WEATHER ANALYSIS

1. Data Preparation with Python
2. Advanced Analysis with Power BI or Tableau
3. Correlation and Regression Analysis
4. Documentation

Weather Analysis

Project Requirements:

1. Data Preparation with Python:

- Use Python to clean and preprocess the weather dataset.

- Handle missing values, outliers, and any other inconsistencies in the data.

2. Advanced Analysis with Power BI or Tableau:

- Leverage Power BI or Tableau for advanced data analysis on the weather dataset.

- Create interactive dashboards and visualizations to highlight trends and patterns.

3. Correlation and Regression Analysis:

- Perform correlation analysis to identify relationships between different weather parameters.

- Implement regression analysis to predict one weather parameter based on others.

4. Documentation:

- Document your data preparation and analysis steps.

- Explain the insights derived from the advanced analysis.

WEATHER FORECASTING APPLICATION USING PYTHON

ABSTRACT

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Ancient weather forecasting methods usually relied on observed patterns of events, also termed pattern recognition.

Here this system will predict weather based on parameters such as temperature, humidity and wind. Weather forecasting system takes parameters such as temperature, humidity, and wind and will forecast weather based on previous record therefore this prediction will prove reliable. This system can be used in Air Traffic, Marine, Agriculture, Forestry, Military, and Navy etc.

KEYWORDS

- ✔ Weather forecasting
- ✔ temperature
- ✔ short range
- ✔ challenging task
- ✔ forecasting

FEATURES WEATHER FORECAST

- ✔ Time to time update weather
- ✔ Temperature Update
- ✔ change weather in every hours as according to weather changes.
- ✔ provide accurate data information about weather.

PROCESSING AND CLEANING DATA FOR TABLEAU

In order to create effective visualizations in Tableau, it is important to ensure that the data is in a format that can be easily consumed by the software.

1.Remove irrelevant or duplicate data:

Processing the data, it is important to remove any data that is irrelevant or duplicated. This can help you reduce the size of your data set.

2.Convert data types:

Tableau works best with certain data types, such as date/time, string, and numeric data. If your data set includes other data types, such as Boolean or object data.

3.Split and merge columns:

Sometimes the data set may contain columns that need to be split into multiple

columns or merged into one. Tableau allows you to do this using the "Split" and "Merge" features.

4.Aggregate data:

Tableau works best with aggregated data, which means that is need to summarize the data by grouping it into categories or calculating summary statistics.

5.Visualization Examples

There are many different types of visualizations that can be used to display weather data in Tableau. Here are some examples of different types of visualizations,

- ✔ **Line chart**
- ✔ **Heat map**
- ✔ **Scatter plot**
- ✔ **Bar chart**

SOURCE CODE :

```
import pandas as pd
import json
df=pd.read_csv("weather.csv")
print(df)
```

OUTPUT

```
     MinTemp  MaxTemp  Rainfall  Evaporation  Sunshine WindGustDir  \
0        8.0     24.3       0.0          3.4       6.3          NW
1       14.0     26.9       3.6          4.4       9.7         ENE
2       13.7     23.4       3.6          5.8       3.3          NW
3       13.3     15.5      39.8          7.2       9.1          NW
4        7.6     16.1       2.8          5.6      10.6         SSE
..       ...      ...       ...          ...       ...         ...
361      9.0     30.7       0.0          7.6      12.1         NNW
362      7.1     28.4       0.0         11.6      12.7           N
363     12.5     19.9       0.0          8.4       5.3         ESE
364     12.5     26.9       0.0          5.0       7.1          NW
365     12.3     30.2       0.0          6.0      12.6          NW
```

	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...	Humidity3pm	\
0	30.0	SW	NW	6.0	...	29	
1	39.0	E	W	4.0	...	36	
2	85.0	N	NNE	6.0	...	69	
3	54.0	WNW	W	30.0	...	56	
4	50.0	SSE	ESE	20.0	...	49	
..	...	...	...	...	...	...	
361	76.0	SSE	NW	7.0	...	15	
362	48.0	NNW	NNW	2.0	...	22	
363	43.0	ENE	ENE	11.0	...	47	
364	46.0	SSW	WNW	6.0	...	39	
365	78.0	NW	WNW	31.0	...	13	

