**ForeWeather**

A Project Report

Submitted for the partial fulfillment for the award of the degree of

### Bachelor in Technology (B.Tech.)

**in**

**Computer Science and Engineering**

Submitted by

Tejashiwini Singh (2016317), Aryma Rawat (2016124), Suhani Chaudhary (2016305), Simple Jain (2016292)

Under the supervision of

Dr. Sneha Asopa Prof. Saurabh Mukherjee, Professor,

Dr. Neelam Sharma, Associate Professor,

Dr. Pooja Asopa, Assistant Professor,

Dr. Sneha Asopa, Assistant Professor



**Department of Mathematics & Computing**

**Banasthali Vidyapith**

**Banasthali - 304022**

**Session: 2022-23**

**Certificate**

### Certified that Aryma Rawat (2016124) Tejashiwini Singh (2016317), Aryma Rawat (2016124), Suhani Chaudhary (2016305), and Simple Jain (2016292) has carried out the project work titled “ForeWeather” from January 2022 to April 2022 for the award of the Bachelor in Technology (B.Tech.) from Banasthali Vidyapith, Tonk, Rajasthan under my supervision. The thesis embodies the result of original work and studies carried out by the student herself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else.

Dr. Sneha Asopa

Designation: Assistant Professor

Place: Banasthali Vidyapith

Date: 12/04/2023

**ABSTRACT**

ForeWeather is an application of science and technology to predict the state of the atmosphere for a given location. Here, this system will predict the weather based on parameters such as temperature, humidity, and wind speed. This system is an application with an effective graphical user interface. To predict the future weather condition, the variation in the conditions in past years must be utilized. The probability that it will match within the span of the adjacent fortnight of the previous year is very high. We have proposed the use of Decision Tree Regression for a weather prediction system with parameters such as temperature, humidity, and wind speed. It will predict weather based on previous records therefore this prediction will prove reliable. This system can be used in Air Traffic, Marine, Agriculture, Forestry, Military, and Navy ECT.

**ACKNOWLEDGEMENT**

Engineering in all disciplines must acquire knowledge of project making. Students, in particular, will find project-making an integral part of their studies that will infuse the spirit of doing practical work in them.

The satisfaction that accompanies a successful completion of any task would be incomplete without the mention of the people who made it possible and whose constant guidance crowned our efforts with success.

We sincerely express our deep gratitude to the management of our college for giving us the liberty to choose and to work on the most relevant project i.e. “ForeWeather”. We are thankful to Dr. Sneha Asopa (designation, Department of Computer Science) who helped us with our project from the very beginning till the end. Her continuous surveillance of our work allowed us to work more efficiently.

Tejashiwini Singh (2016317)

Aryma Rawat (2016124)

Suhani Chaudhary (2016305)

Simple Jain (2016292)

**TABLE OF CONTENTS**

* Introduction
* Scope
* Overall Description
* User Characteristics
* Constraints
* Libraries Used
* Assumptions and Dependencies
* Requirement Analysis (SRS)
  + Requirement specification
  + Hardware and Software Requirements
  + Feasibility Study
  + Detailed Description of Components
  + Product Functions
  + Use-case Diagrams
* System Design (SDS)
  + High-level Design
  + Class Diagram
  + Activity Diagram
  + Sequential Diagram
* Dataset description
* Coding
* Testing
* User Interface
* Appendices
* References

**INTRODUCTION**

Weather forecasting is the task of predicting the state of the atmosphere at a future time and a specified location. Traditionally, this has been done through physical simulations in which the atmosphere is modeled as a fluid. The present state of the atmosphere is sampled, and the future state is computed by numerically solving the equations of fluid dynamics and thermodynamics. However, the system of ordinary differential equations that govern this physical model is unstable under perturbations, and uncertainties in the initial measurements of the atmospheric conditions and an incomplete understanding of complex atmospheric processes restrict the extent or accurate weather forecasting to a 10-day period, beyond which weather forecasts are significantly unreliable. Machine learning, on the contrary, is relatively robust to perturbations and doesn’t require a complete understanding of the physical processes that govern the atmosphere. Therefore, machine learning may represent a viable alternative to physical models in weather forecasting.

Machine learning is the ability of a computer to learn without being explicitly programmed. It allows machines to find hidden patterns and insights. In supervised learning, we build a model based on labeled training data. The model is then used for mapping new examples. So, based, on the observed weather pattern from the past, a model can be built and used to predict the weather.

This project work focuses on solving the weather prediction anomalies and in-efficiency based on linear regression algorithms and formulating an efficient weather prediction model based on the Decision Tree Regression algorithms.

Weather forecasting is the application of science and technology to predict the conditions of the atmosphere for a given location and time. People have attempted to predict the weather informally for millennia and formally since the 19th century. Weather forecasts are made by collecting quantitative data about the current state of the atmosphere, land, and ocean and using meteorology to project how the atmosphere will change at a given place.

For a long time, the researchers had attempted to establish a linear relationship between the input weather data attributes and the corresponding target attribute. But with the discovery of nonlinearity within different attributes of weather data, the focus has shifted towards the nonlinear prediction of the weather.

In order to analyze weather forecasting and learn about it, we shall use K-Means clustering and decision tree as machine learning algorithms in our project for training and forecasting and implement out training model in a user-friendly website, ‘ForeWeather’.

**SCOPE**

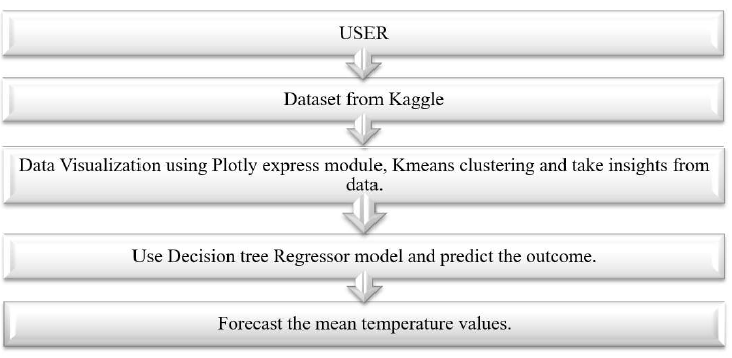
ForeWeather web application has improved and user-efficient ways to eradicate the existing problem. The application will gather weather information, examine predicted readings, display results, and assist users with their day as needed. It will show the temperature, day type, wind speed, humidity, and what the rest of the day will be like. The system displays hourly and daily weather changes, computes a weekly output, and assists the user in choosing clothing that is appropriate for the weather.

The benefit of this application is as follows:

1. Ease of Use: Ease to use is definitely a big advantage of a weather monitoring system. Weather stations like all other weather devices are designed to be efficient and straightforward, therefore, everyone can use them. It is so convenient and comfortable for users to get the most accurate information in the simplest way possible.
2. Accurate Local Forecast: In reality, the meteorological department may be located far from your home, and weather forecasts are made for regions, not specific areas. That’s a reason why in these instances, the weather predictions that they give are not always the most accurate. A home weather station can tell you the exact temperature in the area, what weather today that you are living in, and give you specific weather reports that are centered around your home.
3. Real-Time Alerts: Weather conditions can change in an instant and take a turn for the worse, especially in areas that are prone to storms and hurricanes. Therefore, having a home weather station in these instances can be a breeches-buoy for you and your family members to stay safe the moment the weather changes are detected.
4. Create a Smarter Home: For those who want to make their homes smarter, more modern with full of convenient functions, a weather station is definitely something that should be integrated into the house.

**THE OVERALL DESCRIPTION**

**PRODUCT PERSPECTIVE**



**USER CHARACTERISTICS**

The target audience for ForeWeather won’t need any kind of specialized knowledge or experience. They merely need rudimentary computer literacy and a working understanding of how to navigate a web application.

**CONSTRAINTS**

1. Accurate up to a certain level: Professional weather forecasters are not perfect, but their predictions are typically more accurate than those of this machine learning model. This implies that weather is a non-linear system. Additionally, our predictions were all based on data from one location as opposed to the multiple locations that most forecasters use. Even in our model, we have described the limitation of using linear regression in predicting the weather.
2. Actual motive and functionality of the application: The weather apps are excessively broad and the principal objective of the application basically directs the budget of the development of the application.
3. Design of the application: UI/UX designs always play a key role when designing an application. So consequently, it demands some huge chunks.
4. Security of the application: Security and privacy are always of primary concern these days. Nobody wants to expose the data of their application so it is better to take some measures in order to maintain perfect privacy and secure the information.

**LIBRARIES USED**

In this project, we have imported numpy, pandas, Plotly express, and datetime libraries.

1. NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms basic linear algebra, basic linear algebra, basic statistical operations, random simulation and much more.

At the core of the NumPy package, is the array object. This encapsulates n-dimensional arrays of homogenous data types, with many operations being performed in compiled code for performance.

1. Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.
2. Plotly Express is a built-in part of the Plotly library and is recommended starting point for creating the most common figures. Instead of matplotlib, we have used plotly as it is very helpful for data visualization and understanding the data easily with its easily built API. It is easy to switch from scatter plots to histograms to sunbursts with plotly.
3. DateTime is a class and an object in Python’s DateTime module. The arguments are a combination of date and time attributes, starting from the year and ending in microseconds.

**ASSUMPTIONS AND DEPENDENCIES**

We will be rigorously making use of the libraries in Python, to implement our algorithm and finally structure our website.

The developer’s dependencies that must be installed to run the project are:

1. NumPy to work with Linear Algebra
2. Pandas to work with Data
3. Plotly for visualization
4. Datetime for time series analysis
5. Anaconda with Jupyter or Python 3.2 or above

**REQUIREMENT ANALYSIS (SRS)**

* **SPECIFIC REQUIREMENTS**

**FUNCTIONAL REQUIREMENTS**

**Functional Requirement**

1. Users shall browse our website on Web Browsers such as Microsoft Edge, and Google Chrome.
2. They shall find our landing page and the search bar.
3. Users will be required to enter a valid city name.

**NON-FUNCTIONAL REQUIREMENTS**

**PERFORMANCE**

The landing page shall provide 6 second or less response time in a Chrome desktop browser, including the rendering of text and images and over an LTE connection.

**AVAILABILITY**

1. The internet is required whenever using ForeWeather.
2. ForeWeather will be available to users at any point in time.

**RELIABILITY**

Our model will have a well-structured database and all information regarding all its datasets to have a backup means in need of emergency. We will be storing our data in the cloud which will improve our work efficiency. Our application model will provide and use all sorts of authentic sources that are reliable and will help in smooth functioning and providing desired output.

**MAINTAINABILITY**

We will also use a modular structure to build our application model hence it will be easy to maintain and work upon. We will have a separate folder for screens and assets. Hence the division of modules and making separate folders will help in maintaining. Tackling errors and finding bugs will be easier and faster, hence maintenance will be done easily.

**PORTABILITY**

Our model would be able to function properly in computers with Windows OS 7 or higher without any change in its behavior and performance.

* **HARDWARE AND SOFTWARE REQUIREMENTS**

**Hardware Requirements:**

We will be needing a PC with

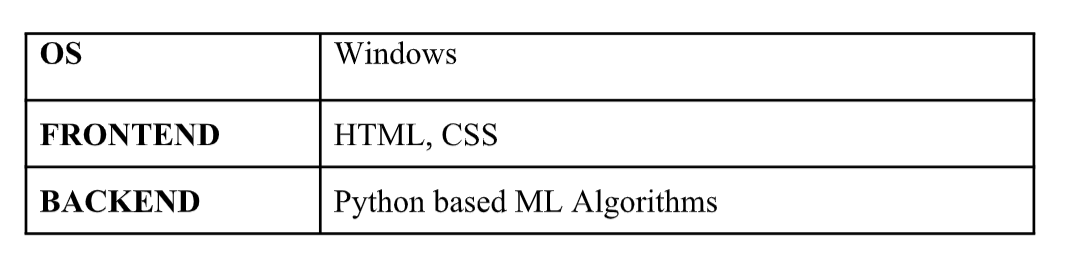
* a processor CORE i5 or above
* RAM 4GB or above
* System type- 32bit Operating System or above

**Software Requirements:**

**Client Side**:



**Developer’s Side**:



**FEASIBILITY STUDY**

Our project is feasible as it will available in form of a web application for free use on operating systems such as MACOS, Windows, IOS, Android with just internet connection support.

