

# Project 1

## Performing Analysis of Meteorological Data

**Objective:** Perform Data Cleaning, Perform Analysis for Testing the (given) Hypothesis and finally put-forth your conclusion by writing a blog article.

Data Source : <https://www.kaggle.com/muthuj7/weather-dataset> (<https://www.kaggle.com/muthuj7/weather-dataset>).

The Meteorological Dataset Consists of:

- Total Number of Data Point: 96453
- Total Number of Attribute/Columns in data set is 11.

Attribute Information:

1. Formatted Date
2. Summary
3. Precip Type
4. Temperature (C)
5. Apparent Temperature (C)
6. Humidity
7. Wind Speed (km/h)
8. Wind Bearing (degrees)
9. Visibility (km)
10. Pressure (millibars)
11. Daily Summary

## Loading the Data

In [1]:

```
#Python Library
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
%matplotlib inline
```

In [2]:

```
#Reading weatherHistory.csv file
df = pd.read_csv("weatherHistory.csv")
df.head(10)
```

Out[2]:

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Vis
0	2006-04-01 00:00:00.000 +0200	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251	15
1	2006-04-01 01:00:00.000 +0200	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259	15
2	2006-04-01 02:00:00.000 +0200	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204	14
3	2006-04-01 03:00:00.000 +0200	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269	15
4	2006-04-01 04:00:00.000 +0200	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259	15
5	2006-04-01 05:00:00.000 +0200	Partly Cloudy	rain	9.222222	7.111111	0.85	13.9587	258	14
6	2006-04-01 06:00:00.000 +0200	Partly Cloudy	rain	7.733333	5.522222	0.95	12.3648	259	9
7	2006-04-01 07:00:00.000 +0200	Partly Cloudy	rain	8.772222	6.527778	0.89	14.1519	260	9
8	2006-04-01 08:00:00.000 +0200	Partly Cloudy	rain	10.822222	10.822222	0.82	11.3183	259	9
9	2006-04-01 09:00:00.000 +0200	Partly Cloudy	rain	13.772222	13.772222	0.72	12.5258	279	9



In [3]:

```
#shape of Dataset
df.shape
```

Out[3]:

(96453, 11)

In [4]:

```
#Statistical Summary of DataFrame
df.describe()
```

Out[4]:

	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	
count	96453.000000	96453.000000	96453.000000	96453.000000	96453.000000	96453.000000	96
mean	11.932678	10.855029	0.734899	10.810640	187.509232	10.347325	1
std	9.551546	10.696847	0.195473	6.913571	107.383428	4.192123	
min	-21.822222	-27.716667	0.000000	0.000000	0.000000	0.000000	
25%	4.688889	2.311111	0.600000	5.828200	116.000000	8.339800	1
50%	12.000000	12.000000	0.780000	9.965900	180.000000	10.046400	1
75%	18.838889	18.838889	0.890000	14.135800	290.000000	14.812000	1
max	39.905556	39.344444	1.000000	63.852600	359.000000	16.100000	1

In [5]:

```
#Concise Summary of the DataFrame
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96453 entries, 0 to 96452
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Formatted Date                        96453 non-null  object
1   Summary                              96453 non-null  object
2   Precip Type                          95936 non-null  object
3   Temperature (C)                      96453 non-null  float64
4   Apparent Temperature (C)             96453 non-null  float64
5   Humidity                             96453 non-null  float64
6   Wind Speed (km/h)                    96453 non-null  float64
7   Wind Bearing (degrees)               96453 non-null  int64
8   Visibility (km)                      96453 non-null  float64
9   Pressure (millibars)                 96453 non-null  float64
10  Daily Summary                        96453 non-null  object
dtypes: float64(6), int64(1), object(4)
memory usage: 8.1+ MB
```

In [6]:

```
#Missing Values on Dataset from String to Date Time
df.isnull().sum()
```

Out[6]:

```
Formatted Date          0
Summary                0
Precip Type            517
Temperature (C)         0
Apparent Temperature (C) 0
Humidity               0
Wind Speed (km/h)       0
Wind Bearing (degrees)  0
Visibility (km)         0
Pressure (millibars)    0
Daily Summary          0
dtype: int64
```

## Observation:

- In 'Precip Type' Column there are 517 missing values.
- 'Wind Bearing (degrees)' has only integer values and Formatted Date is in String.
- Minimum values of [Humidity](https://www.chicagotribune.com/news/ct-xpm-2011-12-16-ct-wea-1216-asktom-20111216-story.html) (<https://www.chicagotribune.com/news/ct-xpm-2011-12-16-ct-wea-1216-asktom-20111216-story.html>), [Wind Speed \(km/h\)](https://www.researchgate.net/post/Can-wind-speed-be-zero) (<https://www.researchgate.net/post/Can-wind-speed-be-zero>), [Wind Bearing \(degrees\)](https://en.wikipedia.org/wiki/Wind_direction#:~:text=Wind%20direction%20is%20usually%20reported,to%20) ([https://en.wikipedia.org/wiki/Wind\\_direction#:~:text=Wind%20direction%20is%20usually%20reported,to%20](https://en.wikipedia.org/wiki/Wind_direction#:~:text=Wind%20direction%20is%20usually%20reported,to%20)), [Visibility \(km\)](https://en.wikipedia.org/wiki/Visibility#:~:text=Visibility%20of%20less%20than%20100,be%20activated%20t) (<https://en.wikipedia.org/wiki/Visibility#:~:text=Visibility%20of%20less%20than%20100,be%20activated%20t> are Zero and they can be Zero.

