Assignment 2

Subject: Machine Learning

Course code: CSE 465

Assignment No: 2

Submitted To	Submitted By
Khan Md. Hasib Lecturer Dept. of CSE BUBT	Ahmed Mahir Shoaib ID: 18192103235 Intake: 41 Section: 7 Dept. of CSE BUBT

> Apply calculating mathematical statistics techniques (such as: mean -
> average value, median - middle value, median - middle value, median -
> middle value) in the following dataset -
> https://www.kaggle.com/datasets/muthuj7/weather-dataset

Once I have a pandas DataFrame that contains numerical data, I can use various methods to compute summary statistics like mean, median, and mode.

The code is given below:

```
import pandas as pd

# read a CSV file into a DataFrame
df = pd.read_csv('filename.csv')

# compute summary statistics
mean = df.mean()
median = df.median()
mode = df.mode()

# display the summary statistics
print('Mean:\n', mean)
print('Median:\n', median)
print('Mode:\n', mode)
```

The mean() method computes the mean (average) of each column in the DataFrame. The median() method computes the median (middle value) of each column. The mode() method computes the mode (most common value) of each column.

Note that if your DataFrame contains non-numerical data, you may need to select the numerical columns first using the select_dtypes() method.

```
# select numerical columns only
numeric_cols = df.select_dtypes(include='number')

# compute summary statistics
mean = numeric_cols.mean()
median = numeric_cols.median()
mode = numeric_cols.mode()
```

```
# display the summary statistics
print('Mean:\n', mean)
print('Median:\n', median)
print('Mode:\n', mode)
```

This selects only the columns that have a numerical data type (e.g., float, int) and then computes the summary statistics.