

| Project Cover Sheet |
| --- |

| **TO BE FILLED BY THE STUDENT** | |
| --- | --- |
| Student Name | *Kelvin Charles Meena* |
| Email ID | *kvoode@gmail.com* |
| Date Submitted | *10-12-2023* |

**ASSESSMENT FEEDBACK**

| **TO BE FILLED BY THE ASSESSOR** | | |
| --- | --- | --- |
| Assessment type | Marks | Marks Awarded |
| Project Documentation | 50 |  |
| Program File | 50 |  |
| Overall Marks achieved |  | |
| GRADE ACHIEVED |  | |
| **Summative Feedback by Assessor for further improvement** | | |
|  | | |
| **Comments for REDO submission (If applicable)** | | |
|  | | |

**GRADE DESCRIPTORS**

| 70% and above  (Distinction) | The Project evaluated is of a high to exemplary standard. The work addresses clearly and articulately the project requirements and thus meets and satisfies all the learning outcomes (either well or in an exemplary way). The work demonstrates: clear knowledge; references to appropriate academic literature; analysis; critical evaluation; and originality of argument. It is structured and presented to a high (or exemplary) standard. Referencing conventions are fully observed. |
| --- | --- |
| 60 to 69%  (Merit) | The project evaluated is of a good to a high standard. Substantial knowledge, comprehension and analysis is evident throughout. Arguments presented are clear and focussed with a logical structure in place. There is clear evidence of critical evaluation of a wide range of theories/perspectives from academic literature and some independent thought. The work is well-written and addresses well all of the learning outcomes. Referencing conventions are fully observed. |
| 50 to 59%  (Pass) | The project evaluated is of a fair to good standard. Adequate knowledge, comprehension and analysis is evident throughout. The arguments presented have a logical structure and show some critical evaluation in places, although there may be limited evidence of an independent perspective. There is evidence of some good engagement with some of the appropriate literature. Learning outcomes have been largely met and to an appropriate degree. Referencing conventions are observed. |
| 40 to 49%  (Fail/Redo) | The project evaluated is of a basic standard. The arguments presented have some logical structure and are supported by academic literature in most cases. The academic literature used is outside of the suggestions made in the module guide but remains limited. Little critical evaluation is evident, and the work tends more widely towards a descriptive style. Learning outcomes have been addressed in a basic but satisfactory way. Referencing conventions are mostly observed. |
| Fail Grades | |
| 30 to 39%  (Module retake) | The project evaluated is of a limited standard. Limited use of academic literature and as such knowledge and argument is very weak. A simple descriptive style with no evidence of critical evaluation throughout. Over-reliance on simplistic, limited sources. Referencing conventions may not be observed. Some learning outcomes met but in a weak and simplistic way. The work is needs to be developed in greater depth and detail to move to a passable standard at this level of study |
| 29% and Below  (Module retake) | The project evaluated is of an unacceptable standard. There is little or no evidence of knowledge and understanding that is required at this level. Referencing is inadequate or non-existent. The learning outcomes have not been addressed fully and the work requires significant modification to bring it to a passable standard. |

**PREDICTING TOMORROW’S TEMPERATURE**

**Abstract**

Weather’s change affect our day-to-day activities performance according to its state, Example if it is a rain day it’ll impact some activities either in positive way or negative way according to the nature of the activity

So knowing or having upcoming weather information helps us to plan their activities accordingly by considering impacts and sets mitigation ways to against impacts for the aim of improving or maintaining performance.

This project aims to develop a predictive model for forecasting tomorrow's temperature based on historical weather data. Leveraging machine learning algorithms and statistical analysis, the model seeks to provide accurate and reliable predictions to enhance short-term weather forecasting. The integration of advanced data analytics techniques will contribute to improved decision-making processes for various industries and individuals dependent on precise weather information

This project utilizes machine learning algorithms, to build predictive models. Data preprocessing techniques are applied to handle missing values, outliers, and feature engineering to extract relevant information from the available data. The dataset includes information such as Maximum Temperature of each day, Average Temperature of each day, Minimum Temperature of each day and Station where measurements taken from.

The results of this analysis contribute to more efficient planning and set decisions of upcoming events or activity, and mitigate potential risks associated with temperature fluctuations. This project serves as a foundation for further research and advancements in predictive analytics for weather forecasting and determining our feature climatic changes.

**Introduction**

Weather prediction has long been a critical aspect of our daily lives, influencing activities ranging from agriculture and transportation to disaster preparedness. This project addresses the need for an efficient and accurate method of predicting tomorrow's temperature by utilizing historical weather data. The integration of predictive modeling into weather forecasting can significantly enhance our ability to plan and adapt to changing climatic conditions.

**Significance and Relevance:**

Accurate temperature predictions play a crucial role in multiple sectors, including agriculture, energy management, and emergency response. By developing a robust model that forecasts tomorrow's temperature, this project aims to provide stakeholders with timely and reliable information, enabling them to make informed decisions and mitigate potential risks associated with temperature fluctuations.

**Objectives:**

Develop a predictive model: Employs various machine learning algorithms, such as regression, time series analysis, or ensemble methods, to develop a predictive model that can forecast weather data accurately. This model will utilize historical weather information.

Evaluate and select the most suitable machine learning algorithm for temperature prediction.: Analyze the predictive model to identify temperature change

Train the model using a comprehensive dataset of historical weather information.

**Scope and Constraints:**

The scope of this project encompasses the development and implementation of a temperature prediction model using historical data. The model's accuracy may be influenced by factors such as the availability and quality of historical data, as well as the complexity of weather patterns. Constraints include potential inaccuracies arising from unpredictable meteorological events and limitations in the scope of historical data.

By addressing these objectives within the defined scope and considering the associated constraints, this project aims to provide actionable insights and solutions to improve weather prediction accuracy and operational efficiency in planning next day routine.

**Methodology**

To implement the online food order prediction project using Python, we can follow the following steps:

**Data Collection**: Gather the historical weather data. This may include meteorological agencies, weather stations, or publicly available datasets, for my case I got data from NOAA ( National Oceanic and Atmospheric Administration).

**Data Preprocessing**:

* Data Cleaning: Address missing or inconsistent data points, ensuring the dataset is of high quality and free from anomalies. Handle missing values, outliers, and perform necessary data transformations. Use Python libraries like Pandas and NumPy for data preprocessing tasks.

