Weather Data for the Years 2020, 2021, and 2022

1. Introduction

This report details a compiled dataset encompassing weather data for the years 2020, 2021, and 2022. The information encompasses various meteorological parameters, including temperature, humidity, wind conditions, pressure, precipitation, and prevailing weather conditions.

2. Data Source

The data originates from weather monitoring stations or reputable meteorological databases for the respective years.

3. Data Description

3.1 Contents

The dataset comprises multiple variables gathered at regular intervals throughout the specified period. Each record signifies a distinct timestamp or date.

3.2 Variables

- **DATE:** Date and time of the recorded data.
- **Temperature:** Temperature in degrees Fahrenheit (°F).
- **Humidity:** Relative humidity percentage.
- **Wind Speed:** Wind speed in miles per hour (mph).
- **Wind Gust:** Maximum wind speed (gust) in miles per hour (mph).
- **Pressure:** Atmospheric pressure measured in inches of mercury (inHg).
- **Precipitation:** Amount of precipitation in inches.
- **Condition:** Weather condition represented by numerical codes (e.g., 1 for Fog, 2 for Smoke).

3.3 Data Structure

Each variable occupies a separate column in the dataset. The data is organized in a tabular format, with each row representing a single observation.

4. Data Preprocessing

The dataset underwent various steps to assure data quality and consistency:

- **Missing value handling:** Appropriate techniques were employed to address missing values.
- **Text removal:** Textual data within cells was removed.
- **Categorical to numeric conversion:** Necessary categorical columns were converted into numerical format for consistent analysis.
- **Date and time formatting:** Date and time columns were formatted consistently across the dataset.

4.1 Preprocessing Specifics

- **2021 data:** Minor cleaning involved eliminating unnecessary rows and formatting specific columns (temperature, wind speed, wind gust, pressure, and precipitation).
- **2020 and 2022 data:** Similar preprocessing steps were applied, including handling missing values, formatting date columns, and cleaning humidity and dew point data.

5. Potential Use Cases

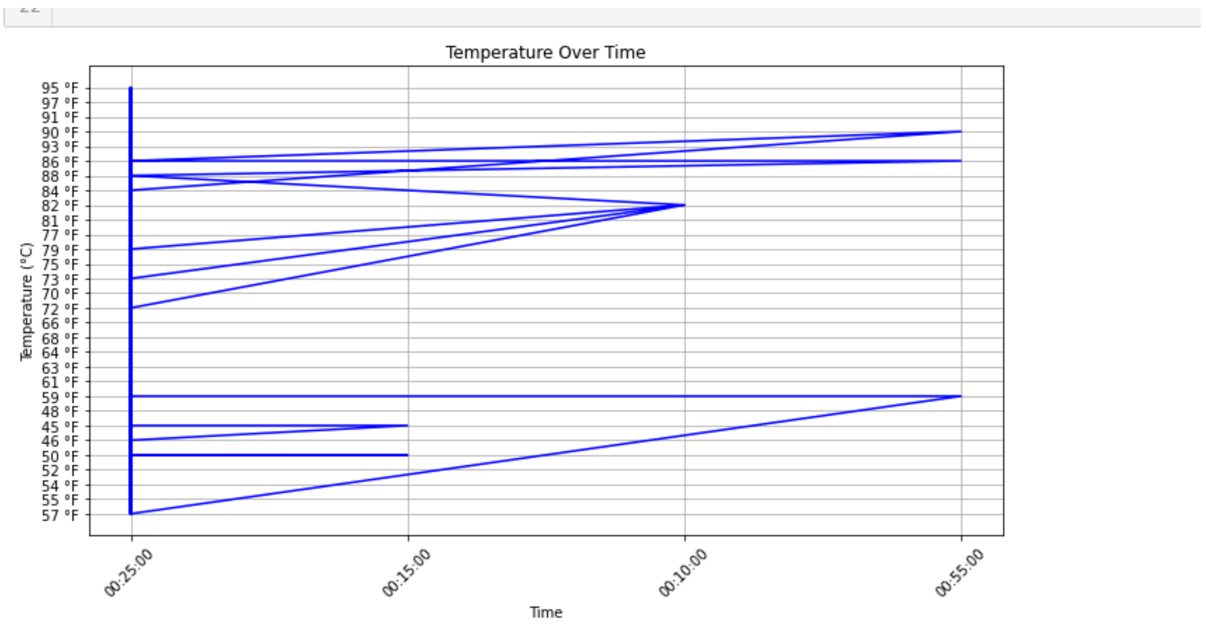
This dataset holds potential for various applications related to weather analysis and forecasting:

