degree      standard
1   female        group C                some college      standard
2   female            NaN             master's degree      standard
3     male        group A          associate's degree  free/reduced
4     male        group C                         NaN      standard
5   female        group B          associate's degree      standard
6   female        group B                some college      standard
7     male        group B                some college  free/reduced
8     male        group D                 high school  free/reduced
9   female        group B                 high school  free/reduced
10    male        group C          associate's degree      standard
11    male        group D          associate's degree      standard
12  female        group B                 high school      standard
13    male        group A                some college      standard
14  female        group A             master's degree      standard
15  female        group C           some high school      standard
16    male        group C                 high school      standard
17  female        group B           some high school  free/reduced
18    male        group C             master's degree  free/reduced
19  female        group C          associate's degree  free/reduced

   test_preparation_course  math digree  reading digree  writing score

0                     none           72            72.0             74

1                completed           69            90.0             88

2                     none           90            95.0             93

3                     none           47            57.0             44

4                     none           76            78.0             75

5                     none           71            83.0             78

6                completed           88            95.0             92

7                     none           40            43.0             39

8                completed           64            64.0             67
```

```
9                    none           38            60.0            50
10                   none           58            54.0            52
11                   none           40            52.0            43
12                   none           65            81.0            73
13              completed           78            72.0            70
14                    NaN           50            53.0            58
15                   none           69            75.0            78
16                   none           88            17.0            86
17                   none           18            32.0            28
18              completed           46            42.0            46
19                   none           54            58.0            61
```

Handle Duplicates Values

```
dataset

          sex race_ethnicity parental_level_of_education         lunch
\
0      female        group B           bachelor's degree      standard
1      female        group C                some college      standard
2      female            NaN             master's degree      standard
3        male        group A          associate's degree  free/reduced
4        male        group C                         NaN      standard
...       ...            ...                         ...           ...
20999  female        group E             master's degree      standard
21000    male        group C                 high school  free/reduced
21001  female        group C                 high school  free/reduced
21002  female        group D                some college      standard
21003  female        group D                some college  free/reduced
```

```
      test_preparation_course  math digree  reading digree  writing
score
0                        none           72            72.0
74
1                   completed           69            90.0
88
2                        none           90            95.0
93
3                        none           47            57.0
44
4                        none           76            78.0
75
...                       ...          ...             ...
...
20999               completed           58            99.0
95
21000                    none           62            55.0
55
21001               completed           49            71.0
65
21002               completed           68            48.0
77
21003                    none           77            89.0
86

[21004 rows x 8 columns]
```

```
dup=dataset[dataset.duplicated()]
dup
```

```
          sex race_ethnicity parental_level_of_education         lunch
\
1000   female        group B           bachelor's degree      standard

1001   female        group C                some college      standard

1002   female            NaN             master's degree      standard

1003     male        group A          associate's degree  free/reduced

1004     male        group C                         NaN      standard

...       ...            ...                         ...           ...

20984    male        group B            some high school      standard

20986    انثی     group A                some college      standard
20987    انثی     group A                some college      standard
20993  female        group D                some college  free/reduced
```

```
21000    male        group C                  high school  free/reduced

      test_preparation_course  math digree  reading digree  writing
score
1000                     none           72            72.0
74
1001                completed           69            90.0
88
1002                     none           90            95.0
93
1003                     none           47            57.0
44
1004                     none           76            78.0
75
...                       ...          ...             ...
...
20984               completed           79            85.0
86
20986               completed           77            87.0
91
20987               completed           77            87.0
91
20993               completed           67            86.0
83
21000                    none           62            55.0
55

[19858 rows x 8 columns]
```

```
dropDu=dataset.drop_duplicates()
dropDu
```

```
          sex race_ethnicity parental_level_of_education         lunch
\
0      female        group B           bachelor's degree      standard

1      female        group C                some college      standard

2      female            NaN             master's degree      standard

3        male        group A          associate's degree  free/reduced

4        male        group C                         NaN      standard

...       ...            ...                         ...           ...

