DATA ANALYSIS OF METROLOGICAL DATA

By Divyam Parhwal (divyamparhwal2002@gmail.com)
Student at JAGAN INSTITUTE OF MANAGEMENT STUDIES,
SECTOR – 5, ROHINI, NEW DELHI – 110085, INDIA
Affiliated to GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY,SECTOR – 16 C, DWARKA, NEW DELHI – 110078, INDIA
Under the Guidance of Dr. Deepti Sharma, Professor (IT), JIMS Rohini, Delhi

***Abstract* — This project report is set to give an interactive visualization and analytical presentation for Meteorological records in Finland. These Meteorological records Data of Finland is recorded by way of integratingthe three current infrastructures for numerical weather prediction, observational information and satellite tv for pc image processing and this is recorded. The Meteorological data used in the study consists of near- floor atmospheric elements including wind direction, apparent temperature, cloud layer(s), ceiling peak, visibility, current weather, wind velocity, cloud cowl and precipitation amount and so on. The data consists of hourly recorded data for the past ten years of Finland from 2006-04- 01 at time 00:00:00.000 to 2016-09-09 at time 23:00:00.000. The analysis had been performed out for 2-m floor temperature. Through this analysis, all the valuable insights into the changing weather patterns and environmental conditions in Finland are analyzed and presented in an interactive visualization, providing a solid foundation for understanding and addressing the impactsof Global Warming in the region. This project main objective is to show the complete analysis of the influences of the Global Warming on the Apparent temperature & humidity in Finland over the course of 10 years from 2006 to 2016.**

***Keywords* — Data Analysis, Data Visualization, Meteorological Data, Climate Change, Global Warning, Apparent Temperature, Humidity**

I. INTRODUCTION

As the capacity of computing technology expands, so too does the volume of data produced by climate simulations. Advanced techniques, such as those used in satellite operation, add to surge of climate related data. However, as the data grows, so does the challenge of analyzing it to uncover inherent patterns and statistical

relationships. This analysis is increasingly becoming a bottleneck. There is also a heightened demand for methodologies that can validate the robustness of conclusions drawn from simulation data. Crafting intuitive and impactful visual representations to convey climate data entails several obstacles.

Firstly, the diversity of meteorological data necessitates a range of visualization approaches, such as 2D maps, chronological graphs, and scatter plots. In addition, there is a need for innovative, yet user-friendly visualization methods to interpret extensive time-series data that are three-dimensional or multivariate, in an interactive manner.

Secondly, the audience for climate data visualizations is diverse, encompassing individuals with varying levels of expertise, qualifications, interests, and professional backgrounds, all of whom undertake critical tasks.

Thirdly, translating scientific data into visualizations is complex due to the assortment of available tools, methodologies, and parameters. Data analytics can yield a wealth of information for evaluating normal climate patterns. Accurate understanding of weather conditions is crucial for the general public & businesses alike, with many industries relying on precise weather statistics for informed decision-making. Weather data, including historical records of various meteorological elements such as temperature, humidity, wind direction and speed, cloudiness, cloud layers, ceiling height, visibility, current weather conditions, & precipitation volumes, is more readily accessible online through numerous platforms.

In this study I have made a report on the premises of data visualizationresults which I got from a similar dataset of historical whether data on various meteorological parameters of Finland.

I've additionally defined the impact of Global warning on Finland's apparent temperature and humidity across a historical time period of ten years from 2006-04-01 at time 00:00:00.000. to 2016-09-09 23:00:00.000.

This report is divided as: first section of report includes the Introduction, second section includes research topic and background of it, the third section discusses about the literature review, the fourth section discusses about the conducted study, fifth section include findings and results of the study, sixth section includes comparison chartmade for the study. Finally, section seventh will include the conclusion and recommendation of future work.

I.I PROJECT OBJECTIVE

This project aims to provide an interactive visualization and analytical presentation of hourly recorded meteorological datafrom Finland, a Northern European country, spanning from 2006-04-01, at 00:00:00.000 to 2016-09-09, at 23:00:00.000.
Project objective is to conduct data cleaning on the given dataset, followed by analysis to test the influences of Global Warming on apparent temperature and humidity. The ultimategoal is to draw conclusions based on this analysis. By examining changing weather patterns and environmental conditions in Finland, valuable insights into the impacts of Global Warming in the region will be presented through interactive visualization. Specifically, the project focuses on analyzing the complete 10-year dataset to understand the influences of Global Warming on apparent temperature and humidity in Finland. At the end, I would like to conclude that this project main objective is to show the complete analysis ofthe Influences of Global Warming on Apparent temperature and humidity in Finland over the course of 10 years from 2006 to 2016.

I.II PROJECT HYPOTHESIS

The Null Hypothesis HO is "Has the Apparent temperature and humidity compared monthly across 10 years of the data indicate an increase due to Global warming".
The HO means we need to find whether the average Apparent temperature for the month of a month says April starting from 2006 to 2016 and the average humidity for the same period has increased or not.

I.III SCOPE OF THE PROJECT

The scope of the project encompasses the analysis of temperature and humidity data in Finland over the specified 10-year period, from 2006-04-01, at 00:00:00.000 to 2016- 09-09, at 23:00:00.000. It involves data preprocessing, including cleaning and formatting, to ensure data integrity. The analysis focuses on exploring the relationship between Apparent Temperature and Humidity through visualizations and statistical techniques. The project aims to identify trendsand patterns in the data and draw conclusions regarding the impact of global warming on Finland's climate. Additionally,the project scope includes generating monthly comparison charts to illustrate the changes in temperature and humidity over time, providing a comprehensive understanding of the effects of climate change in Finland.

II. LITERATURE REVIEW

Meteorology delves into the comprehensive study of Earth's atmosphere and is integral to our daily lives. It covers a broad spectrum of factors that shape the atmosphere's dynamics and appearance. The challenge of grasping meteorological data purely through numerical representations necessitates the development of visualization techniques. Such visualizations help people understand the intricate relationships between various atmospheric elements. These insights are crucial for practical applications such as urban planning, construction, and the strategic placement of solar panels, where long-term impacts of weather conditions on surfaces must be monitored.

Key meteorological variables include temperature, moisture levels, barometric pressure, wind velocity and direction, and precipitation volume. To effectively interpret and utilize this data, it’s essential to comprehend the interactions among these variables and their collective influence on the local environment. Meteorological elements can be categorized based on their visibility; for example, air temperature, which quantifies the kinetic energy in the atmosphere, is not directly observable. However, the consequences of air temperature manifest visibly through phenomena such as thermal shimmer, frost formation, and optical

illusions like mirages, illustrating the tangible effects of temperature variations.

