```
import numpy as np
import pandas as pd

import matplotlib as plt
import matplotlib.pyplot as plt
import os

weather = pd.read_csv("/content/weatherHistory.csv")
weather
weather.head(10)
```

```
                    Formatted Date  ...                          Daily
Summary
0  2006-04-01 00:00:00.000 +0200  ...  Partly cloudy throughout the
day.
1  2006-04-01 01:00:00.000 +0200  ...  Partly cloudy throughout the
day.
2  2006-04-01 02:00:00.000 +0200  ...  Partly cloudy throughout the
day.
3  2006-04-01 03:00:00.000 +0200  ...  Partly cloudy throughout the
day.
4  2006-04-01 04:00:00.000 +0200  ...  Partly cloudy throughout the
day.
5  2006-04-01 05:00:00.000 +0200  ...  Partly cloudy throughout the
day.
6  2006-04-01 06:00:00.000 +0200  ...  Partly cloudy throughout the
day.
7  2006-04-01 07:00:00.000 +0200  ...  Partly cloudy throughout the
day.
8  2006-04-01 08:00:00.000 +0200  ...  Partly cloudy throughout the
day.
9  2006-04-01 09:00:00.000 +0200  ...  Partly cloudy throughout the
day.

[10 rows x 12 columns]
```
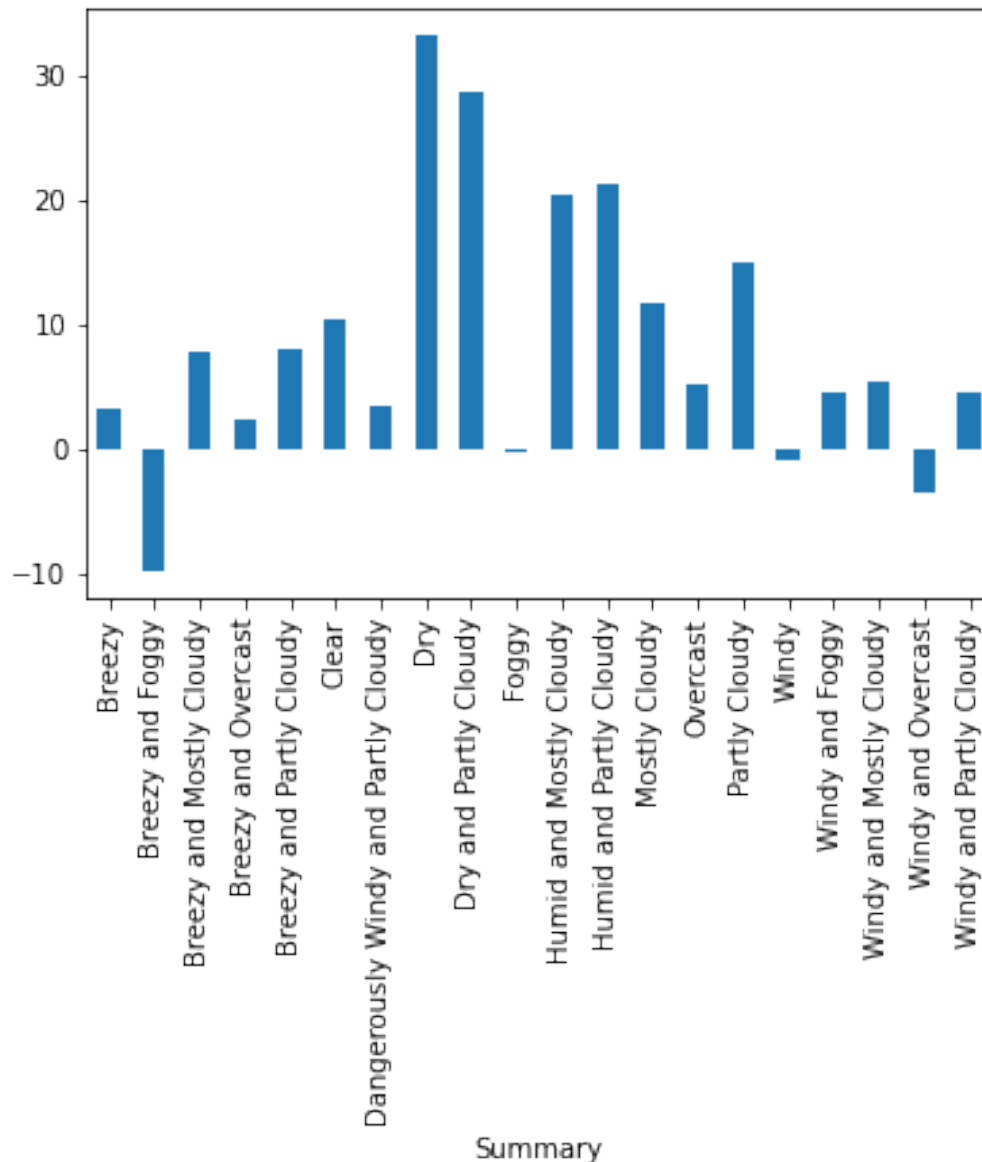
```
weather.groupby('Summary').mean()["Apparent Temperature
(C)"].plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff60a19dc10>
```

```
weather[weather["Summary"] == "Dry"].mean()

Temperature (C)              33.766374
Apparent Temperature (C)     32.451170
Humidity                      0.228421
