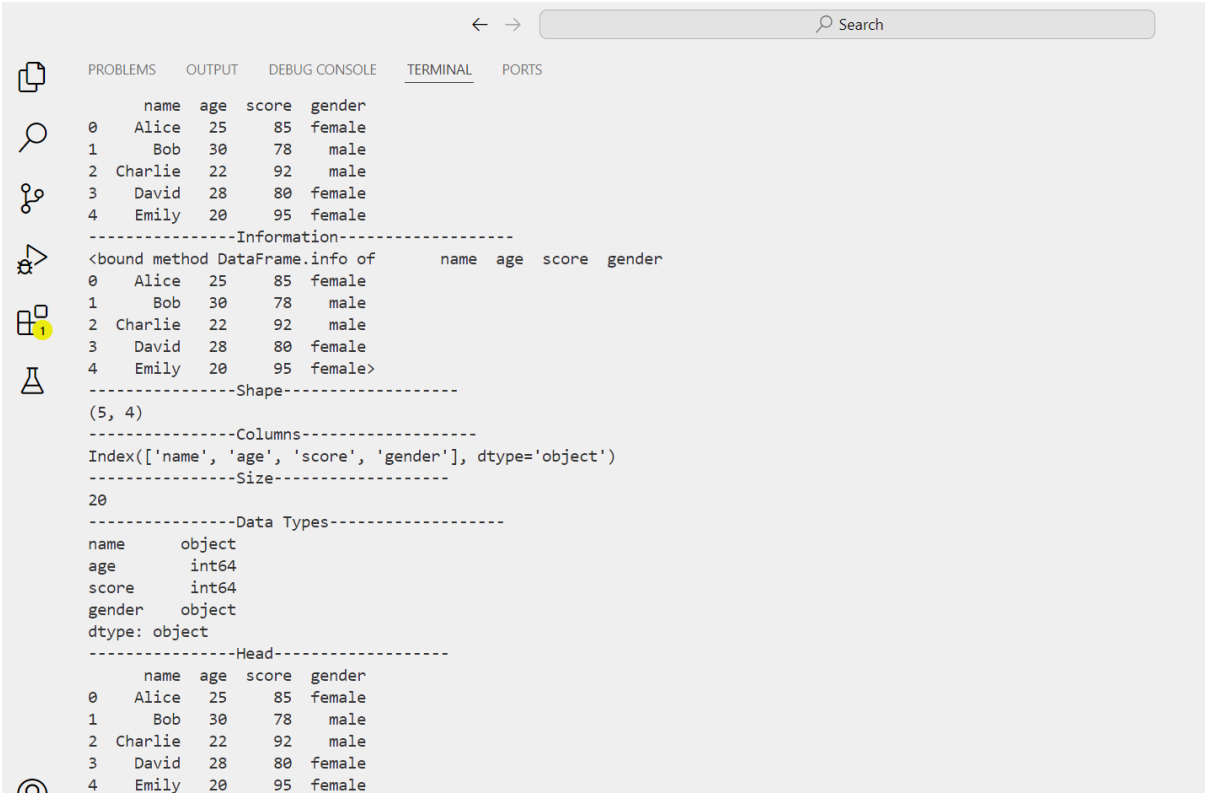
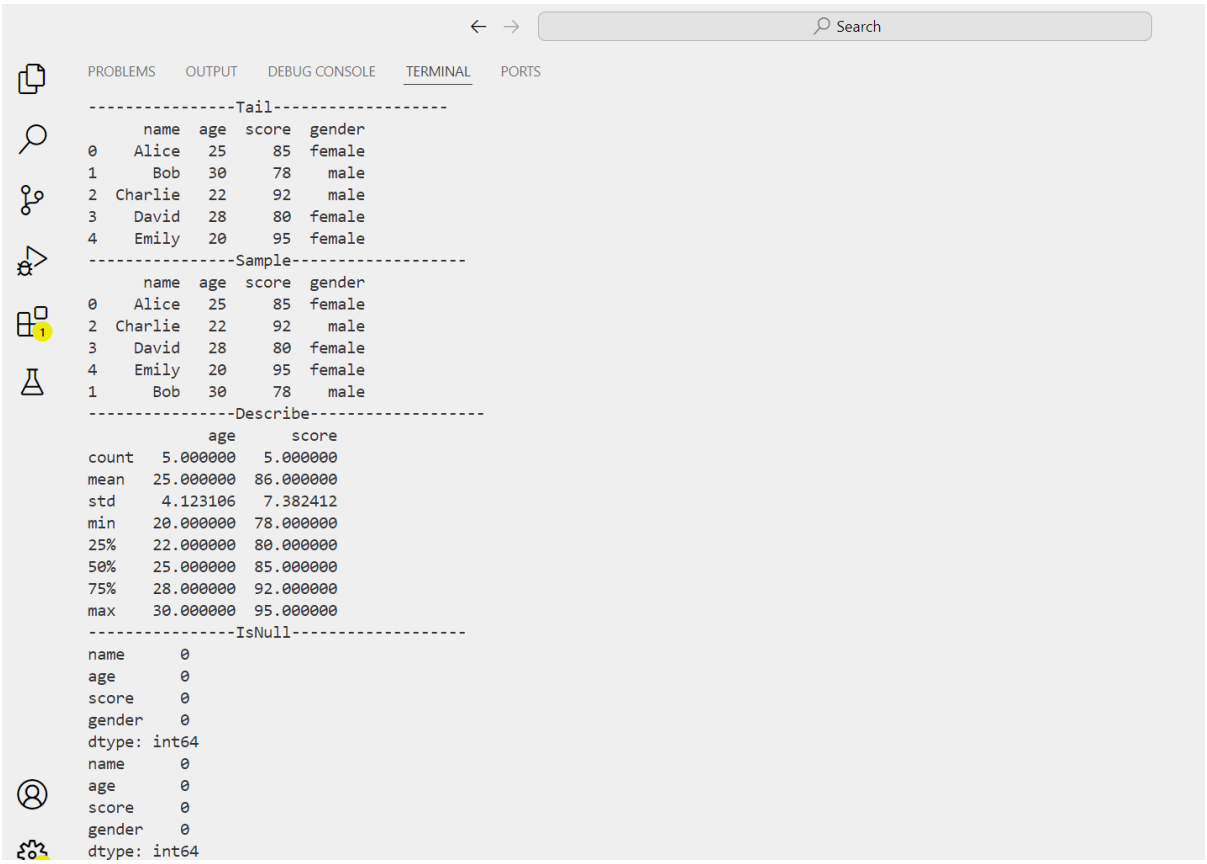
```
import numpy as np
import pandas as pd

df = pd.read_csv('ass1.csv')

print(df)
print("----------"+"Information"+"----------")
print(df.info)
print("----------"+"Shape"+"----------")
print(df.shape)
print("----------"+"Columns"+"----------")
print(df.columns)
print("----------"+"Size"+"----------")
print(df.size)
print("----------"+"Data Types"+"----------")
print(df.dtypes)
print("----------"+"Head"+"----------")
print(df.head(5))
print("----------"+"Tail"+"----------")
print(df.tail(5))
print("----------"+"Sample"+"----------")
print(df.sample(5))
print("----------"+"Describe"+"----------")
print(df.describe())
print("----------"+"IsNull"+"----------")
print(df.isna().sum())
print(df.isnull().sum())
print("----------"+"NotNull"+"----------")
print(df.notnull().sum())


print("----------"+"Data Normalization"+"----------")
```

```
from sklearn.preprocessing import MinMaxScaler
data = [[1,3],[3,1],[4,5],[2,7]]
scaler = MinMaxScaler()
model=scaler.fit(data)
scaled_data=model.transform(data)
print(scaled_data)

from sklearn.preprocessing import StandardScaler
data = [[1,3],[3,1],[4,5],[2,7]]
scaler = StandardScaler()
model=scaler.fit(data)
scaled_data=model.transform(data)
print(scaled_data)

print("----------"+"Fill NA"+"----------")
df['gender'].fillna('female', inplace=True)
print(df)

print("----------"+"Replace"+"----------")
df['gender'].replace(['male','female'],[0,1], inplace=True)
print(df)