	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	\
0	1019.7	1015.0	7	7	14.4	23.6	
1	1012.4	1008.4	5	3	17.5	25.7	
2	1009.5	1007.2	8	7	15.4	20.2	
3	1005.5	1007.0	2	7	13.5	14.1	
4	1018.3	1018.5	7	7	11.1	15.4	
..	...	...	...	...	...	...	
361	1016.1	1010.8	1	3	20.4	30.0	
362	1020.0	1016.9	0	1	17.2	28.2	
363	1024.0	1022.8	3	2	14.5	18.3	
364	1021.0	1016.2	6	7	15.8	25.9	
365	1009.6	1009.2	1	1	23.8	28.6	

	RainToday	RISK_MM	RainTomorrow
0	No	3.6	Yes
1	Yes	3.6	Yes
2	Yes	39.8	Yes
3	Yes	2.8	Yes
4	Yes	0.0	No
..	...	...	...
361	No	0.0	No
362	No	0.0	No
363	No	0.0	No
364	No	0.0	No
365	No	0.0	No

[366 rows x 22 columns]

from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"

```
df.describe()
df[df['MaxTemp'] <= 15]
df.head()
df[df['Sunshine'] == 8.600000]
df.info()
df.dtypes
df.columns
df.describe().transpose()
df.isnull().sum()
df=df.drop_duplicates()
df.info()
cols=df.columns.tolist()
cols
for i, col in enumerate(cols):
    if '(' in col:
        cols[i] = col + ")"
        df.cols= cols
    print(df.columns)
```

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed
count	366.000000	366.000000	366.000000	366.000000	363.000000	364.000000
mean	7.265574	20.550273	1.428415	4.521858	7.909366	39.840659
std	6.025800	6.690516	4.225800	2.669383	3.481517	13.059807
min	-5.300000	7.600000	0.000000	0.200000	0.000000	13.000000
25%	2.300000	15.025000	0.000000	2.200000	5.950000	31.000000
50%	7.450000	19.650000	0.000000	4.200000	8.600000	39.000000
75%	12.500000	25.500000	0.200000	6.400000	10.500000	46.000000
max	20.900000	35.800000	39.800000	13.800000	13.600000	98.000000

WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm
359.000000	366.000000	366.000000	366.000000	366.000000	366.000000
9.651811	17.986339	72.035519	44.519126	1019.709016	1016.810383
7.951929	8.856997	13.137058	16.850947	6.686212	6.469422
0.000000	0.000000	36.000000	13.000000	996.500000	996.800000
6.000000	11.000000	64.000000	32.250000	1015.350000	1012.800000
7.000000	17.000000	72.000000	43.000000	1020.150000	1017.400000
13.000000	24.000000	81.000000	55.000000	1024.475000	1021.475000
41.000000	52.000000	99.000000	96.000000	1035.700000	1033.200000

Cloud9am	Cloud3pm	Temp9am	Temp3pm	RISK_MM
366.000000	366.000000	366.000000	366.000000	366.000000
3.890710	4.024590	12.358470	19.230874	1.428415
2.956131	2.666268	5.630832	6.640346	4.225800
0.000000	0.000000	0.100000	5.100000	0.000000
1.000000	1.000000	7.625000	14.150000	0.000000
3.500000	4.000000	12.550000	18.550000	0.000000
7.000000	7.000000	17.000000	24.000000	0.200000
8.000000	8.000000	24.700000	34.500000	39.800000

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...	H
141	13.0	14.8	0.0	8.2	0.0	SE	30.0	ESE	ESE	9.0	...	
154	7.0	14.3	2.6	9.6	9.7	WNW	63.0	NW	W	26.0	...	
179	4.3	11.3	7.2	4.4	5.6	W	57.0	WNW	WNW	26.0	...	
180	-2.1	13.8	0.2	1.8	9.5	NNW	22.0	NaN	NNW	0.0	...	
181	-1.8	14.8	0.0	1.4	7.0	N	28.0	E	N	2.0	...	
...	...	...	...	...	...	...	...	...	...	...	...	
309	5.8	12.4	0.0	1.6	0.0	SE	26.0	SSE	ESE	13.0	...	
313	-3.7	14.7	0.0	3.4	10.9	SSE	43.0	SE	N	9.0	...	
320	3.9	13.2	3.4	6.6	11.0	WNW	65.0	WNW	WNW	26.0	...	
321	0.7	14.1	0.0	5.6	9.0	ENE	20.0	SSW	NNW	6.0	...	
356	3.4	15.0	0.8	4.8	11.7	S	70.0	S	S	35.0	...	