Wind Speed (km/h)            14.736584
Wind Bearing (degrees)      210.157895
Visibility (km)              10.080295
Loud Cover                    0.000000
Pressure (millibars)       1011.652105
dtype: float64

def convert_summary(col):
    return len(col)
```
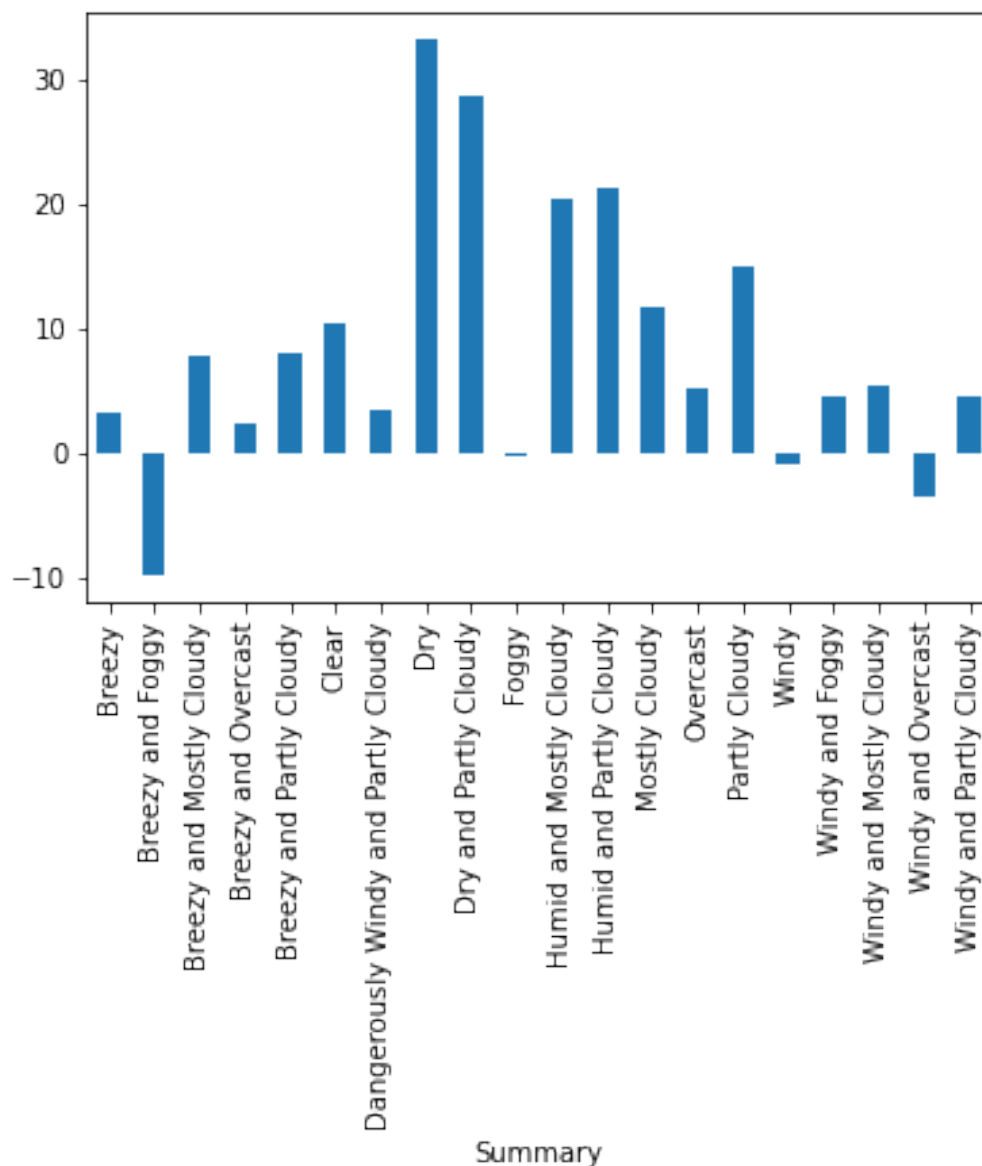
```python
weather_temp = weather[["Humidity","Apparent Temperature (C)"]]

weather_temp.head(12)
```

```
    Humidity  Apparent Temperature (C)
0       0.89                  7.388889
1       0.86                  7.227778
2       0.89                  9.377778
3       0.83                  5.944444
4       0.83                  6.977778
5       0.85                  7.111111
6       0.95                  5.522222
7       0.89                  6.527778
8       0.82                 10.822222
9       0.72                 13.772222
10      0.67                 16.016667
11      0.54                 17.144444
```

```python
weather.groupby('Summary').mean()["Apparent Temperature (C)"].plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff60a073310>
```

```
dummies = pd.get_dummies(weather["Summary"])
dummies.head(122)
```

	Breezy	Breezy and Foggy	...	Windy and Overcast	Windy and Partly Cloudy
0	0	0	...	0	0
1	0	0	...	0	0
2	0	0	...	0	0
3	0	0	...	0	0
4	0	0	...	0	0

```
 ..        ...              ...  ...                          ...
 ...
117        0                0  ...                            0
0
118        0                0  ...                            0
0
119        0                0  ...                            0
0
120        0                0  ...                            0
0
121        0                0  ...                            0
0

[122 rows x 20 columns]

weather_temp2 = pd.concat([weather_temp,dummies],axis=1)
weather_temp2.head(12)

    Humidity   ...   Windy and Partly Cloudy
0       0.89  ...                         0
1       0.86  ...                         0
2       0.89  ...                         0
3       0.83  ...                         0
4       0.83  ...                         0
5       0.85  ...                         0
6       0.95  ...                         0
7       0.89  ...                         0
8       0.82  ...                         0
9       0.72  ...                         0
10      0.67  ...                         0
11      0.54  ...                         0

[12 rows x 22 columns]


Y = weather_temp["Apparent Temperature (C)"]
X = weather_temp2

X

       Humidity   ...   Windy and Partly Cloudy
0          0.89  ...                         0
1          0.86  ...                         0
2          0.89  ...                         0
3          0.83  ...                         0
4          0.83  ...                         0
...         ...  ...                       ...