* **PRODUCT FUNCTION**

The essential ForeWeather application features include adding multiple locations, earth time-lapse, rain predictions, time of sun rise and sun set, UV Weather map, climate map, weather forecast, humidity, visibility, wind predictions, push notifications, API integrations, etc.

1. **Earth time-lapse**: The changes in the weather conditions across the globe are simply displayed by using images and pictures. This feature explains the previous climatic conditions, at the present moment and how it will be in the next consequences.
2. **Predictions about the rainfall:** This is another fundamental attribute that shows the forecasts for the rain. It also showcases the percentage of likelihood of rainfall and it is classified into various elements like cloudy, sunny, semi-cloudy, etc.
3. **Time of sunrise and sunset:** This Feature shows the duration of day and night. It will also mention the sunrise time and the sunset time.
4. **Predictions of Wind:** This feature is an added benefit, for the fishermen, sailors, windsurfers, and paragliders. Also, people who are planning to spend their weekends outside are also profited. For general users, this attribute is not that useful.
5. **Updates about Humidity:** For the people who are planning for a long drive or to have a long journey. It is always essential to monitor the humidity level and to start the journey.
6. **UV Weather Map**: This attribute displays the ultraviolet radiation of the sun across the globe by the Solar UV index. This is one of those unique features of the weather app development, and it is generally most helpful in summer.
7. **Map about Climate Conditions:** You can get a clear picture of the climate data with this feature. It comprises of humidity level, the temperature of the surroundings, and the level of carbon dioxide. This attribute is highly beneficial for scholarly people who are carrying out scientific experiments.
8. **Weather Forecast**: It is a fundamental factor of any weather app. This feature displays the prevailing status of the weather on a weekly, monthly, daily, and hourly basis.

* **DETAILED DESCRIPTION OF COMPONENT**

The data sets are collected by the program and are pre-processed which means the raw data is transformed into a useful and efficient form. Then all the incorrect, corrupt, incorrectly formatted, duplicate, or incomplete data is fixed or removed. After this, the appropriate data type and source, as well as suitable instruments to collect data is determined by the program. Next, the data is transported for further processes.

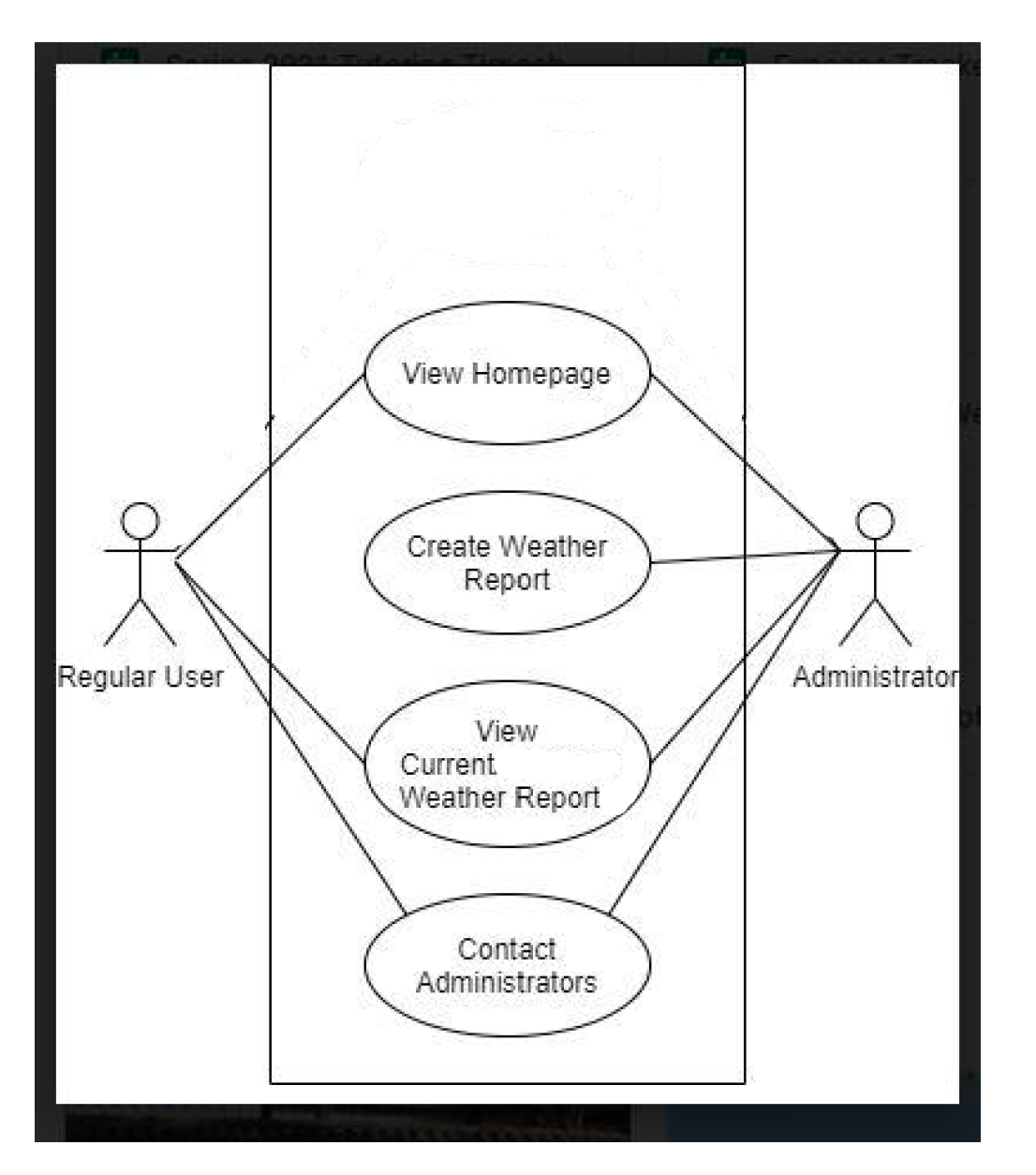
Simultaneously, the raw data sets are used to apply K-means Cluster which uses unlabeled data to predict the weather for the next five hours. The unlabeled data has been labeled and pre-processed for use in the program.

**K-means Clustering** is an unsupervised learning algorithms, which groups the unlabeled dataset into different clusters. The K defines the number of predefined clusters that must be created in the process.

* **USE-CASE DIAGRAM**

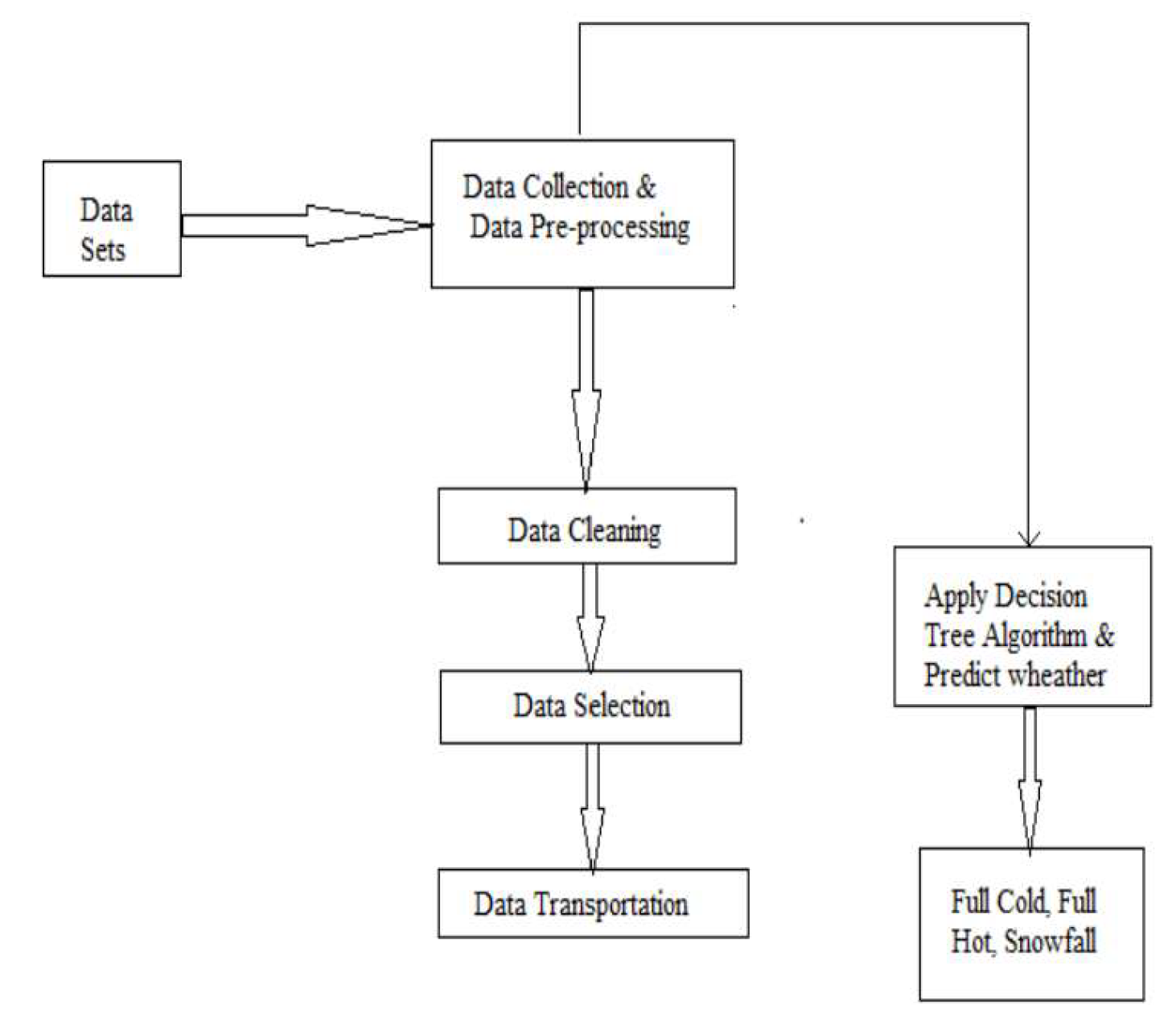
An interaction between a user and a system is described by a use case diagram. Use case diagrams to describe what a system does from the standpoint of an external observer. This emphasis is on what a system does rather than how. Use case diagrams are closely connected to scenarios. A scenario is an example of what happens when someone interacts with the system. A use case diagram is a collection of actors, use cases, and their communications.

For initial development, we can use this use case. In this use case diagram, we can see the following use cases and factors. Use cases are self-explanatory and they represent the main function of ForeWeather.