**Feature Engineering**: Identify relevant features such as temperature, humidity, wind speed, and atmospheric pressure. Remove irrelevant or redundant data that may not contribute to accurate predictions.

**Splitting the Data**: Split the dataset into training and testing sets. The training set will be used to build and train the predictive model, while the testing set will be used to evaluate its performance. We can use the Ridge function from the Scikit-learn library for this purpose.

**Model Selection:** Choose an appropriate machine learning algorithm for tomorrow’s temperature prediction. Consider algorithms like linear regression, decision trees, random forests, gradient boosting, or neural networks. Select the algorithm based on the characteristics of your data and the problem statement.

**Model Training:** Train the selected machine learning model using the training data. Fit the model to the training set using the fit method provided by the chosen Python library (e.g., scikit-learn). Adjust the model's hyperparameters based on experimentation or cross-validation techniques .

**Training and Validation:**

* **Model Evaluation**:Assess the model's performance using the testing set, employing metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), or R-squared.
* **Validation**:Validate the model's accuracy against real-time data to ensure its reliability for predicting future temperatures.

**Interpretation and Recommendations:**

Throughout the implementation process, leverage Python libraries like Pandas, scikit-learn to streamline data manipulation, model training, evaluation, and prediction tasks. Use appropriate visualization libraries, such as Matplotlib, to visualize the data and model outputs for better interpretation and communication of results.

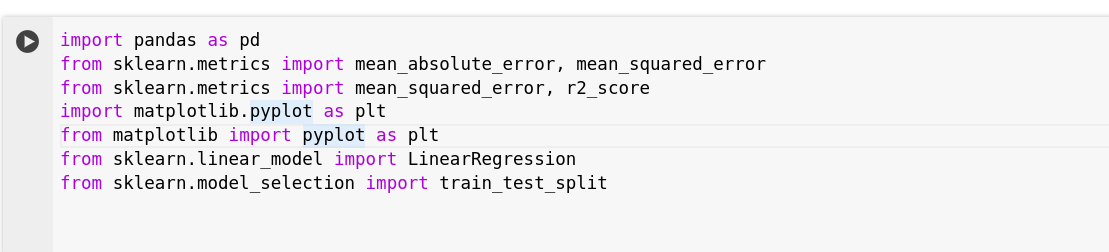
**Project Implementation**

Let’s start the Implementation.Firstly import all the libraries as shown.Here's a brief description of each library:

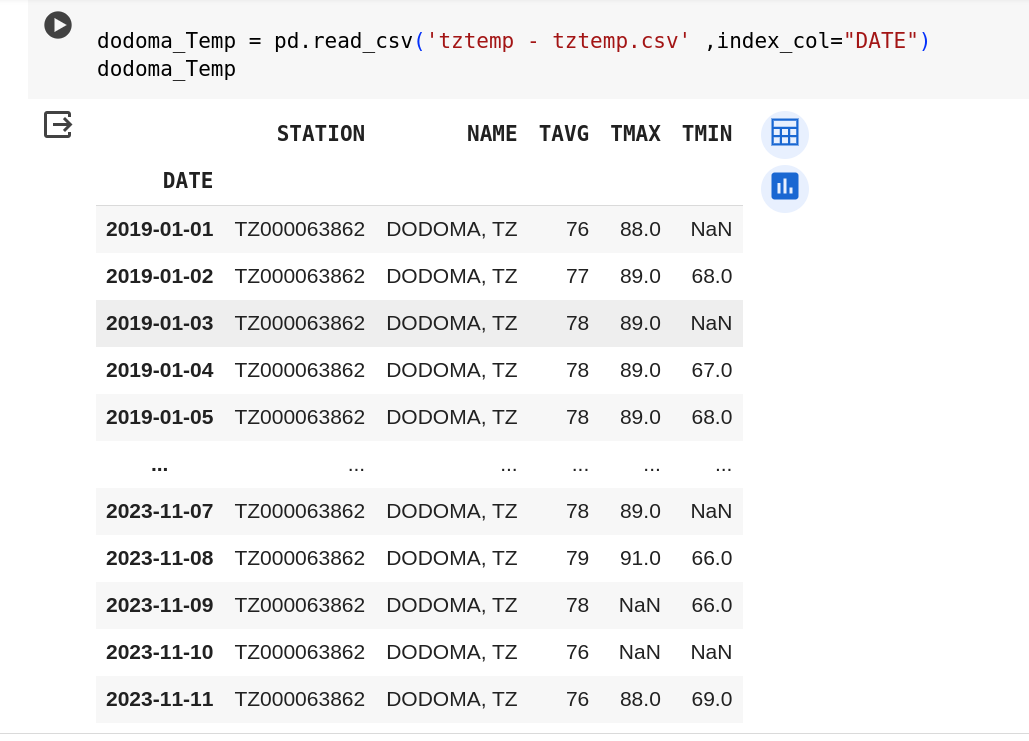
**pandas (imported as pd):** A powerful data manipulation library that provides data structures such as DataFrame for efficient handling and analysis of structured data.

**matplotlib.pyplot (imported as plt):** A widely used plotting library that offers a wide range of static, animated, and interactive visualizations.

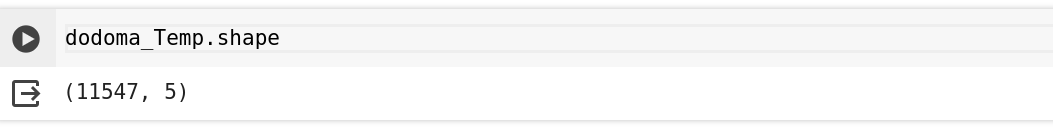
**from sklearn.linear\_model import LinearRegression:** learning the coefficients (weights) that define the linear relationship between input features (like temperature on previous days) and the target variable (e.g., tomorrow's temperature).



