**Performing analysis of meteorological data**

Many people are interested in making calculations and forecasting the weather  The internship assignment I was given is also the same. analysing a weather dataset that has been provided.

Hourly temperature data for the past 10 years, from 2005 to 2016, are presented in the format (2006-04-01 00:00:00.000 +0200). Additionally, it includes a daily summary, pressure, visibility, apparent temperature, humidity, wind speed, wind direction, and kind of precipitation.

This dataset was created using information from the nation of Finland.

Dataset link:( <https://drive.google.com/file/d/1ScF_1a-bkHi1qe8Rn78uxK6_5QwUD9Bu/view>)

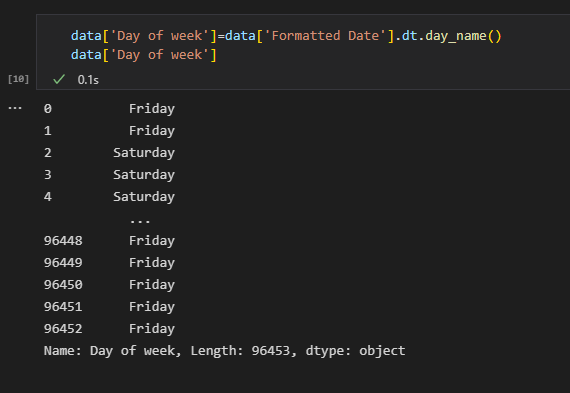
Analyzing the excel sheet and determining if it is in the form of a structured or unstructured form is the first step in any sort of analytical activity. The csv file must then be read in order to determine whether or not all of the columns contain the right datatypes using the Jupytor platform and Python. Formatted date in this csv file has an object data type rather than a date type. So, the datatype of the formatted date has been converted by using the to datatime function.

Then, using the "isnull" function, we must locate the null values and eliminate any that won't have an impact on our output.

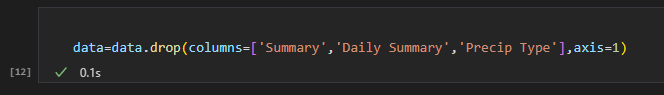
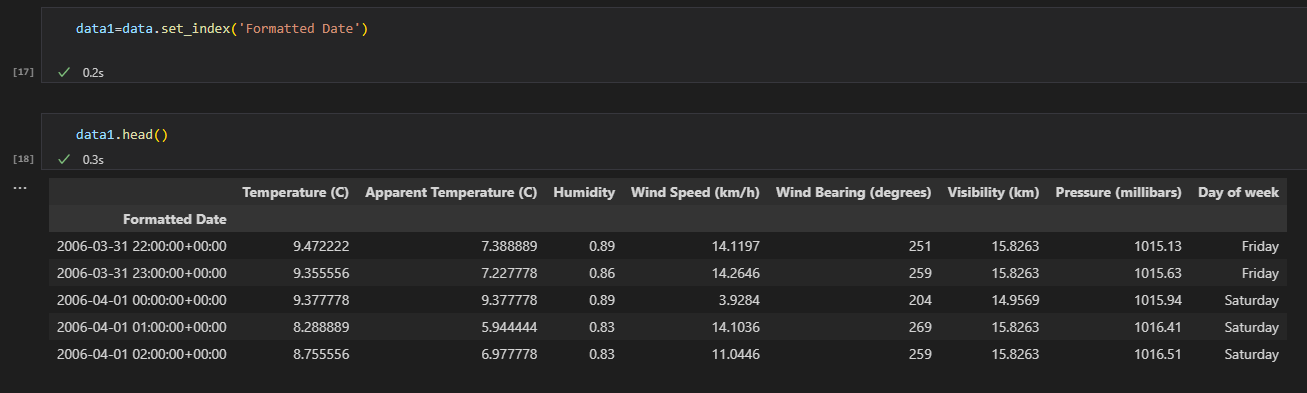
I then choose to determine the mean, standard deviation, lower and higher quartile values after cleaning the data. I discovered all of these numbers for all integer type values using the 'describe' command.

It should be highlighted that categorical variables cannot be defined by mean and standard deviation.

A new column named Day of week has also been added to the dataset. The day variable for each date has been added. This will give me an advantage in correctly assessing the data.

