Roll No: CS6-86

PRN:202401080012

DATASET- WEATHER

EDS ACTIVITY NO.1

Summarize DataSet

```
[8]  df.describe()
```

0s

	MinTemp	MaxTemp	Rainfall	WindSpeed	Humi
count	10.000000	10.000000	10.000000	10.000000	10.0(
mean	18.800000	27.520000	2.290000	17.700000	62.3(
std	4.142463	5.352424	3.207439	5.313505	8.8{
min	12.500000	19.900000	0.000000	12.000000	50.0(
25%	15.575000	22.575000	0.000000	14.250000	55.7{
50%	19.000000	28.750000	1.150000	16.500000	61.5(
75%	21.775000	30.875000	2.975000	19.500000	67.2{
max	25.000000	35.500000	10.000000	30.000000	80.0(

Find Highest Temperature

```
df['MaxTemp'].max()
```

0s

```
35.5
```

Finding Lowest Temperature

```
df['MinTemp'].min()
```

0s

```
12.5
```

How many number of location

```
df['Location'].nunique()
```

0s

```
8
```

Read Data Set

```
[3] import pandas as pd
    df = pd.read_csv('/content/weather_data.csv')
```

Display first 5 rows

```
[4] df.head()
```

	Date	Location	MinTemp	MaxTemp	Rainfall	Wind
0	2025-01-01	Sydney	21.1	30.5	0.0	
1	2025-01-02	Melbourne	14.3	22.1	2.3	
2	2025-01-03	Brisbane	23.4	33.2	5.1	
3	2025-01-04	Adelaide	18.2	27.8	0.0	
4	2025-01-05	Perth	19.8	31.0	0.0	

Show Column Names

```
[5] df.columns
```

```
Index(['Date', 'Location', 'MinTemp', 'MaxTemp',
'Rainfall', 'WindSpeed',
       'Humidity', 'RainToday', 'RainTomorrow'],
      dtype='object')
```

Defining Shape of Dataset

```
[7] df.shape
```

```
(10, 9)
```

Where is max temperature is greater than 30

```
df[df['MaxTemp'] > 30]
```

	Date	Location	MinTemp	MaxTemp	Rainfall	Wind
0	2025-01-01	Sydney	21.1	30.5	0.0	
2	2025-01-03	Brisbane	23.4	33.2	5.1	
4	2025-01-05	Perth	19.8	31.0	0.0	
6	2025-01-07	Darwin	25.0	35.5	10.0	

New Column of Temp Range

```
df['TempRange'] = df['MaxTemp'] - df['MinTemp']
```

```
[19] df
```

Humidity	RainToday	RainTomorrow	TempRange
55	No	No	9.4
65	Yes	No	7.8
70	Yes	Yes	9.8
60	No	No	9.6
50	No	No	11.2
68	Yes	Yes	7.4
80	Yes	Yes	10.5
58	No	No	7.6

```
df['Location'].unique()
```

```
array(['Sydney', 'Melbourne', 'Brisbane',
'Adelaide', 'Perth', 'Hobart',
       'Darwin', 'Canberra'], dtype=object)
```

How Many Days it Rained Tommorow

```
df['RainTomorrow'].value_counts()
```

RainTomorrow	count
No	7
Yes	3

dtype: int64

Where it rained today

```
df[df['RainToday'] == 'Yes']
```

WindSpeed	Humidity	RainToday	RainTomorrow
20	65	Yes	No
18	70	Yes	Yes
22	68	Yes	Yes
30	80	Yes	Yes
17	63	Yes	No

Find Average Rainfall

```
df['Rainfall'].mean()
```

```
np.float64(2.29)
```

Find Null values

```
df.isnull().sum()
```

	0
Date	0
Location	0
MinTemp	0
MaxTemp	0
Rainfall	0
WindSpeed	0
Humidity	0
RainToday	0
RainTomorrow	0
TempRange	0

dtype: int64

Drop Entries with Null

```
df_clean = df.dropna()
```

Replace missing values

```
df['Rainfall'].fillna(0, inplace=True)
```

```
<ipython-input-23-78d7d8d1caf8>:1: FutureWarning:
The behavior will change in pandas 3.0. This inpla

For example, when doing 'df[col].method(value, inp
```

Sort By Max Temperature

```
df.sort_values(by='MaxTemp', ascending=False)
```

0s

	Date	Location	MinTemp	MaxTemp	Rainfall	Wind
6	2025-01-07	Darwin	25.0	35.5	10.0	
2	2025-01-03	Brisbane	23.4	33.2	5.1	
4	2025-01-05	Perth	19.8	31.0	0.0	
0	2025-01-01	Sydney	21.1	30.5	0.0	
8	2025-01-09	Sydney	22.0	29.7	0.0	
3	2025-01-04	Adelaide	18.2	27.8	0.0	
7	2025-01-08	Canberra	16.4	24.0	0.5	
1	2025-01-02	Melbourne	14.3	22.1	2.3	
9	2025-01-10	Melbourne	15.3	21.5	1.8	
5	2025-01-06	Hobart	12.5	19.9	3.2	

Rename Column

```
[26]olumns={'MinTemp': 'MinimumTemperature'}, inplace=True)
```

0s

```
df
```

Location	MinimumTemperature	MaxTemp	Rainfall	Wi
Sydney	21.1	30.5	0.0	
Melbourne	14.3	22.1	2.3	

Weather Report for Sydney

```
[28] df[df['Location'] == 'Sydney']
```

	Date	Location	MinimumTemperature	MaxTemp	Rainfall	WindSpeed	Humidity	RainToday	RainTomorrow	TempRange
0	2025-01-01	Sydney	21.1	30.5	0.0	14	55	No	No	9.4
8	2025-01-09	Sydney	22.0	29.7	0.0	13	54	No	No	7.7