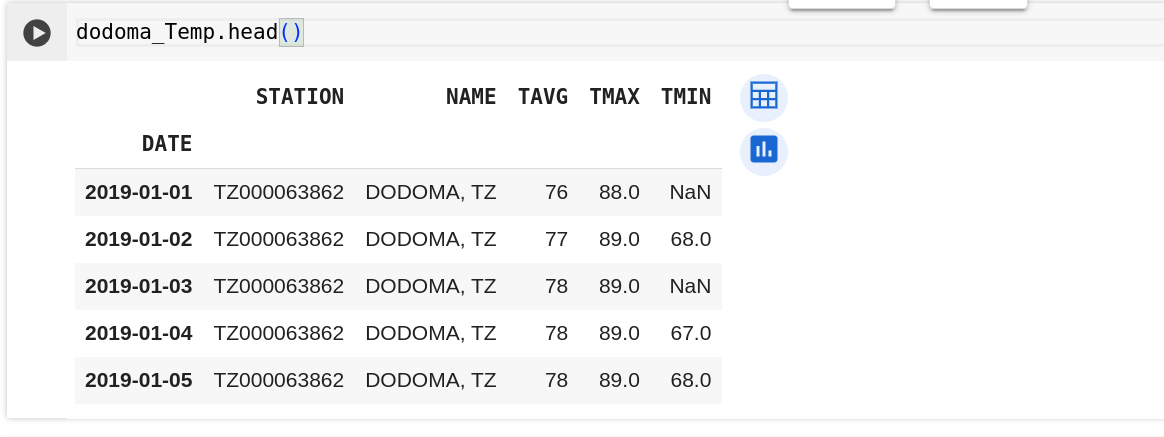
Now reading data from the dataset.



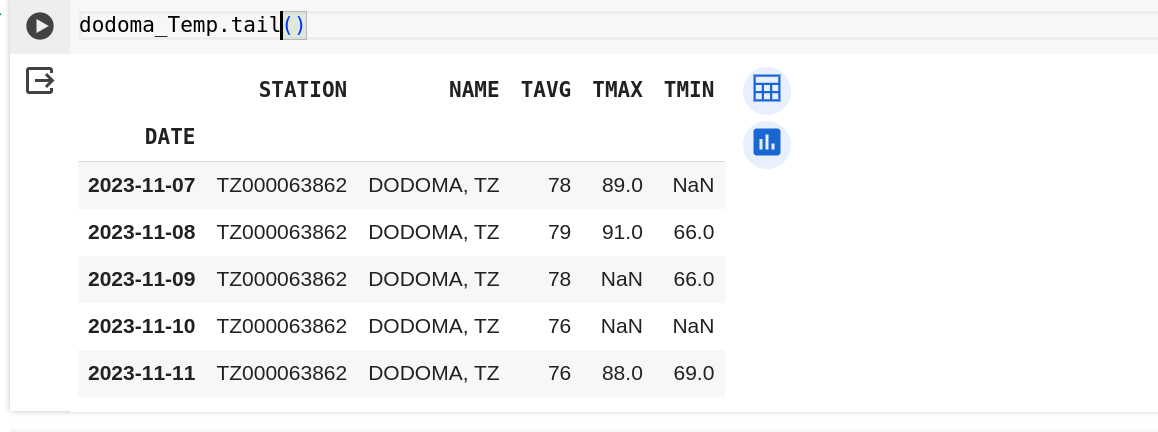
**1.Exploring the Data**

****

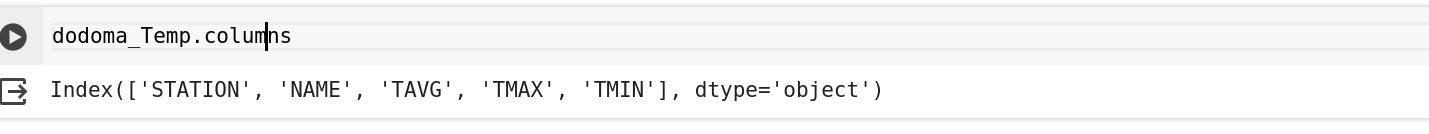
**The data has 11547 rows and 5 columns.**

****

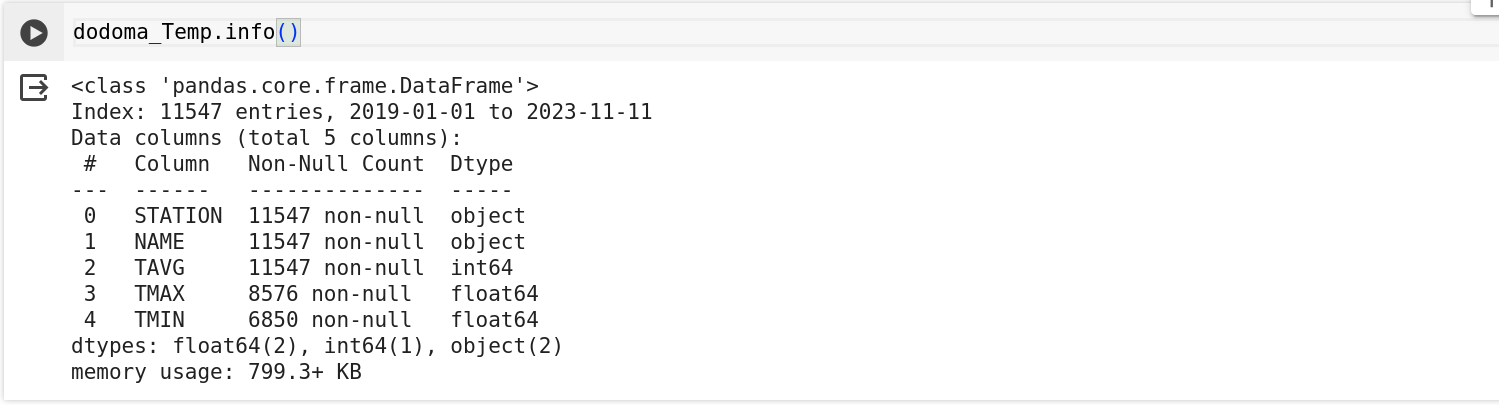
The **data.head()** method returns the first 5 rows of the DataFrame if a number is not specified.

