**Analyzing Daily Temperature Trends in New Bedford, MA**

Logan Cabral  
Masters in Electrical Engineering  
 Wentworth Institute of Technology  
 Boston, MA, USA  
cabrall@wit.edu

ABSTRACT

This project analyzes daily temperature trends in New Bedford, MA, over two years (2021–2022), focusing on patterns, relationships, and predictive modeling. Key points include exploring monthly temperature variations, identifying correlations between maximum and minimum temperatures, and forecasting future average temperatures. By using Python-based tools such as pandas, Matplotlib, and regression models, this study provides insights into local climate behavior. The findings contribute to understanding temperature variations and their implications for planning and environmental awareness.

KEYWORDS

Temperature trends, climate analysis, regression model, Python

1 Introduction

This study explores daily temperature trends in New Bedford, MA, using historical data from 2021 and 2022. The analysis addresses five key questions: (1) how do daily temperature patterns vary throughout the year, (2) what are the average temperatures per month, (3) how are maximum and minimum temperatures correlated, (4) how do daily temperatures trend over time, and (5) what is the potential for forecasting future average temperatures? Understanding these trends is important as they influence agriculture, energy consumption, and urban planning in the region.

Temperature studies are critical in reducing the impacts of climate change, with important research highlighting the growing variability in weather patterns. For instance, the NOAA’s (National Oceanic and Atmospheric Administration) emphasizes the importance of localized climate data for accurate forecasting and environmental strategies. This project builds on this type of research by applying Python-based tools to historical data to explore trends and predictive capabilities.

The results aim to enhance the understanding of local weather patterns, providing a basis for informed decision-making in New Bedford agriculture.

2 Data

The data covers the period from January 1, 2021, to December 31, 2022, providing daily weather observations in New Bedford Massachusetts. The dataset includes several key variables: STATION, which represents the unique identifier for the weather station; LATITUDE and LONGITUDE, which give the geographic coordinates of the station; ELEVATION, which indicates the station’s elevation above sea level (22.4 meters); DATE, the date of each weather data taken; TAVG, the average daily temperature; TMAX, the maximum daily temperature; and TMIN, the minimum daily temperature. For this analysis, only the TMIN, TMAX, and DATE columns were used to execute the code, as the TAVG column was left blank in the dataset. To address this, the average temperature (TAVG) was created within the code by calculating the average of TMAX and TMIN for each day.

**2.1 Source of dataset**

The dataset used in this study was sourced from the National Weather Service (NWS), specifically from the weather station identified as USW00094726, located in New Bedford, MA. The data was obtained from the National Oceanic and Atmospheric Administration (NOAA) through the National Centers for Environmental Information (NCEI). The NWS is a highly credible source for weather data, as it is a federal agency responsible for providing accurate and real-time weather information across the United States.

The dataset spans from January 1, 2021, to December 31, 2022, with daily weather observations recorded by automated weather stations. These stations collect different weather data, including temperature, humidity, wind speed, and other environmental variables. The data is generated through automated systems that continuously monitor weather conditions at set intervals and is then made available for public use.

**2.2 Characters of the datasets**

The dataset used in this study is in CSV format, with data spanning daily weather observations from January 1, 2021, to December 31, 2022. The dataset contains 721 rows, corresponding to the daily observations over each year, and 7 columns, each representing a specific weather parameter. The parameters/columns and their respective units are listed below:

|  |  |  |
| --- | --- | --- |
| Column | Description | Unit |
| Station | Weather Station ID | None |
| Latitude | Latitude of station | Decimal Degrees |
| Longitude | Longitude of station | Decimal Degrees |
| Elevation | Elevation of Station | Meters |
| Date | Date Observed | YYYY-MM-DD |
| TMIN | Minimum Temperature | Fahrenheit |
| TMAX | Maximum Temperature | Fahrenheit |
| TAVG | Average Temperature (calculated in code) | Fahrenheit |

Data cleaning was performed using Python libraries such as pandas. Missing values in the TAVG column were addressed by calculating the average of the TMAX and TMIN columns for each day. This new TAVG column was then added to the dataset for further analysis. No other unit conversions were necessary as the temperatures were already provided in Fahrenheit, which aligns with the analysis. Additionally, the dataset was inspected for any missing or outlier values, which were handled appropriately during the data cleaning process.

