Meteorologists analyze many sources of data in forecasting the weather. This is done by examining a large quantity of observation data including surface observations, satellite imagery, radiosonde data, upper-air data, wind profilers, aircraft observations, river gauges, and simply looking outside.

**Data Collection:**

I collect this  weather dataset from Kaggle.

**Aim:**

 My goal in this project is data analysis and transform the raw data into information and convert it into knowledge using famous data analysis library called Pandas.

**I will be performing some basic tasks to perform our analysis such as following:**

* **Data cleaning**
* **Data normalizing.**
* **Testing the hypothesis**

**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".**

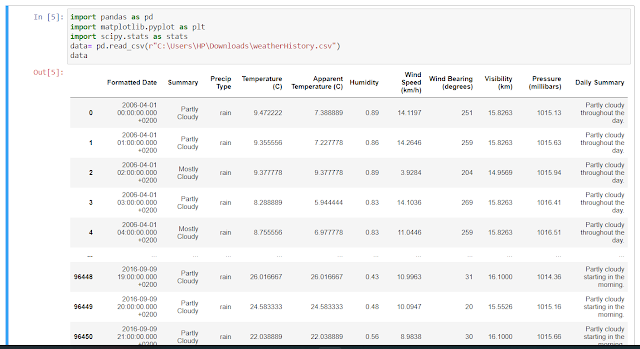
The H0 means we need to find whether the average Apparent temperature for the month of a month say January starting from 2006 and the average humidity for the same period have increased or not. This monthly analysis has to be done for all 12 months over the 10 years period. So you are basically resampling my data from hourly to monthly , then comparing the same month over the10 year period .

So let's started

**Import library and Reading the data:**

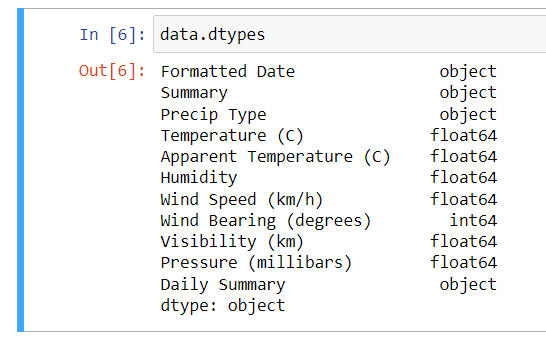
The first thing we need to do is import the all libraries and dataset we will be using such as Pandas ,Matplotlib , Scipy , Pandas is one of the most useful and core libraries for data science projects.

I read the data from my file .I also display 4 rows of  
the dataset .

[](https://blogger.googleusercontent.com/img/b/R29vZ2xl/AVvXsEiA5VS6YFTsDQ3c7L2G0n7ctWc4rzW5WLZLyJHHwQE6gZaKqywHOzZpD59MlFaA3MXO8C7nT-RZkOzjV42S07kDIWcy_ogPzb1fkEMOzcUMFvm7bB9ssZxQUxczO3eNI9QQjouiv-1xCk1WChXXUdA85r2W4OmlQ_OJVJGd_ckG5uVoVdkSGiHAl9bmFg/s1412/111.png)

**Data Types :**

 Here I display the type of  information such as the information is object or numerical values. The following table shows the data types of information .

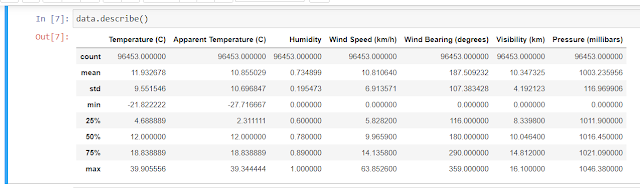
[](https://blogger.googleusercontent.com/img/b/R29vZ2xl/AVvXsEjNtGB-jllPsN-Cr3wdZLLMk3AB_itzpHrboPW1hRPMqUpM3_VNv_x_yIxxn65NAoqcvotxbW9ZDvQCiZpldNxrHT8a62yW_EgVOaqCDed7KRbzOIi0gpalavbxq-OjEUaeZypfozpKUp841bFEy50n4LFELVxN_Pq8oHFnwgF7FiAhncQMcR6KZuQAHQ/s546/112.png)

**Description of the data :**

I  display description of the data in the Data Frame using 'description' method. If the Data Frame contains numerical data, the description contains these information for each column:

count - The number of not-empty values.  
mean - The average (mean) value.  
std - The standard deviation.  
min - the minimum value.  
25% - The 25% percentile\*.  
50% - The 50% percentile\*.  
75% - The 75% percentile\*.  
max - the maximum value.