This project report is about to give an interactive visualizationand analytical platform for Meteorological data in Finland bymeans of integrating the 3 existing infrastructures for observational statistics, numerical climate prediction, andsatellite photograph processing. The analysis has been performed for 2-m floor temperature. This project main objective is to show the complete analysis of the Influences of Global Warming on temperature and humidity in Finland over the course of 10 years from 2006 to 2016. Major Terminologies in the project performed is as follows:

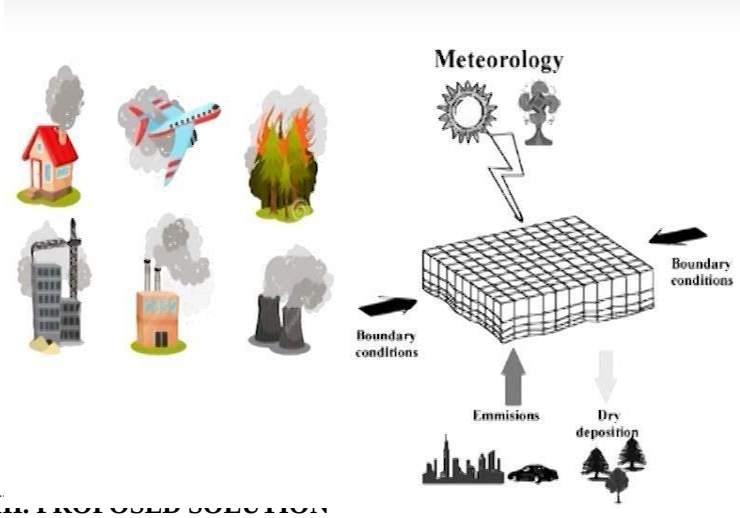
Data Analytics: It is a field centered on deriving meaningful insights from data sets. It encompasses a variety of methodologies, tools, and strategies for analyzing and managing data, which includes the systematic collection, structuring, and preservation of information.

Meteorological Data: It refers to the array of physical measurements recorded using scientific instruments. These measurements capture a range of atmospheric variables such as ambient temperature, moisture level (dew point), wind patterns (both direction and velocity), sky conditions (including cloud formation and density), visibility range, atmospheric pressure, and the quantity and type of precipitation.

Apparent Temperature: It is a concept that describes the temperature humans perceive. This subjective temperature is influenced by a combination of factors including the actual air temperature, the moisture content in the air (humidity), and the rate of air movement (wind speed). This index is most often used to describe how hot or cold it feels outside to a person.

Humidity: It represents the level of water vapor present in the air. It's a measure of the atmosphere's ability to hold moisture, which can affect weather conditions and human comfort.

Dataset: The dataset can be obtained from Kaggle. The dataset has hourly temperature recorded for the 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 Northern Europe country.



II. PROPOSED SOLUTION

In the proposed solution section of the paper, general structure of interactive visualization and analytical platform development process will be clarified. The main tools that were used in the project are ML algorithms for data analysis and visualization and classification purposes. PYTHON version 3.6.4 is used for this project. IDE which used to create solution of this paper is called

Jupyter Notebook. Every monitored ML system requires gooddata setup for training to be adequate enough so that it can predict new results. Data used to train and evaluate the current project were collected from Kaggle website. The dataset has hourly temperature recorded from 2006–04–01 00:00:00.000 +0200 to 2016–09–09 23:00:00.000 +0200 for the 10 years which corresponds to a Northern Europe country, Finland. Second, data was converted into text form to CSV format for easy retrieval. After that preprocessing is completed by cleaning of data by removing all the missing value present in the data and dropping all the unwanted columns. After this pd.to_datetime() functionality of pandas library is used as we have change the 'Formatted Date' column in the dataset into Date-Time format as datatype of this column is 'object' datatype and training of our model on object data type is not possible. Then by Resampling, new DataFrame only for Apparent Temperature and Humidity is created this is done because main objective of the study is to show complete analysis of the influence of Global Warming on the temperature and humidity in Finland over the course of 10 years from 2006 to 2016. Then exploratory Data Analysis is been performed on the new DataFrame in this relationship between Apparent Temperature and Humidity is visual presented and variation between Apparent Temperature and Humidity for the course of 10 years is shown visually. Then Univariate Analysis of the data is been performed by showing
of different weather conditions with different Apparent Temperature in a visual manner and also by showing of different weather conditions with different Humidity in a visual manner. Thenat the end to know exact major impact of Global Warming on temperature and humidity in Finland over the course of 10 years we make monthly comparison charts to show how global warning has affected the temperature and humidity for 10 years. This can be summed up as follows:

- Downloading of Jupyter Notebook
- Generating of Dataset
- Data cleaning
- Changing FormattedDate Column Datatype from Object to Datetime
- By Resampling, creation of the new DataFrame only for ApparentTemperature and Humidity
- Performing of Exploratory Data Analysis
- Performing of Univariate Analysis
- Making of monthly comparison charts to show temperature andhumidity Finland over the course of 10 years on monthly basis
- Make a conclusion as per the results of analysis and comparisonchart

IV. SYSTEM REQUIREMENTS SEPCIFICATIONS

The software being developed utilizes Python, specifically version 3.6, as its primary programming language. The dataset integral to the project has been sourced from the Kaggle platform and is stored locally on the development machine. For the purpose of creating an efficient development workflow, 'Jupyter Notebook' has been chosen as the development environment of choice. Jupyter Notebook is renowned for its robust capabilities tailored towards the scientific community, including but not limited to advanced code editing, in-depth analysis, and an array of debugging and profiling tools. It seamlessly integrates these features with the added advantages of data exploration and interactive computing, providing comprehensive inspection tools and sophisticated visualization options, making it an ideal choice for scientists, engineers, and data analysts working with Python.

IV.I System Configuration:

IV.I.I Hardware Requirements:

- Processor: Intel i3 2.0 Gigahertz or similar processor
- RAM: 4 GB or more
- Hard Disk: 500 GB
- Solid State Drive (SSD): Optional but recommended for improvedperformance.

IV.I.II Software Requirements:

- Operating System: Windows 7 or above

- Programming Language: Python
- Software: Anaconda distribution (includes Jupyter Notebook)

IV.II Python 3:

Python is one of the most common languages used for machine learning because it is easy to learn, enormous packages it supports like scikit learn that contains the implementation of common
machine learning algorithms, built-in library supports, pre-processedmodels, support for larger networks and huge toolset.

IV.III Anaconda Software:

Anaconda is a Python distribution that includes installations and packages managements tools. It offers a wide selection of packages and commercial support [20]. Additionally, Anaconda is an environment manager that allows users to create different Python environments, each without its own settings. This feature enables developers to work on multiple projects with different dependencies and configurations without conflicts [20]. Anaconda's package management toolsprovide access to over 7,500 open-source packages for data science, machine learning, and AI. It includes packages such as NumPy, pandas, and scikit-learn, making it an excellent tool for data science projects. Furthermore, Anaconda provides commercial support for its enterprise users, allowingbusinesses to use Anaconda with the assurance of dedicated support. Overall, Anaconda is a powerful tool for Python developers, especially for those working on data science and machine learning projects. Its package management tools and environment management features provide flexibility and easeof use for working on multiple projects. Anaconda can help with:

- Python can be installed over the multiple platforms
- Different environments can be supported separately
- Distributing with not having correct privileges and
- Support and running with specific packages and libraries

IV.IV Kaggle:

It operates as a hub for data science and machine learning enthusiasts, owned by Google LLC. This platform serves as a

