

Weather Data Integration Pipeline and Dashboard Technical Report

Rohit Sunilkumar Hooda, Rimsha Kayastha, Tri Watanasuparp

November 25, 2024

1 Introduction

Many people start their day by checking the weather forecast to plan activities such as commuting, exercising, or scheduling events. Given how essential accurate weather information is to daily life, it's important to have reliable and comprehensive tools at our disposal. This project seeks to fill that need by integrating data from multiple sources and offering a detailed weather dashboard that provides both current forecasts, for future preparations, and historical data, to be aware of the ongoing weather trends. Additionally, it incorporates interactive, user-friendly visualizations that empower users to analyze trends, draw relevant conclusions, and make informed decisions based on weather patterns. This combination of real-time updates and historical context aims to enhance users' ability to anticipate weather conditions and better plan their activities. This project focuses on Massachusetts, a US state, in particular for work on a more detailed, smaller scale.

2 Literature Review

2.1 Arcgis Dashboard

The [Arcgis weather dashboard](#) is a powerful tool designed to display real-time weather data with an emphasis on USA weather alerts, including storms, heat waves, and hurricane watches. Central to the dashboard is an interactive map, as you can see in Figure 1, that users can easily navigate—dragging, zooming in, or zooming out—to access weather information specific to any location. This interactivity provides an accessible way for users to obtain localized data, making it especially valuable for monitoring conditions and staying informed about potential hazards in targeted areas.

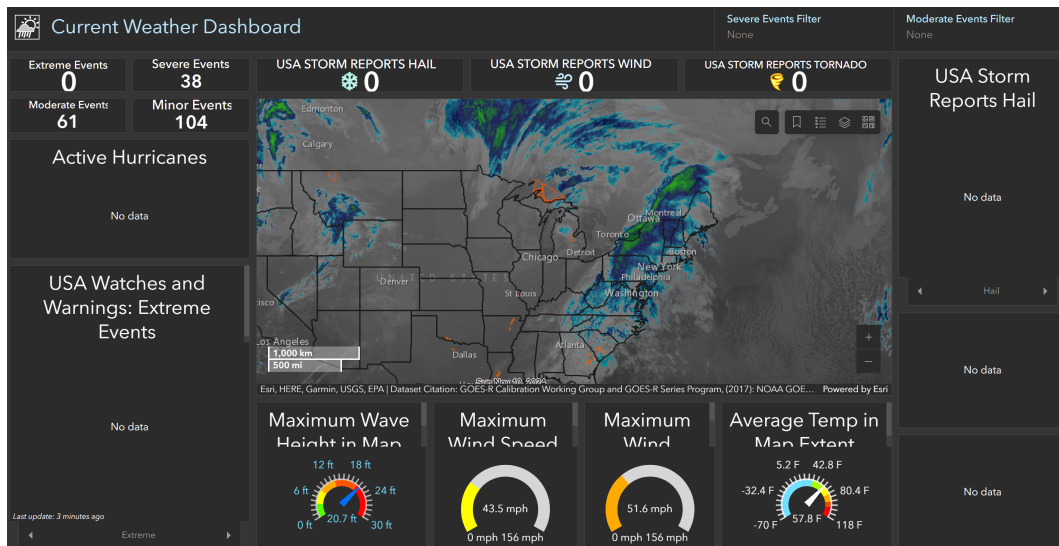


Figure 1: Arcgis Weather Dashboard

2.2 Ambient Weather Dashboard

The [Ambient Weather dashboard](#) provides detailed, real-time weather data sourced from personal weather stations, offering hyper-localized information that is essential for users who need precise, neighborhood-specific updates. As shown in Figure 2, it tracks key weather metrics such as temperature, humidity, wind speed, and rainfall, ensuring a comprehensive view of current conditions. Notably, the dashboard is highly customizable, allowing users to choose the specific weather parameters they wish to monitor and arrange the data widgets to suit their needs, whether for gardening, event planning, or professional purposes. Additionally, it includes historical data analysis capabilities, enabling users to observe weather trends over time. This tailored and interactive approach makes the Ambient Weather dashboard a valuable resource for anyone wanting to stay informed about local weather conditions.