\*Percentile meaning: how many of the values are less than the given percentile

[](https://blogger.googleusercontent.com/img/b/R29vZ2xl/AVvXsEhYTXI5ppicaSC55SAfQ5e8jKASb6at4pxsbp3QhkMpQyrAeguxaul2F2A83fuLRkxMlSI2OXuddMF-jntCk6zDKbdTSvtEldkjqbrx9E3VRJgywaeUdTONAC7jHLPH-CaLnXBqeHSTYj2IrgYyZsx4BDxBt72EXZ91PIB1kiF3kFtNo5EbPuJhGcKn5Q/s1276/113.png)

**Data Cleaning:**

I am  checking for null values . Here we can see there is a total of 517 Null Values are present.

|  |
| --- |
|  |
| Check for null values |

 **Datetime Format:**

We also convert the dates into standard Datetime format.  We can see the following datetime formatted table.

|  |
| --- |
|  |
| In datetime format |

**Resampling  of data:**

 Here we resample minutely data to monthly data .

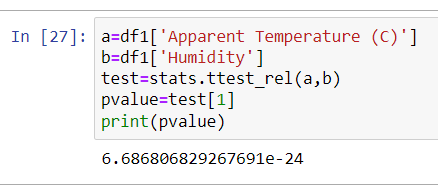
|  |
| --- |
|  |
| Resampling Series |

**Plotting Data :**

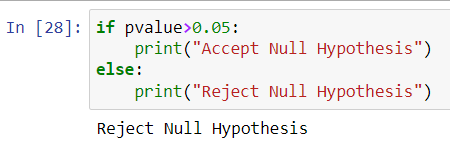
The plotting line show the variation in data yearly. It shows the variation of apparent temperature and Humidity yearly . Apparent temperature is various  but Humidity still same  yearly .

|  |
| --- |
|  |
| Appenrent Temperature and Humidity  variation yearly   **Apparent Temperature and Humidity variation montwise:**  Here we show the variation of apparent temperature and humidity variation monthwise. Following graph show the January month variation .We will also get the variation of all months. |

This is a test for the null hypothesis that two related or repeated samples have identical average (expected) values.

[](https://blogger.googleusercontent.com/img/b/R29vZ2xl/AVvXsEjVn8WR6GolGcEnwWseggy9rXxVN9UYkm9rNW8PpvlBINuyoxunvAo0yL1pGWzgKFAnUU-OQ9HyxkZuH_9wz_kLYVGc-9YOdGKZXcmoq9KMo9uEgVabcxvm5Ilhi0qtmwVems3jLdgYcvzFc17b5hSerniqHrkjQaARsCfifgtJqPEcm2240RpQfHz7gg/s438/120.png)

 Here Pvalue is less than 0.05 so we reject the null hypothesis.

[](https://blogger.googleusercontent.com/img/b/R29vZ2xl/AVvXsEhFwamt9sqP6vjOgYt87_VPYE3XYcmF_bkYSbsPw2gt12y2KtYJKxla2L5SookPXfDN-y14zSZ1dYUI1liILIhs8wOgYfnDEAyjZi7oZaOQGDlMemU3mVGHGU40rDqX2xnycvCj0H1VEskALE4G6wEiPyqgn9NL1Nj1i2g9ukQ1FozOAY9PI1SF3bRqgg/s450/121.png)

**Conclusion:**

From the analysis, we have gathered useful information over the past 10 years. For example Humidity , temperature .

Global warming is deteriorating the climate and is affecting various parameters of the environment.

Hence from the analysis , I infer that there is either sharp falls in temperature or sharp rise in temperature in 10 years .

Global warming is deteriorating the climate and is affecting various parameters of the environment.