Performing Analysis of Meteorological Data

My name is Aarav Chauhan. I am undertaking B.Tech (a pre-final year student) from NIET Greater Noida. Currently, I am working as Intern at Suven Consultant Pvt. Ltd.

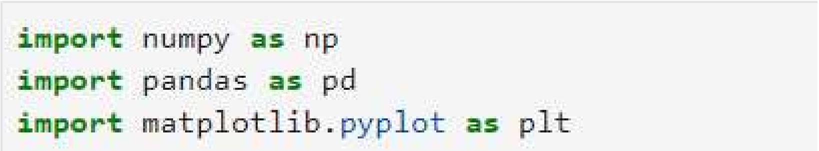
Today, I am here to explain my project i.e., "PERFORMING ANALYSIS OF METEOROLOGICAL DATA".

In this blog we will discuss about a data_analysis which is based on the following dataset.

"Has the Apparent temperature and humidity compared monthly across 10 years of the data indicate an increase due to Global warming" following is the Hypothesis for the analysis.

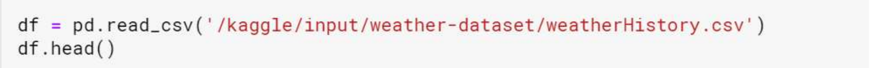
The Hypothesis means we need to find whether the average Apparent temperature for the month of a month say April starting from 2006 to 2016 and the average humidity for the same period have increased or not. This monthly analysis has to be done for all 12 months over the 10 year period. So you are basically resampling your data from hourly to monthly, then comparing the same month over the 10 year period. Support your analysis by appropriate visualizations using matplotlib and / or seaborn library.

> **Step 1: Importing of libraries and Dataset.**



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

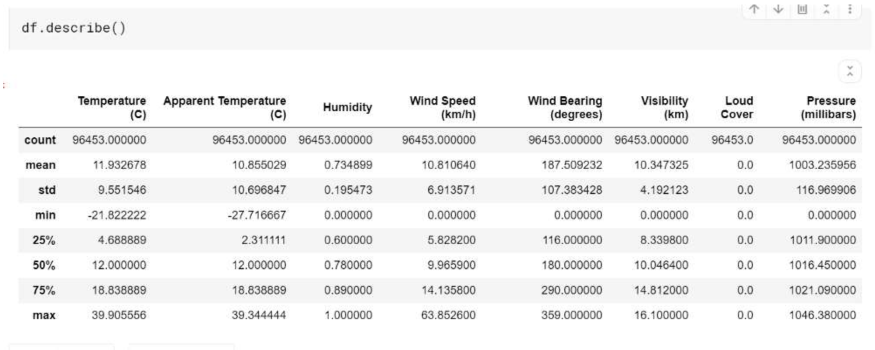
Required libraries



```
df = pd.read_csv('/kaggle/input/weather-dataset/weatherHistory.csv')
df.head()
```

Importing dataset

Step 2: Looking at the dataset.



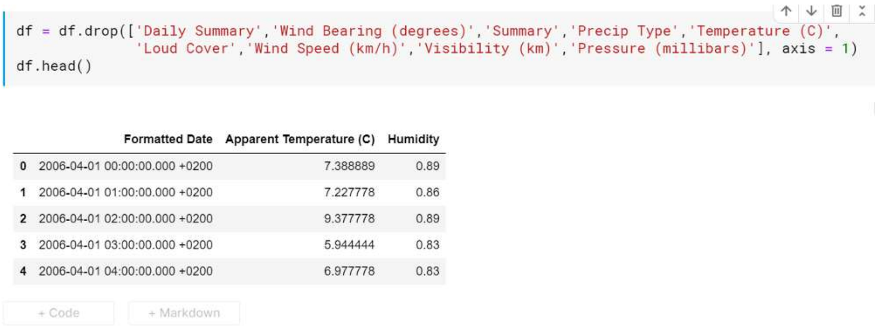
```
df.describe()
```

	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)
count	96453.000000	96453.000000	96453.000000	96453.000000	96453.000000	96453.000000	96453.0	96453.000000
mean	11.932678	10.855029	0.734899	10.810640	187.509232	10.347325	0.0	1003.235956
std	9.551546	10.696847	0.195473	6.913571	107.383428	4.192123	0.0	116.969906
min	-21.822222	-27.716667	0.000000	0.000000	0.000000	0.000000	0.0	0.000000
25%	4.688889	2.311111	0.600000	5.828200	116.000000	8.339800	0.0	1011.900000
50%	12.000000	12.000000	0.780000	9.965900	180.000000	10.046400	0.0	1016.450000
75%	18.838889	18.838889	0.890000	14.135800	290.000000	14.812000	0.0	1021.090000
max	39.905556	39.344444	1.000000	63.852600	359.000000	16.100000	0.0	1046.380000

Details of dataset

Step 3: Cleaning Dataset

In this step we will prepare our data for the plotting , we will first drop the unwanted columns (all except temperature and humidity) .

