**PRACTICAL NO : 02**

**DATA WRANGING 2**

**CODE :**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from scipy import stats

df= pd.read\_csv(r'E:\DSBDA\DSBDA PR 02\StudentPerformance.csv')

display(df)

#isnull

df.isnull()

data = pd.isnull(df['math score'])

display(data)

#notnull

df.notnull()

data = pd.notnull(df['math score'])

display(data)

#fillna

df.fillna(1)

#replace

df.replace(to\_replace=np.nan,value=-99)

#dropna

df.dropna()

df.dropna(axis = 1)

df.dropna(axis=0)

#Detecting outlier using Boxplot

col=['math score','reading score','writing score','placement score']

df.boxplot(col)

print(np.where(df['math score']>90))

print(np.where(df['reading score']<25))

print(np.where(df['writing score']<30))

#Detecting outlier using Scatterplot

fig, ax=plt.subplots(figsize=(18,10))

ax.scatter(df['placement score'],df['placement offer count'])

ax.set\_xlabel('placement score')

ax.set\_ylabel('placement offer count')

ax.set\_title('scatter plot')

plt.show()

#Detecting outlier using Z-score

z=np.abs(stats.zscore(df['math score']))

print(z)

threshold = 0.18

sample\_outliers = np.where(z<threshold)

print(sample\_outliers)

#Histogram

df['math score'].plot(kind='hist')

df['log\_math'] = np.log10(df['math score'])

df['log\_math'].plot(kind='hist')

OUTPUT :

