

```

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

print("=====Load dataset into panda dataframe=====")

df=pd.read_csv(r'ass2.csv')
print(df)
print("=====IS null=====")
print(df.isna().sum())
print("=====Not null=====")
print(df.notnull().sum())
print("=====Fillna=====")
df['Domain'].fillna('pqr',inplace=True)
print(df)
print("=====Fillna=====")
df['UNIVERSITY'].fillna('aaa',inplace=True)
print(df)
print("=====Rename=====")
df.rename(columns={'Domain':'domain'},inplace=True)
print(df)
print("=====Describe=====")
print(df.describe())

print("=====Detect oulier using Z-score=====")
Score=df['GENERAL MANAGEMENT SCORE (OUT of 50)']
mean=np.mean(Score)
std=np.std(Score)
print("Standard deviation :",std)
print("Mean:",mean)

```

```

threshold=2
outlier=[]
for i in Score:
    z=(i - mean)/std
    if z> threshold:
        outlier.append(i)
print("Outlier",outlier)

print("====Detect outlier using IQR====")
Score=df['GENERAL MANAGEMENT SCORE (OUT of 50)']
NScore=sorted(Score)
print(NScore)
q1,q3=np.percentile(Score,[25,75])
print("Q1,Q3:",q1,q3)
iqr=q3-q1
print("IQR:",iqr)

lower_fence=q1-(1.5*iqr)
upper_fence=q3+(1.5*iqr)
print("Lower fence,upper_fence:",lower_fence,upper_fence)
outlier =[]
for x in Score:
    if ((x> upper_fence) or (x<lower_fence)):
        outlier.append(x)
print(' outlier in the dataset is', outlier)

```

```

print("====Before Removing outliers====")
print(df['GENERAL MANAGEMENT SCORE (OUT of 50)'])

```

```

ua=np.where(df['GENERAL MANAGEMENT SCORE (OUT of 50)']>=upper_fence)[0]

la=np.where(df['GENERAL MANAGEMENT SCORE (OUT of 50)']<=lower_fence)[0]

df.drop(index=ua,inplace=True)

df.drop(index=la,inplace=True)

print("=====After removing outliers=====")

print(df['GENERAL MANAGEMENT SCORE (OUT of 50)'])

print("=====Data tranformation=====")

df['GENERAL MANAGEMENT SCORE (OUT of 50)']=np.log(df['GENERAL MANAGEMENT SCORE (OUT
of 50)'])

print('Display dataset after data tranformation')

print(df)

```

OUTPUT -

```

File Edit Selection View Go Run ... ← → Search
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

Name Age UNIVERSITY Domain GENERAL MANAGEMENT SCORE (OUT of 50)
0 Alice 25 University X IT 38
1 Bob 30 University Y Marketing 42
2 Charlie 22 University Z Finance 48
3 David 28 University X Management 35
4 Emily 20 NaN Accounting 45
5 Frank 35 University Z HR 20
6 Grace 40 University X Operations 55

=====IS null=====
Name 0
Age 0
UNIVERSITY 1
Domain 0
GENERAL MANAGEMENT SCORE (OUT of 50) 0
dtype: int64

=====Not null=====
Name 7
Age 7
UNIVERSITY 6
Domain 7
GENERAL MANAGEMENT SCORE (OUT of 50) 7
dtype: int64

=====Fillna=====
Name Age UNIVERSITY Domain GENERAL MANAGEMENT SCORE (OUT of 50)
0 Alice 25 University X IT 38
1 Bob 30 University Y Marketing 42
2 Charlie 22 University Z Finance 48
3 David 28 University X Management 35
4 Emily 20 NaN Accounting 45
5 Frank 35 University Z HR 20
6 Grace 40 University X Operations 55

```

```
File Edit Selection View Go Run ... Search
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

=====Rename=====
   Name  Age  UNIVERSITY  domain  GENERAL MANAGEMENT SCORE (OUT of 50)
0  Alice   25  University X    IT          38
1   Bob    30  University Y  Marketing    42
2  Charlie 22  University Z    Finance    48
3   David  28  University X  Management    35
4   Emily  20          aaa  Accounting    45
5   Frank  35  University Z    HR         20
6   Grace  40  University X  Operations    55
=====Describe=====
      Age  GENERAL MANAGEMENT SCORE (OUT of 50)
count  7.000000          7.000000
mean   28.571429         40.428571
std     7.114706         11.148350
min    20.000000         20.000000
25%    23.500000         36.500000
50%    28.000000         42.000000
75%    32.500000         46.500000
max    40.000000         55.000000
=====Detect outlier using Z-score=====
Standard deviation : 10.321366781821967
Mean: 40.42857142857143
Outlier []
```

```
File Edit Selection View Go Run ... Search
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

=====Detect outlier using IQR=====
[20, 35, 38, 42, 45, 48, 55]
Q1,Q3: 36.5 46.5
IQR: 10.0
Lower fense,upper_fense: 21.5 61.5
outlier in the dataset is [20]
=====Before Removing outliers=====
0  38
1  42
2  48
3  35
4  45
5  20
6  55
Name: GENERAL MANAGEMENT SCORE (OUT of 50), dtype: int64
=====After removing outliers=====
0  38
1  42
2  48
3  35
4  45
6  55
Name: GENERAL MANAGEMENT SCORE (OUT of 50), dtype: int64
=====Data tranformation=====
Display dataset after data tranformation
   Name  Age  UNIVERSITY  domain  GENERAL MANAGEMENT SCORE (OUT of 50)
0  Alice   25  University X    IT          3.637586
1   Bob    30  University Y  Marketing    3.737670
2  Charlie 22  University Z    Finance    3.871201
3   David  28  University X  Management    3.555348
4   Emily  20          aaa  Accounting    3.806662
6   Grace  40  University X  Operations    4.007333
```