	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millib
count	96453.000000	96453.000000	96453.000000	96453.000000	96453.000000	96453.000000	96453.0	96453.000
mean	11.932678	10.855029	0.734899	10.810640	187.509232	10.347325	0.0	1003.235
std	9.551546	10.696847	0.195473	6.913571	107.383428	4.192123	0.0	116.969
min	-21.822222	-27.716667	0.000000	0.000000	0.000000	0.000000	0.0	0.000
25%	4.688889	2.311111	0.600000	5.828200	116.000000	8.339800	0.0	1011.900
50%	12.000000	12.000000	0.780000	9.965900	180.000000	10.046400	0.0	1016.450
75%	18.838889	18.838889	0.890000	14.135800	290.000000	14.812000	0.0	1021.090
max	39.905556	39.344444	1.000000	63.852600	359.000000	16.100000	0.0	1046.380

repository where individuals can discover and share datasets, as well as a workspace where they can develop models through a browser-based interface. It fosters collaboration among its community of users, who range from data scientists to machine learning engineers. Additionally, Kaggle is known for hosting competitions that challenge participants to come up with solutions to various data science problems. Our project leverages this resource to obtain the necessary dataset for execution.

V. RESULTS AND FINDINGS

The project starts by importing of all the libraries which functions willbe used for cleaning of the data, data analysis & visualisation. Then dataset used for the project is been loaded after that dimensions and datatype of the dataset is been known with the help of pandas library and also statistical information of the dataset is known by writing codefor it this was all done to have a complete overview of the dataset before doing the data cleaning. Below figure 1 is showing the datatypeof the DataFrame and figure 2 shows statistical information of the DataFrame.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96453 entries, 0 to 96452
Data columns (total 12 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  float64
 8   Visibility (km)           96453 non-null  float64
 9   Loud Cover                96453 non-null  float64
 10  Pressure (millibars)      96453 non-null  float64
 11  Daily Summary             96453 non-null  object
dtypes: float64(8), object(4)
memory usage: 8.8+ MB
```

Figure 1: Datatypes of the dataframe