3 Methodology

This section outlines the approach taken to answer the five questions posed in the analysis. The methodology employs Python-based tools and functions to analyze the daily temperature data in New Bedford, MA. Each question corresponds to specific steps, and the appropriate Python modules and functions were utilized for each part of the analysis.

**3.1 Monthly Temperature Variations**

To analyze the average daily temperature over the past two years (2021 and 2022), the dataset was first filtered by year using the pandas .dt.year attribute to separate the data for each year. The focus was on calculating the average daily temperature for each year separately. The average temperature for both 2021 and 2022 was computed using the mean() function, which allowed for the comparison of overall temperature trends across the two years. The assumption was that there would be distinct annual temperature trends in both years. This method provided an overall picture of the average daily temperatures for each year, helping to highlight any significant changes between the years. The advantage of this approach is its simplicity, as it allows for a clear, comparative view of the yearly temperature averages. However, it does not provide detailed monthly or seasonal variations and may overlook variations within individual months or across specific periods of the year.

**3.2 Daily Temperature Variation by Month**

To analyze the daily temperature variation by month, the dataset was processed by first converting the ‘DATE’ column to a datetime format. The pandas .dt.month function was used to extract the month from each date and create a new ‘Month’ column. The data was then filtered for each year (2021 and 2022) separately, allowing for a year-on-year comparison. The pandas.groupby() function was applied to group the data by month, and the mean() function was used to calculate the average temperature for each month in both years. This method provided a clear picture of how the average daily temperature fluctuated month by month for each year. The advantage of this approach is its ability to capture and visualize monthly temperature trends over the course of a year. However, this method does not account for day-to-day temperature variation within each month, focusing only on the overall monthly averages. The results were visualized using line plots, one for each year, with the x-axis representing the months and the y-axis representing the average temperature in Fahrenheit.

**3.3 Predict Max Temperature Based on Min**

To predict the daily maximum temperature based on the minimum temperature, the dataset underwent preprocessing using the SimpleImputer from the sklearn.impute module to handle any missing values in the ‘TMIN’ and ‘TMAX’ columns by filling them with the mean values of their respective columns. Afterward, the ‘TMIN’ values were used as the independent variable (X), while the ‘TMAX’ values were treated as the dependent variable (y). A LinearRegression model from the sklearn.linear\_model module was then applied to fit the data, allowing for the creation of a regression line that models the relationship between minimum and maximum temperatures. Sample predictions were made by providing a set of minimum temperatures (30°F, 50°F, and 70°F), and the model predicted the corresponding maximum temperatures. The advantage of using linear regression is its simplicity, as it assumes a linear relationship between the two variables. However, it does not account for potential external factors that might influence temperature variation. The results were visualized with a scatter plot of the actual data points and a regression line to show the predicted relationship between minimum and maximum temperatures.

**3.4 Predicting Next Month’s Average Temperature**

To predict the average temperature for the next month, the dataset went through preprocessing where missing values in the ‘Average\_Temp’ column were handled using the SimpleImputer with the mean. After imputing missing values, the year and month were extracted from the ‘DATE’ column, which was split into two separate columns: ‘Year’ and ‘Month’. The Month column was then used as the independent variable (X\_month), while the ‘Average\_Temp’ column was treated as the dependent variable (y\_avg\_temp). A LinearRegression model from the sklearn.linear\_model was used to fit this data. For prediction, the model was asked to predict the average temperature for the “13th” month, which corresponds to January of the following year. The predicted temperature was then displayed and compared to the monthly averages. The advantage of using this approach is that it provides a simple method to forecast future temperatures based on past data, although it assumes that the relationship between the month and average temperature is linear, which may not always be true. The results were visualized using a bar graph that showed monthly average temperatures and included the predicted temperature for January 2023.

