PRACTICAL NO : 03

DESCRIPTIVE STATISTICS : MEASURES OF CENTRAL TENDANCY AND VARIABILITY

CODE :

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from scipy import stats

from sklearn import preprocessing

df1=pd.read_csv(r'E:\DSBDA\DSBDA Datasets\cal_cities_lat_long.csv')

print(df1)


df1.columns


df1.mean()

df1.loc[:,'Latitude'].mean()

df1.mean(axis=1)[0:4]


df1.median()

df1.loc[:,'Latitude'].median()

df1.median(axis=1)[0:4]


df1.mode()

df1.loc[:,'Latitude'].mode()
```

```
df1.min()
df1.loc[:,'Latitude'].min(skipna=False)


df1.max()
df1.loc[:,'Latitude'].max(skipna=False)



df1.std()
df1.loc[:,'Latitude'].std()


df1.groupby(['Latitude'])['Longitude'].mean()
enc=preprocessing.OneHotEncoder()
enc_df=pd.DataFrame(enc.fit_transform(df1[['Latitude']]).toarray())
enc_df
df_encode=df1.join(enc_df)
df_encode
```

OUTPUT :

```
df1.columns
Out[2]: Index(['Name', 'Latitude', 'Longitude'], dtype='object')


df1.loc[:,'Latitude'].mean()
Out[3]: 35.99163177995642


df1.loc[:,'Latitude'].median()
Out[4]: 35.489417
```

```
df1.loc[:,'Latitude'].mode()
```

Out[5]:

```
0    32.991156
1    33.787794
2    34.003903
3    34.068622
4    34.090008
5    34.106400
6    34.107231
7    37.797428
8    37.977978
Name: Latitude, dtype: float64
```

```
df1.loc[:,'Latitude'].min(skipna=False)
```

Out[6]: 32.583944

```
df1.loc[:,'Latitude'].max(skipna=False)
```

Out[7]: 41.967369

```
df1.loc[:,'Latitude'].std()
```

Out[8]: 2.3159098613708125

```
df1.groupby(['Latitude'])['Longitude'].mean()
```

Out[9]:

Latitude

32.583944   -117.113086

32.640053   -117.084197

32.678108   -117.099197

32.678947   -115.498883

32.685886   -117.183089

41.728197   -122.527800

41.735419   -122.634472

41.755947   -124.201747

41.955989   -121.477492

41.967369   -121.918061

Name: Longitude, Length: 450, dtype: float64

OUTPUT :

```
In [2]: df1.columns
Out[2]: Index(['Name', 'Latitude', 'Longitude'], dtype='object')

In [3]: df1.loc[:,'Latitude'].mean()
Out[3]: 35.99163177995642

In [4]: df1.loc[:,'Latitude'].median()
Out[4]: 35.489417

In [5]: df1.loc[:,'Latitude'].mode()
Out[5]:
0    32.991156
1    33.787794
2    34.003903
3    34.068622
4    34.090008
5    34.106400
6    34.107231
7    37.797428
8    37.977978
Name: Latitude, dtype: float64

In [6]: df1.loc[:,'Latitude'].min(skipna=False)
Out[6]: 32.583944

In [7]: df1.loc[:,'Latitude'].max(skipna=False)
Out[7]: 41.967369

In [8]: df1.loc[:,'Latitude'].std()
Out[8]: 2.3159098613708125

In [9]: df1.groupby(['Latitude'])['Longitude'].mean()
Out[9]:
```

```
In [6]: df1.loc[:,'Latitude'].min(skipna=False)
Out[6]: 32.583944

In [7]: df1.loc[:,'Latitude'].max(skipna=False)
Out[7]: 41.967369

In [8]: df1.loc[:,'Latitude'].std()
Out[8]: 2.3159098613708125

In [9]: df1.groupby(['Latitude'])['Longitude'].mean()
Out[9]:
Latitude
32.583944    -117.113086
32.640053    -117.084197
32.678108    -117.099197
32.678947    -115.498883
32.685886    -117.183089
                 ...
41.728197    -122.527800
41.735419    -122.634472
41.755947    -124.201747
41.955989    -121.477492
41.967369    -121.918061
Name: Longitude, Length: 450, dtype: float64

In [10]:
```