```
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
Loud Cover                   0
Pressure (millibars)         0
Daily Summary                0
dtype: int64
```

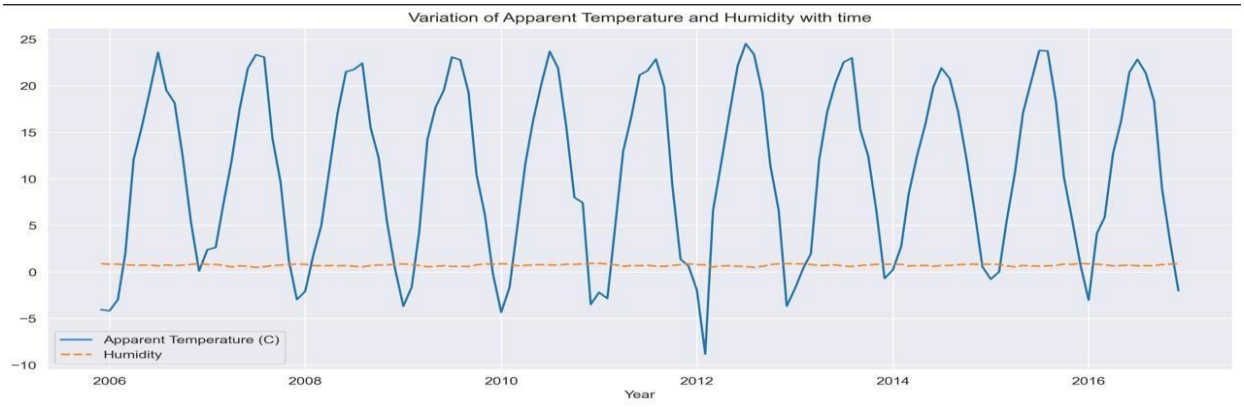
Figure 2: Statistical details of the dataframe

Then data cleaning process of dataframe starts for this firstly a check for any missing value present in the dataframe is done. This was done because if during analysis any missing value is found then visualisation of it will not be possible. Then isnull().sum() function is used to know the null values if any present in the dataframe. The picture of all null values present in the dataframe as per column name is as follows:

As it is clearly visible in the above figure that the feature — 'Precip Type' has 517 null values.Thankfully this feature is of no use to us, so we can simply drop this feature later. After this all unnecessary columns will be dropped which ends the data cleaning process. For analysing data further pd.to_datetime() functionality of pandas library is used. It is used to change the 'Formatted Date' columnin the dataset into Date-Time format as datatype of this column is 'object' datatype and training of our modelon object data type is not possible. After this a heatmap is made which shows the pairwise correlationbetween all columns in the dataframe it is shown in the below figure.

HEATMAP SHOWING PAIRWISE CORRELATION OF ALL COLUMNS IN THE DATAFRAME

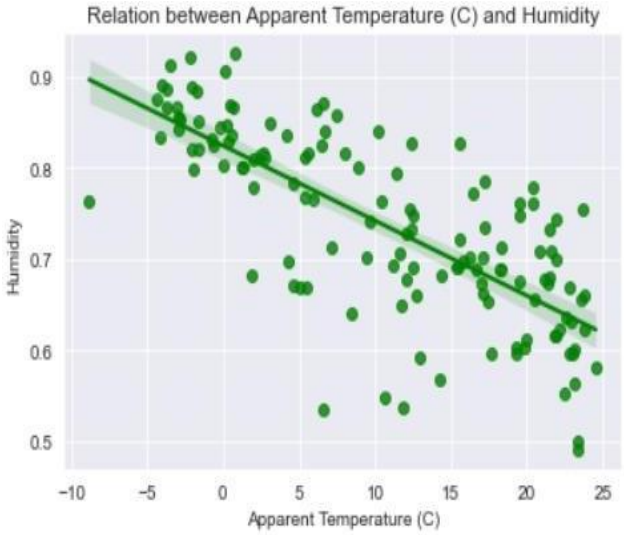
After this I firstly used pd.set_index() of pandas library to set the index of very element in dataframe asper the Formatted Date column and then I created a new dataframe only for Apparent Temperature & Humidity by using pd.resample() function of the pandas library.



The head of new dataframe is shown in the below figure.

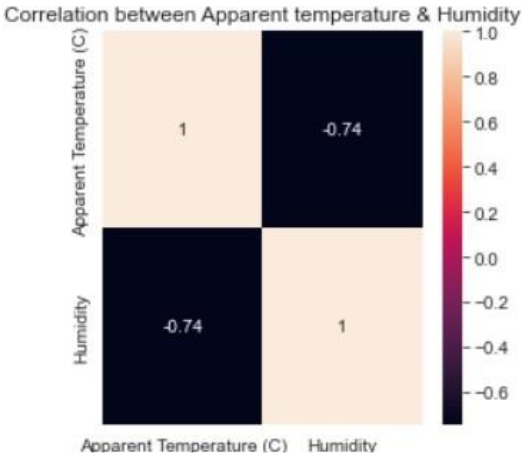
Formatted Date	Apparent Temperature (C)	Humidity