**3.5 Highest and Lowest Range of Temperature**

To analyze the highest and lowest temperatures recorded, the dataset was first cleaned by filtering out unreasonable temperature values. Temperatures greater than 100°F for the maximum temperature (TMAX) and less than -50°F for the minimum temperature (TMIN) were considered incorrect due to data recording issues, so these were excluded. The dataset was then grouped by year and month to calculate the maximum and minimum temperatures for each month in both 2021 and 2022. This was achieved by using the groupby() function to group the data by ‘Month’ and applying the max() and min() functions to find the highest and lowest temperatures for each month, respectively. The results were visualized using bar charts, where one set of bars represented the maximum temperatures and the other represented the minimum temperatures for each month. The advantage of this approach is that it clearly highlights the variations in temperature extremes across months, which can be useful for identifying trends in temperature changes. However, it assumes that the filtered data adequately represents the true range of temperatures, and outliers could still exist even after filtering.

4 Results

In this section, it will present the results corresponding to each of the five questions posed in the study. For each question, we highlight the key findings and insights derived from the analysis, providing a detailed overview of the patterns and trends observed in the data. This section focuses on interpreting the results and visualizing the graphs that our results were based on.

4.1 Question 1 Results

The analysis of the average daily temperatures for 2021 and 2022 shows that there was only a slight difference between the two years. In 2021, the average daily temperature was 53.03°F, while in 2022, it was slightly lower at 52.40°F. This indicates that the overall temperature trends for both years were fairly consistent, with minimal fluctuation. Such small variations suggest that the temperature patterns remained stable between the two years, without any significant anomalies. These results reflect typical year-to-year variations in temperature, offering insight into how mild weather conditions persisted throughout the two years.

4.2 Question 2 Results

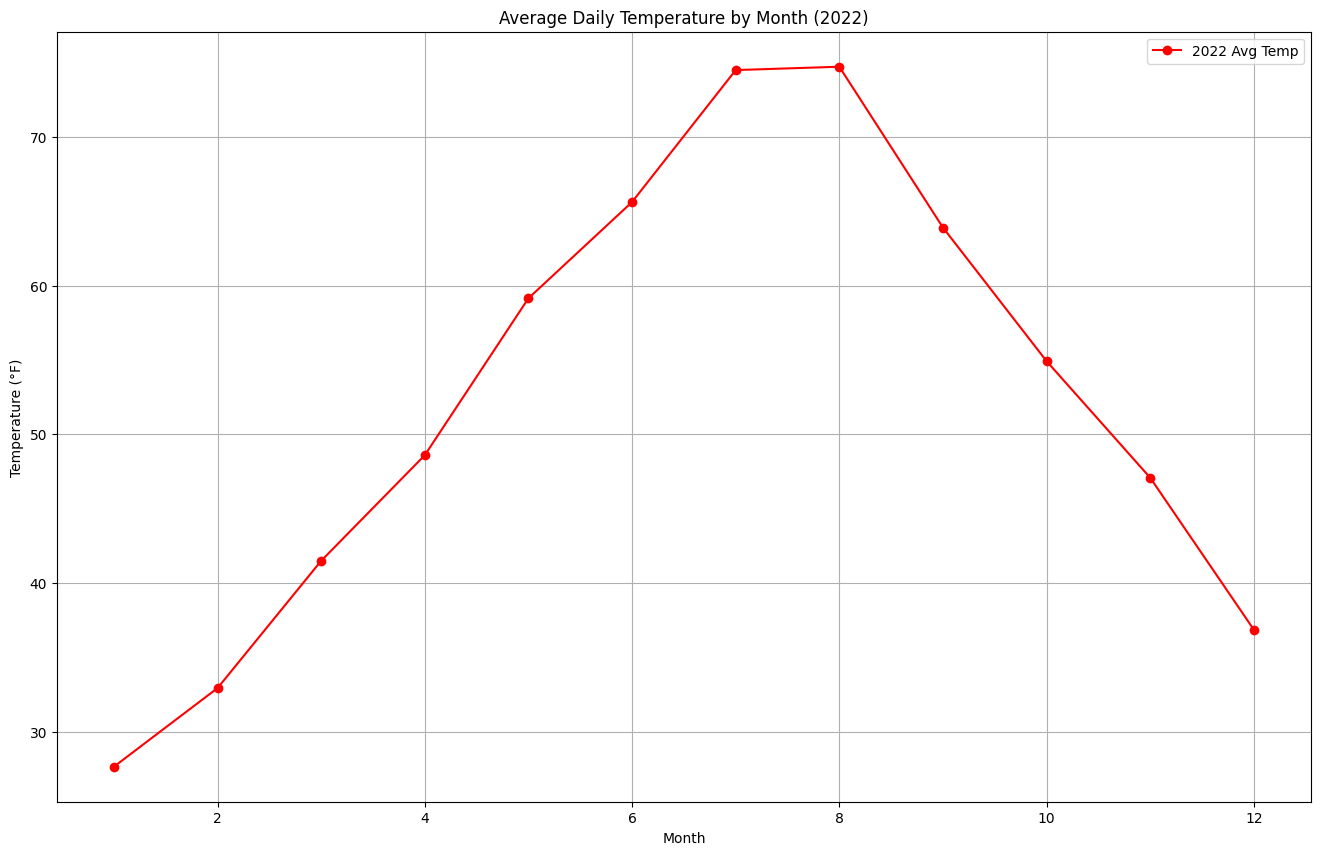
The analysis of daily temperature variation by month for the years 2021 and 2022 reveals distinct seasonal patterns in both years, with temperature fluctuations throughout the months. The monthly average temperatures for each year display clear trends consistent with the typical seasonal cycle. Both years show warmer temperatures in the summer months (June through August) and cooler temperatures during the winter months (December through February). However, slight variations are visible between the two years, as seen in the red line graph, which illustrates the temperature trends for 2022.