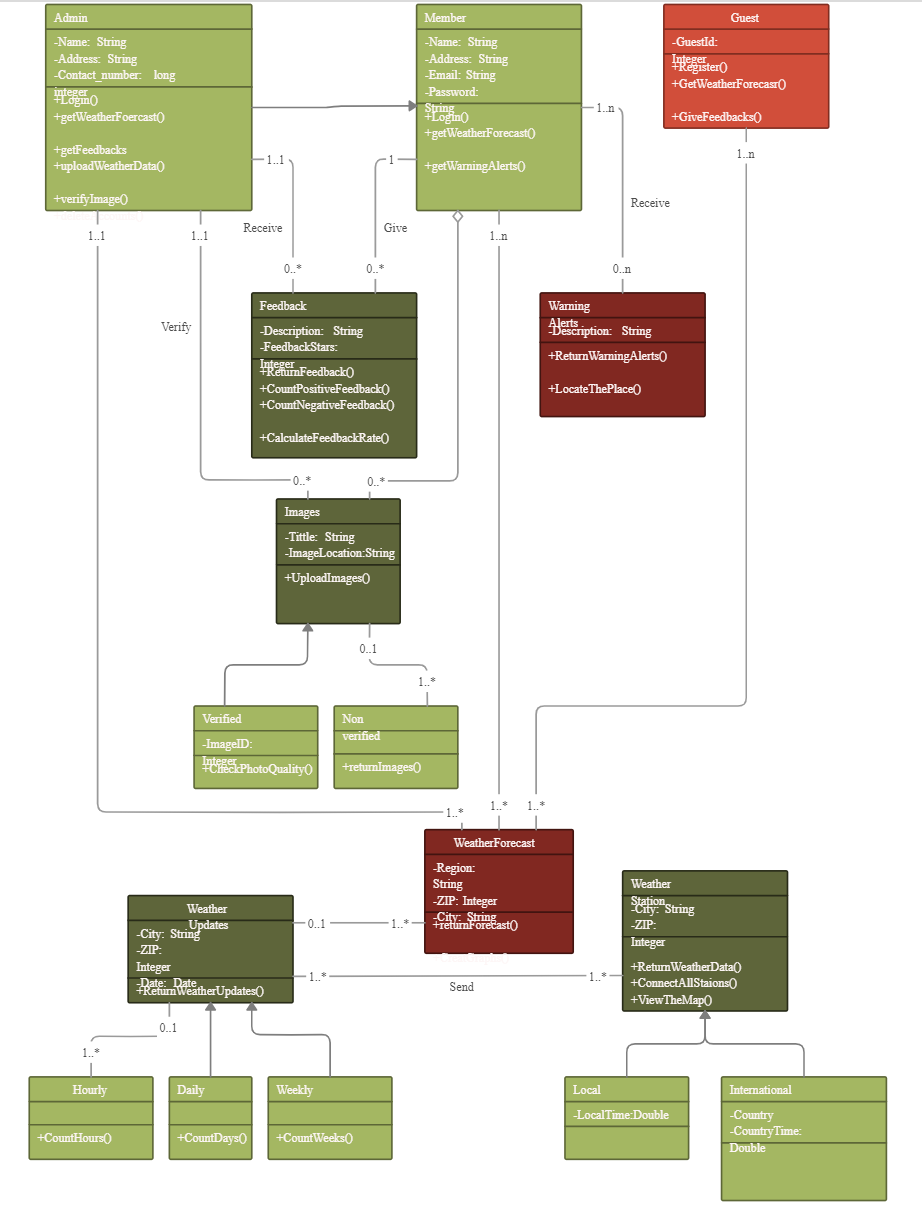


**SYSTEM DESIGN (SDS)**

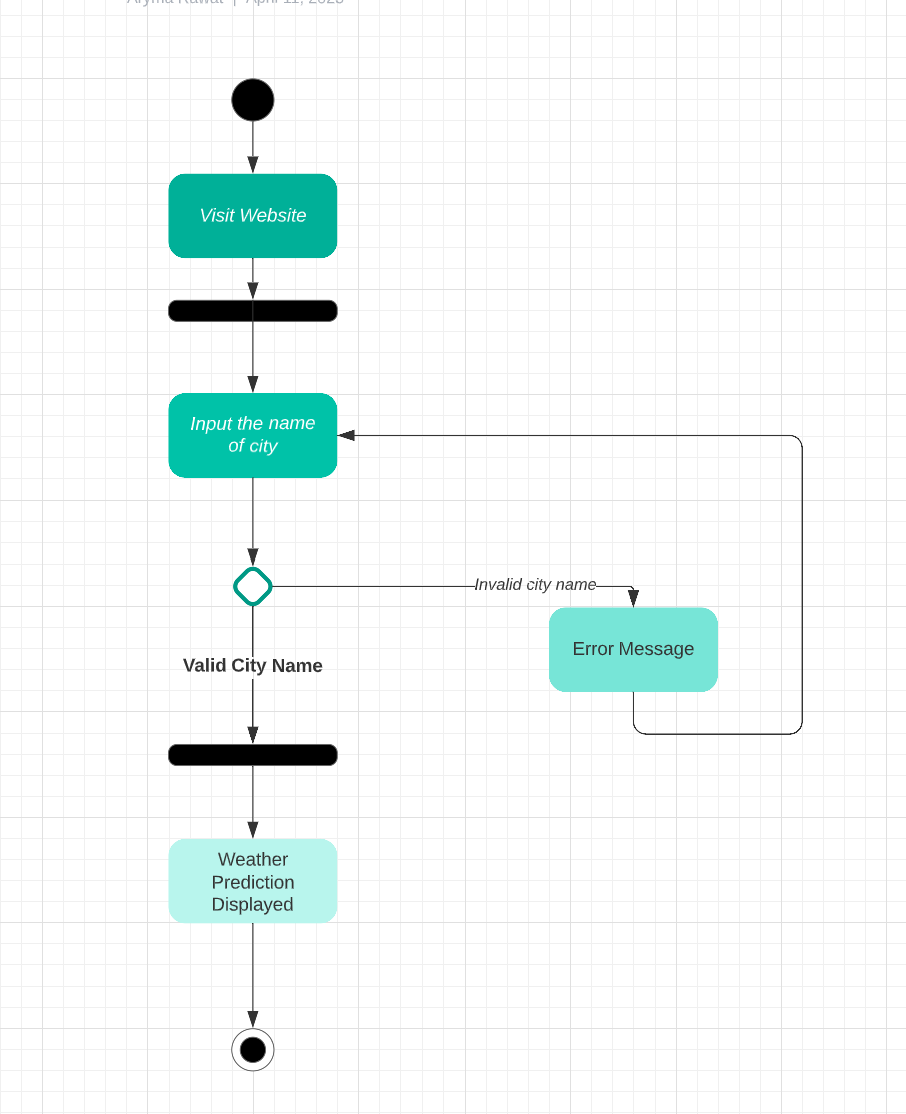
* **HIGH LEVEL DESIGN**



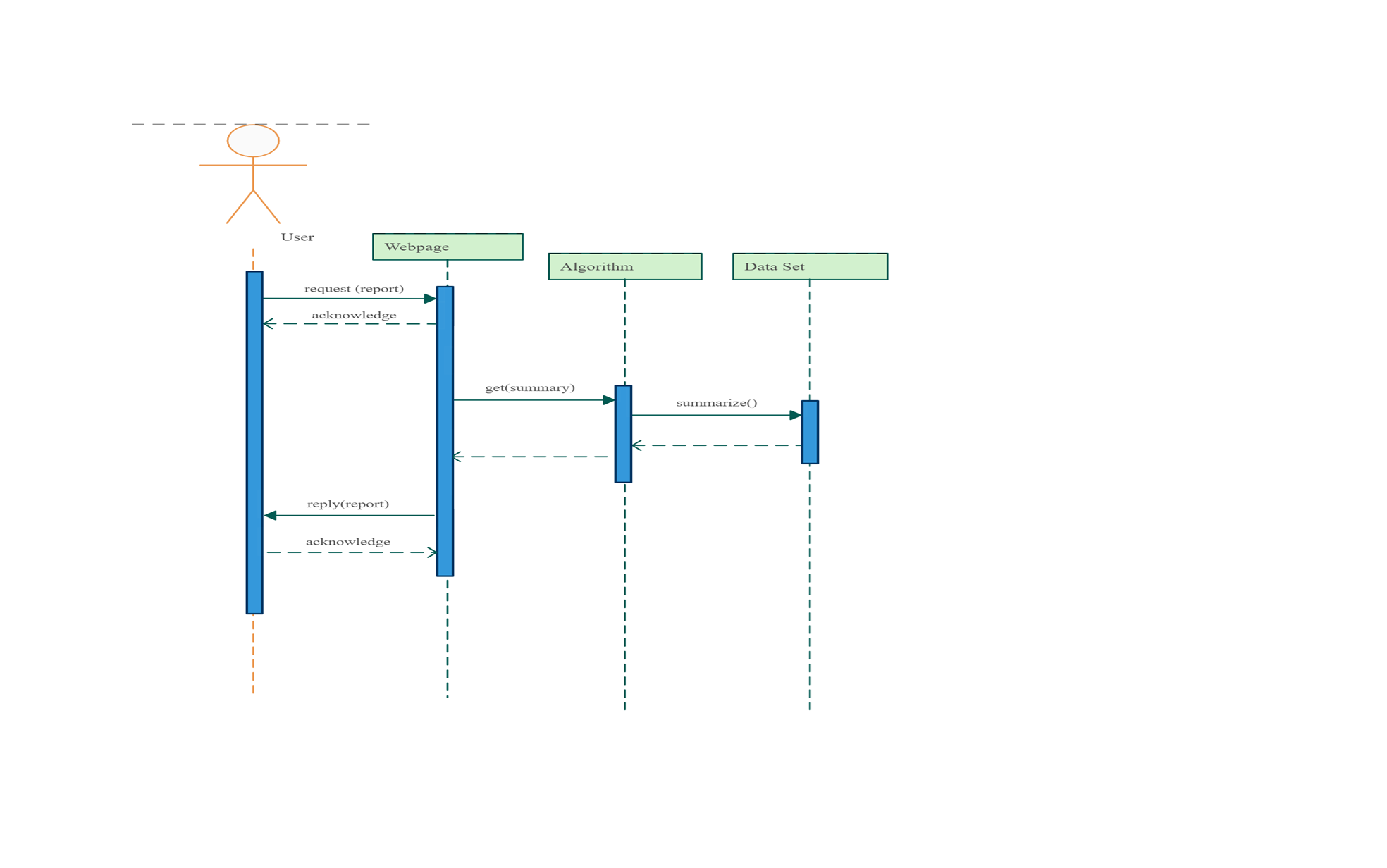
* **CLASS DIAGRAM**



* **ACTIVITY DIAGRAM**



* **SEQUENTIAL DIAGRAM**



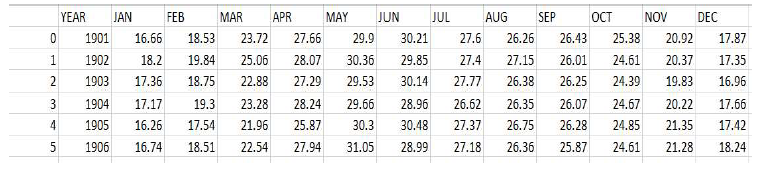
**DATASET DESCRIPTION**

The Dataset is derived from Kaggle. The dataset contains the monthly mean temperature of India (Jaipur) from 1901 to 2021 in CSV format, It contains only numerical input values with no missing or undefined values.

The dataset has a total of 1680 values in 120 rows and 14 columns.

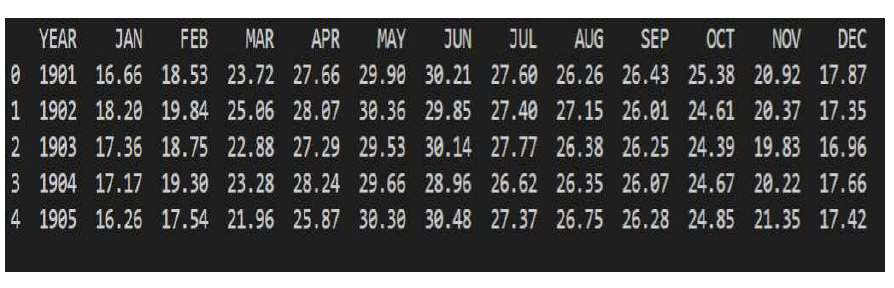
**SCREENSHOT OF THE DATASET**

CSV Format of Dataset

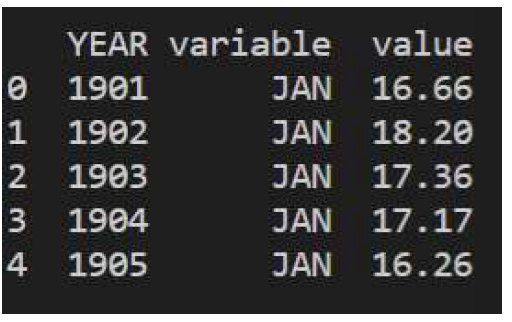


**DATA PRE-PROCESSING**

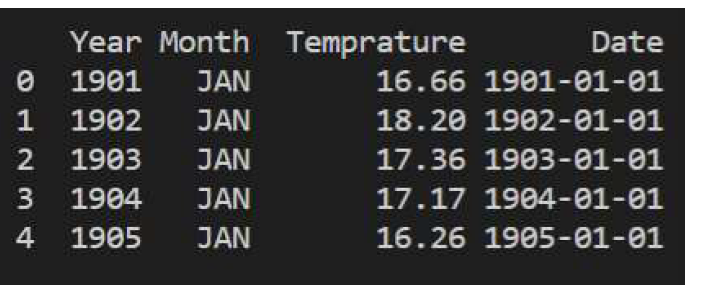
The dataset we have taken has an unnamed column that is from the data as it is of no use. So, the column must be removed from the dataset. So, we define the index column as zero.