24503      0.78  ...                         0
24504      0.79  ...                         0
24505      0.72  ...                         0
24506      0.75  ...                         0
```

```
24507       0.75  ...                        0

[24508 rows x 22 columns]


x = (X >=0)
x.sum()

print(x.sum(),X.shape)
non_zero_humidity = X >= 0
print(non_zero_humidity.sum())

Humidity                              24508
Apparent Temperature (C)              19787
Breezy                                24508
Breezy and Foggy                      24508
Breezy and Mostly Cloudy              24508
Breezy and Overcast                   24508
Breezy and Partly Cloudy              24508
Clear                                 24508
Dangerously Windy and Partly Cloudy   24508
Dry                                   24508
Dry and Partly Cloudy                 24508
Foggy                                 24508
Humid and Mostly Cloudy               24508
Humid and Partly Cloudy               24508
Mostly Cloudy                         24508
Overcast                              24508
Partly Cloudy                         24508
Windy                                 24508
Windy and Foggy                       24508
Windy and Mostly Cloudy               24508
Windy and Overcast                    24508
Windy and Partly Cloudy               24508
dtype: int64 (24508, 22)
Humidity                              24508
Apparent Temperature (C)              19787
Breezy                                24508
Breezy and Foggy                      24508
Breezy and Mostly Cloudy              24508
Breezy and Overcast                   24508
Breezy and Partly Cloudy              24508