****

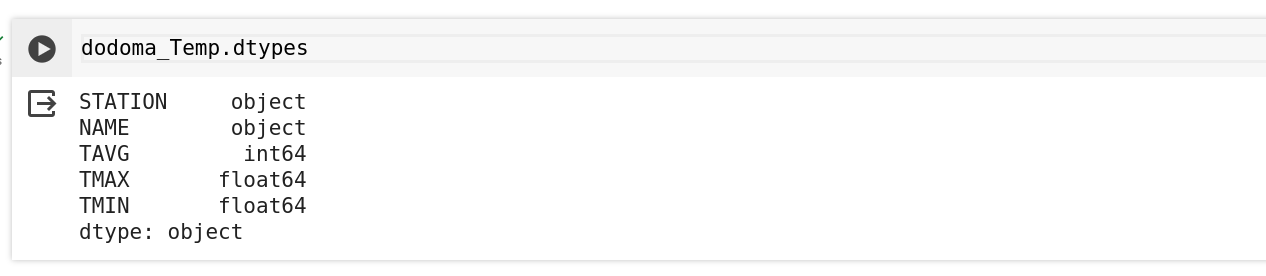
Likewise, **data.tail()** method returns the last 5 rows of the Data Frame if a number is not specified.



The **data. Columns** returns column names of the Data Frame. Here the columns are Station, Region Name (NAME), Average Temperature (TAVG), Maximum Temperature (TMAX),Minimum Temperature(TMIN)



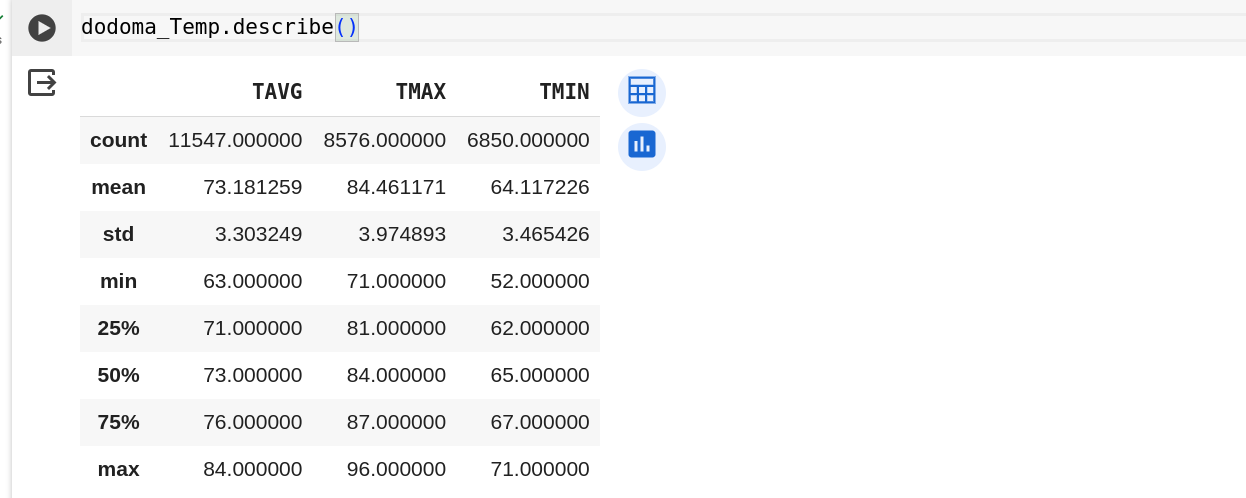
The **data.info()** returns the DataFrame's column names, data types, memory usage, and the number of non-null values in each column.



The **data.dtypes** returns the data types used in the Data Frame. The data types shown in Data Frame are int64, object, float64.

**2.Descriptive Statistics:**

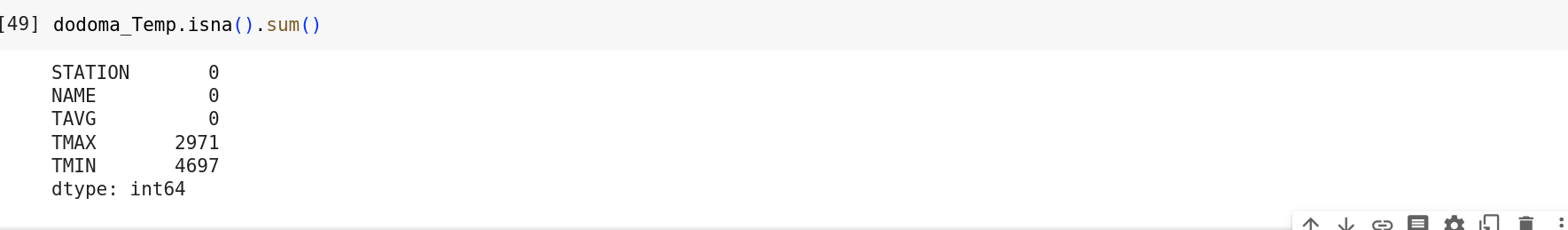
Use **data.describe()** to generate summary statistics of the numeric columns in the Data Frame.

****

It shows the various statistical measures for each numerical column in the Data Frame, including count, mean, standard deviation (std), minimum (min), 25th percentile (25%), median (50%), 75th percentile (75%), and maximum (max). Non-numeric columns are excluded from the result.

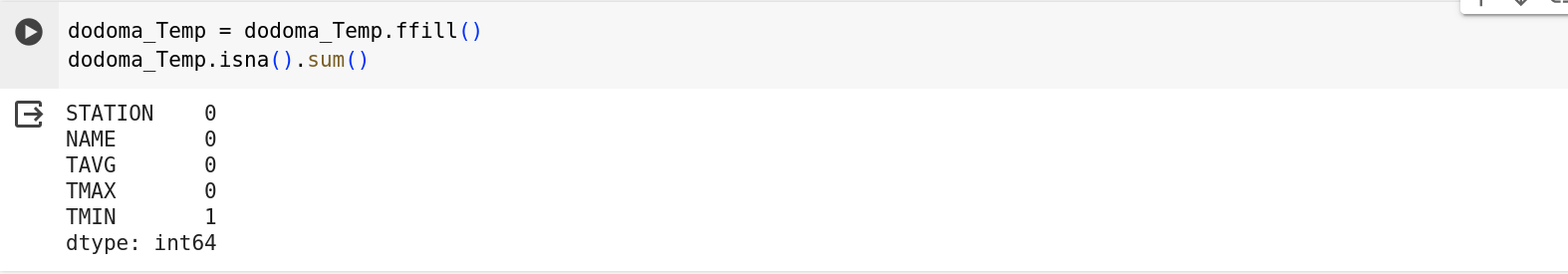
1. **Data Cleaning and Preparation:**

Next step to check any null values in the data set. **data.isna().sum()** is used to count the number of missing or NaN (Not a Number) values in a Data Frame.

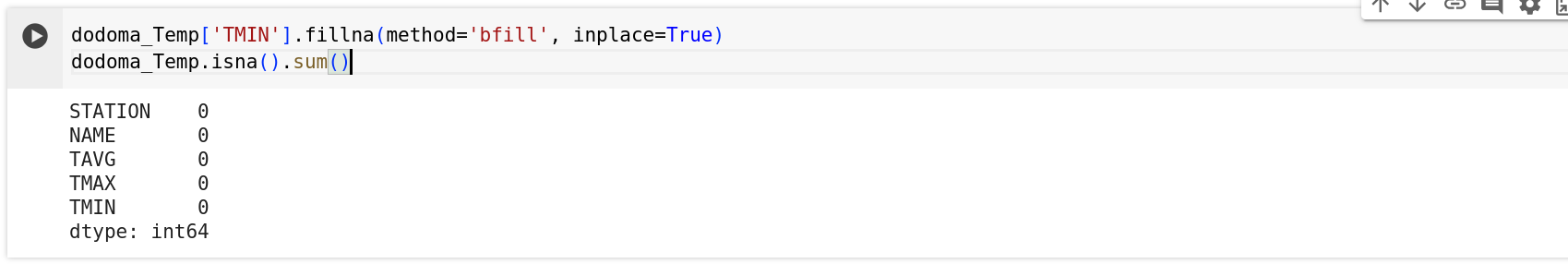


From the above result we came to know that there are null values in the data frame.

Lets filling missing values



We remain with single null value from TMIN column, let's try to fix it

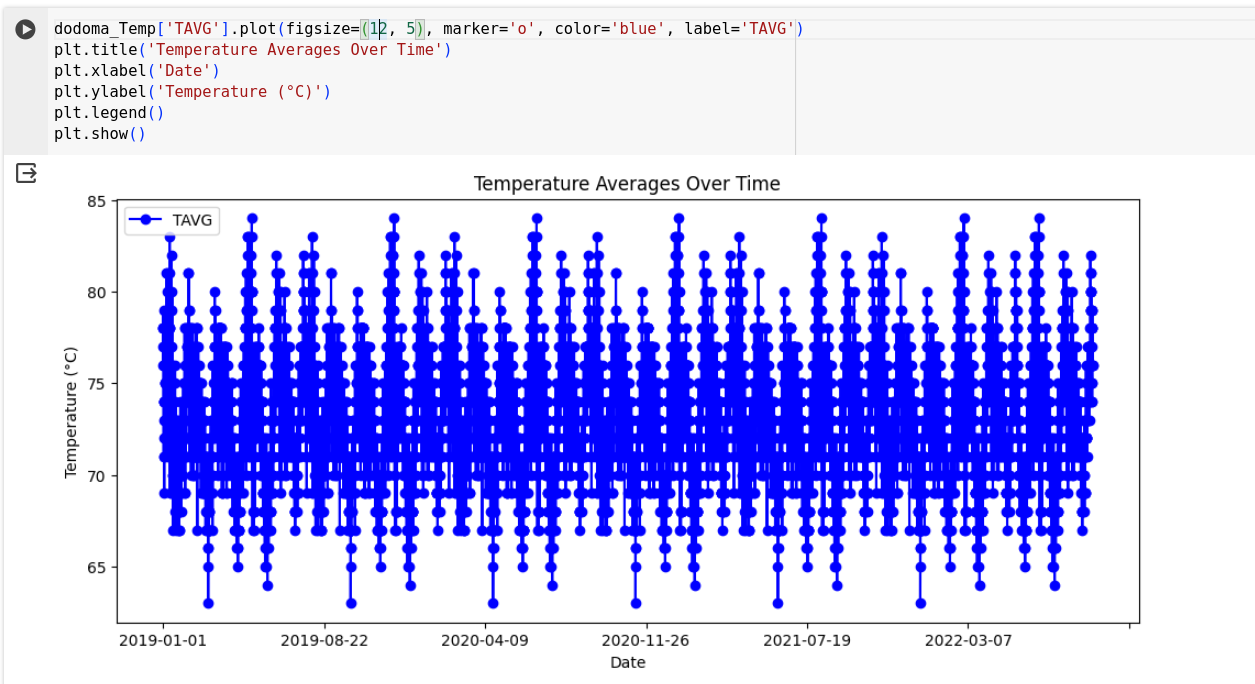


From the above result we came to know that there are no null values in the dataframe.