2005-12-01 00:00:00+00:00	-4.050000	0.890000
2006-01-01 00:00:00+00:00	-4.173708	0.834610
2006-02-01 00:00:00+00:00	-2.990716	0.843467
2006-03-01 00:00:00+00:00	1.969780	0.778737
2006-04-01 00:00:00+00:00	12.098827	0.728625

After this I Exploratory Data Analysis of the dataframe is started in this firstly a scatter-plot is made which shows the correlation between Humidity and Apparent temperature. In this scatter-plot it is observed that there might be Linear Relationship between Humidity and Apparent Temperature with negative slope. The scatter-plot which shows the correlation between Humidity and Apparent temperature is shown in the below figure.



Then a line diagram is made which shows the overall variationof Humidity and Apparent Temperature over the course of 10 years from 2006 to 2016. From this line graph two things are observed firstly Humidity remained same from 2006 to 2016 and secondly Apparent Temperature frequently changed from2006 to 2016. The diagram which shows the overall variation of Humidity and Apparent Temperature over the course of 10 years is shown in the below figure.

After this a heatmap is made which shows correlation between Humidity and Apparent Temperature through which it was observedthat there is a correlation of negative 0.74 between Apparent Temperature & Humidity. The below figure shows heatmap showingcorrelation between Humidity and Apparent Temperature.



HEATMAP SHOWS CORRELATION BETWEEN APPARENT TEMPERATURE & HUMIDITY

Then a 2D Color Coding Scatter-Plot for each Summary type(Summery type column Included the type of whether on a particular day) on the premises of Humidity and Apparent Temperature is made. By this it was observed that mostly weather was partly cloudy/rainy or clear in Finland and hardly a few days had a light rain or dry or dangerously windy & partly cloudy. This 2D Scatter Plot is shown in the below figure.