20998    male        group A                 high school      standard

20999  female        group E             master's degree      standard
```

```
21001  female        group C                 high school  free/reduced

21002  female        group D               some college      standard

21003  female        group D               some college  free/reduced


      test_preparation_course  math digree  reading digree  writing
score
0                        none           72            72.0
74
1                   completed           69            90.0
88
2                        none           90            95.0
93
3                        none           47            57.0
44
4                        none           76            78.0
75
...                       ...          ...             ...
...
20998                    none           63            63.0
65
20999               completed           58            99.0
95
21001               completed           49            71.0
65
21002               completed           68            48.0
77
21003                    none           77            89.0
86

[1146 rows x 8 columns]
```

Pandas Handling Wrong Format

```
# create dataframe
data = {
    'Country': ['USA', 'Canada', 'Australia', 'Germany', 'Japan'],
    'Date': ['2023-07-20', '2023-07-21', '2023-07-22', '2023-07-23',
'2023-07-24'],
    'Temperature': [25.5, '28.0', 30.2, 22.8, 26.3]
}
df = pd.DataFrame(data)

print(df.Temperature.unique())
print('-------------------------')
# convert temperature column to float
df['Temperature'] = df['Temperature'].astype(float)
```

```
print(df.Temperature.unique())
print('-------------------------')
print(df.info())
print('-------------------------')
# calculate the mean temperature
mean_temperature = df['Temperature'].mean()

print(mean_temperature)
```

```
[25.5 '28.0' 30.2 22.8 26.3]
-------------------------
[25.5 28.  30.2 22.8 26.3]
-------------------------
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 3 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Country      5 non-null      object
 1   Date         5 non-null      object
 2   Temperature  5 non-null      float64
dtypes: float64(1), object(2)
memory usage: 252.0+ bytes
None
-------------------------
26.560000000000002
```

```
# create a sample dataframe with mixed date formats
df = pd.DataFrame({'date': ['2022-12-01', '01/02/2022', '2022-03-23',
'03/02/2022', '3 4 2023', '2023.9.30']})

# convert the date column to datetime format
df['date'] = pd.to_datetime(df['date'], format='mixed', dayfirst=True)

print(df)
```

```
        date
0 2022-12-01
1 2022-01-02
2 2022-03-23
3 2022-03-02
4 2023-03-04
5 2023-09-30
```

Pandas Handling Wrong Data

```
data = {
    'Name': ['John', 'Michael', 'Tom', 'Alex', 'Ryan'],
    'Age': [8, 9, 7, 80, 100],
    'Gender': ['M', 'M', 'M', 'F', 'M'],
```

```
    'Standard': [3, 4, 12, 3, 5]
}
df = pd.DataFrame(data)

# replace F with M
df.loc[3, 'Gender'] = 'M'

print(df)
```

```
      Name  Age Gender  Standard
0     John    8      M         3
1  Michael    9      M         4
2      Tom    7      M        12
3     Alex   80      M         3
4     Ryan  100      M         5
```

```
data = {
    'Name': ['John', 'Michael', 'Tom', 'Alex', 'Ryan'],
    'Age': [8, 9, 7, 80, 100],
    'Gender': ['M', 'M', 'M', 'M', 'M'],
    'Standard': [3, 4, 12, 3, 5]
}
df = pd.DataFrame(data)

# replace values based on conditions
for i  in df.index:
    age_val = df.loc[i, 'Age']
    if (age_val > 14) and (age_val%10 == 0):
        df.loc[i, 'Age'] = age_val/10

print(df)
```

```
      Name  Age Gender  Standard
0     John    8      M         3
1  Michael    9      M         4
2      Tom    7      M        12
3     Alex    8      M         3
4     Ryan   10      M         5
```

```
len(dataset)
```

```
21004
```

```
dataset.sex.unique()
```

```
array(['female', 'male', 'انثی','ذکر', nan], dtype=object)
```

```
for i in range(0,len(dataset)):
    if dataset.loc[i,'sex']=='انثی':
         dataset.loc[i,'sex']='female'
```

```
dataSex=dataset[dataset['sex']=='انثی']
dataSex
```

```
      sex race_ethnicity parental_level_of_education         lunch  \
2081  ذكر     group B                 high school  free/reduced
2091  ذكر     group C                 high school  free/reduced

     test_preparation_course  math digree  reading digree  writing
score  \
2081                    none           49            45.0
45
2091                    none           27            34.0
36

       18728   18729   18731   20985   20986   20987
2081  female  female  female  female  female  female
2091  female  female  female  female  female  female
```

```
dataSex=dataset[dataset['sex']=='ذكر']
dataSex
```

```
Empty DataFrame
Columns: [sex, race_ethnicity, parental_level_of_education, lunch,
test_preparation_course, math digree, reading digree, writing score,
18728, 18729, 18731, 20985, 20986, 20987]
Index: []
```

```
for i in range(0,len(dataset)):
    if dataset.loc[i,'sex']=='ذكر':
         dataset.loc[i,'sex']='Male'
```

Write To New DataSets

```
dataset.to_csv('datanew2.csv')
```