92 rows × 22 columns

Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
62	1017.4	1019.3	7	8	13.3	13.9	No	0.0	No
40	1010.3	1012.7	1	3	9.1	13.7	Yes	0.0	No
60	1006.3	1008.1	3	6	7.1	8.8	Yes	0.2	No
40	1020.6	1019.6	0	1	6.3	13.2	No	0.0	No
40	1024.2	1020.5	1	7	5.3	13.9	No	0.0	No
...	...	...	...	...	...	...	...	...	...
73	1027.3	1023.5	8	8	8.7	10.8	No	0.0	No
25	1025.8	1020.9	0	0	5.5	13.4	No	0.0	No
33	1017.0	1017.6	3	1	7.1	12.2	Yes	0.0	No
43	1026.7	1022.1	7	1	7.4	13.7	No	0.0	No
24	1023.4	1023.1	1	5	8.3	14.3	No	0.0	No

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...
0	8.0	24.3	0.0	3.4	6.3	NW	30.0	SW	NW	6.0	...
1	14.0	26.9	3.6	4.4	9.7	ENE	39.0	E	W	4.0	...
2	13.7	23.4	3.6	5.8	3.3	NW	85.0	N	NNE	6.0	...
3	13.3	15.5	39.8	7.2	9.1	NW	54.0	WNW	W	30.0	...
4	7.6	16.1	2.8	5.6	10.6	SSE	50.0	SSE	ESE	20.0	...

5 rows × 22 columns

Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
29	1019.7	1015.0	7	7	14.4	23.6	No	3.6	Yes
36	1012.4	1008.4	5	3	17.5	25.7	Yes	3.6	Yes
69	1009.5	1007.2	8	7	15.4	20.2	Yes	39.8	Yes
56	1005.5	1007.0	2	7	13.5	14.1	Yes	2.8	Yes
49	1018.3	1018.5	7	7	11.1	15.4	Yes	0.0	No

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm
113	12.7	28.6	6.6	3.2	8.6	W	50.0	NaN	W
211	-0.1	18.0	0.0	2.0	8.6	NaN	NaN	N	NNW
271	0.8	12.2	0.0	1.8	8.6	NaN	NaN	S	S
311	-0.9	16.7	0.0	2.8	8.6	NW	35.0	WNW	NW
322	1.1	18.0	0.0	1.6	8.6	NNW	39.0	NNW	NNW
352	6.7	24.7	0.0	5.4	8.6	NW	43.0	N	NW

6 rows × 22 columns

WindSpeed9am	...	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
NaN	...	42	1008.3	1002.3	5	5	16.5	27.4	Yes	0.0	No
2.0	...	46	1033.2	1028.7	1	1	5.6	17.4	No	0.0	No
19.0	...	49	1020.9	1016.8	3	1	6.5	11.2	No	0.0	No
7.0	...	33	1018.1	1013.9	7	7	6.4	15.5	No	1.0	No
2.0	...	36	1021.0	1015.6	4	1	7.0	17.8	No	0.0	No
4.0	...	31	1025.7	1022.2	1	7	12.7	23.7	No	0.0	No

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 22 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   MinTemp        366 non-null    float64
 1   MaxTemp        366 non-null    float64
 2   Rainfall       366 non-null    float64
 3   Evaporation    366 non-null    float64
 4   Sunshine       363 non-null    float64
 5   WindGustDir    363 non-null    object 
 6   WindGustSpeed  364 non-null    float64
 7   WindDir9am     335 non-null    object 
 8   WindDir3pm     365 non-null    object 
 9   WindSpeed9am   359 non-null    float64
 10  WindSpeed3pm   366 non-null    int64  
 11  Humidity9am    366 non-null    int64  
 12  Humidity3pm    366 non-null    int64  
 13  Pressure9am    366 non-null    float64
 14  Pressure3pm    366 non-null    float64
 15  Cloud9am       366 non-null    int64  
 16  Cloud3pm       366 non-null    int64  
 17  Temp9am        366 non-null    float64
 18  Temp3pm        366 non-null    float64
 19  RainToday      366 non-null    object 
 20  RISK_MM        366 non-null    float64
 21  RainTomorrow   366 non-null    object 
dtypes: float64(12), int64(5), object(5)
memory usage: 63.0+ KB
```

```
MinTemp          float64
MaxTemp          float64
Rainfall         float64
Evaporation      float64
Sunshine         float64
WindGustDir       object
WindGustSpeed    float64
WindDir9am        object
WindDir3pm        object
WindSpeed9am     float64
WindSpeed3pm       int64
Humidity9am        int64
Humidity3pm        int64
Pressure9am      float64
Pressure3pm      float64
Cloud9am           int64
Cloud3pm           int64
Temp9am          float64
Temp3pm          float64
RainToday         object
RISK_MM          float64
RainTomorrow      object
dtype: object
```