In [7]:

```
#Changing Formatted Date from String to Datetime
df['Formatted Date'] = pd.to_datetime(df['Formatted Date'], utc=True)
```

In [8]:

```
#Now Formatted Date is in Date Time Format  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 96453 entries, 0 to 96452  
Data columns (total 11 columns):  
#   Column                                Non-Null Count  Dtype  
---  -  
0   Formatted Date                        96453 non-null  datetime64[ns, UTC]  
1   Summary                              96453 non-null  object  
2   Precip Type                           95936 non-null  object  
3   Temperature (C)                      96453 non-null  float64  
4   Apparent Temperature (C)             96453 non-null  float64  
5   Humidity                             96453 non-null  float64  
6   Wind Speed (km/h)                    96453 non-null  float64  
7   Wind Bearing (degrees)               96453 non-null  int64  
8   Visibility (km)                      96453 non-null  float64  
9   Pressure (millibars)                 96453 non-null  float64  
10  Daily Summary                        96453 non-null  object  
dtypes: datetime64[ns, UTC](1), float64(6), int64(1), object(3)  
memory usage: 8.1+ MB
```

In [9]:

```
df.sample(20)
```

Out[9]:

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degree)
59633	2012-11-25 17:00:00+00:00	Foggy	rain	4.916667	4.916667	0.93	4.7978	90
5369	2006-03-02 16:00:00+00:00	Foggy	snow	-0.944444	-5.644444	0.98	15.4399	180
74149	2014-07-23 14:00:00+00:00	Overcast	rain	21.227778	21.227778	0.78	12.4614	350
83001	2015-07-27 10:00:00+00:00	Mostly Cloudy	rain	22.783333	22.783333	0.56	0.3703	180
96164	2016-09-25 21:00:00+00:00	Mostly Cloudy	rain	12.950000	12.950000	0.81	5.9731	180
44704	2011-08-14 14:00:00+00:00	Partly Cloudy	rain	29.444444	29.244444	0.42	6.0858	310
53739	2012-08-25 01:00:00+00:00	Clear	rain	20.538889	20.538889	0.57	0.0000	180
28965	2009-02-25 20:00:00+00:00	Mostly Cloudy	rain	0.555556	0.555556	0.75	3.2200	350
46207	2011-02-15 06:00:00+00:00	Mostly Cloudy	snow	-2.627778	-7.388889	0.70	13.8299	90
77683	2014-10-18 21:00:00+00:00	Foggy	rain	8.427778	8.427778	0.99	1.3846	190
9970	2007-08-27 08:00:00+00:00	Clear	rain	26.161111	26.161111	0.39	8.7101	90
31929	2009-03-29 08:00:00+00:00	Overcast	rain	12.222222	12.222222	0.57	24.1500	160
85371	2015-05-05 05:00:00+00:00	Mostly Cloudy	rain	15.972222	15.972222	0.88	9.5634	110

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degree)
7617	2006-10-21 08:00:00+00:00	Mostly Cloudy	rain	12.755556	12.755556	0.69	11.1251	14
75292	2014-03-11 06:00:00+00:00	Partly Cloudy	snow	-0.138889	-2.411111	0.77	6.6332	34
24151	2008-11-01 07:00:00+00:00	Partly Cloudy	rain	17.038889	17.038889	0.68	24.4720	13
43578	2010-09-26 16:00:00+00:00	Mostly Cloudy	rain	16.833333	16.833333	0.82	22.7815	22
71889	2014-12-02 08:00:00+00:00	Foggy	rain	2.177778	2.177778	0.93	0.8694	7
63062	2013-12-17 13:00:00+00:00	Overcast	rain	0.050000	-2.661111	0.85	7.9856	13
5863	2006-05-10 06:00:00+00:00	Partly Cloudy	rain	14.855556	14.855556	0.78	9.8532	3



In [10]:

```
#Checking Wheather this dataset has Duplicate Values or not
sum(df.duplicated())
```

Out[10]:

24

## Explortary Data Analysis(EDA):

### Data Cleaning

In [11]:

```
#Number of Distinct Observation  
df.nunique()
```

Out[11]:

Formatted Date	96429
Summary	27
Precip Type	2
Temperature (C)	7574
Apparent Temperature (C)	8984
Humidity	90
Wind Speed (km/h)	2484
Wind Bearing (degrees)	360
Visibility (km)	949
Pressure (millibars)	4979
Daily Summary	214

dtype: int64



In [12]:

```
#DataFrame for Duplicate Values
df_duplicated = df[df.duplicated()]
df_duplicated
```

