# Assignment No 1

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



from google.colab import files
uploaded=files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving StudentsPerformance.csv to StudentsPerformance.csv
```

```
df=pd.read_csv(io.BytesIO(uploaded['StudentsPerformance.csv']))
```

```
print(df)
```

```
     gender race/ethnicity parental level of education         lunch  \
0       NaN        group B           bachelor's degree      standard
1    female        group C                some college      standard
2    female        group B            master's degree      standard
3      male        group A          associate's degree  free/reduced
4      male        group C                some college      standard
..      ...            ...                         ...           ...
995  female        group E            master's degree      standard
996    male        group C                 high school  free/reduced
997  female        group C                 high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

    test preparation course  math score  reading score  writing score
0                      none       120.0             72             74
1                 completed       150.0             90             88
2                      none         NaN             95             93
3                      none         NaN             57             44
4                      none         NaN             78             75
..                      ...         ...            ...            ...
995               completed        88.0             99             95
996                    none        62.0             55             55
997               completed        59.0             71             65
998               completed        68.0             78             77
999                    none        77.0             86             86

[1000 rows x 8 columns]
```

```
df.describe()
```

```
       math score  reading score  writing score
count  997.000000    1000.000000    1000.000000
```

```
mean    66.203611      69.169000      68.054000
std     15.475906      14.600192      15.195657
min      0.000000      17.000000      10.000000
25%     57.000000      59.000000      57.750000
50%     66.000000      70.000000      69.000000
75%     77.000000      79.000000      79.000000
max    150.000000     100.000000     100.000000
```

```
print("dim",df.ndim)
```

```
dim 2
```

```
print("dim",df.size)
print("dim",df.shape)
```

```
dim 8000
dim (1000, 8)
```

```
print(df['gender'])
```

```
0         NaN
1      female
2      female
3        male
4        male
        ...
995    female
996      male
997    female
998    female
999    female
Name: gender, Length: 1000, dtype: object
```

```
df.gender.str.isdigit()
```

```
0        NaN
1      False
2      False
3      False
4      False
       ...
995    False
996    False
997    False
998    False
999    False
Name: gender, Length: 1000, dtype: object
```

```
df.gender.str.isalpha()
```

```
0       NaN
1      True
2      True
```

```
3      True
4      True
       ...
995    True
996    True
997    True
998    True
999    True
Name: gender, Length: 1000, dtype: object
```

```
df.gender.str.isnumeric()
```

```
0        NaN
1      False
2      False
3      False
4      False
       ...
995    False
996    False
997    False
998    False
999    False
Name: gender, Length: 1000, dtype: object
```

```
df.isnull()
```

```
     gender  race/ethnicity  parental level of education  lunch  \
0      True           False                        False  False
1     False           False                        False  False
2     False           False                        False  False
3     False           False                        False  False
4     False           False                        False  False
..      ...             ...                          ...    ...
995   False           False                        False  False
996   False           False                        False  False
997   False           False                        False  False
998   False           False                        False  False
999   False           False                        False  False

     test preparation course  math score  reading score  writing score
0                      False       False          False          False
1                      False       False          False          False
2                      False        True          False          False
3                      False        True          False          False
4                      False        True          False          False
..                       ...         ...            ...            ...
995                    False       False          False          False
996                    False       False          False          False
997                    False       False          False          False
998                    False       False          False          False
```

```
999                   False       False          False          False

[1000 rows x 8 columns]
```

```
print(df[1:3])
```

```
   gender race/ethnicity parental level of education     lunch  \
1  female        group C                some college  standard
2  female        group B             master's degree  standard

  test preparation course  math score  reading score  writing score
1               completed       150.0             90             88
2                    none         NaN             95             93
```

```
df['St'].isnull()
```

```
0        False
1        False
2        False
3        False
4        False
         ...
37075    False
37076    False
37077    False
37078    False
37079    False
Name: Industry_aggregation_NZSIOC, Length: 37080, dtype: bool
```

```
df=pd.get_dummies(df['gender'])
```

```
print(df)
```

```
     female  male
0         0     0
1         1     0
2         1     0
3         0     1
4         0     1
..      ...   ...
995       1     0
996       0     1
997       1     0
998       1     0
999       1     0

[1000 rows x 2 columns]
```

```
print(df.head)
```

```
<bound method NDFrame.head of      female  male
0         0     0
```

```
1         1     0
2         1     0
3         0     1
4         0     1
..      ...   ...
995       1     0
996       0     1
997       1     0
998       1     0
999       1     0

[1000 rows x 2 columns]>
```

```
print(df.tail)
```

```
<bound method NDFrame.tail of      female  male
0         0     0
1         1     0
2         1     0
3         0     1
4         0     1
..      ...   ...
995       1     0
996       0     1
997       1     0
998       1     0
999       1     0

[1000 rows x 2 columns]>
```