```
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
```

	count	mean	std	min	25%	50%	75%	max
MinTemp	366.0	7.265574	6.025800	-5.3	2.300	7.45	12.500	20.9
MaxTemp	366.0	20.550273	6.690516	7.6	15.025	19.65	25.500	35.8
Rainfall	366.0	1.428415	4.225800	0.0	0.000	0.00	0.200	39.8
Evaporation	366.0	4.521858	2.669383	0.2	2.200	4.20	6.400	13.8
Sunshine	363.0	7.909366	3.481517	0.0	5.950	8.60	10.500	13.6
WindGustSpeed	364.0	39.840659	13.059807	13.0	31.000	39.00	46.000	98.0
WindSpeed9am	359.0	9.651811	7.951929	0.0	6.000	7.00	13.000	41.0
WindSpeed3pm	366.0	17.986339	8.856997	0.0	11.000	17.00	24.000	52.0
Humidity9am	366.0	72.035519	13.137058	36.0	64.000	72.00	81.000	99.0
Humidity3pm	366.0	44.519126	16.850947	13.0	32.250	43.00	55.000	96.0
Pressure9am	366.0	1019.709016	6.686212	996.5	1015.350	1020.15	1024.475	1035.7
Pressure3pm	366.0	1016.810383	6.469422	996.8	1012.800	1017.40	1021.475	1033.2
Cloud9am	366.0	3.890710	2.956131	0.0	1.000	3.50	7.000	8.0
Cloud3pm	366.0	4.024590	2.666268	0.0	1.000	4.00	7.000	8.0
Temp9am	366.0	12.358470	5.630832	0.1	7.625	12.55	17.000	24.7
Temp3pm	366.0	19.230874	6.640346	5.1	14.150	18.55	24.000	34.5
RISK_MM	366.0	1.428415	4.225800	0.0	0.000	0.00	0.200	39.8

```
MinTemp          0
MaxTemp          0
Rainfall         0
Evaporation      0
Sunshine         3
WindGustDir      3
WindGustSpeed    2
WindDir9am      31
WindDir3pm       1
WindSpeed9am     7
WindSpeed3pm     0
Humidity9am      0
Humidity3pm      0
Pressure9am      0
Pressure3pm      0
Cloud9am         0
Cloud3pm         0
Temp9am          0
Temp3pm          0
RainToday        0
RISK_MM          0
RainTomorrow     0
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 22 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   MinTemp        366 non-null    float64
 1   MaxTemp        366 non-null    float64
 2   Rainfall       366 non-null    float64
 3   Evaporation    366 non-null    float64
 4   Sunshine       363 non-null    float64
 5   WindGustDir    363 non-null    object 
 6   WindGustSpeed  364 non-null    float64
 7   WindDir9am     335 non-null    object 
 8   WindDir3pm     365 non-null    object 
 9   WindSpeed9am   359 non-null    float64
 10  WindSpeed3pm   366 non-null    int64  
 11  Humidity9am    366 non-null    int64  
 12  Humidity3pm    366 non-null    int64  
 13  Pressure9am    366 non-null    float64
 14  Pressure3pm    366 non-null    float64
 15  Cloud9am       366 non-null    int64  
 16  Cloud3pm       366 non-null    int64  
 17  Temp9am        366 non-null    float64
 18  Temp3pm        366 non-null    float64
 19  RainToday      366 non-null    object 
 20  RISK_MM        366 non-null    float64
 21  RainTomorrow   366 non-null    object 
dtypes: float64(12), int64(5), object(5)
memory usage: 63.0+ KB
```

```
['MinTemp',
 'MaxTemp',
 'Rainfall',
 'Evaporation',
 'Sunshine',
 'WindGustDir',
 'WindGustSpeed',
 'WindDir9am',
 'WindDir3pm',
 'WindSpeed9am',
 'WindSpeed3pm',
 'Humidity9am',
 'Humidity3pm',
 'Pressure9am',
 'Pressure3pm',
 'Cloud9am',
 'Cloud3pm',
 'Temp9am',
 'Temp3pm',
 'RainToday',
 'RISK_MM',
 'RainTomorrow']
```

```
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
```

```
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
```

```
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
```

```
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
```

df.columns
df.to_csv('cln_weather_data.csv', encoding='utf-8', index=False)
df['MaxTemp'].value_counts()

```
Index(['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine',
       'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm',
       'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',
       'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am',
       'Temp3pm', 'RainToday', 'RISK_MM', 'RainTomorrow'],
      dtype='object')
```