Out[12]:

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Be (deg
36072	2010-08-01 22:00:00+00:00	Clear	rain	18.800000	18.800000	0.93	6.2790	
36073	2010-08-01 23:00:00+00:00	Clear	rain	18.222222	18.222222	0.97	6.2790	
36074	2010-08-02 00:00:00+00:00	Clear	rain	18.072222	18.072222	0.98	11.2700	
36075	2010-08-02 01:00:00+00:00	Clear	rain	16.622222	16.622222	0.99	6.4400	
36076	2010-08-02 02:00:00+00:00	Clear	rain	16.094444	16.094444	0.99	3.0751	
36077	2010-08-02 03:00:00+00:00	Clear	rain	15.955556	15.955556	0.99	3.8801	
36078	2010-08-02 04:00:00+00:00	Clear	rain	17.088889	17.088889	1.00	6.4400	
36079	2010-08-02 05:00:00+00:00	Clear	rain	20.822222	20.822222	0.87	3.2200	

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Be (deg)
36080	2010-08-02 06:00:00+00:00	Clear	rain	23.405556	23.405556	0.74	1.8837	
36081	2010-08-02 07:00:00+00:00	Clear	rain	26.050000	26.050000	0.59	1.5939	
36082	2010-08-02 08:00:00+00:00	Clear	rain	27.688889	28.077778	0.50	0.2254	
36083	2010-08-02 09:00:00+00:00	Clear	rain	28.561111	29.588889	0.54	3.2039	
36084	2010-08-02 10:00:00+00:00	Clear	rain	28.816667	29.338889	0.49	1.6100	
36085	2010-08-02 11:00:00+00:00	Partly Cloudy	rain	28.866667	29.044444	0.46	2.8175	
36086	2010-08-02 12:00:00+00:00	Partly Cloudy	rain	29.827778	30.338889	0.47	8.0339	
36087	2010-08-02 13:00:00+00:00	Partly Cloudy	rain	30.072222	30.527778	0.46	2.5921	
36088	2010-08-02 14:00:00+00:00	Partly Cloudy	rain	31.066667	31.627778	0.44	1.3846	
36089	2010-08-02 15:00:00+00:00	Partly Cloudy	rain	30.861111	31.361111	0.44	2.2540	

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Be (deg)
36090	2010-08-02 16:00:00+00:00	Partly Cloudy	rain	29.950000	30.416667	0.46	3.0751	
36091	2010-08-02 17:00:00+00:00	Clear	rain	28.811111	30.616667	0.59	3.2039	
36092	2010-08-02 18:00:00+00:00	Clear	rain	25.250000	25.250000	0.75	2.4955	
36093	2010-08-02 19:00:00+00:00	Clear	rain	22.172222	22.172222	0.87	1.6100	
36094	2010-08-02 20:00:00+00:00	Partly Cloudy	rain	21.061111	21.061111	0.90	0.0000	
36095	2010-08-02 21:00:00+00:00	Clear	rain	20.255556	20.255556	0.92	1.0787	

In [13]:

```
df_duplicated.shape
```

Out[13]:

(24, 11)

## Observation:

- There are no Duplicate Values in this Dataset.

In [14]:

```
#DataFrame for only NaN Values for exploration.  
df_null = df[df.isna().any(axis=1)]  
df_null.head(20)
```

Out[14]:

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Be (deg)
52672	2012-04-11 14:00:00+00:00	Mostly Cloudy	NaN	19.016667	19.016667	0.26	14.8764	
52674	2012-04-11 16:00:00+00:00	Mostly Cloudy	NaN	17.850000	17.850000	0.28	13.7977	
52675	2012-04-11 17:00:00+00:00	Mostly Cloudy	NaN	16.322222	16.322222	0.32	10.8192	
52677	2012-04-11 19:00:00+00:00	Mostly Cloudy	NaN	12.566667	12.566667	0.43	9.0160	
52678	2012-04-11 20:00:00+00:00	Mostly Cloudy	NaN	12.927778	12.927778	0.47	17.6295	
52680	2012-04-11 22:00:00+00:00	Mostly Cloudy	NaN	10.100000	10.100000	0.61	11.3666	
52681	2012-04-11 23:00:00+00:00	Partly Cloudy	NaN	10.183333	10.183333	0.64	7.2128	
52683	2012-04-12 01:00:00+00:00	Mostly Cloudy	NaN	10.700000	10.700000	0.70	14.6993	
52684	2012-04-12 02:00:00+00:00	Partly Cloudy	NaN	9.744444	8.022222	0.73	11.9623	
52687	2012-04-12 05:00:00+00:00	Mostly Cloudy	NaN	10.055556	10.055556	0.76	13.9748	
52692	2012-04-12 10:00:00+00:00	Partly Cloudy	NaN	17.177778	17.177778	0.42	25.3414	
52693	2012-04-12 11:00:00+00:00	Partly Cloudy	NaN	16.183333	16.183333	0.41	28.2877	
52695	2012-04-12 13:00:00+00:00	Partly Cloudy	NaN	16.183333	16.183333	0.47	24.2466	
52696	2012-04-12 14:00:00+00:00	Partly Cloudy	NaN	16.244444	16.244444	0.43	29.8816	
52698	2012-04-12 16:00:00+00:00	Partly Cloudy	NaN	15.183333	15.183333	0.50	26.8548	

