

1. Find and treat all the missing values. Rows or columns with missing values must not be dropped.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
"""
1. Find and treat all the missing values. Rows
or columns with missing values must not be
dropped.
"""
df = pd.read_csv('Dataset_Day4.csv')
print(df.info())
skewness = df.skew()
print(skewness)
missing_value_percent = df.isna().sum() / len(df)
* 100
print(missing_value_percent)
df["TakeHome"].fillna(df["TakeHome"].median(),
inplace=True)
print(df.info())
df["Final"].fillna(df["Final"].median(),
inplace=True)
print(df.info())
```

```

1 C:\Users\tejas\PycharmProjects\pythonProject\venv\Scripts\python.exe C:\Users\
  tejas\PycharmProjects\pythonProject\START\mdh.py
2 <class 'pandas.core.frame.DataFrame'>
3 RangeIndex: 99 entries, 0 to 98
4 Data columns (total 6 columns):
5 #   Column      Non-Null Count  Dtype
6 ---  ---
7 0   Prefix      99 non-null    int64
8 1   Assignment  99 non-null    float64
9 2   Tutorial    99 non-null    float64
10 3   Midterm     99 non-null    float64
11 4   TakeHome    98 non-null    float64
12 5   Final       96 non-null    float64
13 dtypes: float64(5), int64(1)
14 memory usage: 4.8 KB
15 None
16 Prefix      -1.746493
17 Assignment  -1.676609
18 Tutorial     -1.472330
19 Midterm     -0.063458
20 TakeHome    -1.175746
21 Final       0.096656
22 dtype: float64
23 Prefix      0.000000
24 Assignment   0.000000
25 Tutorial     0.000000
26 Midterm     0.000000
27 TakeHome    1.010101
28 Final       3.030303
29 dtype: float64
30 <class 'pandas.core.frame.DataFrame'>
31 RangeIndex: 99 entries, 0 to 98
32 Data columns (total 6 columns):
33 #   Column      Non-Null Count  Dtype
34 ---  ---
35 0   Prefix      99 non-null    int64
36 1   Assignment  99 non-null    float64
37 2   Tutorial    99 non-null    float64
38 3   Midterm     99 non-null    float64
39 4   TakeHome    99 non-null    float64
40 5   Final       96 non-null    float64
41 dtypes: float64(5), int64(1)
42 memory usage: 4.8 KB
43 None
44 <class 'pandas.core.frame.DataFrame'>
45 RangeIndex: 99 entries, 0 to 98
46 Data columns (total 6 columns):
47 #   Column      Non-Null Count  Dtype
48 ---  ---
49 0   Prefix      99 non-null    int64
50 1   Assignment  99 non-null    float64
51 2   Tutorial    99 non-null    float64
52 3   Midterm     99 non-null    float64
53 4   TakeHome    99 non-null    float64
54 5   Final       99 non-null    float64
55 dtypes: float64(5), int64(1)
56 memory usage: 4.8 KB
57 None
58

```

2. Make an in-depth report on the variables using descriptive statistics and also portray the relationship between each variables using visualizations.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

"""2. Make an in-depth report on the variables
using descriptive statistics and also portray the
relationship
between each variables using visualizations"""
df = pd.read_csv('Dataset_Day4.csv')

skewness = df.skew()
print(skewness)

missing_value_percent = df.isna().sum() / len(df) *
100
print(missing_value_percent)
df["TakeHome"].fillna(df["TakeHome"].median(),
inplace=True)
df["Final"].fillna(df["Final"].median(),
inplace=True)
print(df.info())
descriptive_statistics = df.describe()
print(descriptive_statistics)
# box-plot
Viz_data = df
sns.boxplot(data=Viz_data["Final"], orient='v')
# sns.boxplot(data=Viz_data["Tutorial"],
orient='v')
# sns.boxplot(data=Viz_data["Midterm"], orient='v')
plt.show()
# pair-plot(since we're doing it for the entire
dataframe)
sns.pairplot(Viz_data)
plt.show()
# heatmap
sns.heatmap(Viz_data.corr())
plt.show()
```

```

1 C:\Users\tejas\PycharmProjects\pythonProject\venv\Scripts\python.exe C:\Users\
  tejas\PycharmProjects\pythonProject\START\viz.py
2 Prefix      -1.746493
3 Assignment   -1.676609
4 Tutorial     -1.472330
5 Midterm      -0.063458
6 TakeHome     -1.175746
7 Final        0.096656
8 dtype: float64
9 Prefix        0.000000
10 Assignment   0.000000
11 Tutorial     0.000000
12 Midterm      0.000000
13 TakeHome     1.010101
14 Final        3.030303
15 dtype: float64
16 <class 'pandas.core.frame.DataFrame'>
17 RangeIndex: 99 entries, 0 to 98
18 Data columns (total 6 columns):
19 #   Column      Non-Null Count  Dtype
20 ---  ---
21 0    Prefix      99 non-null    int64
22 1    Assignment  99 non-null    float64
23 2    Tutorial    99 non-null    float64
24 3    Midterm     99 non-null    float64
25 4    TakeHome    99 non-null    float64
26 5    Final       99 non-null    float64
27 dtypes: float64(5), int64(1)
28 memory usage: 4.8 KB
29 None
30      Prefix  Assignment  Tutorial  Midterm  TakeHome  Final
31 count  99.000000   99.000000   99.000000   99.000000   99.000000   99.000000
32 mean    7.313131   85.491717   89.731111   68.049495   80.900505   68.348788
33 std     0.932918   12.597694   15.071556   19.376074   23.697863   18.514834
34 min     4.000000   28.140000   34.090000   28.120000   16.910000   28.060000
35 25%     7.000000   80.875000   83.350000   52.810000   67.960000   54.165000
36 50%     8.000000   89.940000   93.100000   69.380000   87.960000   66.250000
37 75%     8.000000   95.000000  100.550000   82.810000   98.425000   83.610000
38 max     8.000000  100.830000  112.580000  110.000000  108.890000  108.890000
39
40 Process finished with exit code 0
41

```

— 0 — X



Figure 1