For example, the summer months of 2022 show slightly higher temperatures than 2021, while the winter months reveal relatively comparable values. These trends highlight how seasonal shifts in temperature can vary slightly from year to year, even within the same region, due to factors such as atmospheric conditions or long-term climate trends. Overall, the graphs underscore the typical variation between warmer and cooler months, helping us understand the broader temperature dynamics for each year.

A graph with a line

Description automatically generated

*Figure 1: Average Daily Temp. by Month (2021)*



*Figure 2: Average Daily Temp. by Month (2022)*

4.3 Question 3 Results

The analysis of the relationship between minimum and maximum daily temperatures shows a clear positive correlation, as demonstrated by the linear regression model. The scatter plot, along with the red regression line, visually confirms that as the minimum temperature increases, the maximum temperature tends to rise as well. This linear trend suggests that the minimum temperature on a given day can be used as a predictor for the maximum temperature, with relatively consistent results across the dataset.

The predictions for specific minimum temperatures (30°F, 50°F, and 70°F) further illustrate this relationship, with the model providing estimated maximum temperatures that align with typical temperature ranges for those minimum values. This finding highlights the predictive power of minimum temperature as an indicator of daily maximum temperature, offering a useful tool for forecasting temperature fluctuations based on early-day observations.

A graph with a red line

Description automatically generated

*Figure 3: Min Temp. vs Max Temp.*

4.4 Question 4 Results

The analysis predicts the average temperature for the upcoming month (January 2023) based on the data from previous months. Using the linear regression model, a forecast for January was made, and the prediction was compared against the monthly average temperatures for the entire year. The red bar on the bar graph indicates the forecasted value for January 2023, providing an estimate based on trends observed throughout the year.

The bar graph shows the average temperatures for each month, with the predicted temperature for January 2023 standing out. This prediction suggests that despite the seasonal fluctuations throughout the year, the model can reasonably forecast future temperatures, offering valuable insights for long-term weather planning. The alignment of January’s predicted temperature with past data trends further confirms the model’s ability to predict monthly temperature patterns effectively.