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Be (deg)
52699	2012-04-12 17:00:00+00:00	Partly Cloudy	NaN	12.455556	12.455556	0.52	17.6134	
52701	2012-04-12 19:00:00+00:00	Partly Cloudy	NaN	7.983333	7.983333	0.70	1.7388	
52702	2012-04-12 20:00:00+00:00	Partly Cloudy	NaN	6.905556	6.905556	0.77	2.1091	
52704	2012-04-12 22:00:00+00:00	Partly Cloudy	NaN	5.283333	5.283333	0.86	3.7996	
52705	2012-04-12 23:00:00+00:00	Partly Cloudy	NaN	4.688889	3.133333	0.83	6.7781	

In [15]:

```
df_null.tail(20)
```

Out[15]:

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearin (degree:
95569	2016-10-31 03:00:00+00:00	Clear	NaN	1.905556	1.905556	0.88	4.7656	24
95570	2016-10-31 04:00:00+00:00	Clear	NaN	1.894444	0.088889	0.89	6.2146	24
95571	2016-10-31 05:00:00+00:00	Clear	NaN	1.983333	-0.133333	0.91	7.1645	25
95572	2016-10-31 06:00:00+00:00	Clear	NaN	2.494444	0.227778	0.90	7.9373	25
95573	2016-10-31 07:00:00+00:00	Clear	NaN	3.766667	1.561111	0.86	8.5169	25
95574	2016-10-31 08:00:00+00:00	Clear	NaN	5.688889	3.783333	0.80	8.7262	25
95575	2016-10-31 09:00:00+00:00	Clear	NaN	7.688889	6.177778	0.75	8.5813	26
95576	2016-10-31 10:00:00+00:00	Clear	NaN	9.594444	7.972222	0.71	11.1251	27
95577	2016-10-31 11:00:00+00:00	Clear	NaN	11.494444	11.494444	0.69	14.0553	28
95578	2016-10-31 12:00:00+00:00	Clear	NaN	12.766667	12.766667	0.67	22.5400	28

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degree)
95579	2016-10-31 13:00:00+00:00	Partly Cloudy	NaN	13.022222	13.022222	0.68	14.1680	295
95580	2016-10-31 14:00:00+00:00	Partly Cloudy	NaN	12.538889	12.538889	0.71	10.9641	295
95581	2016-10-31 15:00:00+00:00	Mostly Cloudy	NaN	11.755556	11.755556	0.74	8.3237	295
95582	2016-10-31 16:00:00+00:00	Mostly Cloudy	NaN	10.688889	10.688889	0.78	7.5187	295
95583	2016-10-31 17:00:00+00:00	Mostly Cloudy	NaN	9.394444	8.377778	0.82	7.6797	295
95584	2016-10-31 18:00:00+00:00	Mostly Cloudy	NaN	8.322222	7.044444	0.85	8.0339	295
95585	2016-10-31 19:00:00+00:00	Mostly Cloudy	NaN	7.627778	6.183333	0.87	8.2271	295
95586	2016-10-31 20:00:00+00:00	Mostly Cloudy	NaN	7.111111	5.511111	0.89	8.5008	295
95587	2016-10-31 21:00:00+00:00	Partly Cloudy	NaN	6.672222	4.961111	0.90	8.6457	295
95588	2016-10-31 22:00:00+00:00	Mostly Cloudy	NaN	6.322222	4.588889	0.91	8.4686	295

In [16]:

```
#Dropping NaN(Not a Number)
df_target = df.dropna()
df_target.shape
```

Out[16]:

(95936, 11)

In [17]:

```
df_target.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 95936 entries, 0 to 96452
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Formatted Date                        95936 non-null  datetime64[ns, UTC]
1   Summary                              95936 non-null  object
2   Precip Type                           95936 non-null  object
3   Temperature (C)                       95936 non-null  float64
4   Apparent Temperature (C)              95936 non-null  float64
5   Humidity                             95936 non-null  float64
6   Wind Speed (km/h)                    95936 non-null  float64
7   Wind Bearing (degrees)                95936 non-null  int64
8   Visibility (km)                       95936 non-null  float64
9   Pressure (millibars)                  95936 non-null  float64
10  Daily Summary                         95936 non-null  object
dtypes: datetime64[ns, UTC](1), float64(6), int64(1), object(3)
memory usage: 8.8+ MB
```

In [18]:

```
df_target.columns
```

Out[18]:

```
Index(['Formatted Date', 'Summary', 'Precip Type', 'Temperature (C)',
      'Apparent Temperature (C)', 'Humidity', 'Wind Speed (km/h)',
      'Wind Bearing (degrees)', 'Visibility (km)', 'Pressure (millibars)',
      'Daily Summary'],
      dtype='object')
```



