



Bangladesh University of Business and Technology

Department of CSE

Assignment -02

Course Title : Data Mining

Course code : CSE-476

Submitted By

Joy Adhikary

ID : 18192103062

Intake: 41

Section: 03

Submitted To

Khan Md. Hasib

Lecturer

Computer Science & Engineering

Submission Date : 22.02.23

01. 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>

1. Import library:

```
import pandas as pd
import statistics
```

```
[ ] from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

2. Upload the dataset & Viewing the data:

```
weather=pd.read_csv("/content/drive/MyDrive/Colab Notebooks/weatherHistory.csv")
weather
```

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)	Daily Summary
0	2006-04-01 00:00:00.000+0200	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251.0	15.8263	0.0	1015.13	Partly cloudy throughout the day.
1	2006-04-01 01:00:00.000+0200	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259.0	15.8263	0.0	1015.63	Partly cloudy throughout the day.
2	2006-04-01 02:00:00.000+0200	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204.0	14.9569	0.0	1015.94	Partly cloudy throughout the day.
3	2006-04-01 03:00:00.000+0200	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269.0	15.8263	0.0	1016.41	Partly cloudy throughout the day.
4	2006-04-01 04:00:00.000+0200	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259.0	15.8263	0.0	1016.51	Partly cloudy throughout the day.
...
96448	2016-09-09 19:00:00.000+0200	Partly Cloudy	rain	26.016667	26.016667	0.43	10.9963	31.0	16.1000	0.0	1014.36	Partly cloudy starting in the morning.
96449	2016-09-09 20:00:00.000+0200	Partly Cloudy	rain	24.583333	24.583333	0.48	10.0947	20.0	15.5526	0.0	1015.16	Partly cloudy starting in the morning.
96450	2016-09-09 21:00:00.000+0200	Partly Cloudy	rain	22.038889	22.038889	0.56	8.9638	30.0	16.1000	0.0	1015.66	Partly cloudy starting in the morning.
96451	2016-09-09 22:00:00.000+0200	Partly Cloudy	rain	21.522222	21.522222	0.60	10.5294	20.0	16.1000	0.0	1015.95	Partly cloudy starting in the morning.
96452	2016-09-09 23:00:00.000+0200	Partly Cloudy	rain	20.438889	20.438889	0.61	5.8765	39.0	15.5204	0.0	1016.16	Partly cloudy starting in the morning.

3. View the top 10 rows of the dataset:

```
weather.head(5)
```

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)	Daily Summary
0	2006-04-01 00:00:00.000+0200	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251.0	15.8263	0.0	1015.13	Partly cloudy throughout the day.
1	2006-04-01 01:00:00.000+0200	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259.0	15.8263	0.0	1015.63	Partly cloudy throughout the day.
2	2006-04-01 02:00:00.000+0200	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204.0	14.9569	0.0	1015.94	Partly cloudy throughout the day.
3	2006-04-01 03:00:00.000+0200	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269.0	15.8263	0.0	1016.41	Partly cloudy throughout the day.
4	2006-04-01 04:00:00.000+0200	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259.0	15.8263	0.0	1016.51	Partly cloudy throughout the day.

4. Showing the mean value of the Humidity Column:

✓
0s

```
[14] import statistics
```

✓
0s



```
mean=statistics.mean(weather["Humidity"])  
  
print("Mean of Humidity is:", mean)
```

Mean of Humidity is: 0.7312608593566565

5. Showing the median value:

✓
0s



```
median=statistics.median(weather["Humidity"])  
  
print("Median of Humidity is:", median)
```

Median of Humidity is: 0.78

6. Showing the mode value:

✓
0s



```
mode=statistics.mode(weather["Humidity"])  
  
print("Mode of Humidity is:", mode)
```

Mode of Humidity is: 0.93

7. Showing the Standard deviation value:

✓
0s



```
stdev=statistics.stdev(weather["Humidity"])  
  
print("Standard deviation of Humidity is:", stdev)
```

Standard deviation of Humidity is: 0.19565322439944888