Clear                                 24508
Dangerously Windy and Partly Cloudy   24508
Dry                                   24508
Dry and Partly Cloudy                 24508
Foggy                                 24508
Humid and Mostly Cloudy               24508
Humid and Partly Cloudy               24508
Mostly Cloudy                         24508
```

```
Overcast                                24508
Partly Cloudy                           24508
Windy                                   24508
Windy and Foggy                         24508
Windy and Mostly Cloudy                 24508
Windy and Overcast                      24508
Windy and Partly Cloudy                 24508
dtype: int64
```

```python
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,Y,test_size=0.3)
```

```python
X_train.shape, y_train.shape
```

```
((17155, 22), (17155,))
```

```python
y_train.head(10)
```

```
18234    11.083333
10223    22.644444
2754     -6.550000
13499    27.066667
20093    -0.005556
22983     3.950000
9420     16.138889
14340    12.288889
154      12.777778
20952    -4.561111
Name: Apparent Temperature (C), dtype: float64
```

```python
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
```

```python
X_train.shape,y_train.shape
```
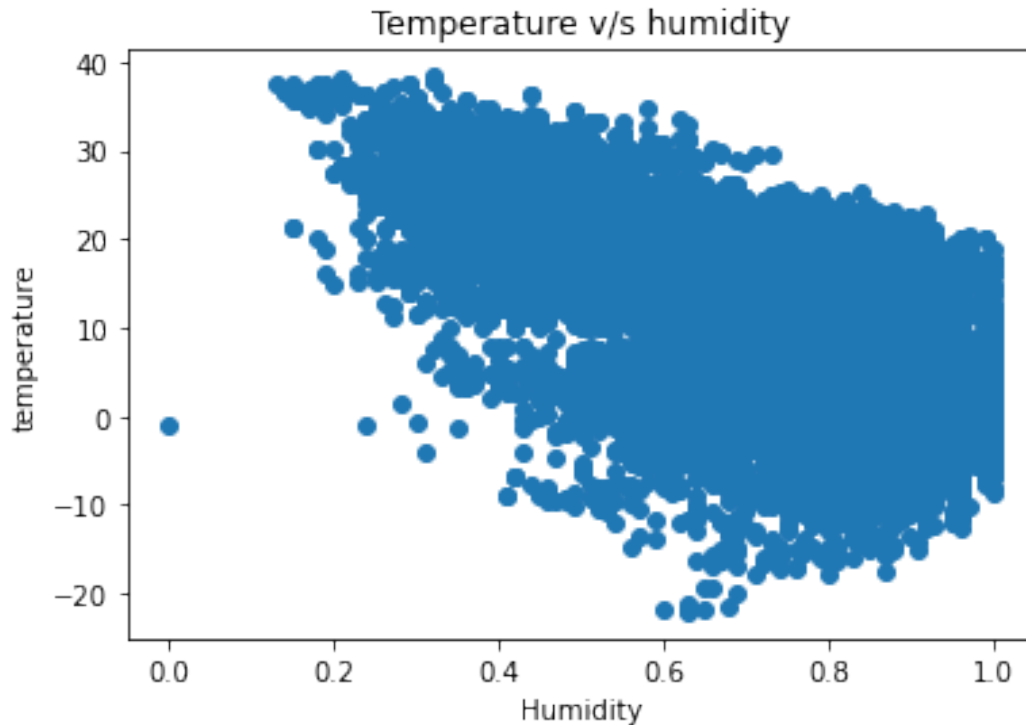
```
((17155, 22), (17155,))
```

```python
plt.scatter(X_train["Humidity"], y_train)
```

```python
plt.title("Temperature v/s humidity")
plt.xlabel("Humidity")
plt.ylabel("temperature")
```

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

Temperature v/s humidity

```
pred = model.predict(X_test)
pred
```

```
array([27.98888889, 22.22222222,  5.86666667, ...,  5.01666667,