As we must improve the performance of our analysis the data can be melted. To do that the YEAR column is made an identifier column and the other two columns are melted into two non-identifier columns namely variable and value.



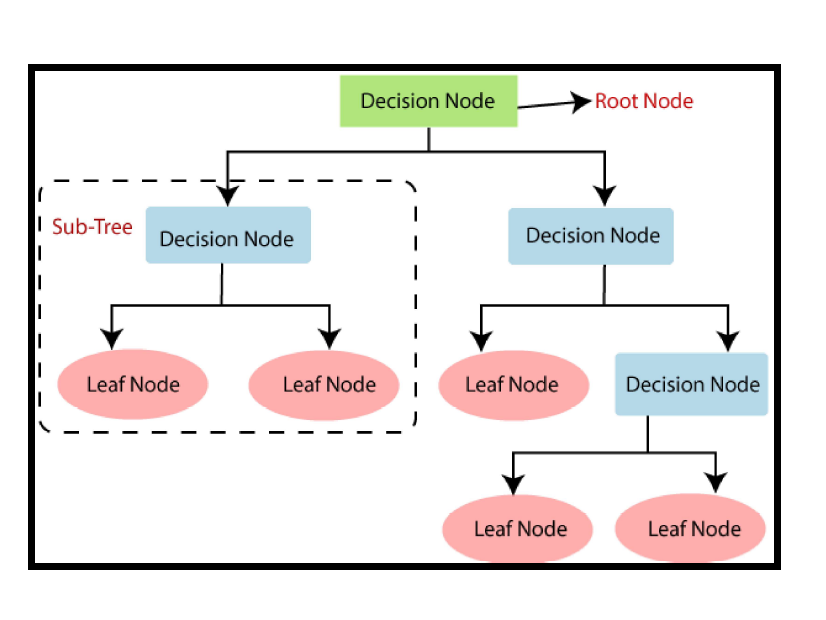
Then a new column named ‘Date’ is added to the dataframe as an attribute for analysis. Using the Date Time library for time series analysis lambda function and loc[] method are used and strptime() is used to format and return a string representation of the date and time.



**IMPLEMENTATION OF K-MEANS CLUSTER ON THE TEMPERATURE DATASET**

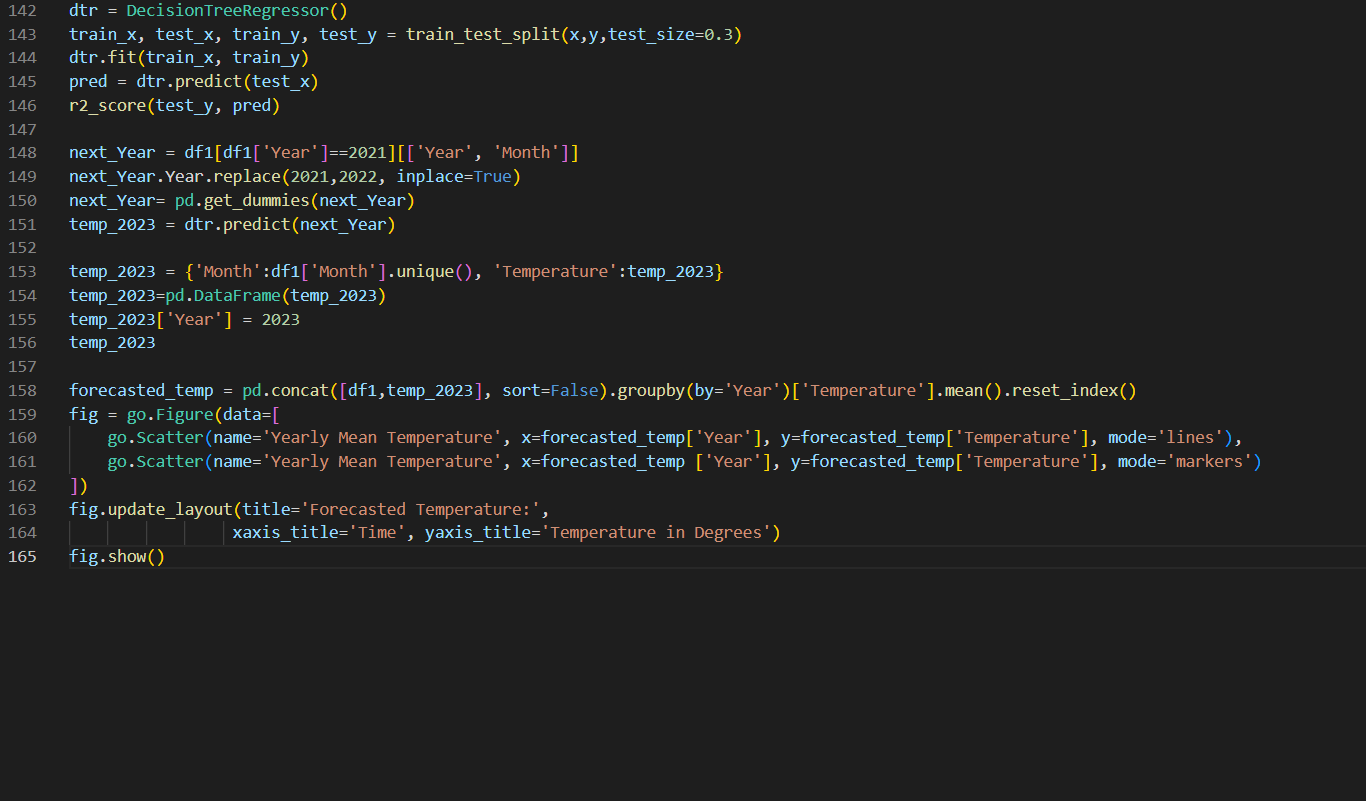
K-Means clustering is used to classify the months into three clusters based on the temperature. The Three clusters are coldest, hottest, and neither too hot nor too cold.

Before proceeding with K-Means Clustering we need to select the better K value, for which we will use Decision Tree with Algorithms such as ID3, CART, etc.

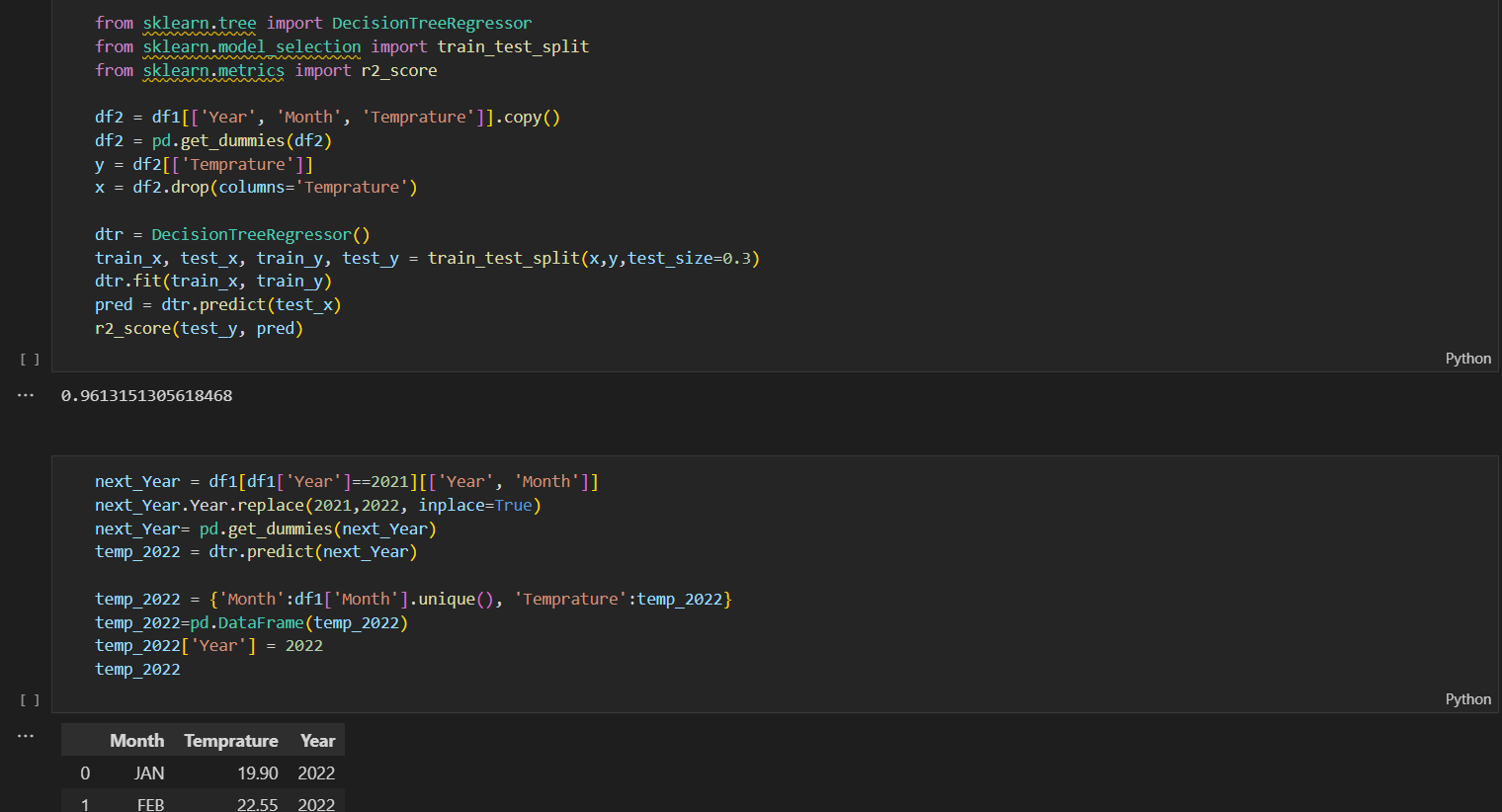
We will further prune our Decision Tree to optimize the accuracy of the Dataset

**CODING**

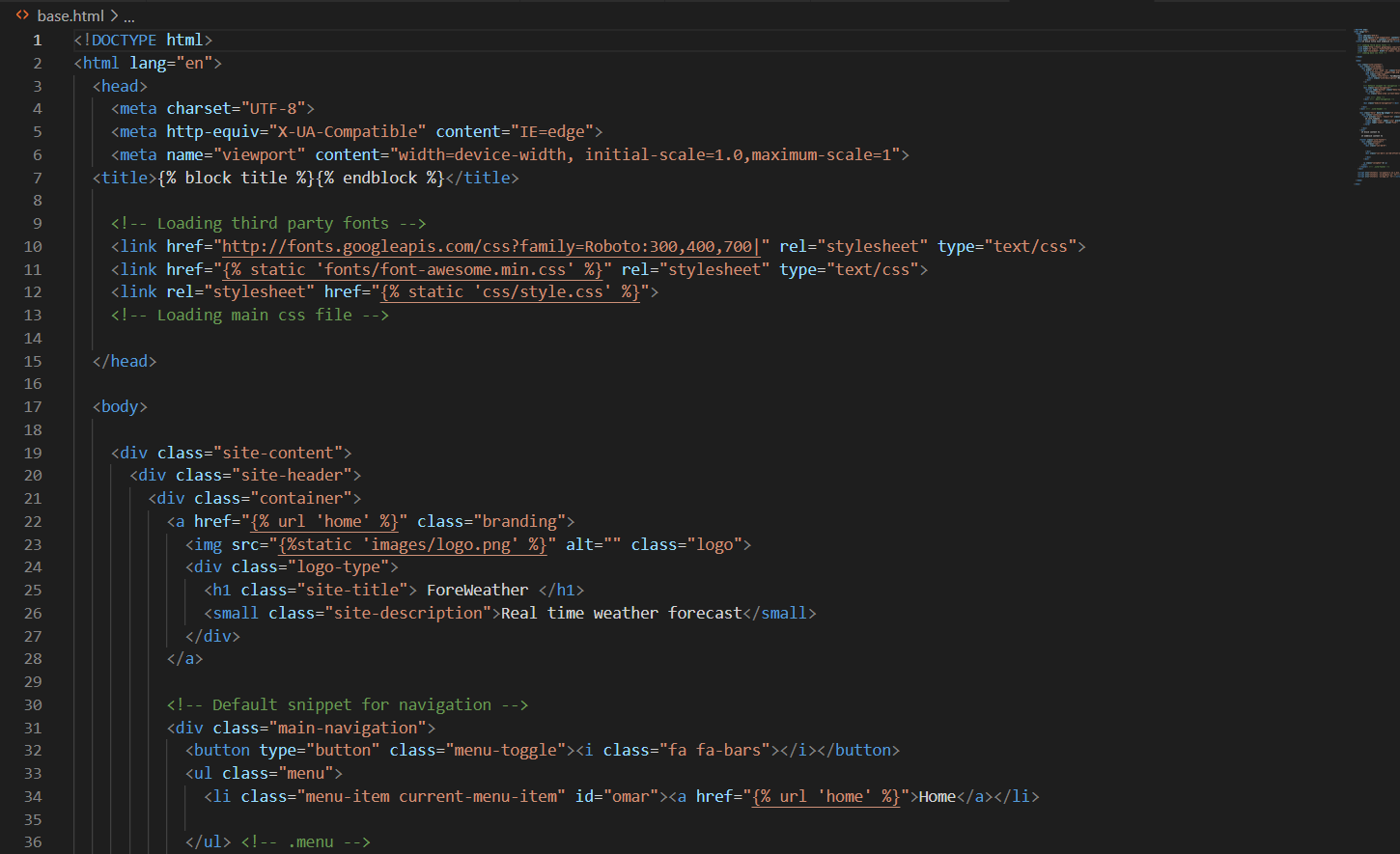
1. **BACKEND-Decision Tree Prediction Module**

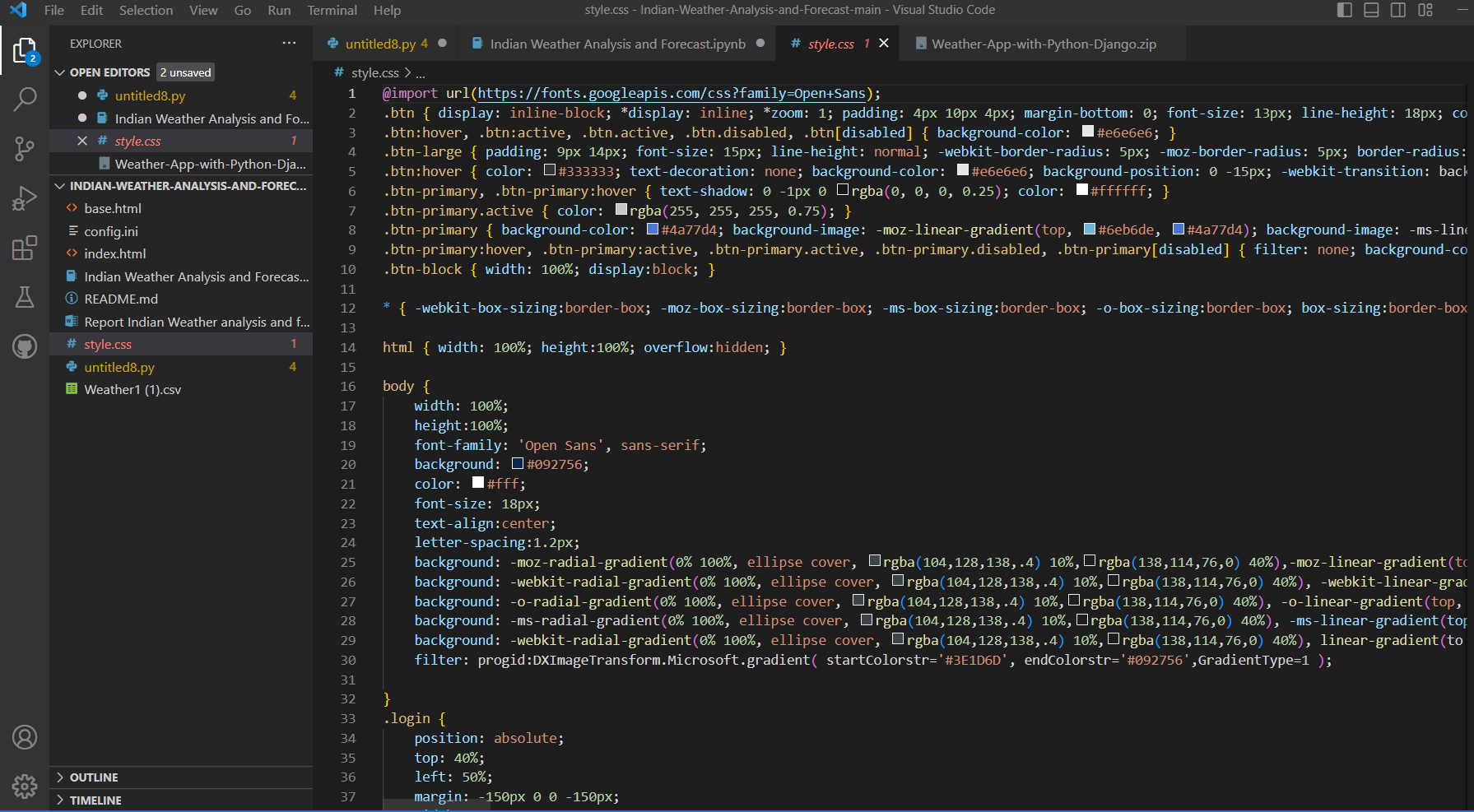
****

1. **Accuracy of prediction- r2 score module**

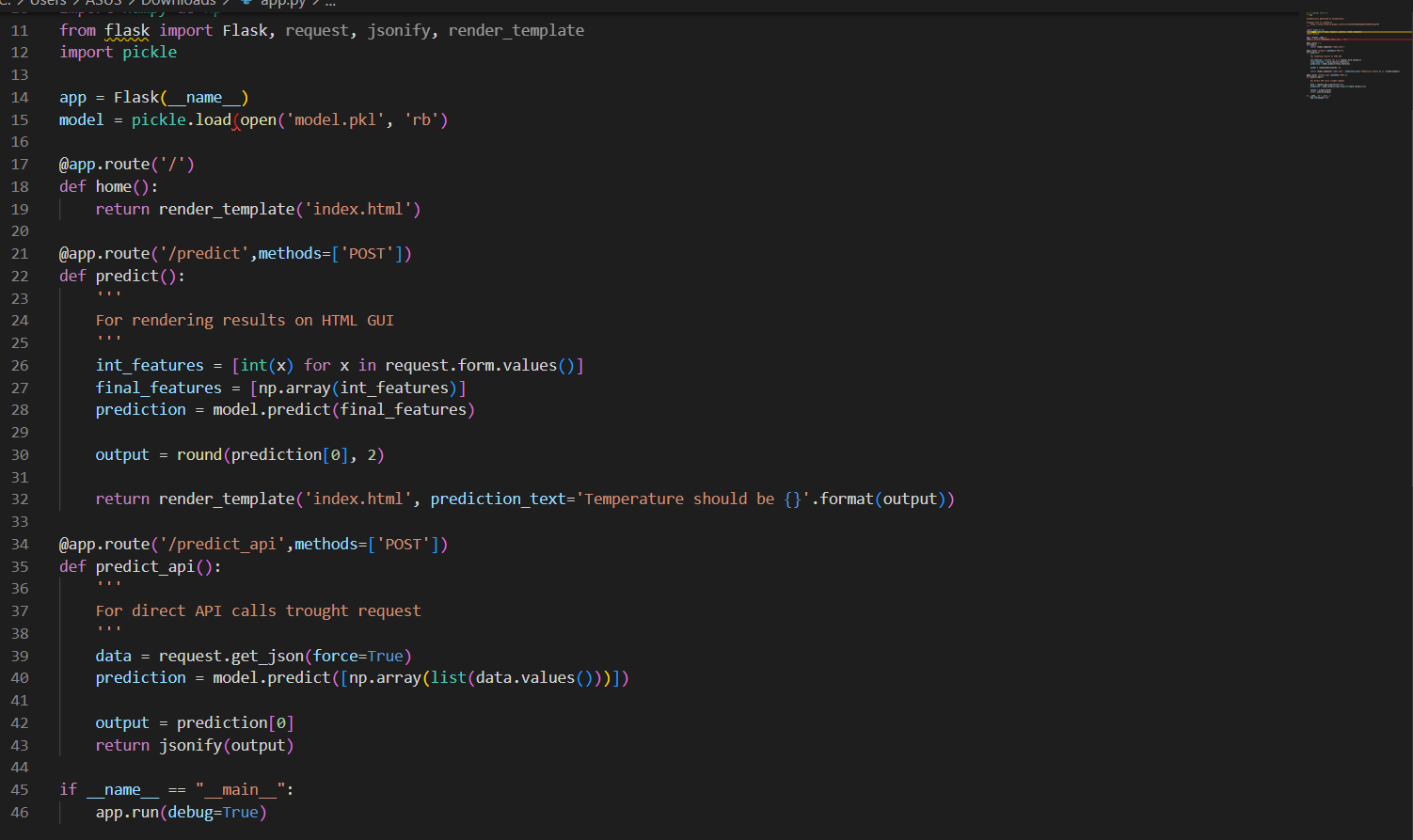
****

1. **FRONTEND- HTML code for web layout**

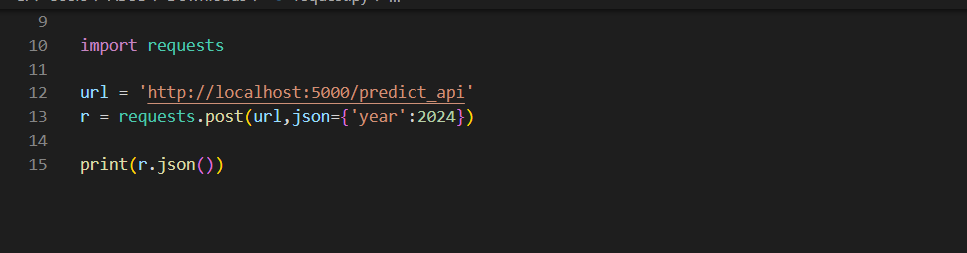


****

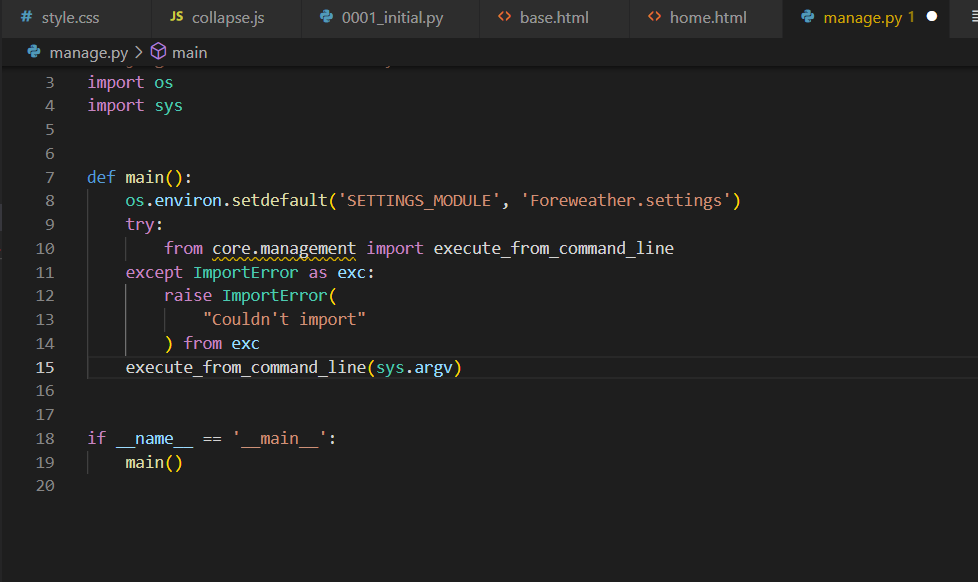
1. **Flask integration code for deployment**

****

1. **Post requests:**

****

1. **Main Module:**

****

**TESTING**

We test our application for its accuracy by predicting the weather of our locality i.e. Tonk. The application generates the weather for the upcoming five days for the city of Tonk.

Then we verify the predictions using Accuweather API, which turns out to be almost exactly the same as our application.

Our r2\_score has also come out to be 0.9566170306063694

Similarly, the application has been tested by predicting the weather for several small cities, as there are not much accurate data available for such cities. This has been repeated several times and any required changes have been made to increase the accuracy further.

**UNIT TESTING**

Unit testing focuses verification efforts on the smaller unit of software design. Using the detailed design description as a guide, important control paths are tested to uncover errors within the boundary of the module. The relative complexity of the test and the error detected as a result is limited by the constraint scope established for unit testing. The unit test is always white box oriented, and the step can be conducted in parallel for multiple modules

* Tested individual Python files by debugging and using print statement
* Individual Component rendering

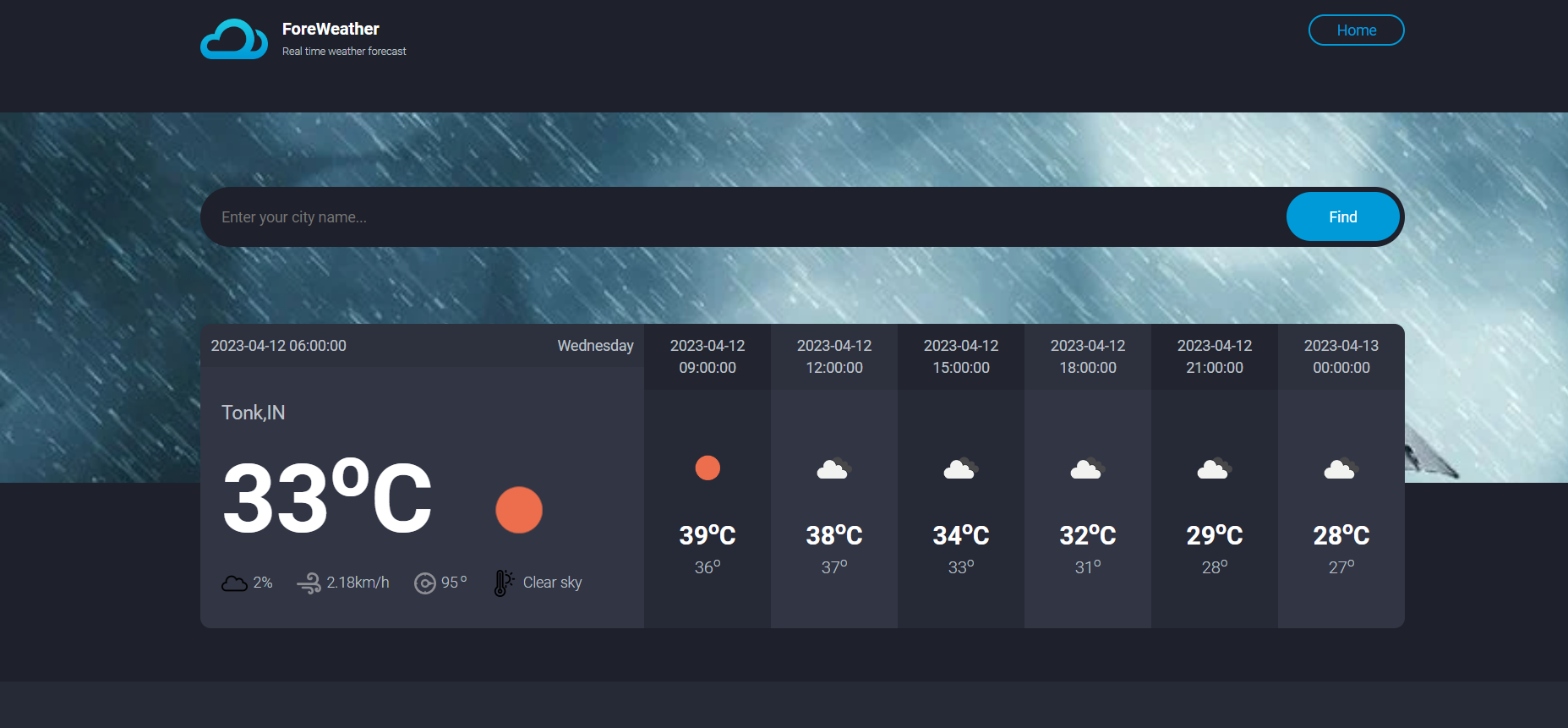
**USER INTERFACE**

The project will be available to the users in a website form. On the home page, the users would be asked to enter the name of the city they want to know the weather of. The users would enter the name of their desired city and would click on the confirm button. The website would then be redirected to another web page which would show the users the predicted weather of the city for the upcoming five days, in a compact, visually pleasing format. The web pages would be designed with the addition of graphics to make them look more attractive to the users. The web pages would be designed using HTML.

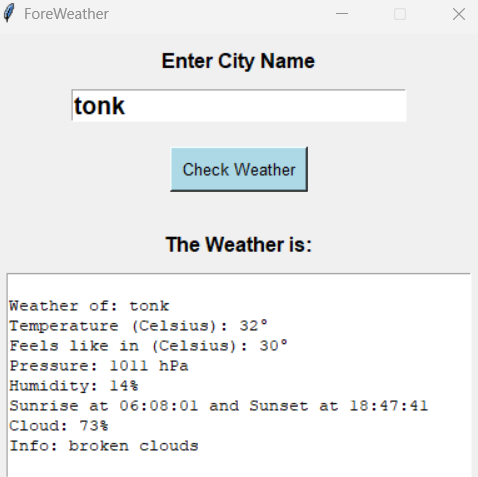
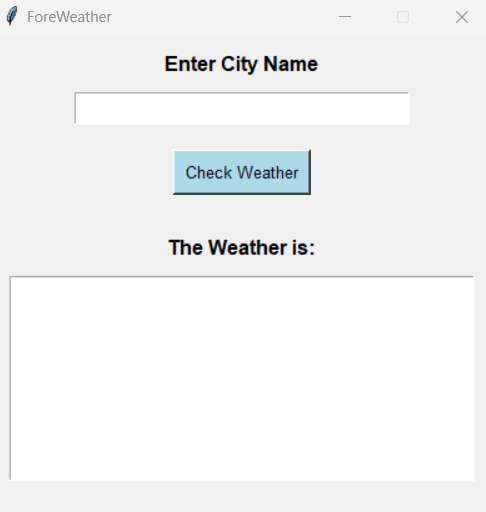
Some of its components are:

1. **Home Page**: users will access our home page where a navigation search bar will be found. On clicking this, the user can enter the name of their city, to get access to its weather at the current time, and predictions for the next 5 days.
2. **Search Bar**: Here, the user would be required to enter a valid name of the desired city to get its forecast.

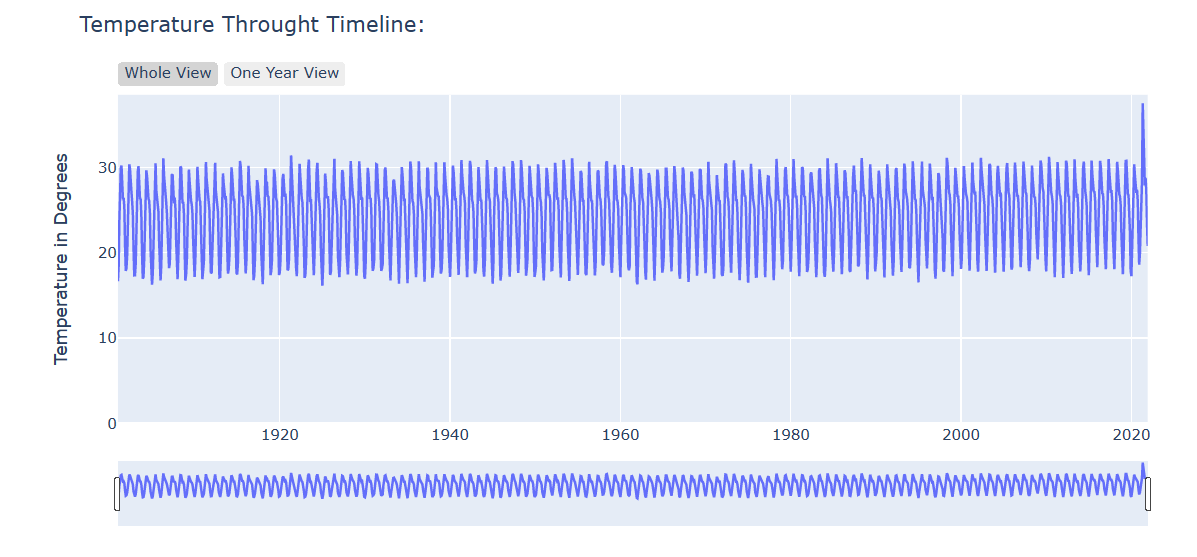
**LANDING PAGE**



**TKINTER INTERFACE**

****

**RESULTS AND CONCLUSION**



**Inferences from the above plot:**

* May 1921 has been the hottest month in India in history.
* Dec, Jan, and Feb are the coldest months. We can group them

together as “Winter”

* Apr, May, Jun, July, and Aug are the hottest months. We could

group them together as “Summer”

* But since this is not how seasons work. We have four main

seasons in India, and this is how they are grouped:

Winter: December, January, and February.

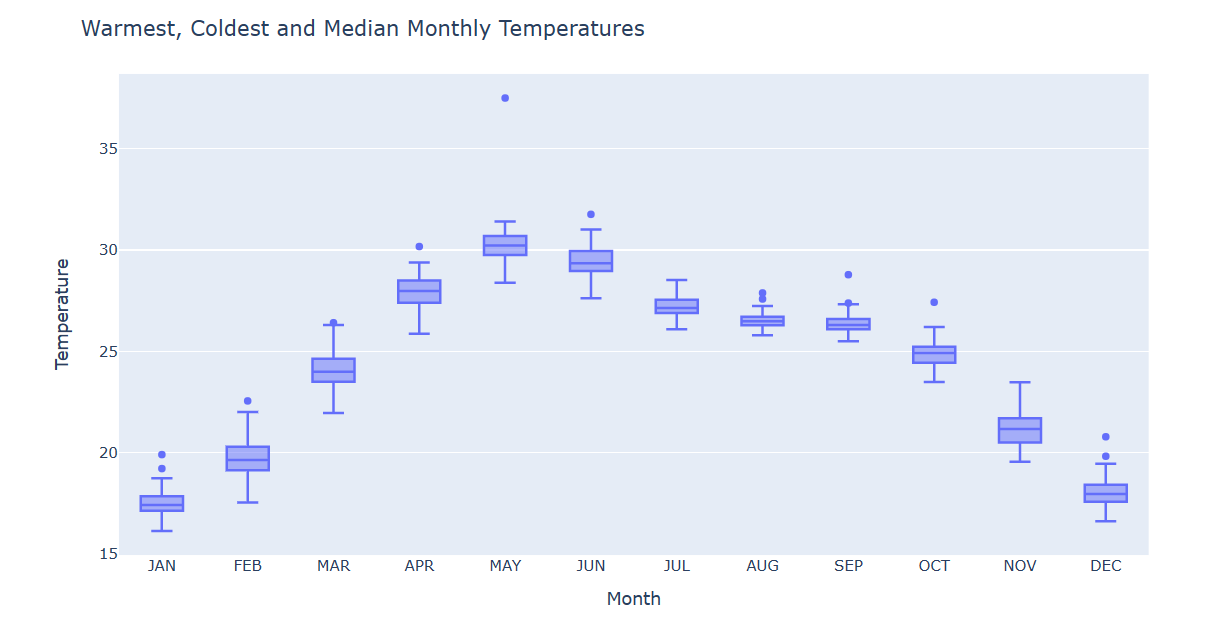
Summer(Also called, “Pre-Monsoon Season”): March, April,

and May.

Monsoon: June, July, August and September.

Autumn(Also called “Post Monsoon Season”): October and

November.

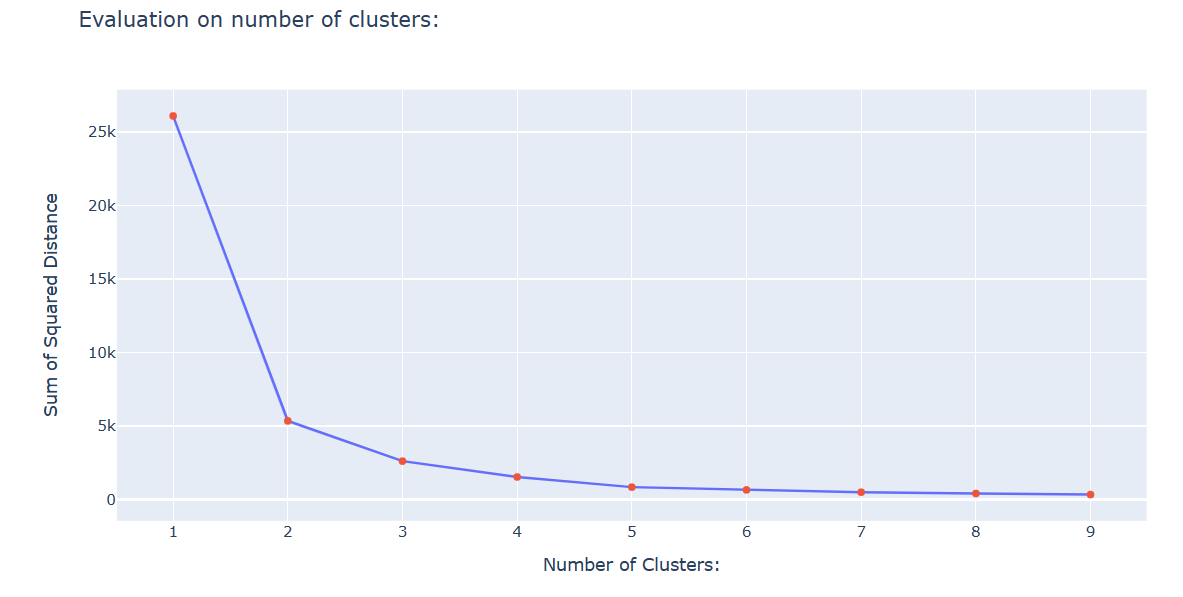


**Inferences from the above box plot:**

* January has the coldest Days in a Year.
* May has the hottest days of the Year.
* July is the month with the least Standard Deviation which means,

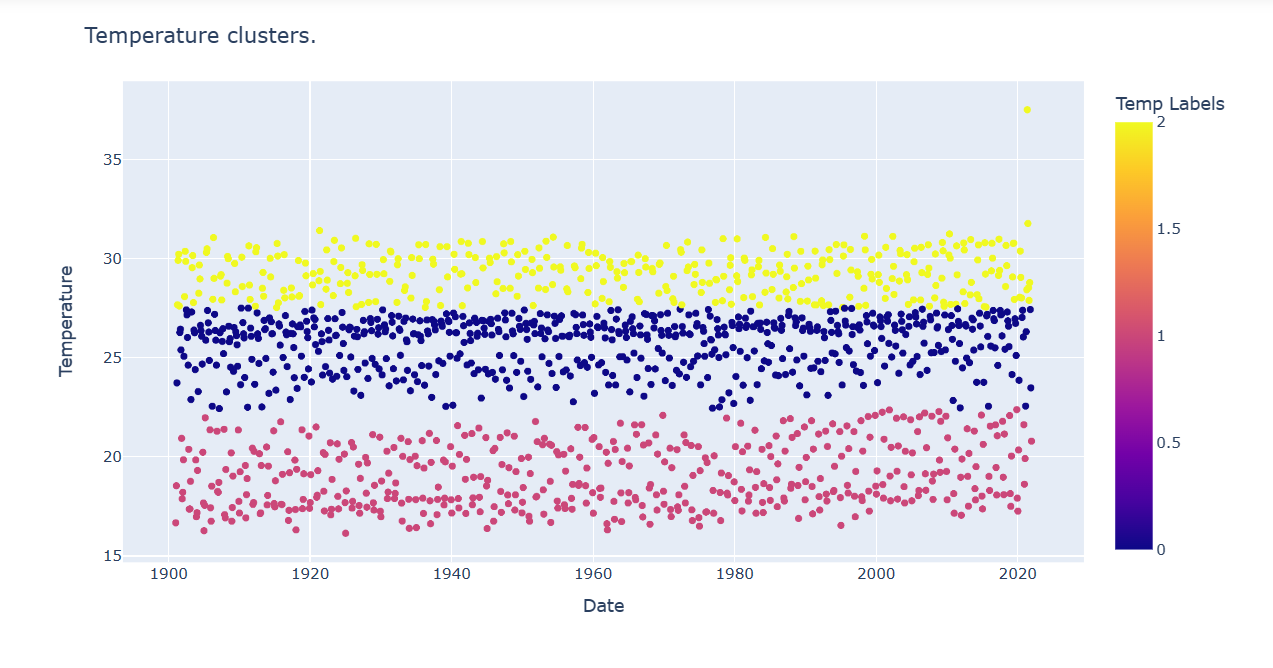
temperatures in July vary the least. We can expect any day in July to

be a warm day.



**Inferences from the above plot:**

* Sum of Squared Errors are calculated.
* Using the Elbow Criterion, the optimized K value is chosen.
* Here cluster size of 3 seems to be a good choice.

****

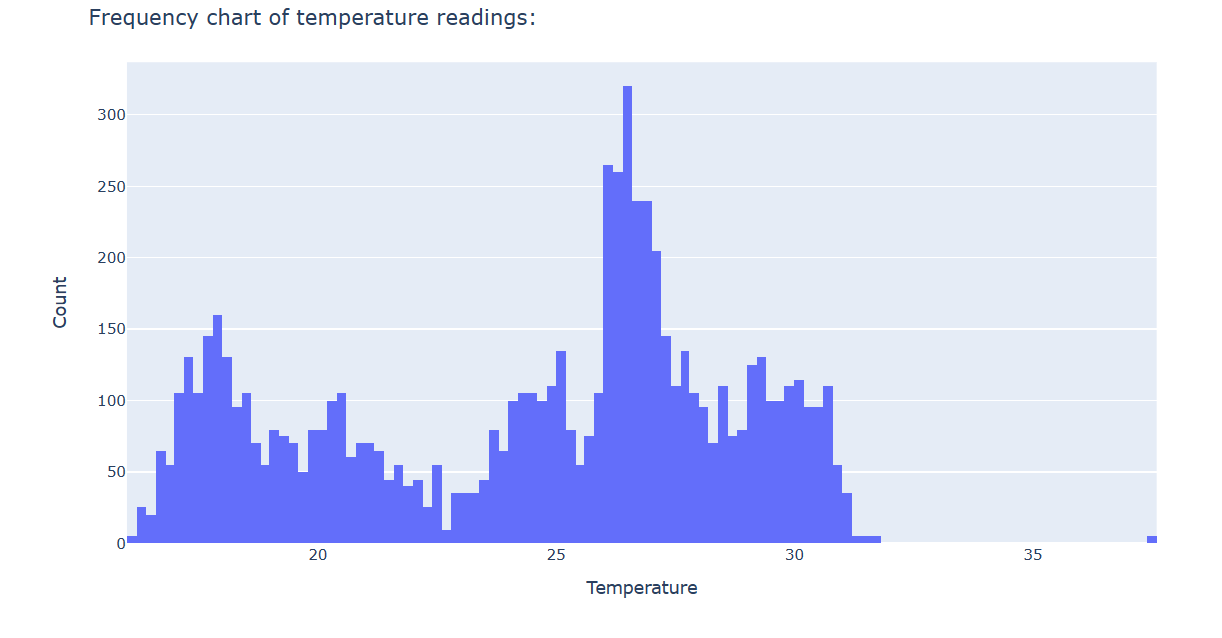
**Inferences from the above scatter plot:**

* Despite having 4 seasons we can see 3 main clusters based on

temperatures.

* Jan, Feb, and Dec are the coldest months.
* Apr, May, Jun, Jul, Aug, and Sep; all have hotter temperatures.
* Mar, Oct, and Nov are the months that have temperatures neither

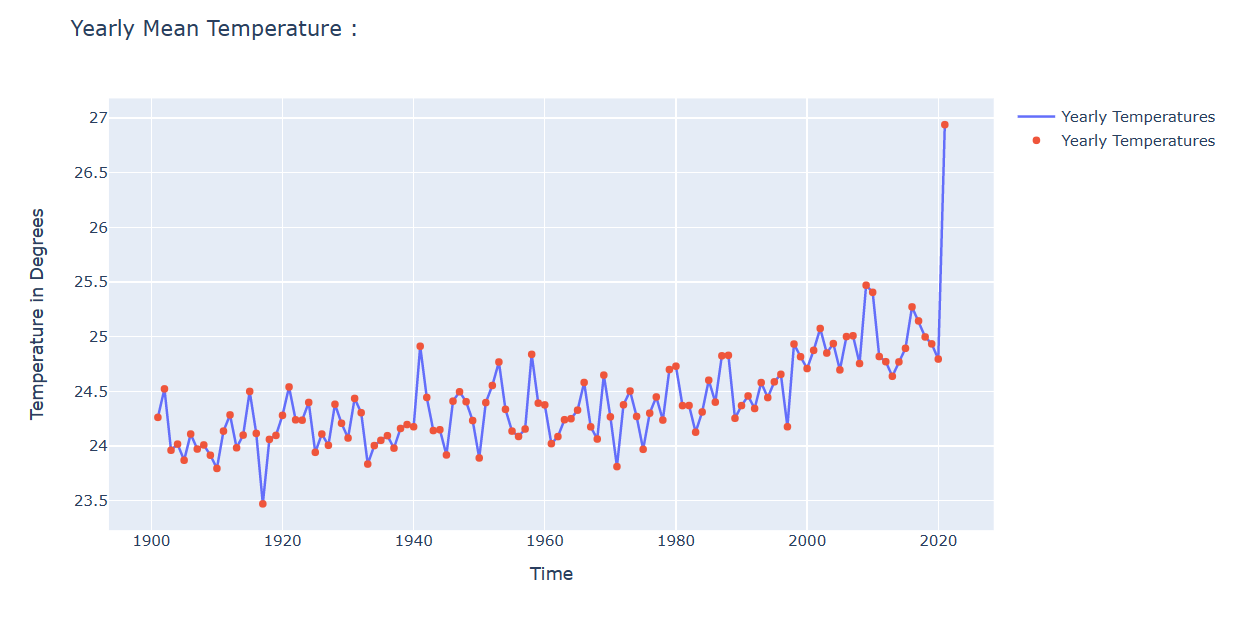
too hot nor too cold.