In [19]:

```
df_target = df_target.set_index("Formatted Date")
df_target
```

Out[19]:

	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visi
Formatted Date								
2006-03-31 22:00:00+00:00	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251	15.
2006-03-31 23:00:00+00:00	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259	15.
2006-04-01 00:00:00+00:00	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204	14.
2006-04-01 01:00:00+00:00	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269	15.
2006-04-01 02:00:00+00:00	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259	15.
...	...	...	...	...	...	...	...	...
2016-09-09 17:00:00+00:00	Partly Cloudy	rain	26.016667	26.016667	0.43	10.9963	31	16.
2016-09-09 18:00:00+00:00	Partly Cloudy	rain	24.583333	24.583333	0.48	10.0947	20	15.
2016-09-09 19:00:00+00:00	Partly Cloudy	rain	22.038889	22.038889	0.56	8.9838	30	16.
2016-09-09 20:00:00+00:00	Partly Cloudy	rain	21.522222	21.522222	0.60	10.5294	20	16.
2016-09-09 21:00:00+00:00	Partly Cloudy	rain	20.438889	20.438889	0.61	5.8765	39	15.

95936 rows × 10 columns



Given:

The **Null Hypothesis H0** is "Has the Apparent temperature and humidity compared monthly across 10 years of the data indicate an increase due to Global warming".

The **Alternative Hypothesis H1** is "Has the Apparent temperature and humidity compared monthly across 10 years of the data not indicate an increase due to Global warming".

In [20]:

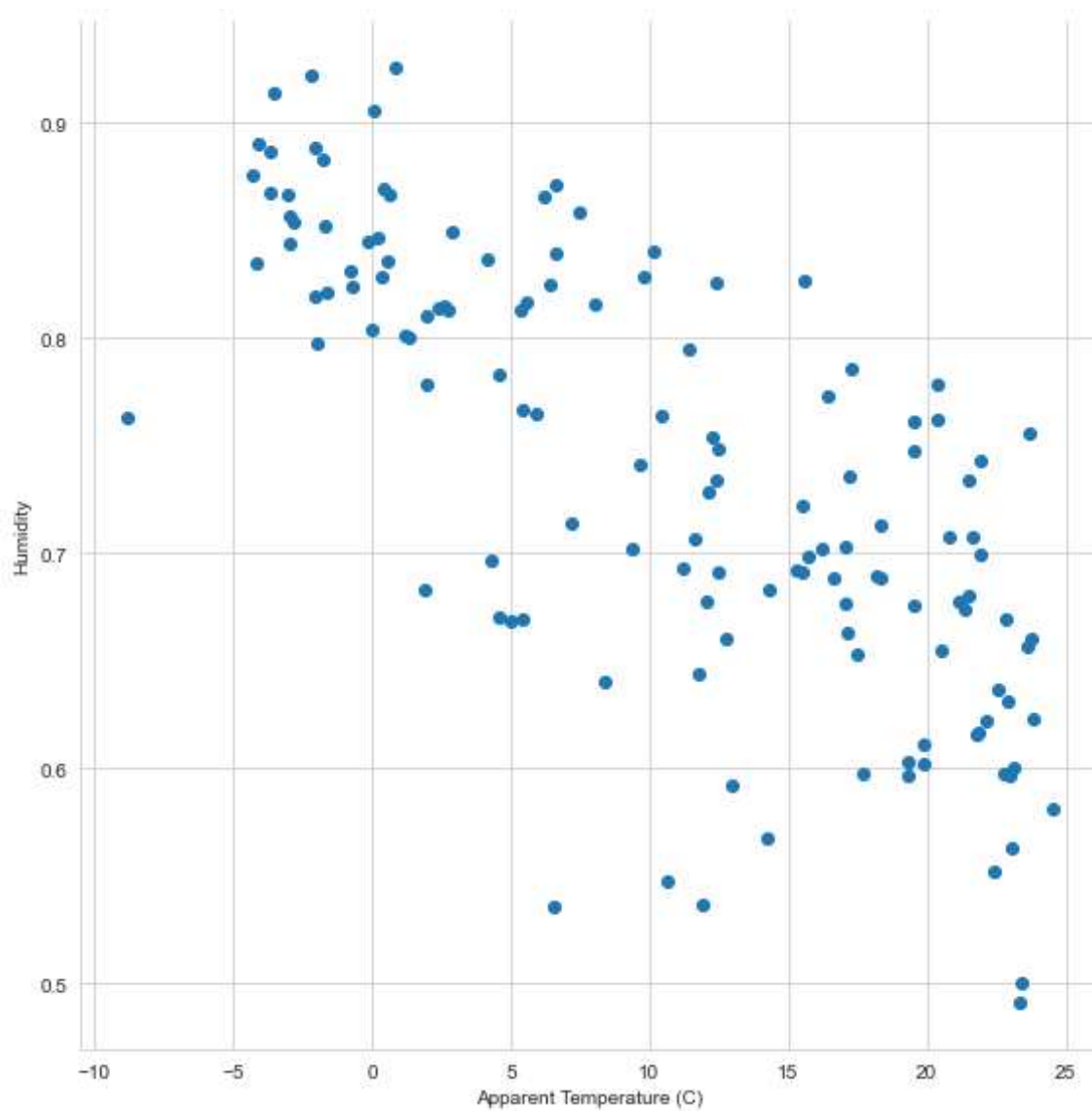
```
#Creating new DataFrame only for Apparent Temperature and Humidity
df_column = ['Apparent Temperature (C)', 'Humidity']
df_monthly_mean = df_target[df_column].resample("MS").mean() #MS-Month Starting
df_monthly_mean.head()
```