```
MaxTemp
14.8    5
20.9    5
12.2    5
15.5    5
11.6    5
       ..
17.1    1
30.3    1
35.8    1
21.2    1
30.7    1
Name: count, Length: 187, dtype: int64
```

df[df['Rainfall'] <= 3.6]

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...
0	8.0	24.3	0.0	3.4	6.3	NW	30.0	SW	NW	6.0	...
1	14.0	26.9	3.6	4.4	9.7	ENE	39.0	E	W	4.0	...
2	13.7	23.4	3.6	5.8	3.3	NW	85.0	N	NNE	6.0	...
4	7.6	16.1	2.8	5.6	10.6	SSE	50.0	SSE	ESE	20.0	...
5	6.2	16.9	0.0	5.8	8.2	SE	44.0	SE	E	20.0	...
...	...	...	...	...	...	...	...	...	...	...	...
361	9.0	30.7	0.0	7.6	12.1	NNW	76.0	SSE	NW	7.0	...
362	7.1	28.4	0.0	11.6	12.7	N	48.0	NNW	NNW	2.0	...
363	12.5	19.9	0.0	8.4	5.3	ESE	43.0	ENE	ENE	11.0	...
364	12.5	26.9	0.0	5.0	7.1	NW	46.0	SSW	WNW	6.0	...
365	12.3	30.2	0.0	6.0	12.6	NW	78.0	NW	WNW	31.0	...

323 rows × 22 columns

Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
29	1019.7	1015.0	7	7	14.4	23.6	No	3.6	Yes
36	1012.4	1008.4	5	3	17.5	25.7	Yes	3.6	Yes
69	1009.5	1007.2	8	7	15.4	20.2	Yes	39.8	Yes
49	1018.3	1018.5	7	7	11.1	15.4	Yes	0.0	No
57	1023.8	1021.7	7	5	10.9	14.8	No	0.2	No
...	...	...	...	...	...	...	...	...	...
15	1016.1	1010.8	1	3	20.4	30.0	No	0.0	No
22	1020.0	1016.9	0	1	17.2	28.2	No	0.0	No
47	1024.0	1022.8	3	2	14.5	18.3	No	0.0	No
39	1021.0	1016.2	6	7	15.8	25.9	No	0.0	No
13	1009.6	1009.2	1	1	23.8	28.6	No	0.0	No

df.groupby('WindGustDir').get_group('NW')

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...
0	8.0	24.3	0.0	3.4	6.3	NW	30.0	SW	NW	6.0	...
2	13.7	23.4	3.6	5.8	3.3	NW	85.0	N	NNE	6.0	...
3	13.3	15.5	39.8	7.2	9.1	NW	54.0	WNW	W	30.0	...
13	12.1	30.9	0.0	6.2	12.4	NW	44.0	WNW	W	7.0	...
14	10.1	31.2	0.0	8.8	13.1	NW	41.0	S	W	6.0	...
...	...	...	...	...	...	...	...	...	...	...	...
352	6.7	24.7	0.0	5.4	8.6	NW	43.0	N	NW	4.0	...
353	8.3	28.5	0.0	5.8	9.8	NW	46.0	W	NW	2.0	...
354	11.3	27.4	0.2	7.6	12.1	NW	52.0	SE	NW	6.0	...
364	12.5	26.9	0.0	5.0	7.1	NW	46.0	SSW	WNW	6.0	...
365	12.3	30.2	0.0	6.0	12.6	NW	78.0	NW	WNW	31.0	...

73 rows × 22 columns

Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
29	1019.7	1015.0	7	7	14.4	23.6	No	3.6	Yes
69	1009.5	1007.2	8	7	15.4	20.2	Yes	39.8	Yes
56	1005.5	1007.0	2	7	13.5	14.1	Yes	2.8	Yes
20	1017.3	1013.1	1	4	19.7	30.7	No	0.0	No
16	1018.2	1013.7	0	1	18.7	30.4	No	0.0	No
...	...	...	...	...	...	...	...	...	...
31	1025.7	1022.2	1	7	12.7	23.7	No	0.0	No
30	1024.1	1019.8	1	6	16.8	27.4	No	0.2	No
20	1021.4	1017.5	1	1	16.4	26.3	No	0.0	No
39	1021.0	1016.2	6	7	15.8	25.9	No	0.0	No
13	1009.6	1009.2	1	1	23.8	28.6	No	0.0	No

```
import matplotlib
import matplotlib.pyplot as plt
fig = plt.figure()
#ax = fig.add_axes([0,1,1,1])
plt.bar(df['Cloud3pm'],df['RainTomorrow'],color = ['maroon','red'])
plt.title('Bar Chart')
plt.xlabel("Clouds")
plt.ylabel("RainTommorrow")
plt.show()
```

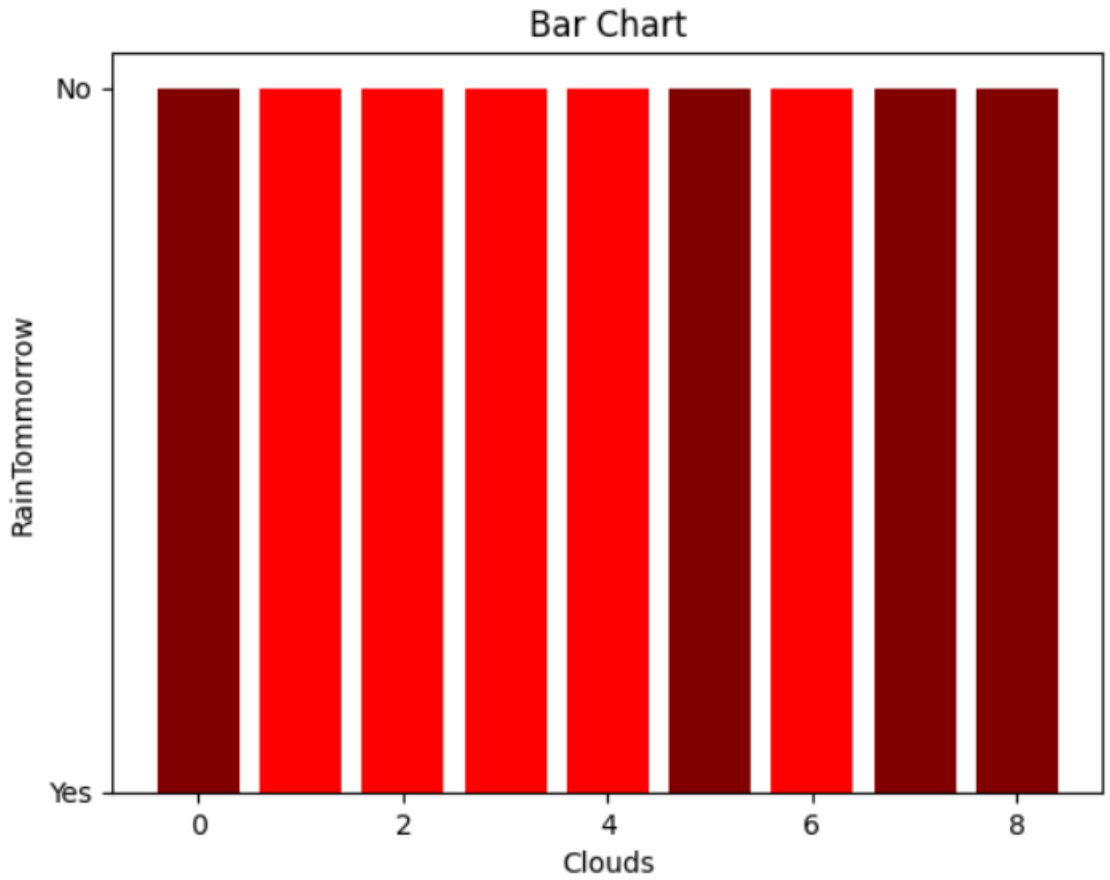
```
<BarContainer object of 366 artists>

Text(0.5, 1.0, 'Bar Chart')

Text(0.5, 0, 'Clouds')

Text(0, 0.5, 'RainTommorrow')
```

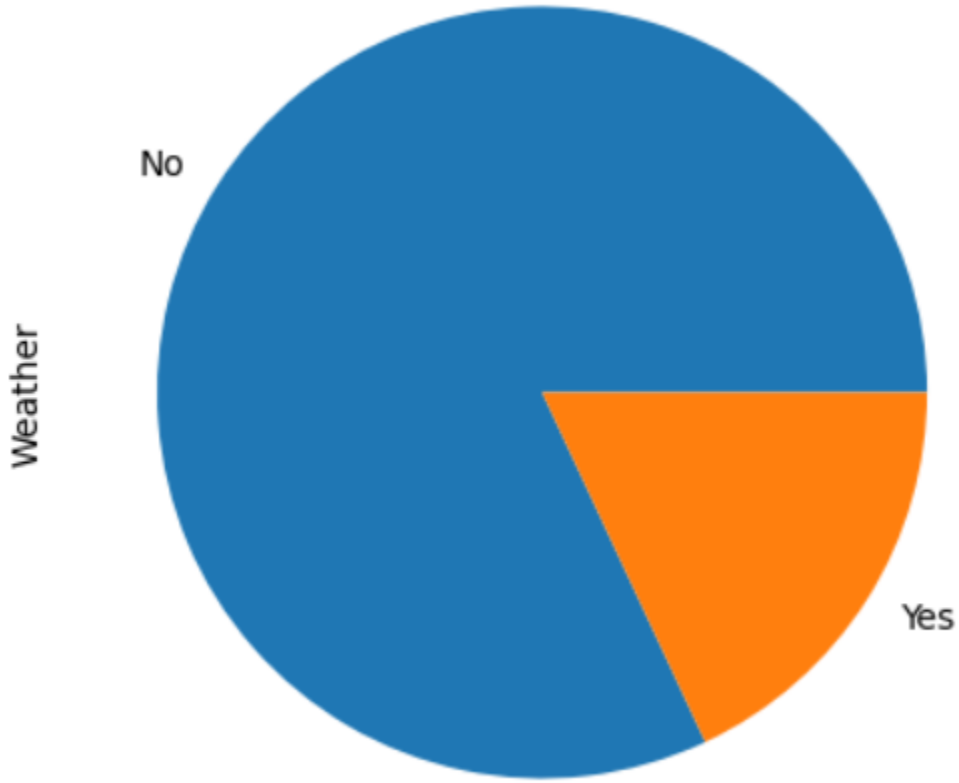
```
weather = dict(df['RainTomorrow'].value_counts())
series = pd.Series(weather.values(), index=weather.keys(), name="Weather")
series.plot.pie(figsize=(5, 5))
```

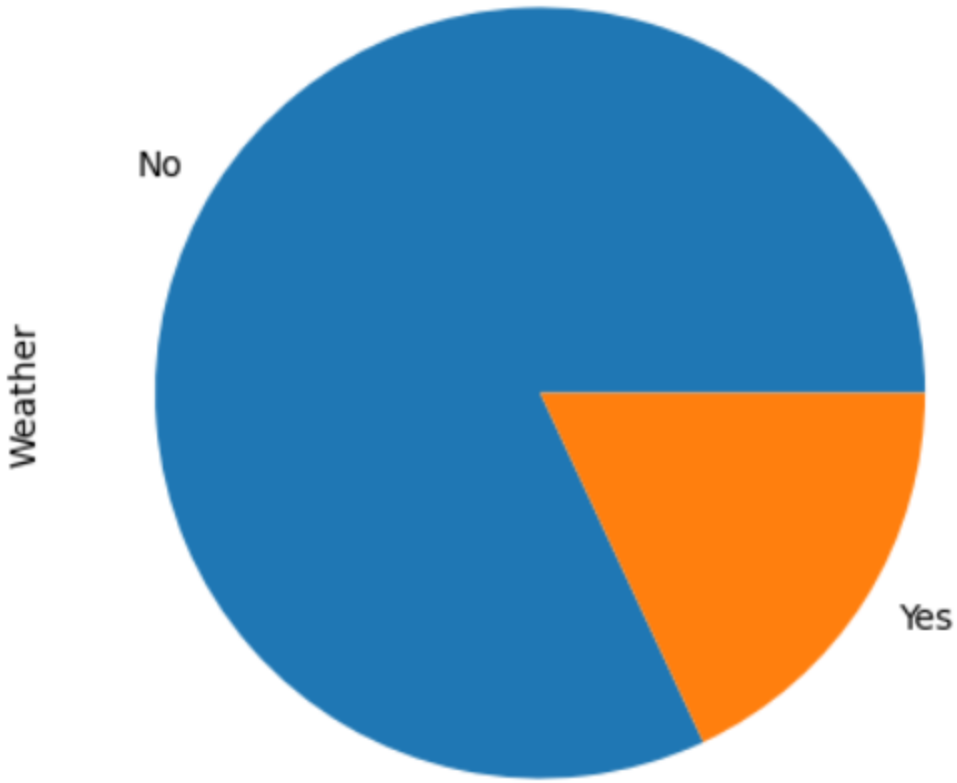
```
<Axes: ylabel='Weather'>
```

```
weather = dict(df['RainToday'].value_counts())
series = pd.Series(weather.values(), index=weather.keys(), name="Weather")
series.plot.pie(figsize=(5, 5))
```

```
<Axes: ylabel='Weather'>
```


Bar Chart
No
Yes
RainTommorrow
0
2
4
6
8
Clouds


No
Weather
Yes


No
Weather
Yes

df[(df['WindDir9am'] == 'E')& (df['WindDir3pm'] == 'E')]

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...
25	15.6	26.9	0.0	6.8	8.9	E	41.0	E	E	6.0	...
26	13.3	22.2	0.2	6.6	2.3	ENE	39.0	E	E	20.0	...

2 rows × 22 columns

Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
48	1019.7	1016.5	2	4	19.8	25.1	No	0.2	No
55	1021.0	1018.6	7	7	16.5	21.2	No	0.0	No

df[(df['RainToday'] == 'No')& (df['WindDir3pm'] == 'E')]

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...
5	6.2	16.9	0.0	5.8	8.2	SE	44.0	SE	E	20.0	...
7	8.3	17.0	0.0	5.6	4.6	E	41.0	SE	E	11.0	...
21	16.4	19.4	0.4	9.2	0.0	E	26.0	ENE	E	6.0	...
25	15.6	26.9	0.0	6.8	8.9	E	41.0	E	E	6.0	...
26	13.3	22.2	0.2	6.6	2.3	ENE	39.0	E	E	20.0	...
34	14.5	21.8	0.0	8.4	9.8	ENE	43.0	ESE	E	11.0	...
40	11.8	18.5	0.6	4.8	2.3	ENE	35.0	ESE	E	9.0	...
41	11.7	21.5	0.0	4.2	7.3	ENE	41.0	ESE	E	15.0	...
42	9.6	20.3	0.0	5.0	3.6	SE	39.0	ESE	E	22.0	...
47	7.5	20.9	0.0	6.6	8.7	ENE	39.0	SE	E	13.0	...
65	12.1	27.5	0.0	7.4	11.7	NE	35.0	SSE	E	7.0	...
83	9.9	24.4	0.0	5.8	10.8	NE	28.0	SE	E	7.0	...
106	10.8	25.2	0.0	5.6	12.6	ENE	35.0	SE	E	7.0	...
107	11.2	26.1	0.0	7.2	12.6	ENE	39.0	SE	E	7.0	...
173	5.6	18.9	0.0	3.8	8.0	SE	33.0	SSE	E	13.0	...
203	2.4	14.7	0.0	2.6	9.2	ESE	20.0	SSE	E	7.0	...
266	-1.0	12.2	0.0	1.6	8.4	ESE	30.0	SE	E	7.0	...

17 rows × 22 columns

Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
57	1023.8	1021.7	7	5	10.9	14.8	No	0.2	No
57	1026.2	1024.2	6	7	12.1	15.5	No	0.0	No
72	1010.7	1008.9	8	8	16.5	18.3	No	25.8	Yes
48	1019.7	1016.5	2	4	19.8	25.1	No	0.2	No
55	1021.0	1018.6	7	7	16.5	21.2	No	0.0	No
63	1015.0	1015.3	7	1	18.6	20.5	No	0.0	No
63	1018.0	1017.4	8	6	13.8	16.4	No	0.0	No
51	1021.1	1019.6	7	5	15.7	19.4	No	0.0	No
49	1021.9	1019.3	7	7	15.8	19.2	No	0.0	No
43	1020.6	1018.3	6	7	15.3	19.8	No	0.0	No
40	1011.8	1007.4	7	2	18.6	26.8	No	0.0	No
38	1021.6	1019.2	1	4	16.2	22.7	No	0.0	No
32	1020.2	1016.7	1	1	15.0	24.3	No	0.0	No
40	1022.9	1020.0	1	1	15.3	24.7	No	0.0	No
49	1025.8	1023.5	7	6	12.5	17.7	No	0.0	No
43	1030.4	1026.0	1	1	5.9	13.8	No	0.0	No
51	1032.2	1027.4	3	5	3.7	10.8	No	0.0	No

CONCLUSION

We successfully predicted the rainfall using the linear regression but here this is not very accurate only sometimes any way it depends upon the climate changes to season to season.

Here we are taking only summer season weather data set it only useful to predict rainfall in summer season. Finally, it is agreed that we made an attempt on the following points:

- ✓ **The description of the purpose the scope and applicability of this project.**
- ✓ **We specify the system's necessary specs as well as the actions that can be performed on these objects.**
- ✓ **We define the system's required specifications and the actions that can be taken on these objects.**

- ✓ **We comprehend the problem domain and create a system model that represents the operations that can be performed on the system.**
- ✓ **We went into great lengths about the features and processes, providing a lot of important information.**
- ✓ **We created the user interface as well as system security issues.**
- ✓ **Finally, the system is built and tested in accordance with the test cases.**