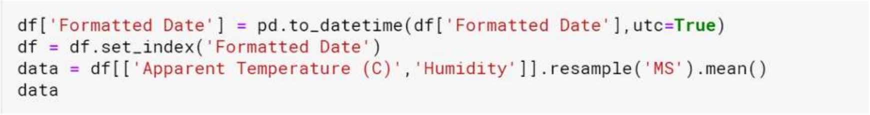


```
df = df.drop(['Daily Summary','Wind Bearing (degrees)','Summary','Precip Type','Temperature (C)',
              'Loud Cover','Wind Speed (km/h)','Visibility (km)','Pressure (millibars)'], axis = 1)
df.head()
```

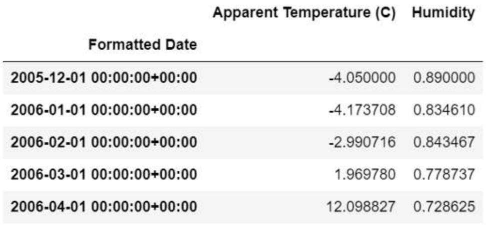
	Formatted Date	Apparent Temperature (C)	Humidity
0	2006-04-01 00:00:00.000 +0200	7.388889	0.89
1	2006-04-01 01:00:00.000 +0200	7.227778	0.86
2	2006-04-01 02:00:00.000 +0200	9.377778	0.89
3	2006-04-01 03:00:00.000 +0200	5.944444	0.83
4	2006-04-01 04:00:00.000 +0200	6.977778	0.83

Dropping the unwanted columns

Then ,we will convert the Timezone to +00:00 UTC .



```
df['Formatted Date'] = pd.to_datetime(df['Formatted Date'],utc=True)
df = df.set_index('Formatted Date')
data = df[['Apparent Temperature (C)','Humidity']].resample('MS').mean()
data
```



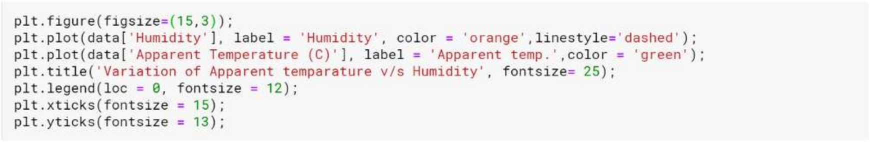
	Apparent Temperature (C)	Humidity
Formatted Date		
2005-12-01 00:00:00+00:00	-4.050000	0.890000
2006-01-01 00:00:00+00:00	-4.173708	0.834610
2006-02-01 00:00:00+00:00	-2.990716	0.843467
2006-03-01 00:00:00+00:00	1.969780	0.778737
2006-04-01 00:00:00+00:00	12.098827	0.728625

Converting the Timezone to +00:00 UTC

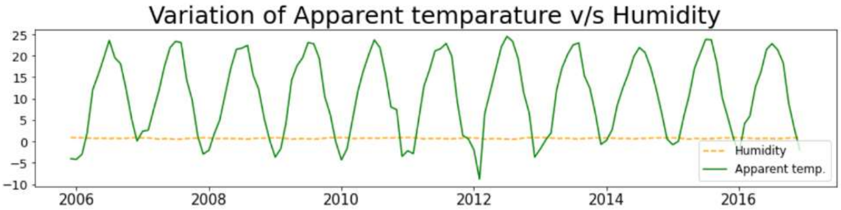
Step 4: Plotting of Data

In is final step we will plot the data to for the analysis ,

> Firstly we will plot the whole dataset for all months .

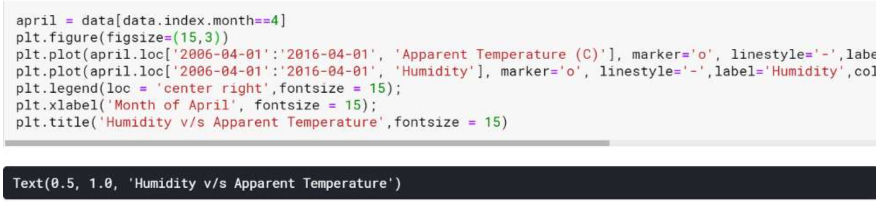


```
plt.figure(figsize=(15,3));
plt.plot(data['Humidity'], label = 'Humidity', color = 'orange',linestyle='dashed');
plt.plot(data['Apparent Temperature (C)'], label = 'Apparent temp.',color = 'green');
plt.title('Variation of Apparent temparature v/s Humidity', fontsize= 25);
plt.legend(loc = 0, fontsize = 12);
plt.xticks(fontsize = 15);
plt.yticks(fontsize = 13);
```



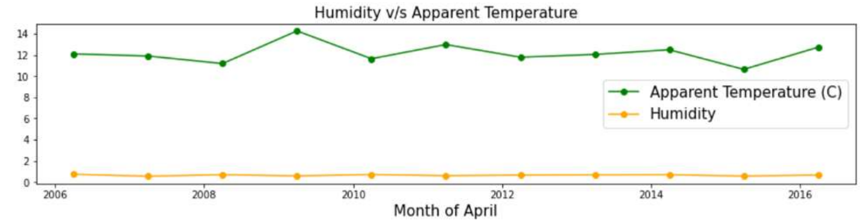
Graph for all months

> Now we will plot graph for a specific month(April).



```
april = data[data.index.month==4]
plt.figure(figsize=(15,3))
plt.plot(april.loc['2006-04-01':'2016-04-01', 'Apparent Temperature (C)'], marker='o', linestyle='-',labe
plt.plot(april.loc['2006-04-01':'2016-04-01', 'Humidity'], marker='o', linestyle='-',label='Humidity',col
plt.legend(loc = 'center right',fontsize = 15);
plt.xlabel('Month of April', fontsize = 15);
plt.title('Humidity v/s Apparent Temperature',fontsize = 15)
```

```
Text(0.5, 1.0, 'Humidity v/s Apparent Temperature')
```



Graph for month of April

Conclusion

As we can analyze there isn’t any change in humidity in past 10 years(2006–2010) for the month of April. where as , temperature increases sharply in 2009 and drops in 2015 for rest of the years there isn’t any sharp change in the temperature.

Note: This conclusion statement is for only month of April , please refer to the model for all the months.