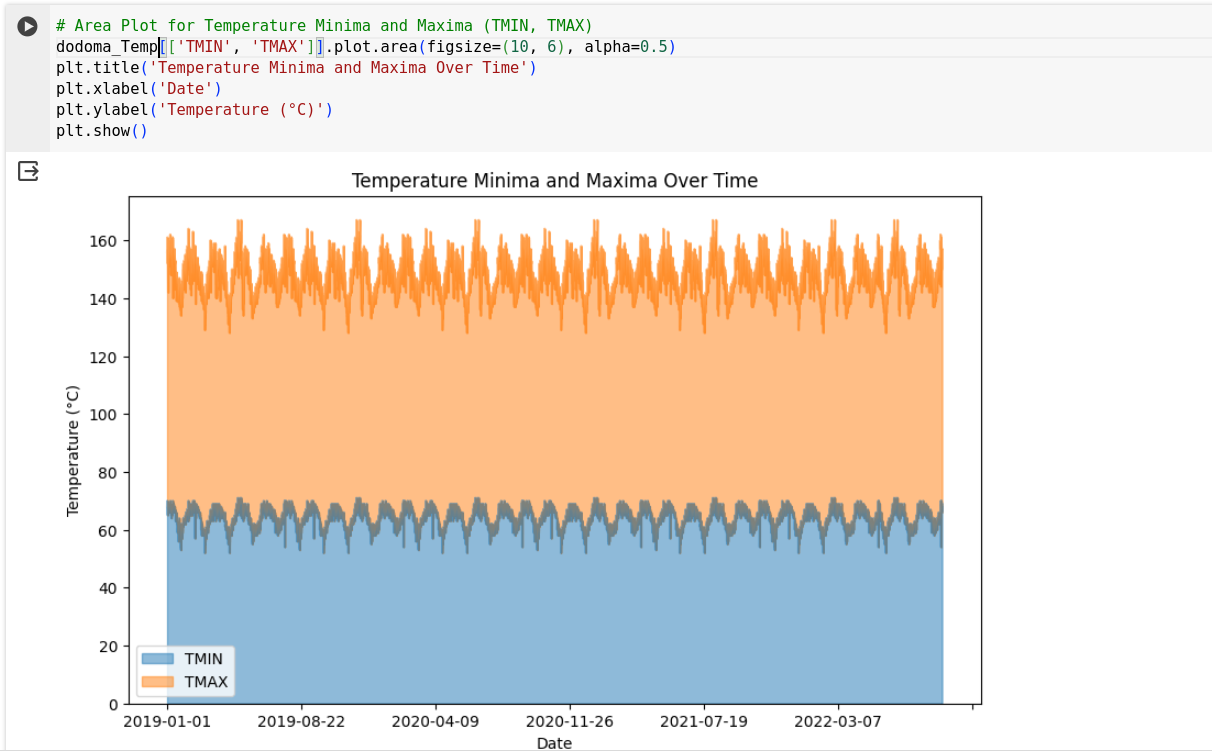
1. **Data Visualization:**

Let’s start the analysis of the data.

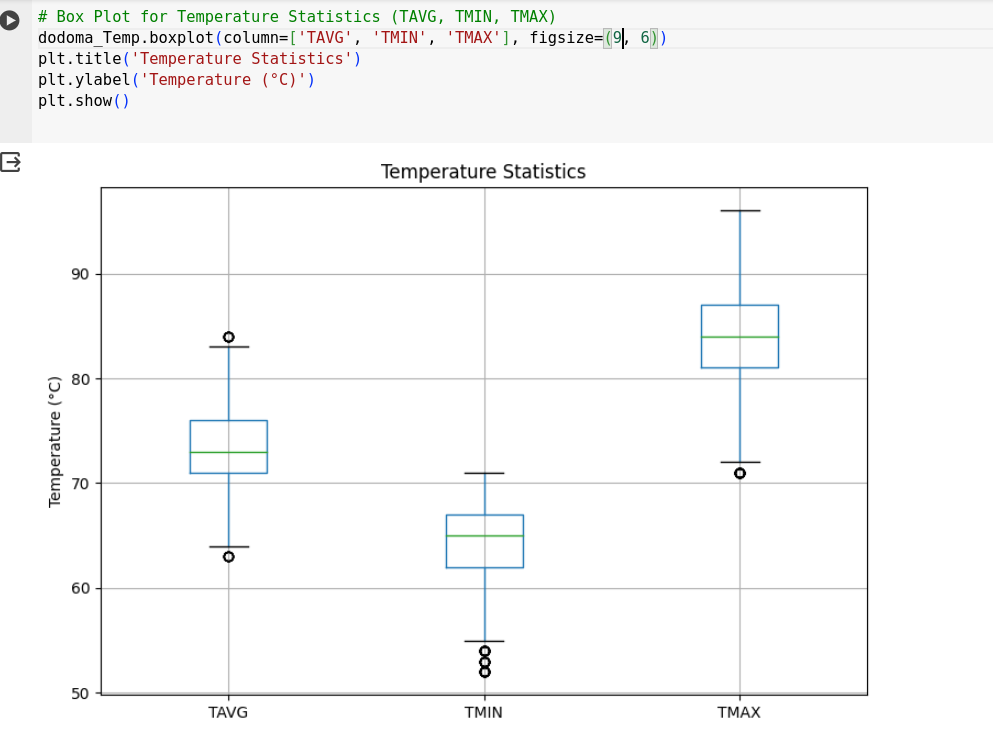
Line Plot for Temperature Averages (TAVG) values using a suitable visualization library like Matplotlib.



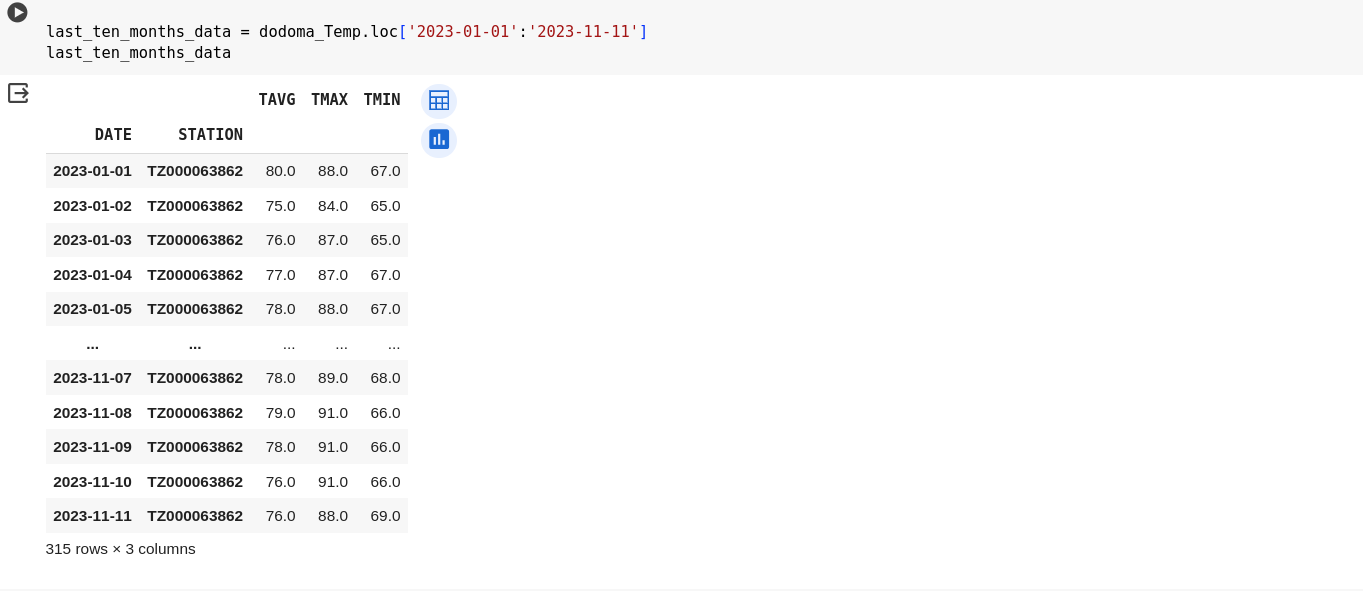
Area Plot for Temperature Minima and Maxima (TMIN, TMAX)