        23.35555556, -2.42777778])
```

```
y_predict = model.predict(X_test)
```

```
y_pred = y_predict.reshape(1,-1)
y_pred
```

```
array([[27.98888889, 22.22222222,  5.86666667, ...,  5.01666667,
        23.35555556, -2.42777778]])
```

```
y_test.head()
```
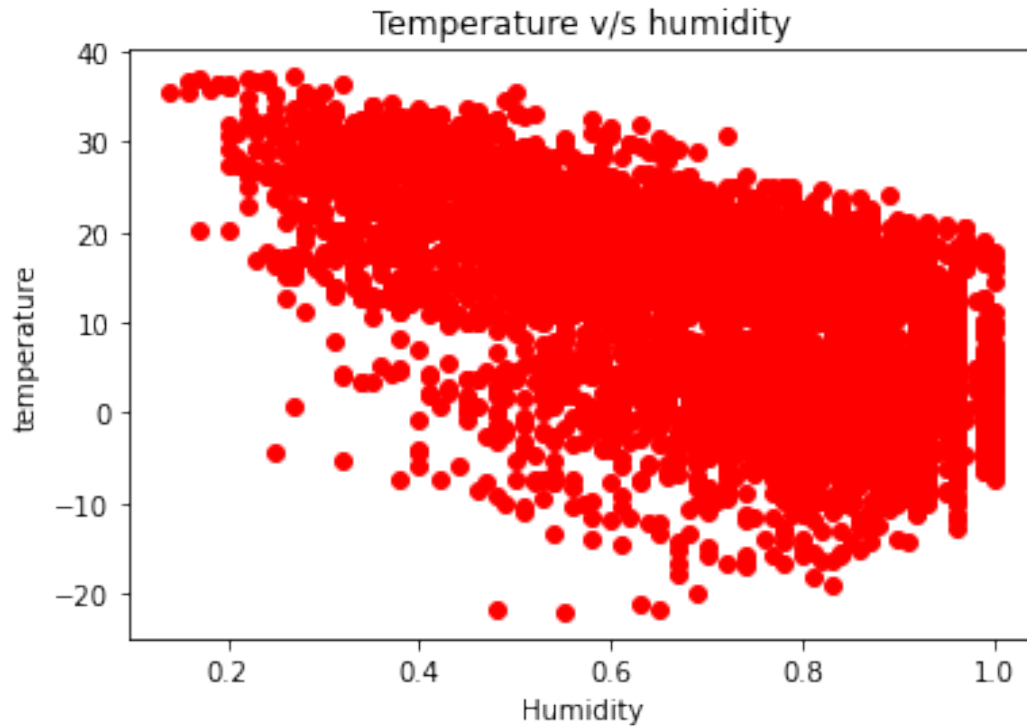
```
12572     27.988889
7667      22.222222
14575      5.866667
16829     15.600000
20326      3.416667
Name: Apparent Temperature (C), dtype: float64
```

```
plt.scatter(X_test["Humidity"],y_predict,color='red')
```

```
plt.title("Temperature v/s humidity")
plt.xlabel("Humidity")
plt.ylabel("temperature")
```

Text(0, 0.5, 'temperature')

Temperature v/s humidity


```
plt.scatter(X_test["Humidity"],y_test,color="green")

plt.title("Temperature v/s humidity")
plt.xlabel("Humidity")
plt.ylabel("temperature")
```
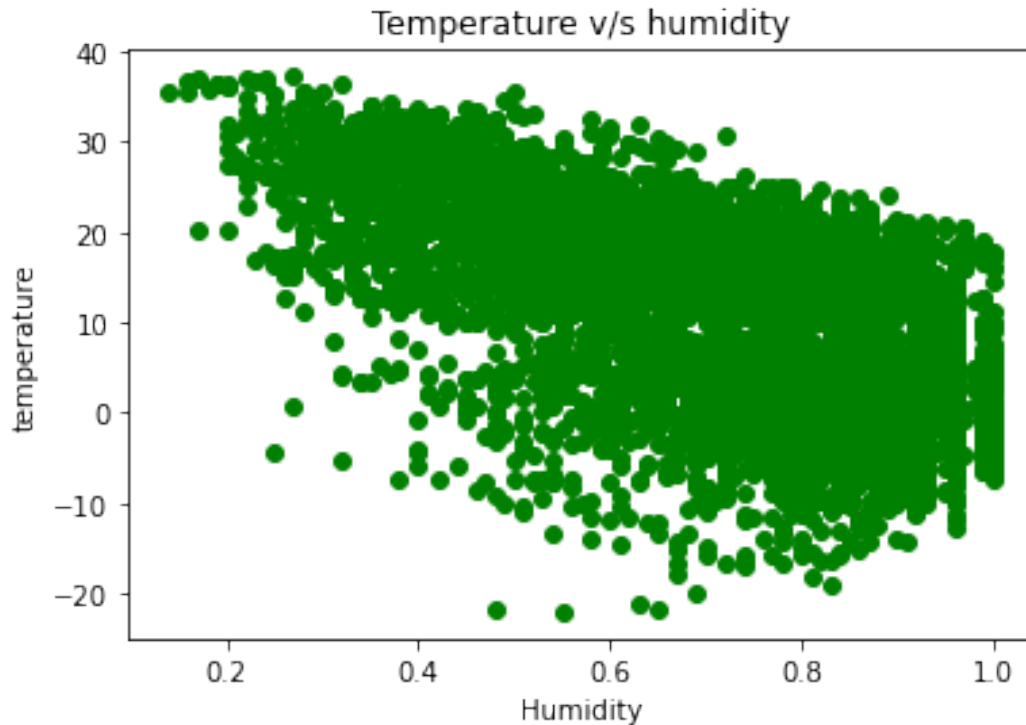
Text(0, 0.5, 'temperature')

Temperature v/s humidity

weather.iloc[3703]

```
Formatted Date              2006-07-12 07:00:00.000 +0200
Summary                                    Partly Cloudy
Precip Type                                         rain
Temperature (C)                                  21.0222
Apparent Temperature (C)                         21.0222
Humidity                                             0.8
Wind Speed (km/h)                                 3.1234
Wind Bearing (degrees)                               127
Visibility (km)                                   9.9015
Loud Cover                                             0
Pressure (millibars)                             1020.42
Daily Summary               Partly cloudy until night.
Name: 3703, dtype: object
```