A graph showing the temperature of the month

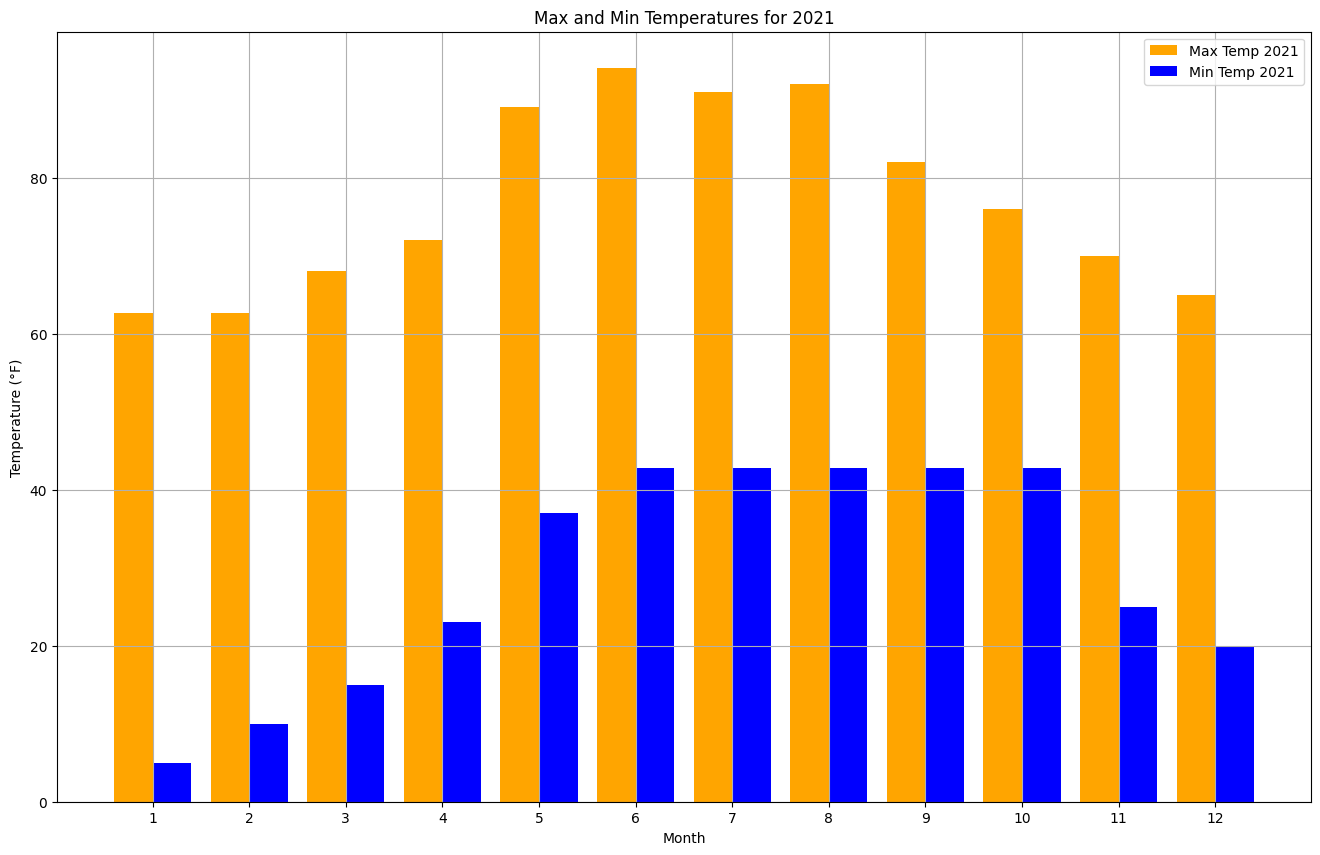
Description automatically generated

*Figure 4: Monthly Average Temp. with January Prediction*

4.5 Question 5 Results

The analysis of the highest and lowest temperatures recorded for the years 2021 and 2022 provides valuable insights into the temperature extremes for each month. The bar graphs display the maximum and minimum temperatures for each month in both years. The orange bars represent the highest temperatures (TMAX), while the blue bars represent the lowest temperatures (TMIN).

