**PRACTICAL 3**

**Aim:** Implementing of data pre-processing.

**Five Number Summary**

import pandas as pd

import numpy as np

d = pd.read\_csv("xyz.csv")

df=pd.DataFrame(d)

min=df['Total'].min()

max=df['Total'].max()

q1,q3=np.percentile(df['Total'],[25,75])

q2=np.percentile(df['Total'],50)

print("Five Number Sumary:")

print('Minimum:',min)

print('Maximum:',max)

print("Quartile(q1):",q1)

print("Median(q2):",q2)

print("Quartile(q3):",q3)

**Handling Missing Data**

import pandas as pd

import numpy as np

df1 = pd.read\_csv("Xyz1.csv")

display(df1)

df1.isna().sum()

df1['Round 1'] = df['Round 1'].fillna(df['Round 1'].mean())

df1['Round 2'] = df['Round 2'].fillna(df['Round 2'].mean())

df1['Round 3'] = df['Round 3'].fillna(df['Round 3'].mean())

df1['Round 4'] = df['Round 4'].fillna(df['Round 4'].mean())

df1['Round 5'] = df['Round 5'].fillna(df['Round 5'].mean())

df1['Round 6'] = df['Round 6'].fillna(df['Round 6'].mean())

df1['Round 7'] = df['Round 7'].fillna(df['Round 7'].mean())

df1['Round 8'] = df['Round 8'].fillna(df['Round 8'].mean())

df1['Round 9'] = df['Round 9'].fillna(df['Round 9'].mean())

df1['Round 10'] = df['Round 10'].fillna(df['Round 10'].mean())

print(df1)

**Binning by frequency (Depth)**

bin1=df1['Round 1'].iloc[0:4]

print(bin1)

bin2=df1['Round 1'].iloc[4:8]

print(bin2)

bin3=df1['Round 1'].iloc[8:12]

print(bin3)

bin4=df1['Round 1'].iloc[12:16]

print(bin4)

a=bin1.mean()

print(a)

b=bin2.mean()

print(b)

c=bin2.mean()

print(c)

d=bin3.mean()

print(d)

**Smoothening** **(by Mean)**

bin1.iloc[0:4]=bin1.mean()

s1=bin1.copy()

print(s1)

bin2.iloc[0:4]=bin2.mean()

s1=bin2.copy()

print(s1)

bin3.iloc[0:4]=bin3.mean()

s1=bin3.copy()

print(s1)

bin4.iloc[0:4]=bin4.mean()

s1=bin4.copy()

print(s1)

**Normalization**

import pandas as pd

df1 = pd.read\_csv("Xyz.csv")

df3 = pd.DataFrame(df1)

print(df3)

del df3["Name of the Contestant"]

print(df3)

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

normalized\_data = scaler.fit\_transform(df3)

print(normalized\_data)

normalized\_df = pd.DataFrame(normalized\_data, columns=df3.columns)

print(normalized\_df)

import scipy.stats as stats

df3['Round1\_scores\_val']=stats.zscore(df3['Round 1'])

df3

**Decimal Scaling**

max\_b = str(int(df3['Round 1'].max()))

df3['Decimal\_scaling'] = df3['Round 1']/(10\*\*len(max\_b))

df3