print("----------"+"Categorical Variable"+"----------")
df['gender']=df['gender'].astype('category')
print("Data Types of gender: ")
print(df.dtypes['gender'])
df['gender']=df['gender'].cat.codes
print("Data Types of gender: ")
print(df.dtypes['gender'])
print(df)
```

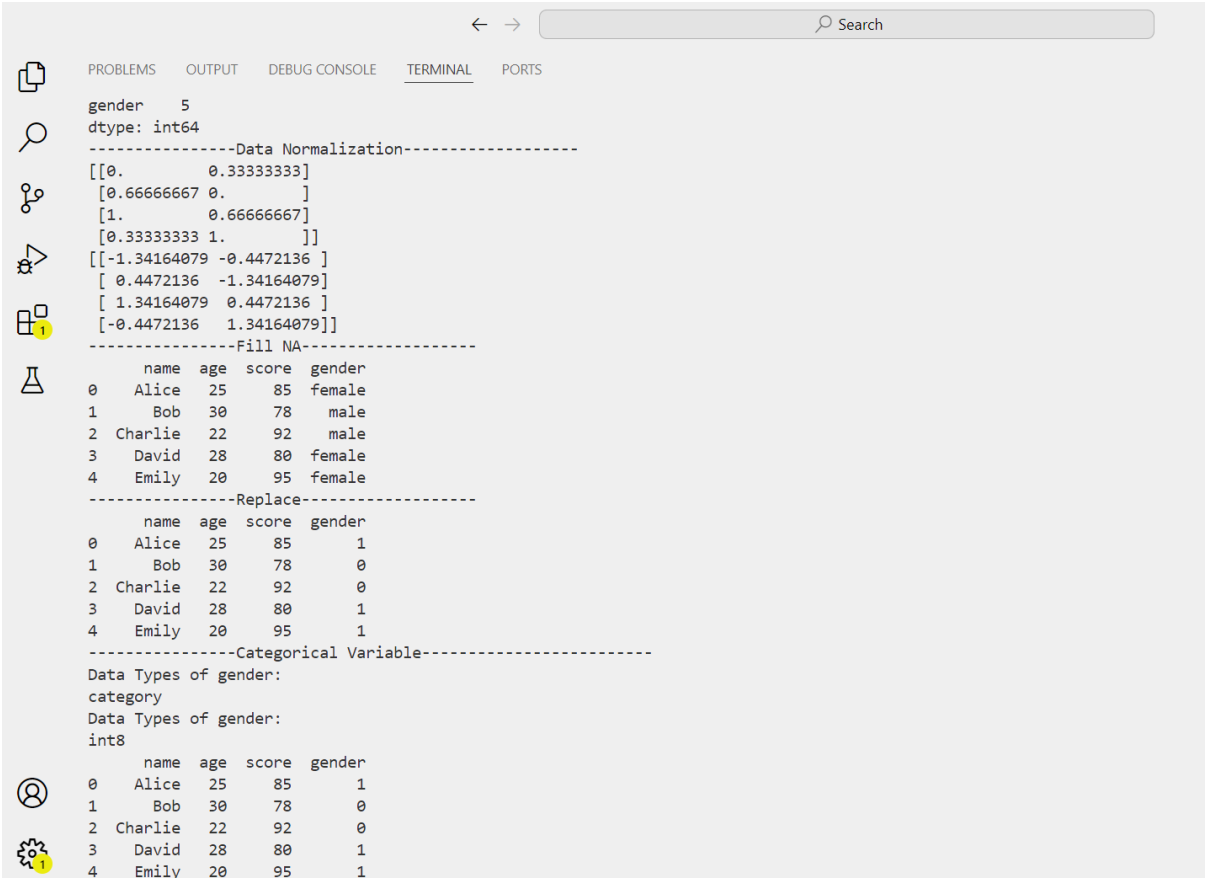
OUPUT -




```
      name  age  score  gender
0    Alice   25     85  female
1      Bob   30     78    male
2  Charlie   22     92    male
3    David   28     80  female
4    Emily   20     95  female
----------------Information-------------------
<bound method DataFrame.info of       name  age  score  gender
0    Alice   25     85  female
1      Bob   30     78    male
2  Charlie   22     92    male
3    David   28     80  female
4    Emily   20     95  female>
----------------Shape-------------------
(5, 4)
----------------Columns-------------------
Index(['name', 'age', 'score', 'gender'], dtype='object')
----------------Size-------------------
20
----------------Data Types-------------------
name      object
age        int64
score      int64
gender    object
dtype: object
----------------Head-------------------
      name  age  score  gender
0    Alice   25     85  female
1      Bob   30     78    male
2  Charlie   22     92    male
3    David   28     80  female
4    Emily   20     95  female
```




```
----------------Tail-------------------
      name  age  score  gender
0    Alice   25     85  female
1      Bob   30     78    male
2  Charlie   22     92    male