After this Univariate Analysis of the dataframe is started in this, firstly a histogram is made which shows Analysis of different WeatherConditions with Apparent Temperature and then secondly a histogram is made which shows Analysis of different Weather Conditions with Humidity. Through this it was observed that Humidity has turned out be a better Feature than Apparent Temperature. Histogram showing Analysis of different Weather Conditions with Apparent Temperature is shown in the below figure 1 and histogram showing Analysis of different Weather Conditions with Humidity is shown in the below figure 2.

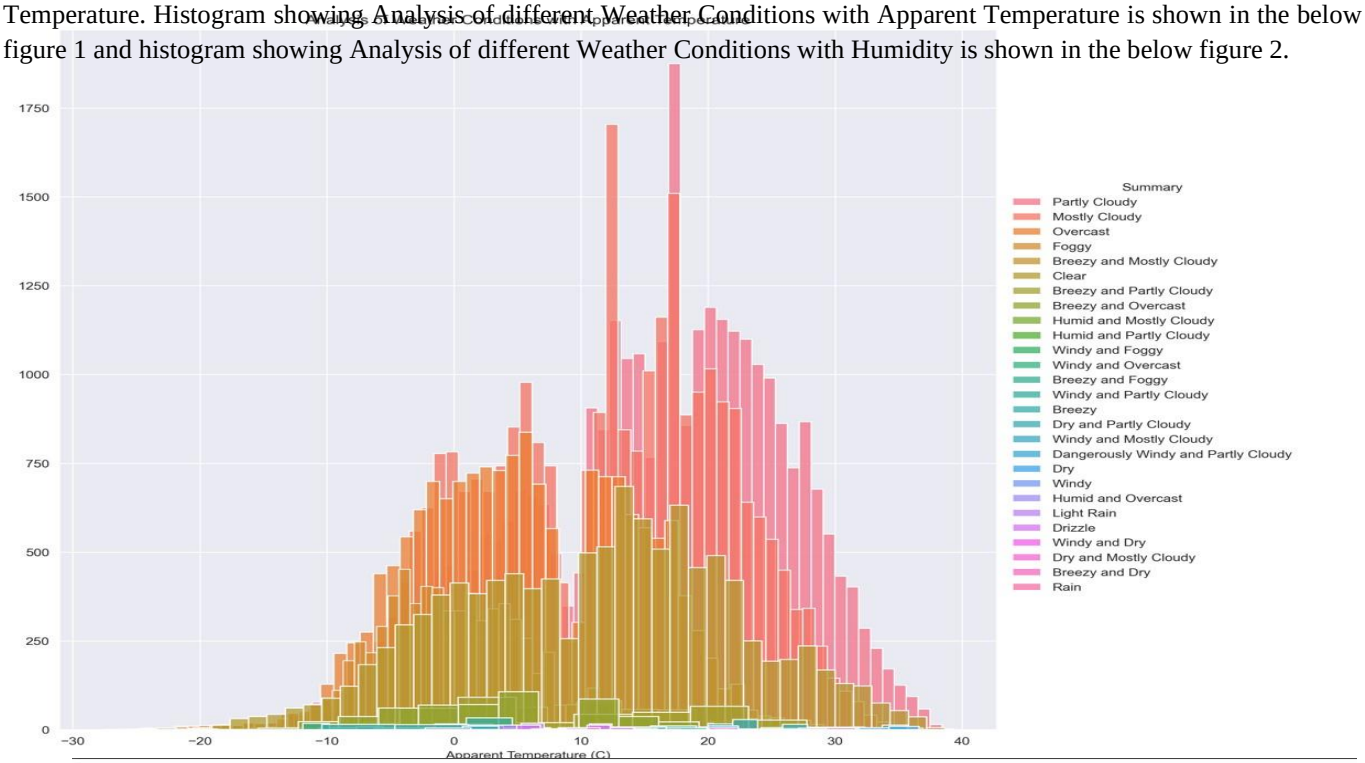


Figure 1

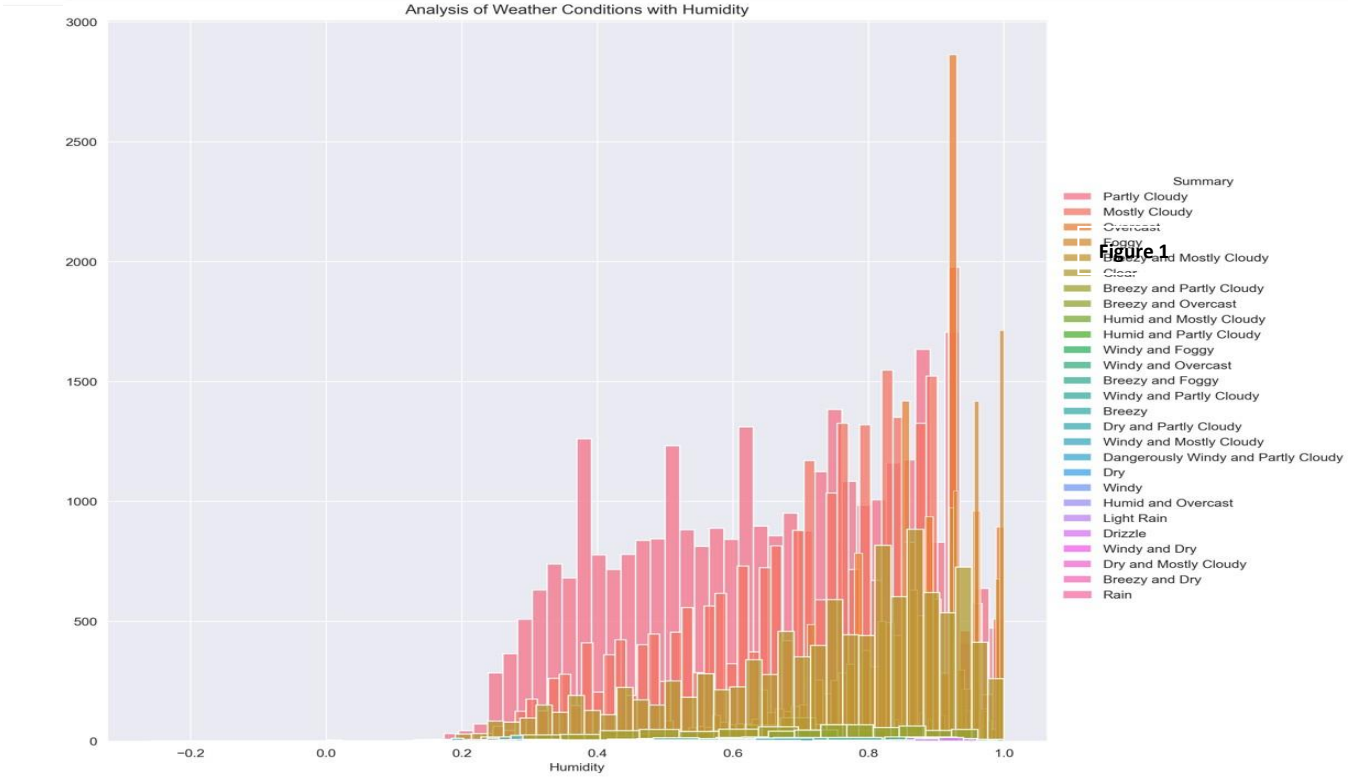
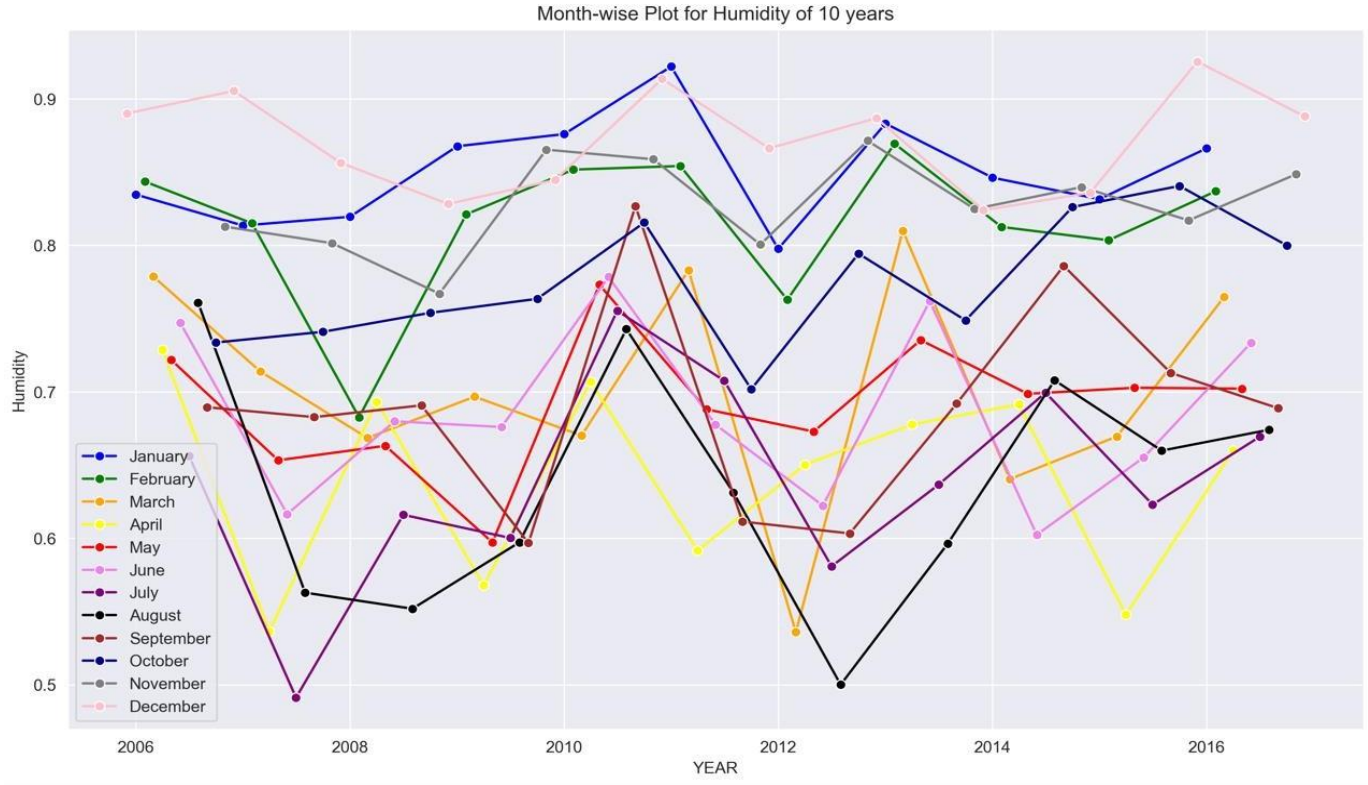
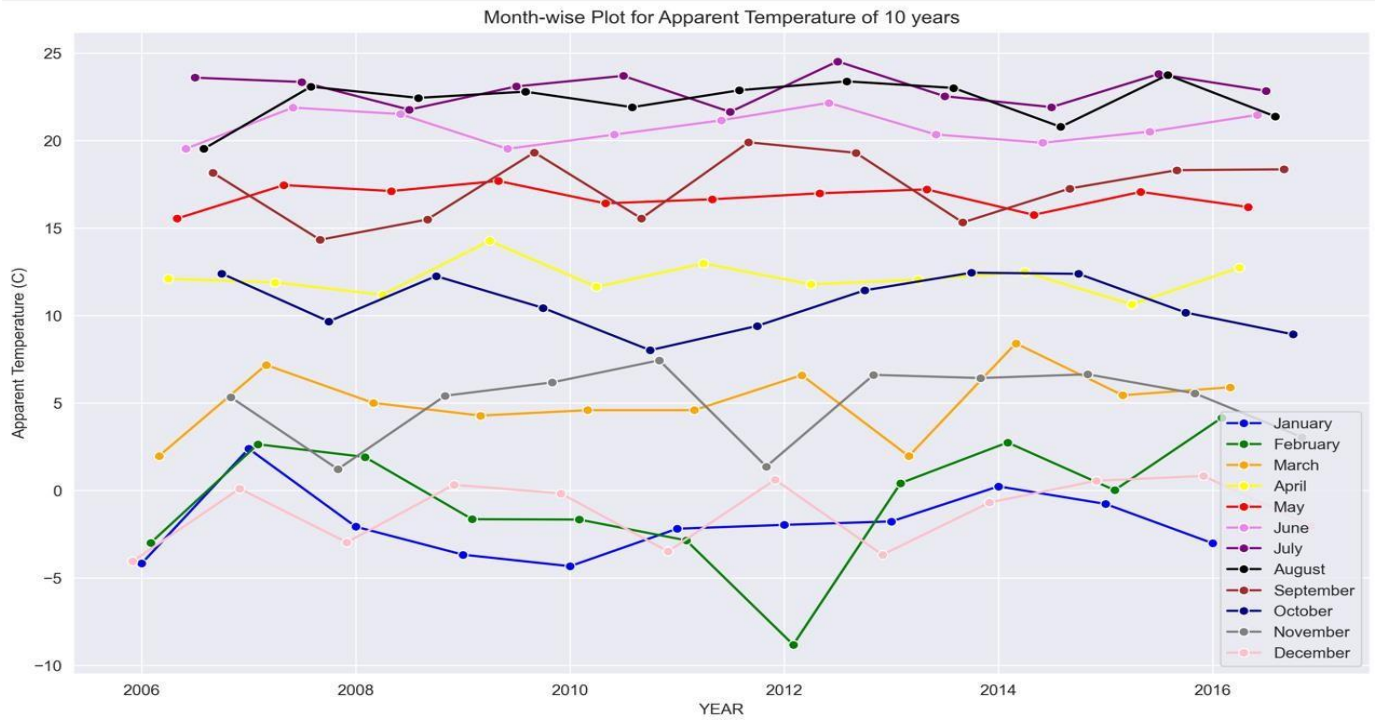


Figure 2

Then a function is made for plotting the Humidity & Apparent Temperature for all the months. After that two lineplots was made first one shows month wise plot for Apparent Temperature of the 10 years and the second one shows month wise plot for humidity of the 10 years. The lineplots is shown in the below figures.

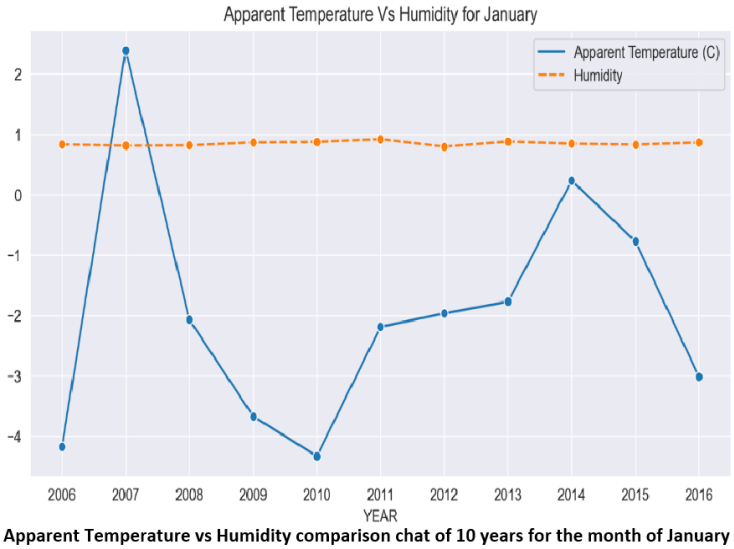




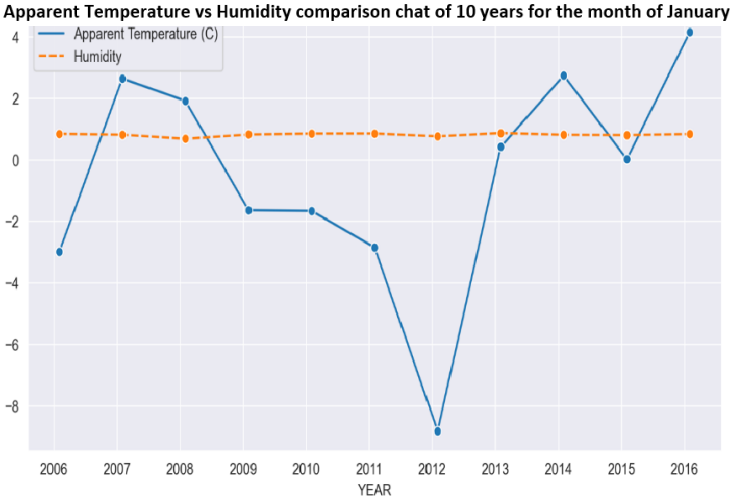
After that month-wise comparison lineplot between the Humidity and Apparent Temperature over the course of 10 years is made to make my final verdict on how Global Warning has affected the humidity and temperature in Finland over the course of 10 years from2006 to 2016. The month-wise comparison lineplot is shown in the comparison chart section of this research paper.

VI. COMPARISON CHART

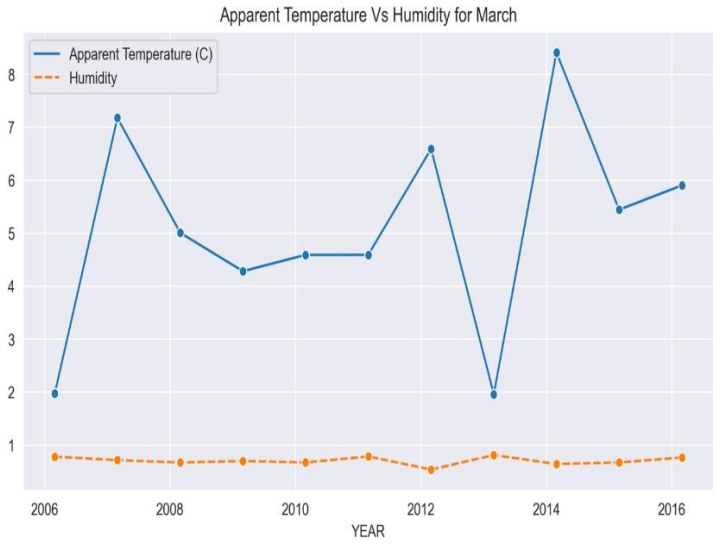
In this section of the project report, I will show the month- wise over the course of 10 years. Comparison chartsare as follows :



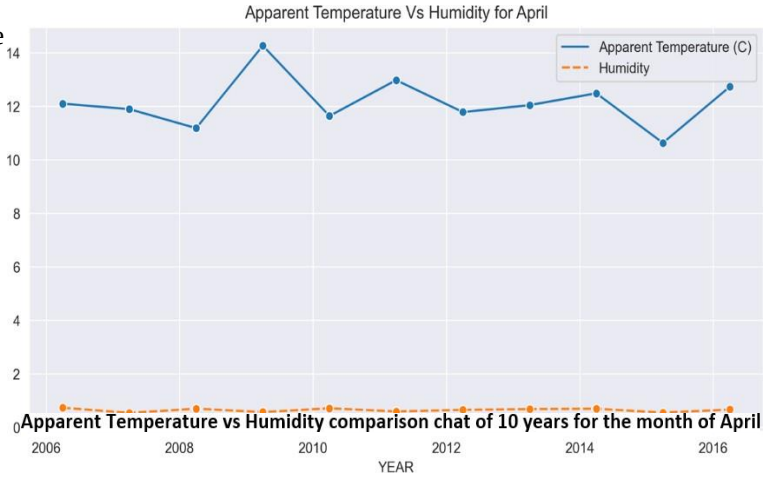
Apparent Temperature vs Humidity comparison chat of 10 years for the month of January