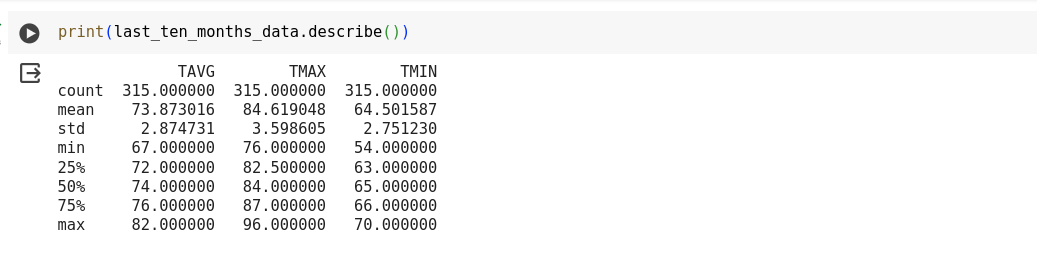


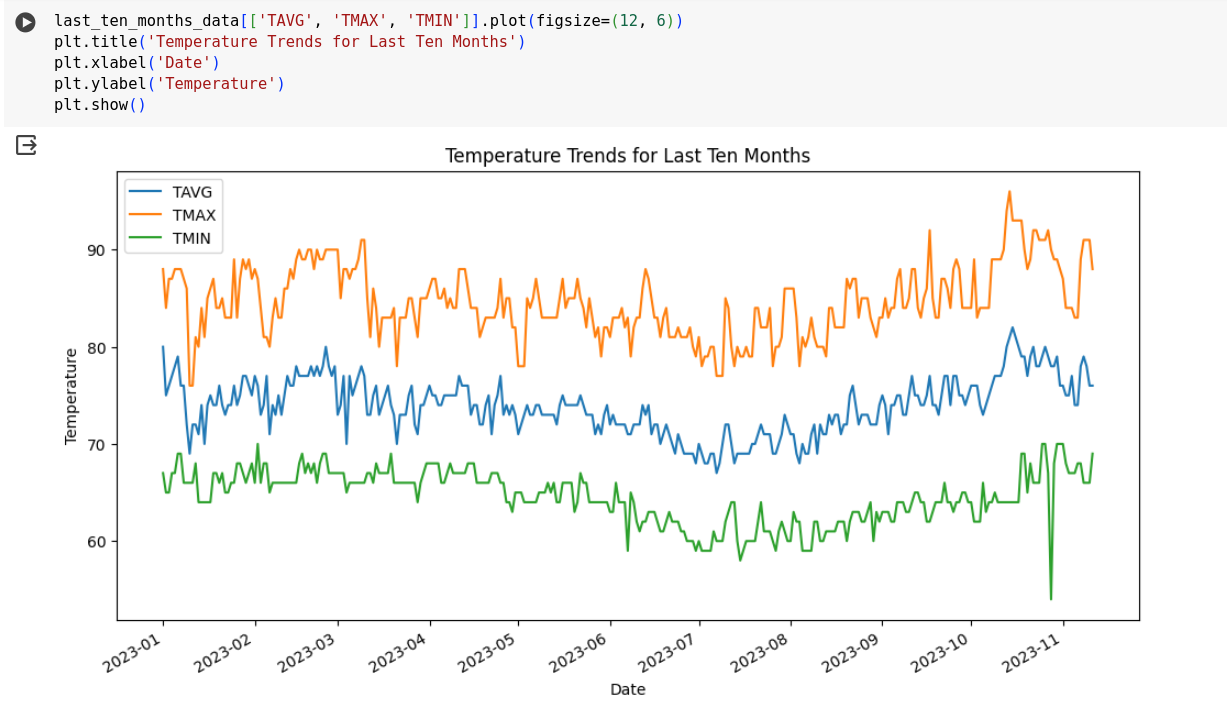
Box Plot for Temperature Statistics (TAVG, TMIN, TMAX)

