

Performing Analysis of Meteorological Data

One type of data that's easier to find on the net is **Weather data**. Many sites provide historical data on many meteorological parameters such as pressure, temperature, humidity, wind_speed, visibility, etc. Our SuvenML team has downloaded one such weather dataset from kaggle. (Source URL: <https://www.kaggle.com/muthuj7/weather-dataset>)

The dataset has hourly temperature recorded for last 10 years starting from 2006-04-01 00:00:00.000 +0200 to 2016-09-09 23:00:00.000 +0200. It corresponds to Finland, a country in the Northern Europe. You can download the dataset from this Google drive link: https://drive.google.com/open?id=1ScF_1a-bkHi1qe8Rn78uxK6_5QwUD9Bu

Your goal in this Internship (of data analysis) is to transform the raw data into information and then convert it into knowledge.

In this Internship assignment, you would be responsible for perform **data cleaning**, perform analysis for testing the (given) **Hypothesis** and finally put-forth your conclusion by **writing a blog article**. Your blog should be as descriptive as possible with relevant visualizations to prove your point.

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** 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.

All the best.

Kindly note there is another project under this **domain : Data Analytics**, which also has to be completed to achieve your **Internship Certification**.