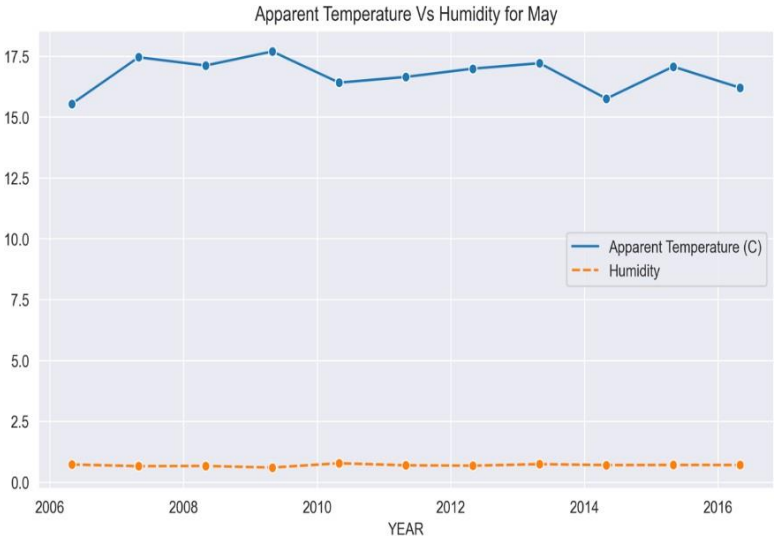
Apparent Temperature vs Humidity comparison chat of 10 years for the month of February



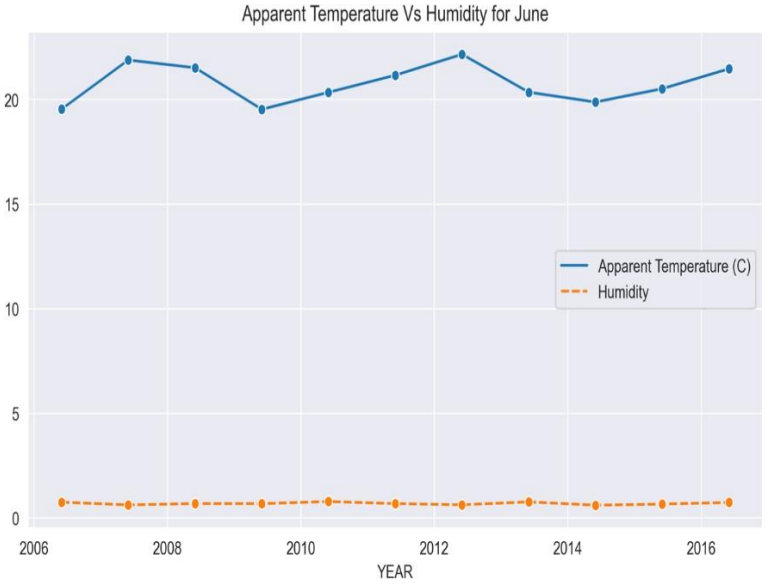
Apparent Temperature vs Humidity comparison chat of 10 years for the month of March



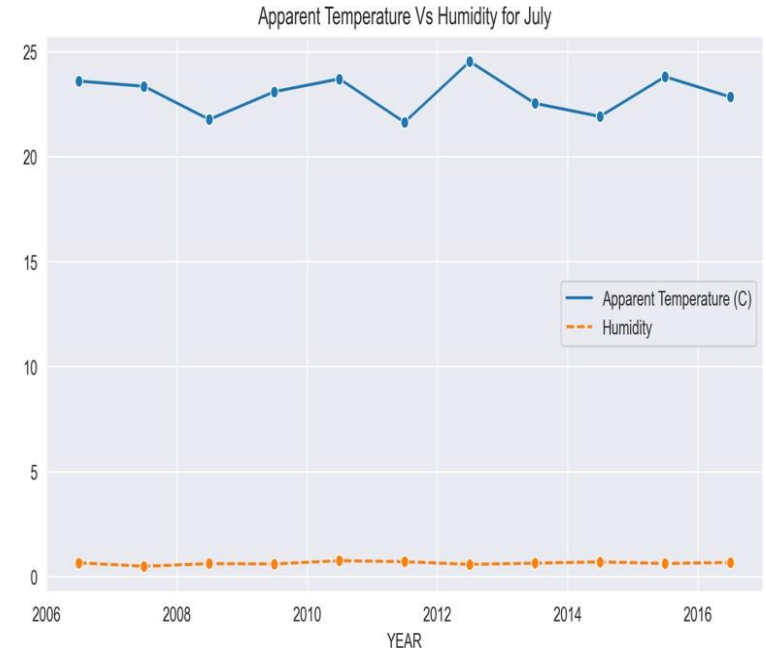
Apparent Temperature vs Humidity comparison chat of 10 years for the month of April



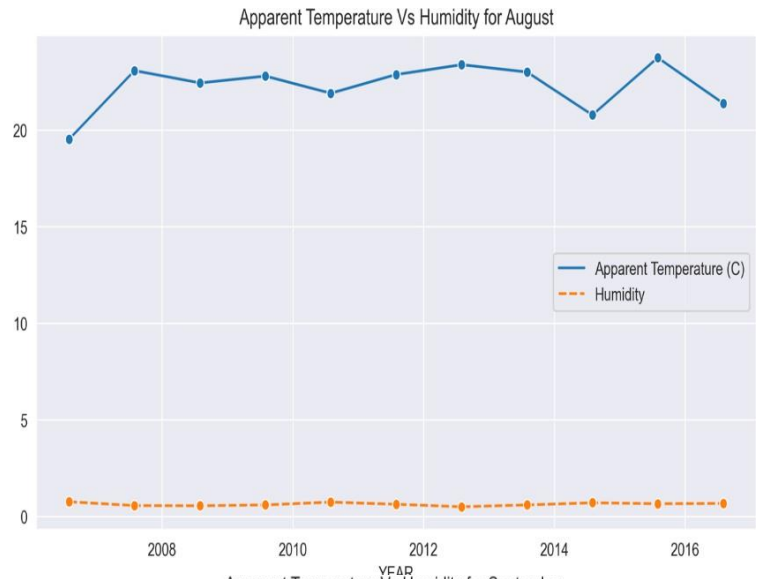
Apparent Temperature vs Humidity comparison chat of 10 years for the month of May



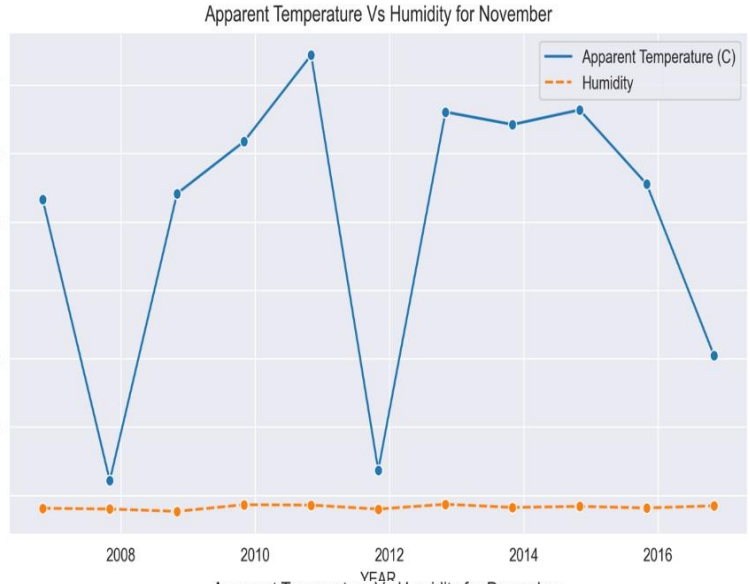
Apparent Temperature vs Humidity comparison chat of 10 years for the month of June

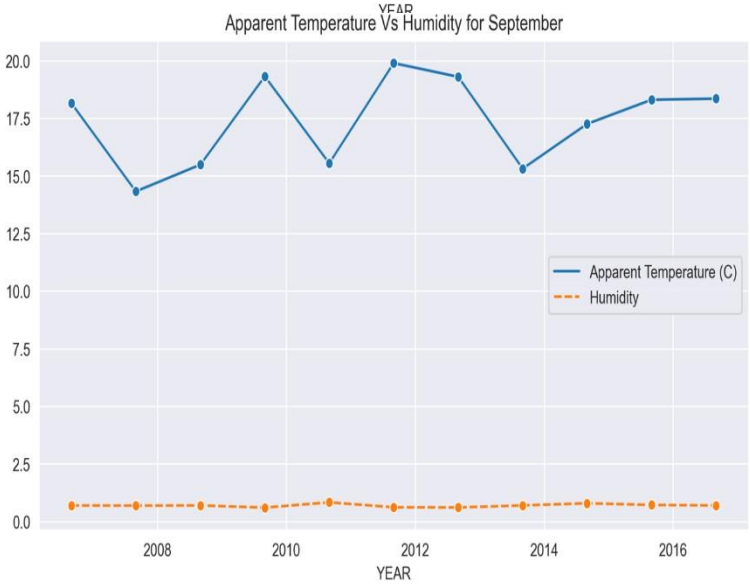


Apparent Temperature vs Humidity comparison chat of 10 years for the month of July

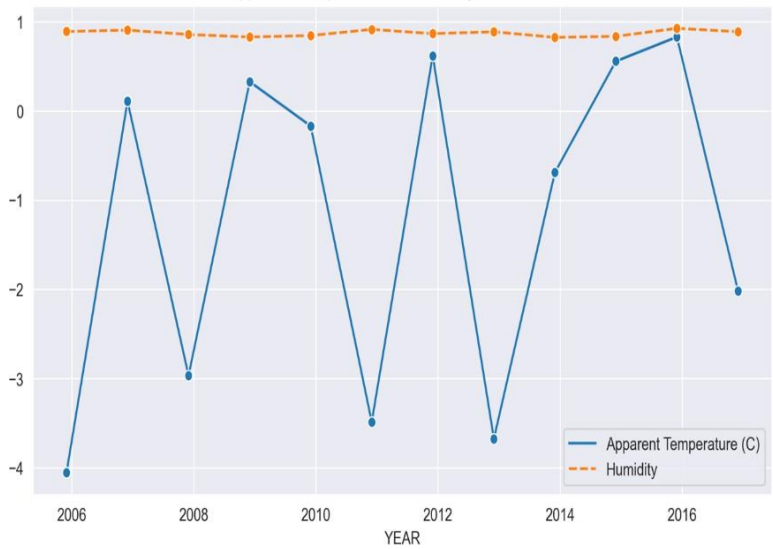


Apparent Temperature vs Humidity comparison chat of 10 years for the month of October





Apparent Temperature vs Humidity comparison chat of 10 years for the month of September



Apparent Temperature vs Humidity comparison chat of 10 years for the month of December

 DOI: 10.55041/IJSREM34399

Apparent Temperature vs Humidity comparison chat of 10 years for the month of August

As it is clearly visible in the above month-wise comparison charts, there are several ups & downs in temperature & the average humidity has remained same over the course of 10 years in the Finland. So, it can be concluded that because of the global warming there has been an uncertainty in the temperature over last 10 years in Finland from 2006 to 2016.

VII. CONCLUSION

From this analysis, it can be concluded that the Apparent temperature & humidity is compared monthly across 10 years of the metrological data indicate a fluctuating increase and decrease in temperature due to Global warming. Our environment is highly affected by Global Warming. From the conducted analysis there has been a clear