Out[20]:

	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

In [21]:

```
sns.set_style("whitegrid")
sns.FacetGrid(df_monthly_mean, height=8).map(plt.scatter, "Apparent Temperature (C)", "Humidity")
plt.show()
```

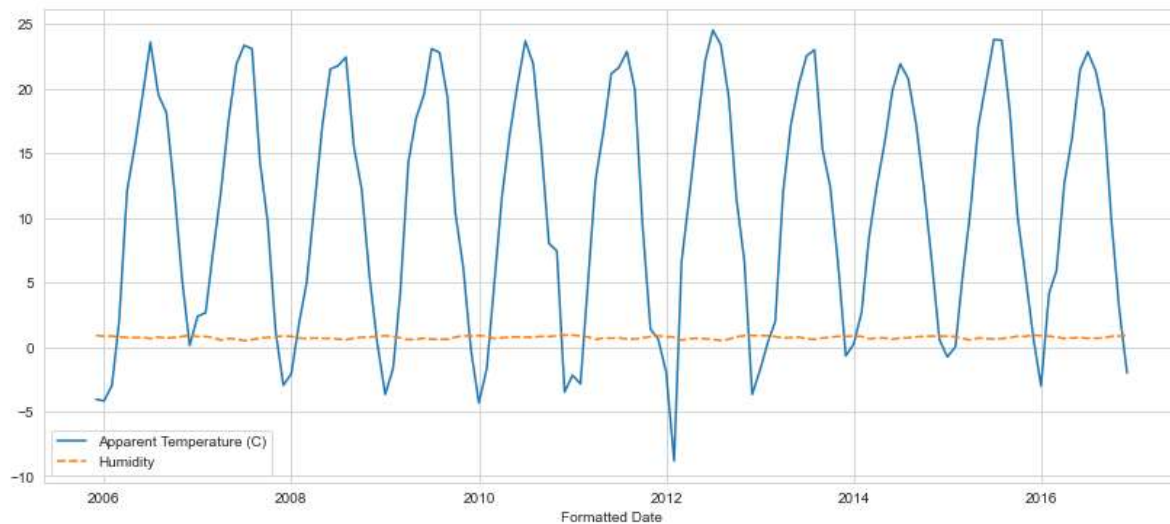


Observation:

- There can be or may be Linear Relationship between "Humidity" and "Apparent Temperature (C)" with negative slope.

In [22]:

```
plt.figure(figsize=(14,6))
sns.lineplot(data = df_monthly_mean)
plt.show()
```



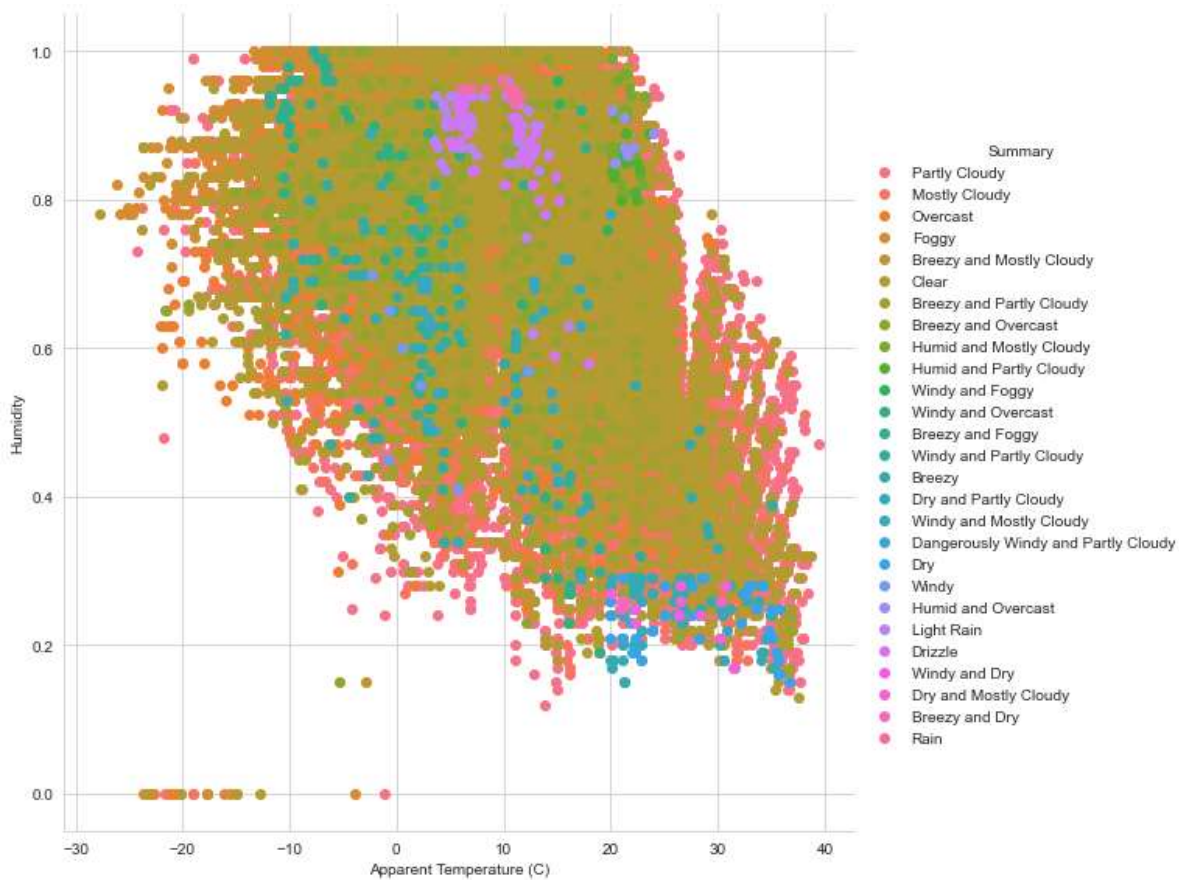
Observation:

- "Humidity" is remain constant from 2006 - 2016
- But "Apparent Temperature (C)" is frequently changes from 2006 - 2016.

**2D Scatter Plot with Color Coding for each Summary type.**

In [23]:

```
sns.set_style("whitegrid")
sns.FacetGrid(df_target, hue="Summary", height=8).map(plt.scatter, "Apparent Temperature (C)",
plt.show())
```



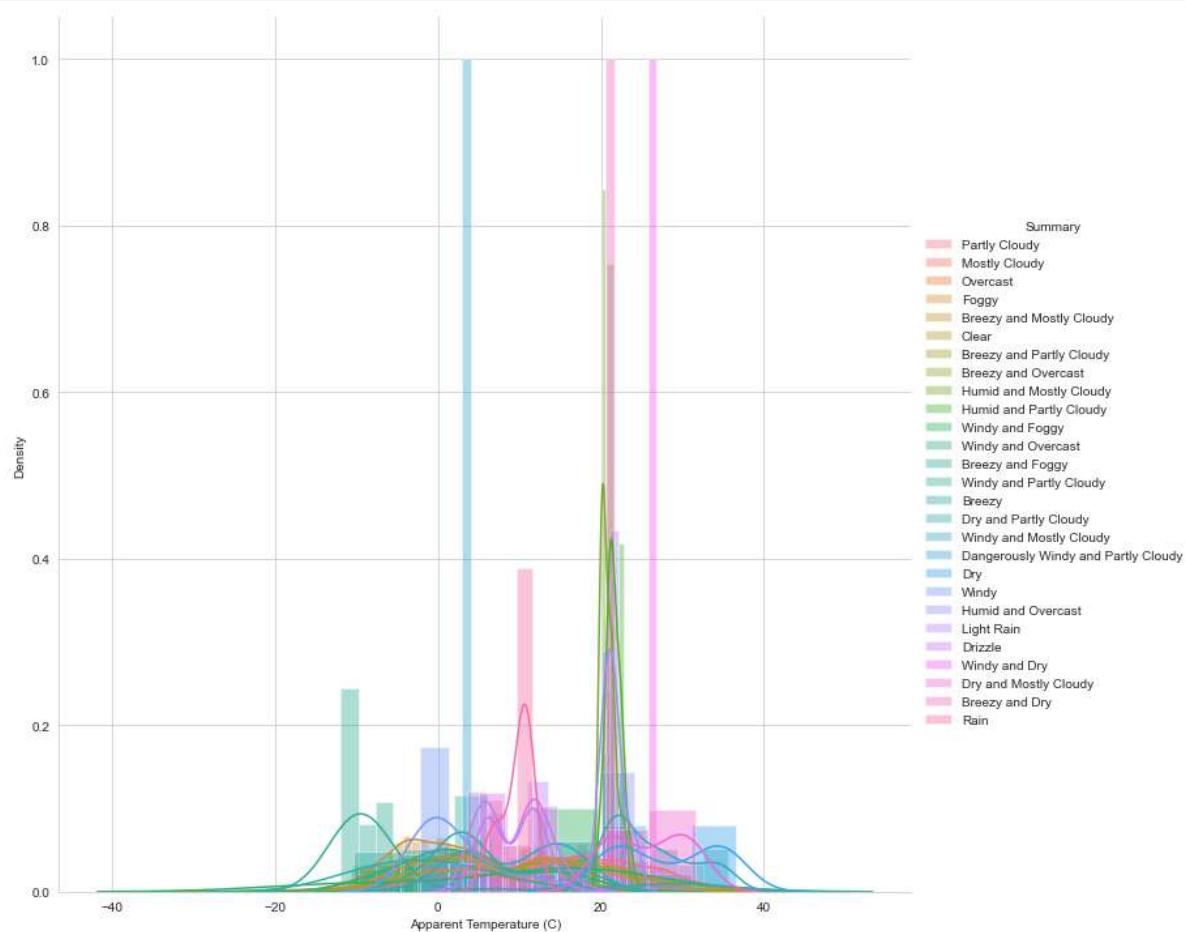
Observation:

- There are very few outlier.
- Mostly Weather is Clear or Partly Cloudy/Rain in Finland.
- Only few days there has a Light Rain or Dry or Dangerously Windy and Partly Cloudy.