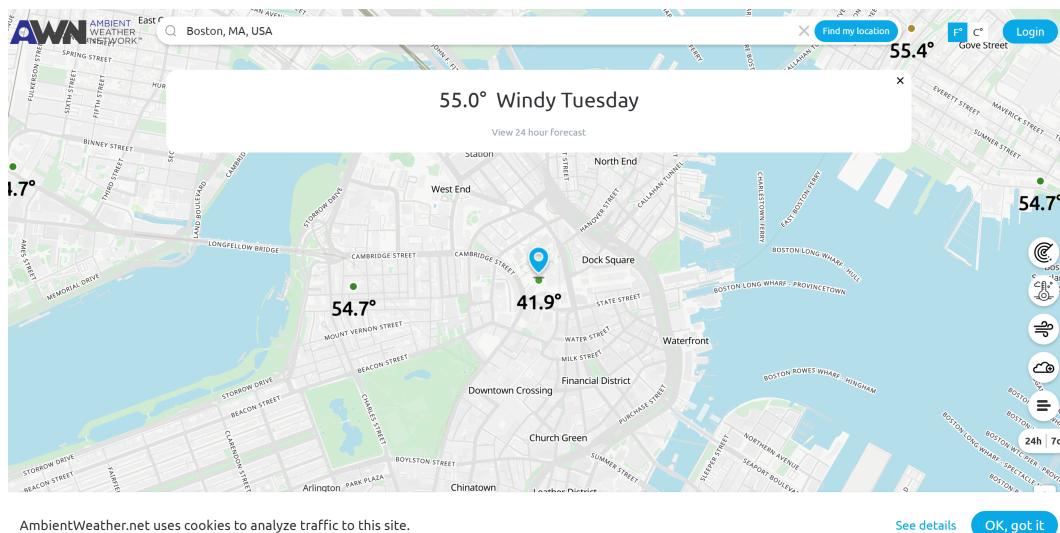


Figure 2: Ambient Weather Dashboard

3 Project Structure

As shown in Figure 3, the flow of this project has three sections: Data sources, transformation, and deliverables. The first section involves retrieving data from multiple sources and integrating them into a singular pipeline and feeding them into the next section of data processing and transformation. Data transformation refers to the cleaning of data in order to address null values and outliers, and pre-processing refers to standardizing the data into a format that makes it more readable for users. Finally, the final section of the project is to produce readable visualizations that helps draw conclusions and portray them in distributable formats.

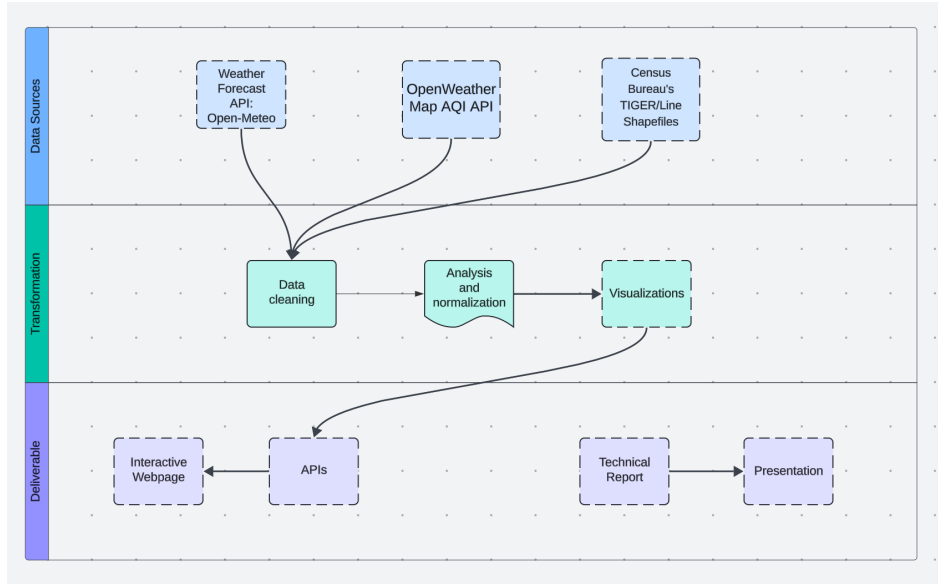


Figure 3: Project Process

4 Data Pre-processing

4.1 Data retrieval

Our weather data is retrieved from three sources through API calls, Mateo, AQI, and Census Bureau's TIGER/Line Shapefiles, and pipe-lined into a single data frame that is then cleaned, analyzed, and plotted for inference. The raw, combined data includes the following features: latitude, longitude, weather code, daylight duration, sunshine duration, max temperature, min temperature, max uv index reached, max precipitation probability, and AQI value. These features are then filtered down to the 12 features we are focusing on for the dashboard: date, county, latitude, longitude, weather code, daylight duration, sunshine duration, max temperature, min temperature, max uv index, max precipitation probability, and aqi value.

4.2 Data Analysis

Gathering data from the three sources resulted in obtaining more than 25 features, from which they're funneled down into the more relevant 12 features mentioned in the previous section. Consequently, analysis was done on the spread of data for those features. For analysis, the features were divided into three different groups based on their data scale. The bar chart shown in Figure 4 for spread of data for sunshine and daylight duration where each unit is two hours.

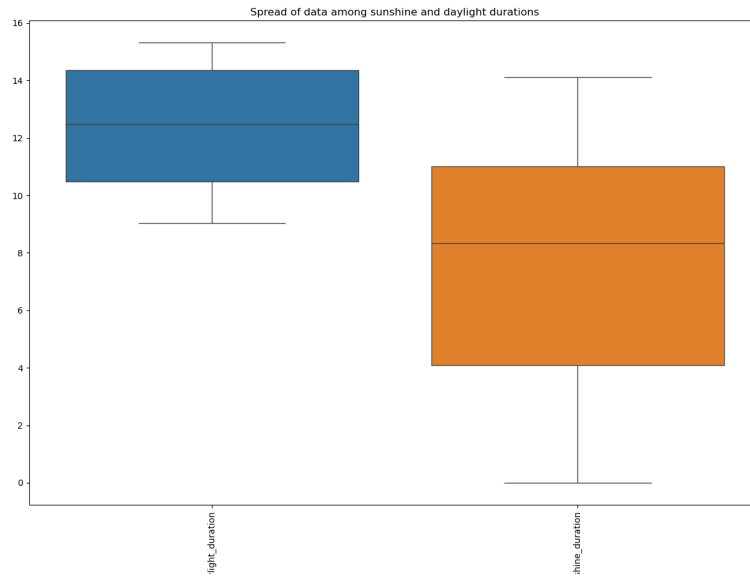


Figure 4: Sunshine-daylight duration data spread

Figure 5 shows the spread of minimum and maximum temperature data. Figure 6 shows the spread of the remaining features.

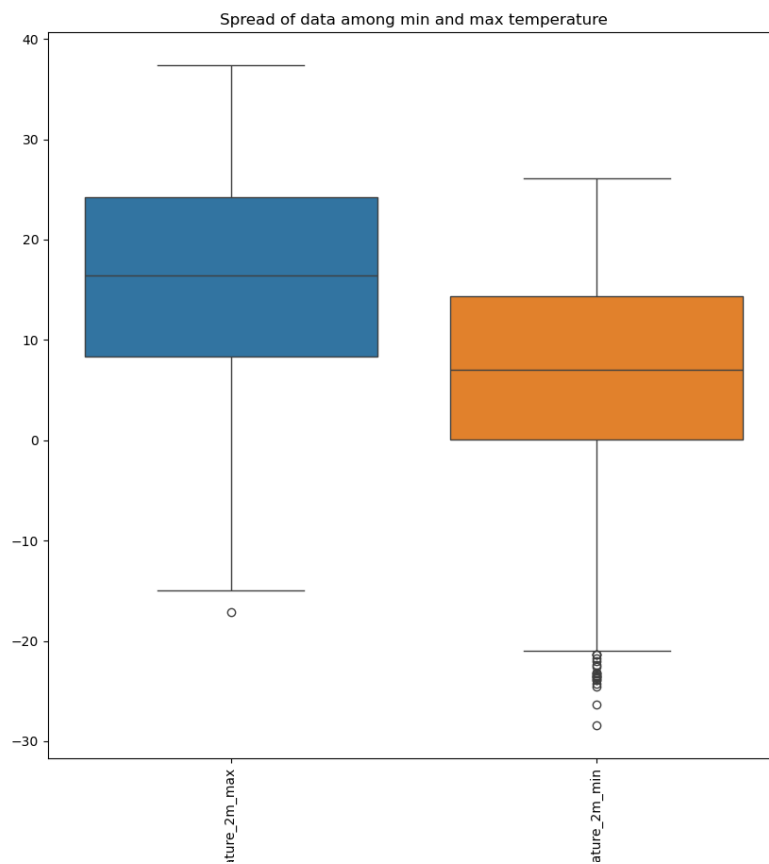


Figure 5: Min and max temperature data spread

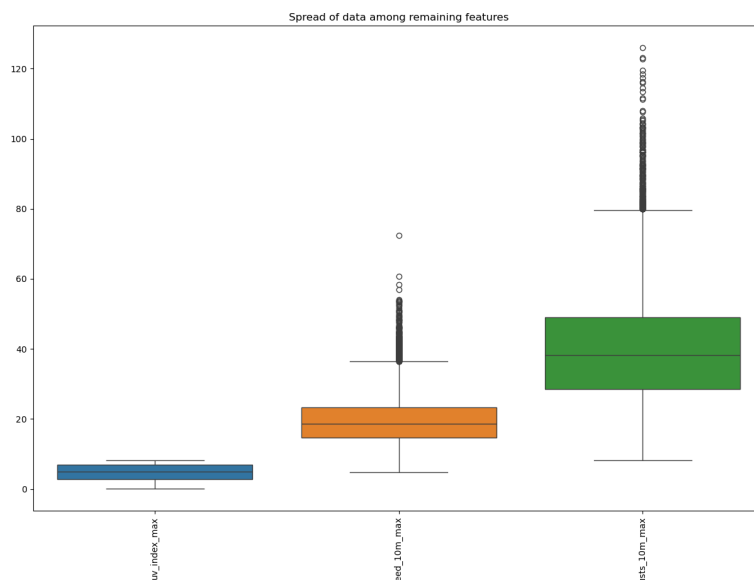


Figure 6: Data spread among the remaining features

4.3 Data Cleaning

The bar charts in the previous section show many outliers that need to be adjusted. Using the quartiles from there data spread, the data points are clipped to an upper and lower bound, such that lower bound is $q1 - 1.5 * IQR$ and upper bound is $q3 + 1.5 * IQR$, where $q1$ is the first quartile, $q3$ is the third quartile, and the IQR is the interquartile range.

5 Website Architecture

6 Dashboard design and visualizations

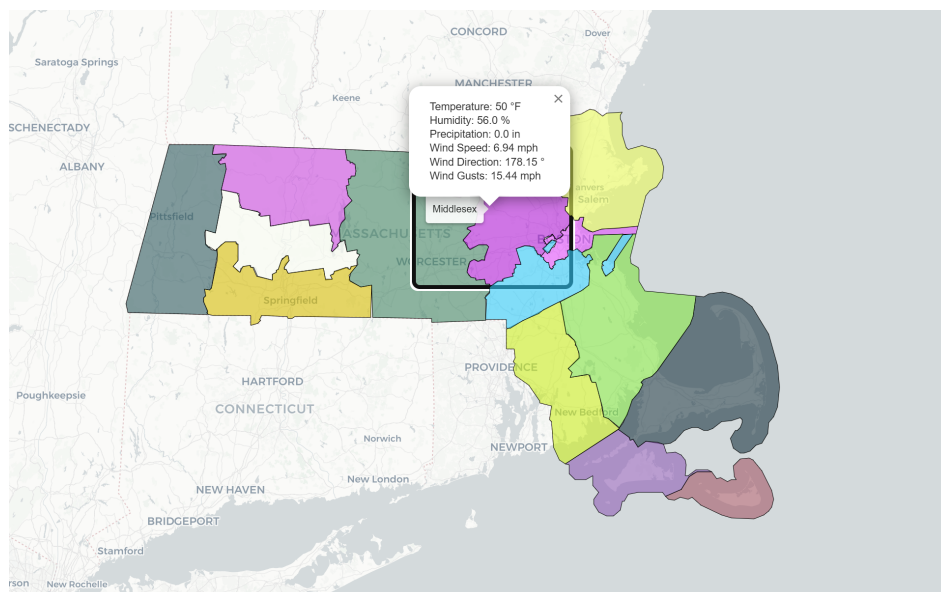


Figure 7: Chloropeth Map - Massachussetts

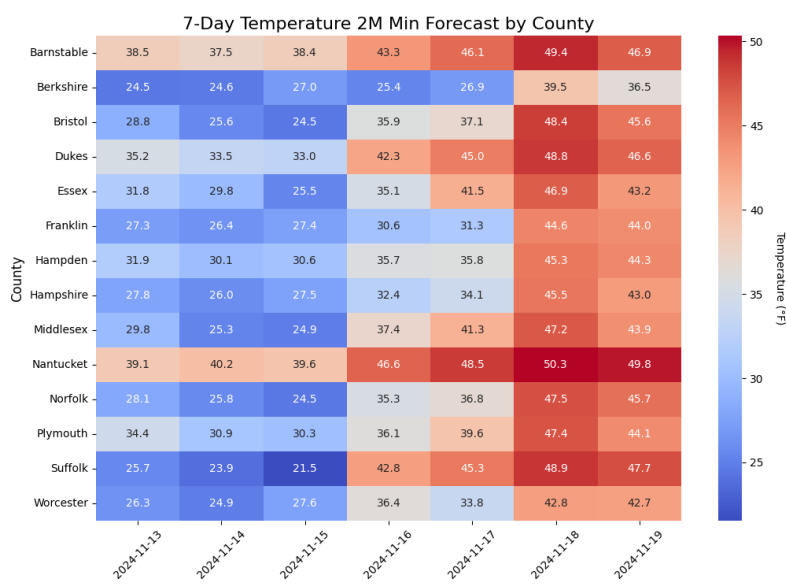


Figure 8: 7-day temperature forecast heatmap

References