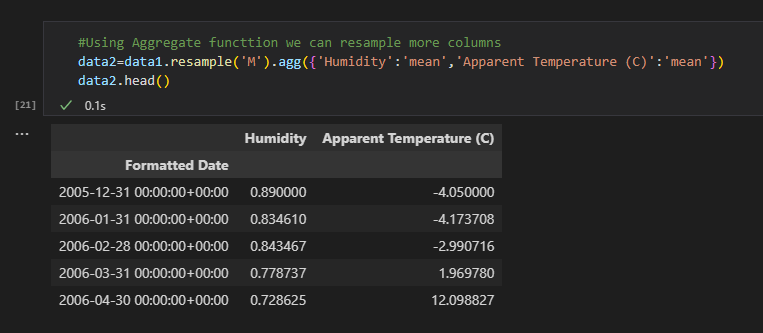


We must eliminate all the irrelevant data that aren't harming our primary aim after completing all pre-processing processes and data cleaning..

I've placed formatted date as an index for easier use and a more effective manner of handling the data.

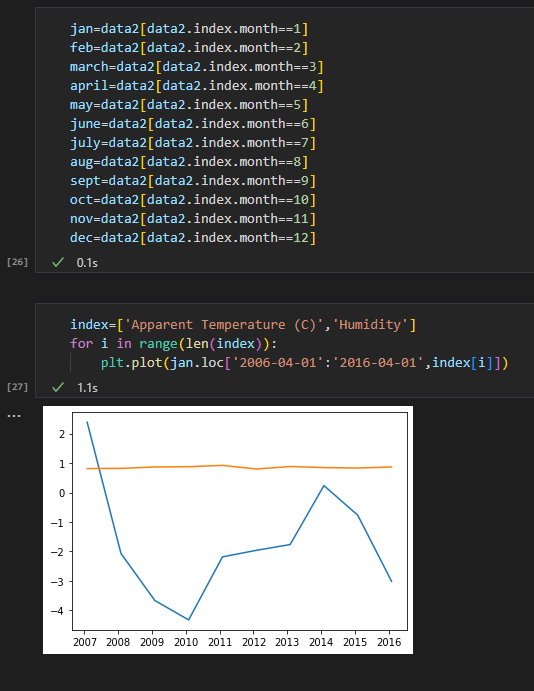
Therefore, we must determine whether or not the average apparent temperature and humidity for each month from 2006 to 2016 and the average humidity for the same time have grown in accordance with the tasks and goals that have been set. Over a ten-year period, this monthly analysis must be completed for each of the 12 months.

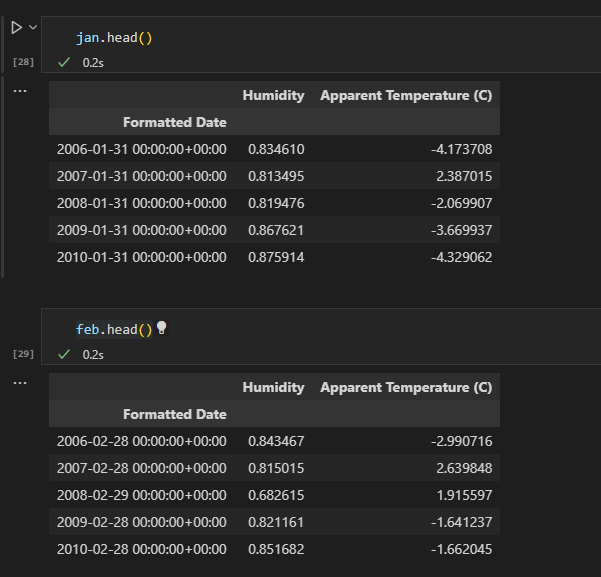
Therefore, I need to calculate and save the year mean for perceived temperature and humidity. Every year, the same procedure must be followed.

I have thus resampled all the data onto one row using the resample technique.

The data must now be visualised with the help of the matplotlib and seaborn libraries.

Therefore, I collected the data from each month and put it all in a single variable. Therefore, the data has been displayed using those variables. Here are a few of the visalizations:





On final stage, using sweetviz tool, I have made brief visualizations of entire dataset.

Link: [analyze.html](file:///E:\Performing%20Analysis%20of%20Meteorological%20Data\analyze.html)

Link: https://colab.research.google.com/drive/1-PrVzbuVHtR\_HXtpb0sHYOpoU3TX55eM

**Conclusion:**

We discovered that perceived temperatures and humidity are increasing in the months of April and May by examining the provided dataset. In December and January, the same were below average. This demonstrates unequivocally that Finland experiences extreme temperatures in the months of May and April.

In addition to this, the fact that temperatures and humidity have been rising year after year supports the conclusion that there has been a consistent increase in global warming. According to this analysis, global warming must be limited and appropriate action must be done globally.

I am thankful to mentors at **https://internship.suvenconsultants.com** for providing awesome problem statements and giving many of us a Coding Internship Exprience. Thank you www.suvenconsultants.com