## Univariate Analysis using PDF:

In [24]:

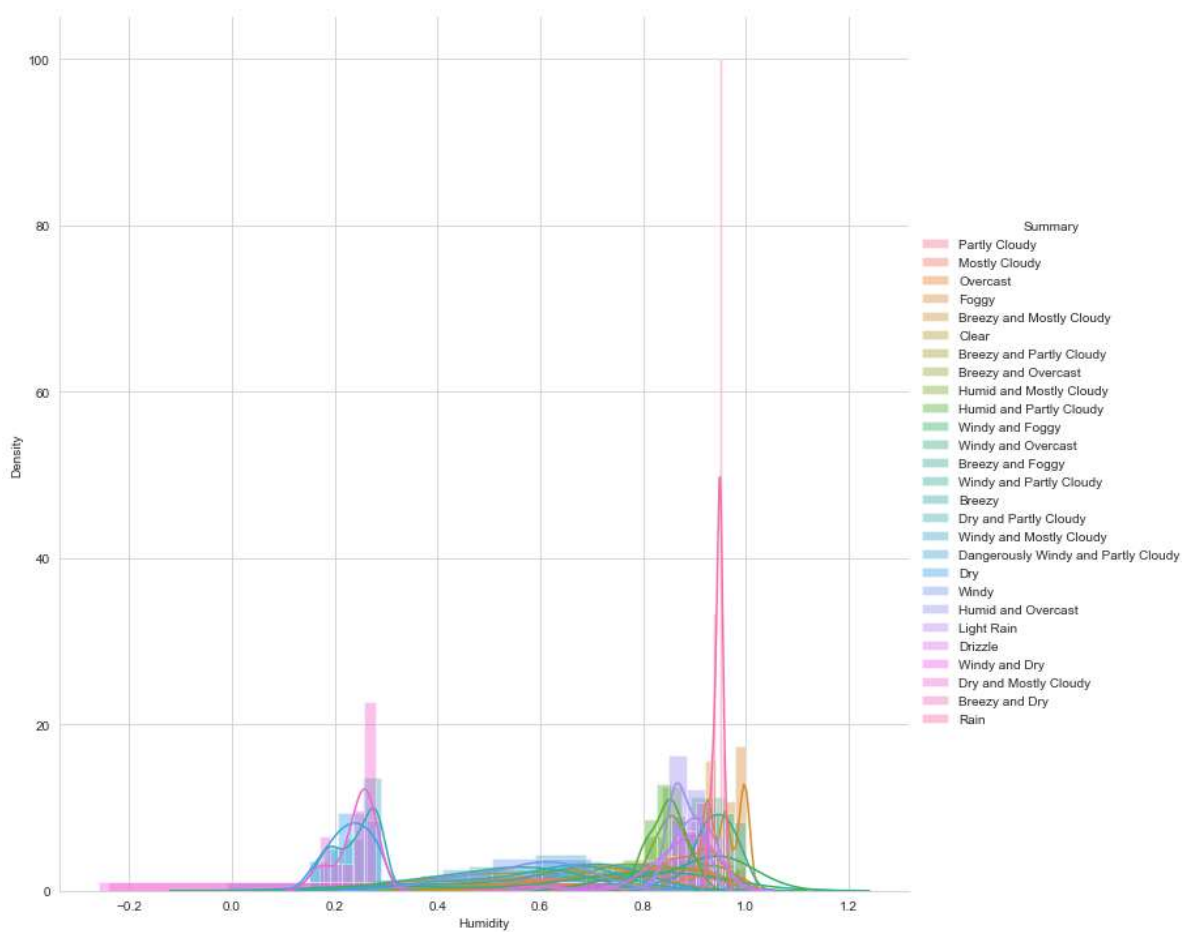
```
# For Apparent Temperature (C)
sns.FacetGrid(df_target, hue="Summary", height=10).map(sns.distplot, "Apparent Temperature")
plt.show()
```



In [25]:

```
# For Humidity
```

```
sns.FacetGrid(df_target, hue="Summary", height=10).map(sns.distplot, "Humidity").add_legend()  
plt.show()
```



Observation:

- "Humidity" is better Feature than "Apparent Temperature (C)".

**Conclusion:**

H0 is not accepted because there is no change in Humidity from 2006–2016. So, we will accept the H1.

In [ ]: