Practical 3

Tutorial

In the 3rd practical we will be performing various statistical operations on the data set.

Steps:

1. **Download a suitable dataset from kaagle.**
2. **Open anaconda and launch spyder.**
3. **Create a new file and import the following libraries:**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
from sklearn import preprocessing

df = pd.read_csv(r"C:\Users\comp\Documents\sushant\movies edited.csv")
df
```

4. **Be sure to give the right pathway for the dataset location.**
5. **Now we will be performing various statistical operations**

- **Mean :**

```
df.mean()

df.loc[:,' name_of_the_column '].mean()

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

Note : here name of the column specifies the column name in t=your data set that has integer values. Keep in mind to perform these operations on the columns which have integer values only.

- **Median :**

df.median()

df.loc[:,'name_of_column'].median()

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

- **Mode :**

df.mode()

df.loc[:,'name_of_column'].mode()

- **Minimum:**

df.min()

df.loc[:,'name_of_column'].min(skipna = False)

- **Maximum :**

df.max()

df.loc[:,'name_of_column'].max(skipna = False)

- **Standard deviation**

df.std()

df.loc[:,'name_of_column'].std()

- **Group by :**

```
df.groupby(['name_of_column_1'])[‘name_of_column_2’].mean()
enc = preprocessing.OneHotEncoder()
enc_df = pd.DataFrame(enc.fit_transform(df[[‘name_of_column’]]).toarray())
enc_df

df_encode =df.join(enc_df)
df_encode
```

6. **After the successful execution of all the operations save the file.**
7. **Create a text file with the final code and out put.**