observation that there is fluctuating increase and decrease in temperature over ten years. But, in case of humidity, conducted analysis shows that there is negligible amount of rise and negligible amount of decrease in humidity so we can say that humidity remained kind-of same over thecourse 10 years in Finland. So, we can say that globalwarming has caused an uncertainty in the temperature over the past 10 years from 2006 to 2016 in Finland. In this analysis we also observed that Mostly Weather is partly cloudy/rainy or clear in Finland and hardly a few days had a light rain or dry or dangerously windy & partly cloudy.

BIBLIOGRAPHY

[1] https://www.kaggle.com/muthuj7/weather-dataset

[2] http://vis.computer.org/vis2004contest/vrvis

[3] https://www.sciencedirect.com/topics/social- sciences/meteorological-data

[4] https://www.semanticscholar.org/paper/Methods-for- the-visualization-of-clustered-climate-Nocke-Schumann/f08db0000a273803f5179a31217ae9001c66 9e74

[5] Visualization of Climate and Climate Change Data -

Potsdam Institute ... - MOAM.INFO

[6] https://www.academia.edu/64444679/Visualization_of

_Climate_and_Climate_Change_Data_An_Overview

[7] https://www.researchgate.net/publication/313869701_ Real-Time_Meteorological_Data_Analysis_and_Mapping

[8] https://ieeexplore.ieee.org/abstract/document/398900

[9] https://digitalcommons.fiu.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=2605&context=e td

[10] https://www.youtube.com/watch?v=E_n3Ft4WozM

[11] Y. LeCun, Y. Bengio and G. Hinton, "“Deep learning,” ,,"

Nature, vol. 521, no. 7553, p. 436–444, 2015

[12] https://www.jstor.org/stable/26230641

[13] Han, Jiawei, M. Kamber and J. Pei, Data mining: concepts andtechniques, 3rd Edition. Morgan, 2013

[14] Treinish, L. A. (1994), Severe Rainfall Events in NorthwesternPeru: Visualization of Scattered Meteorological Data. Proceedings of Visualization’94