****

**Inferences from the above histogram plot:**

* The mean temperature for most months during history has been

between 26.8-26.9.

****

**Inferences from the above plot:**

* We can see that the issue of global warming is true.
* The yearly mean temperature was not increasing till 1980. It was

only after 1979 that we can see the gradual increase in yearly mean

temperature.

* After 2015, the yearly temperature has increased drastically.
* But, there are some problems in this figure.
* We are seeing a monthly up-down pattern in yearly

temperatures as well.

* This is not understandable. Because with months, we have a

phenomenon of seasons and the earth revolving around the sun in

an elliptic path. But this pattern is not expected in yearly

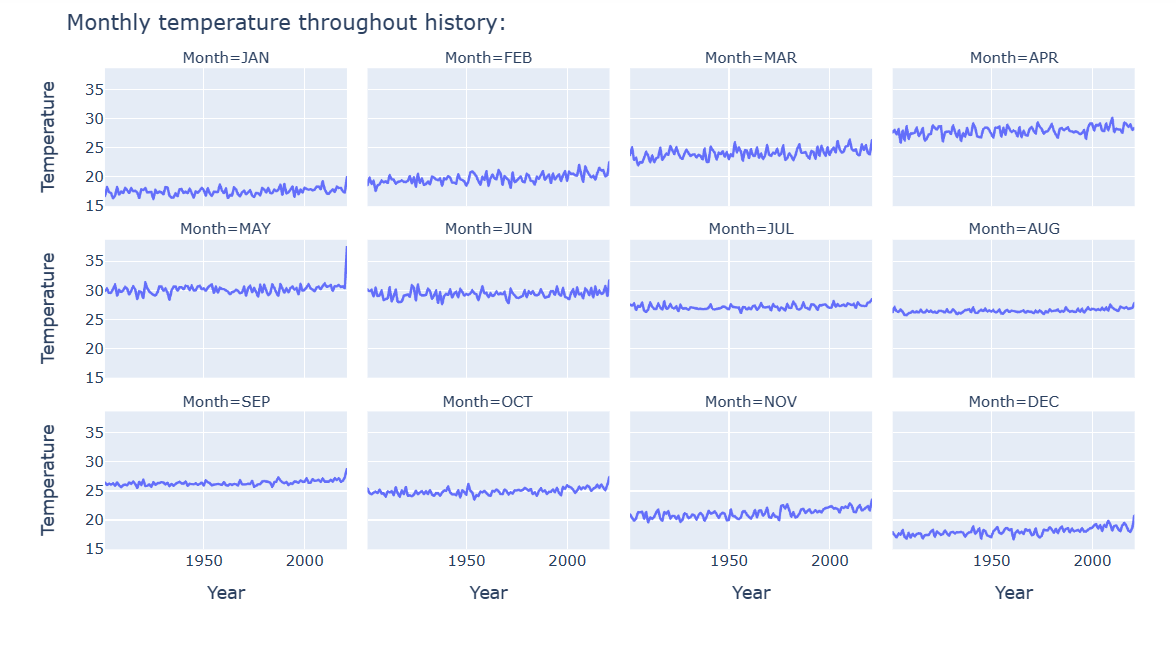
temperature.

* Even from the plot we can infer that the sample points are not

linearly separable.

* Hence decision tree regression algorithm is implemented in our

project.

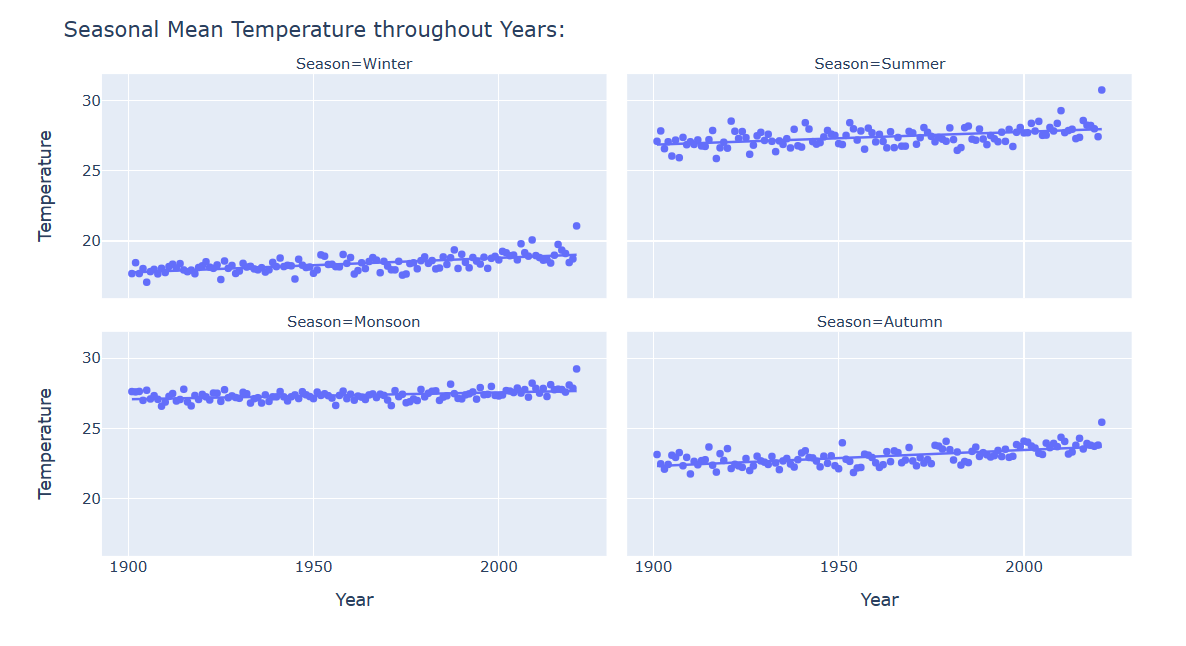
****

**Inferences from the above plot:**

* From this plot we can predict the temperature range for a month

in a year.

* We can also see clearly positive trend lines.

****

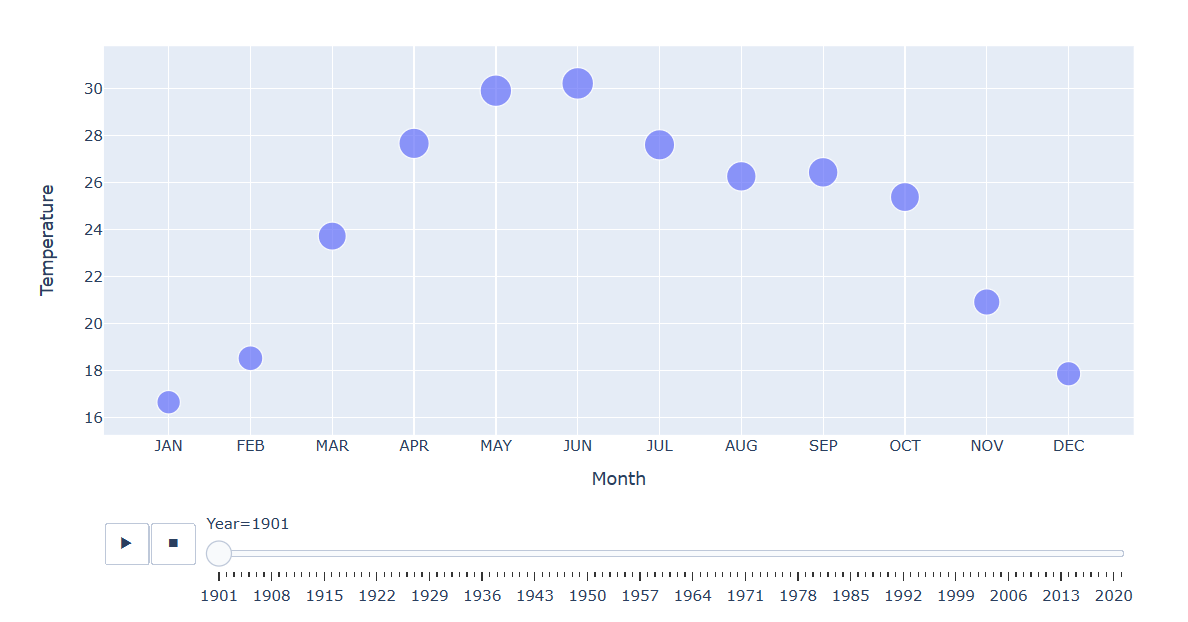
**Inferences from the above plot:**

* From the above visualization, we get an idea about the seasonal

mean temperature.

* The trendline does not have a very high positive correlation with

years, still it is not negligible.

****

**Inferences from the above animation:**

* At first, we can see some fluctuations but it doesn’t give us much of

insight for us.

* If we analyze by adjusting the bar below to early years and late
* years we can notice the change.

**CONCLUSION**

* Data cleaning and data reshaping have been done.
* K-means clustering algorithm has been used to classify the data

samples and the decision tree regression algorithm has been used to

forecast the weather.

* Decision trees prove as an effective method of decision-making

in Weather prediction. As decision trees are ideal for multiple

variable analyses, it is particularly important in the current problem-solving task, like weather forecasting.

* The dataset has been classified according to the mean

temperature of every month in each year and the weather

prediction has been done.

* Professional weather forecasters are not perfect, but their

predictions are typically more accurate than those of this

machine learning model. This implies that weather is a non-linear system. Additionally, our predictions were all based on

data from one location as opposed to the multiple locations that most

forecasters use. Even in our model, we have described the

limitation of using linear regression in predicting the weather.

* This mini project is helpful in providing information to people

and organizations in order to reduce weather-related losses and

enhance societal benefits, including protection of life and

property, public health and safety, and support of economic

prosperity and quality of life.

**APPENDICES**

**DEFINITIONS, ACRONYMS, ABBREVIATIONS**

1. Machine Learning (ML): Machine learning is a branch of artificial intelligence (AI) and computer science that focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.
2. Decision Trees: According to the IEEE Standard Definitions, a decision tree is a tree whose internal nodes can be taken as tests (on input data patterns) and whose leaf nodes can be taken as categories (of these patterns). These tests are filtered down through the tree to get the right output to the input pattern.
3. K-Means Cluster Algorithm: is a method of vector quantization, originally from signal processing, that aims to partition observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.
4. User interface (UI): The user interface (UI) is the point of human-computer interaction and communication in a device. This can include display screens, keyboards, a mouse, and the appearance of a desktop. It is also the way through which a user interacts with an application or a website.
5. API: An application programming interface is a way for two or more computer programs to communicate with each other. It is a type of software interface, that offers a service to other pieces of software. A document or standard that describes how to build or use such a connection or interface is called an API specification. A computer system that meets this standard is said to implement or expose an API. The term API may refer either to the specification or to the implementation.
6. Plotly: The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographical, scientific, and 3-dimensional use-cases.

**REFERENCES**

1. Mark Holmstrom, Dylan Liu, Christopher Vo, “Machine Learning Applied to Weather Forecasting, Stanford University, 2016
2. Piyush Kapoor and Sarabjeet Singh Bedi “Weather Forecasting Using Sliding Window Algorithm”, Kvantum Inc., Gurgaon 122001, India MJP Rohilkhand University, Bareilly 243006, India,2013.
3. Divya Chauhan, Jawahar Thakur “ Data Mining Techniques for Weather Prediction: A Review Shimla 5, India: ISSN, 2013.
4. Siddharth S. Bhatkandel, Roopa G. Hubballi2 “Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques”, Belgaum India: International Journal of Advanced Research in Computer and Communication Engineering, 2016.