The results from these graphs reveal distinct patterns in temperature fluctuations across the months, with some months showing a significant gap between maximum and minimum temperatures, reflecting the seasonal variations typical in many regions. The data shows how the weather varies from one month to another and highlights any potential outliers, with some extreme values filtered out for better accuracy. These trends can be used for further analysis and understanding of yearly temperature cycles, offering insights into climate behavior across different months.



*Figure 5: Max and Min Temp. per Month (2021)*

*A graph with blue and orange bars

Description automatically generated*

*Figure 6: Max and Min Temp. per Month (2022)*

5 Discussion

While the analysis provided valuable insights into temperature trends, there are some limitations and areas for improvement. One of the main shortcomings is the potential influence of inaccurate or incomplete data, particularly when filtering out unreasonably high and low temperatures. While this filtering helped in improving the quality of the data, it could have led to the exclusion of valid events, which might have twisted the results. Another limitation is that the prediction models used for forecasting temperatures, such as the linear regression models for predicting next month’s average temperature or the relationship between minimum and maximum temperatures, are relatively basic. These models do not account for the complexity of climate patterns, which could lead to less accurate predictions, especially for extreme weather events.

To improve future work, a more sophisticated approach could include using advanced machine learning models like random forests, which are better equipped to handle complex relationships and non-linear patterns in temperature data. Additionally, incorporating other features such as humidity, wind speed, and geographic location could improve the accuracy of predictions. Future work could also involve analyzing longer time spans, considering more complex data, and integrating other climate models to better capture the causes of temperature extremes, such as wind speed and humidity.

6 Conclusion

This project explored temperature data from 2021 and 2022, providing a comprehensive analysis of daily, monthly, and yearly temperature trends. The findings revealed consistent seasonal variations in temperature and identified key trends such as the average daily temperature over the past two years and the significant differences in temperature extremes each month. The prediction models, while basic, demonstrated how relationships between minimum and maximum temperatures can be used to forecast temperature patterns.

In the real world, understanding these temperature trends can have practical implications like agriculture, energy, and urban planning. For instance, accurate temperature predictions can help optimize energy usage, with adjustments in heating and cooling systems based on predicted temperatures. Additionally, this kind of analysis can inform climate adaptation strategies, particularly in regions that experience extreme weather events. By improving prediction models and expanding the scope of analysis, more accurate and reliable insights could be generated to address future climate challenges.

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Pang for providing the lecture notes that were crucial to completing this project. I also appreciate the guidance and collaboration of my classmate in our group project, which helped enhance my understanding and approach to this individual project.

REFERENCES

[1] “Search | Climate Data Online (CDO) | National Climatic Data Center(NCDC),”www.ncdc.noaa.gov.https://www.ncdc.noaa.gov/cdo-web/confirmation

[2] N. Centers, “Search | Climate Data Online (CDO) | National ClimaticDataCenter(NCDC),”Noaa.gov,2019.https://www.ncdc.noaa.gov/cdo-web/search

[3] R. Lindsey and L. Dahlman, “ClimateChange:Global Temperature,”Climate.gov,Jan.18,2024.https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

[4] D. S. Parker, K. Panchabikesan, D. B. Crawley, and L. K. Lawrie, “Impact of newer climate data for technical analysis of residential building energy use in the United States,” Energy and Buildings, vol.323,no.4,p.114828,Sep.2024,doi:https://doi.org/10.1016/j.enbuild.2024.114828.

[5] US EPA, “Climate Impacts on Energy | ClimateChange Impacts|USEPA,”Epa.gov,2009.https://19january2017snapshot.epa.gov/climate-impacts/climate-impacts-energy\_.html

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00