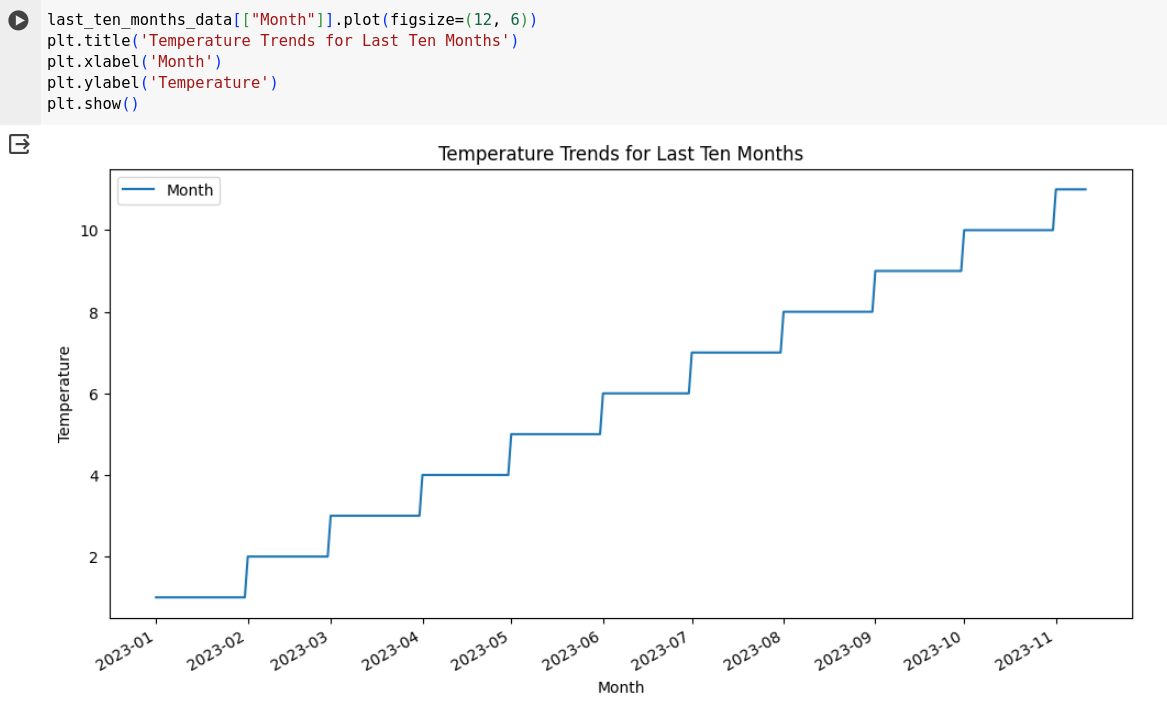


Exploring Ten Months later statistics









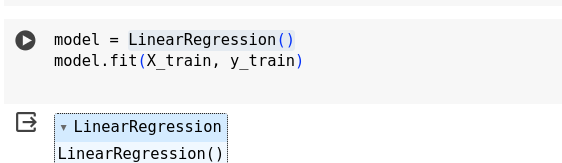
**Model**

LinearRegression is a popular machine learning algorithm that can be used for Tomorrow’s Temperature prediction. It is an ensemble learning method that combines multiple decision trees to make predictions.

Train-Test Split: Split the preprocessed dataset into a training set and a testing set. The training set will be used to train the sklearn.linear\_mode, while the testing set will be used to evaluate its performance on unseen data.

### Create,Initialize and train the Linear Regression Model

Use the training data (X\_train and y\_train) to fit the model:



At this point, the model has learned the relationships between the features and the target variable based on the training data.

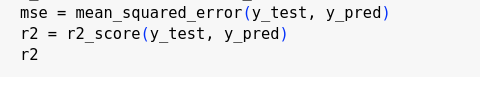
### Make Predictions

Use the trained model to make predictions on the test set:



### Evaluate the Model

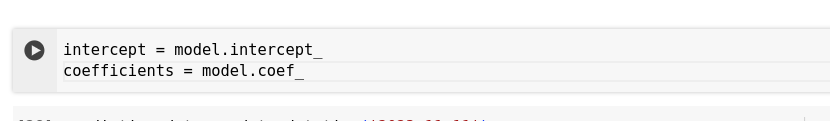
Assess the model's performance using evaluation metrics like mean squared error and R-squared:



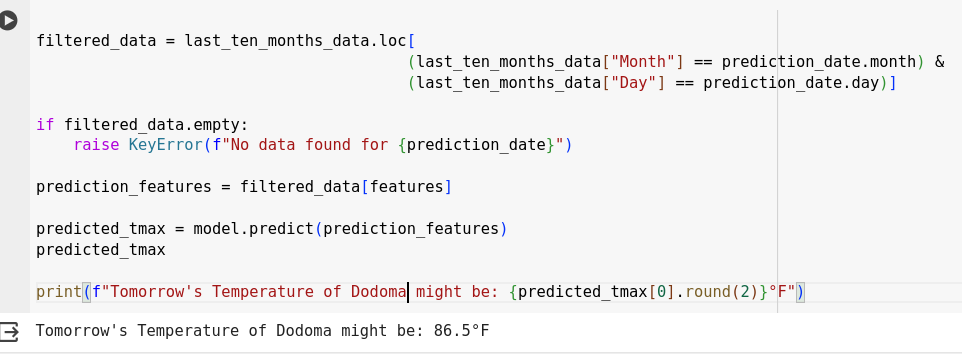
### Interpret the Model

Retrieve the model parameters, such as coefficients and intercept:

These coefficients represent the weights assigned to each feature, and the intercept is the predicted value when all features are zero.



Final we got tomorrow’s Temperature predicted



**Results and Discussion**

The results and discussion of the Tomorrow’s Temperature Prediction with Machine Learning can be summarized as follows:

Model Training and Evaluation: The LinearRegression was trained on the collected data, which included various Weather measurements.

Model Performance: The model demonstrated good performance in predicting whether a customer will order food again or not. The evaluation metrics, especially the model score, indicated the accuracy of the predictions made by the LinearRegression from sklearn.linear\_model .

Further Analysis: While the model score indicates the overall accuracy of the predictions, it is important to delve deeper into other evaluation metrics like precision, recall, and F1 score to gain a comprehensive understanding of the model's performance. These metrics provide insights into the model's ability to correctly classify both positive and negative instances and its balance between precision (correctly predicting positive instances) and recall (capturing all positive instances).

Future Considerations: To further enhance the model's performance, additional steps can be taken, such as feature engineering, hyperparameter tuning, or exploring other algorithms. Regular monitoring and retraining of the model using updated data will help maintain its accuracy and effectiveness over time.

**Conclusion**

In conclusion, the project "Predicting Tomorrow’s Temperature" has successfully addressed the need for a reliable and efficient model in the domain of short-term weather forecasting. By leveraging historical weather data and employing advanced predictive modeling techniques, the project aimed to enhance the accuracy of temperature predictions for the upcoming day. The development and implementation of the predictive model have provided valuable insights into temperature trends and variations, contributing to the improvement of short-term weather forecasting.

Throughout the project, we encountered and overcame various challenges related to data preprocessing, model selection, and validation. The iterative optimization process allowed us to fine-tune the model, ensuring its robustness and reliability in predicting tomorrow's temperature. The inclusion of features such as time series plots, actual vs. predicted temperature comparisons, and residual analysis facilitated a comprehensive understanding of the model's performance.

The successful implementation of this project holds significant implications for diverse industries and individuals who rely on accurate and timely weather information. The predictive model provides a valuable resource for decision-making processes, enabling better preparation and planning in response to changing climatic conditions.

As with any project, there is room for continuous improvement. Future iterations may involve expanding the dataset, incorporating more features, and exploring advanced machine learning algorithms to further enhance the model's predictive capabilities. Additionally, ongoing validation and user feedback will be crucial for maintaining the model's accuracy and relevance over time

In summary, "Predicting Tomorrow’s Temperature" stands as a successful endeavor in the realm of weather forecasting, contributing to the advancement of predictive modeling techniques and offering a practical solution for those seeking reliable temperature predictions for the short term.

**References**

* https://www.ncdc.noaa.gov/cdo-web/results
* https://www.weather-forecast.com/
* https://www.accuweather.com/en/tz/dodoma