3    David   28     80  female
4    Emily   20     95  female
----------------Sample-------------------
      name  age  score  gender
0    Alice   25     85  female
2  Charlie   22     92    male
3    David   28     80  female
4    Emily   20     95  female
1      Bob   30     78    male
----------------Describe-------------------
             age      score
count   5.000000   5.000000
mean   25.000000  86.000000
std     4.123106   7.382412
min    20.000000  78.000000
25%    22.000000  80.000000
50%    25.000000  85.000000
75%    28.000000  92.000000
max    30.000000  95.000000
----------------IsNull-------------------
name      0
age       0
score     0
gender    0
dtype: int64
name      0
age       0
score     0
gender    0
dtype: int64
```


Search
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
gender 5
dtype: int64
----------Data Normalization----------
[[0. 0.33333333]
[0.66666667 0.]
[1. 0.66666667]
[0.33333333 1.]]
[[-1.34164079 -0.4472136]
[0.4472136 -1.34164079]
[1.34164079 0.4472136]
[-0.4472136 1.34164079]]
----------Fill NA----------
name age score gender
0 Alice 25 85 female
1 Bob 30 78 male
2 Charlie 22 92 male
3 David 28 80 female
4 Emily 20 95 female
----------Replace----------
name age score gender
0 Alice 25 85 1
1 Bob 30 78 0
2 Charlie 22 92 0
3 David 28 80 1
4 Emily 20 95 1
----------Categorical Variable----------
Data Types of gender:
category
Data Types of gender:
int8
name age score gender
0 Alice 25 85 1
1 Bob 30 78 0
2 Charlie 22 92 0
3 David 28 80 1
4 Emily 20 95 1