- **Research:** Studying weather patterns, trends, and correlations.
- **Meteorological applications:** Weather forecasting and analysis.
- **Data science:** Utilizing the data for data-driven weather-related projects.

6. Data Quality

While the dataset has been cleaned and preprocessed to address identified issues and ensure consistency, users are advised to conduct further data validation and quality checks according to their specific needs.

7. Data Visualization



The image represents temperature variations over time. It consists of three temperature curves, each corresponding to different conditions or scenarios. The x-axis represents

time (e.g., hours, minutes), and the y-axis represents temperature in Fahrenheit (°F). The graph provides insights into how temperature changes throughout the specified time period.

Key Points:

Data Source: The temperature data is collected from observations or measurements.

Time Intervals: Each point on the x-axis corresponds to a specific time (e.g., hourly intervals).

Temperature Ranges: The y-axis shows temperature values ranging from approximately 57°F to 95°F.

Multiple Curves: The three curves represent different scenarios, conditions, or locations (e.g., indoor vs. outdoor temperature, different days).

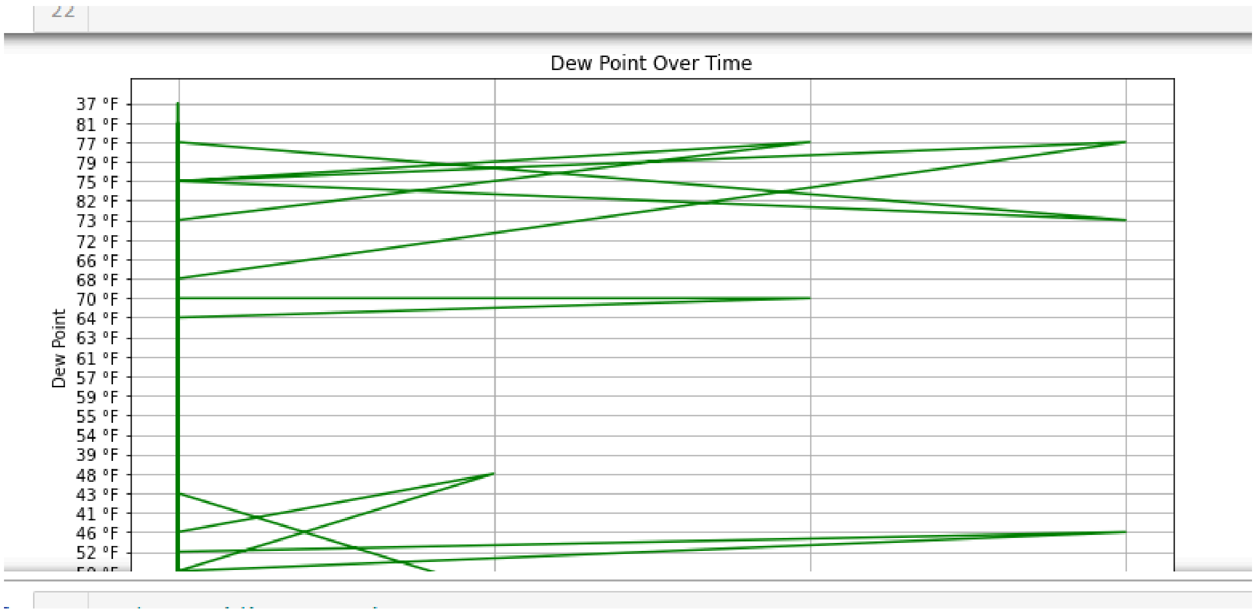
Interpretation:

The graph allows us to analyze trends, patterns, and fluctuations in temperature.

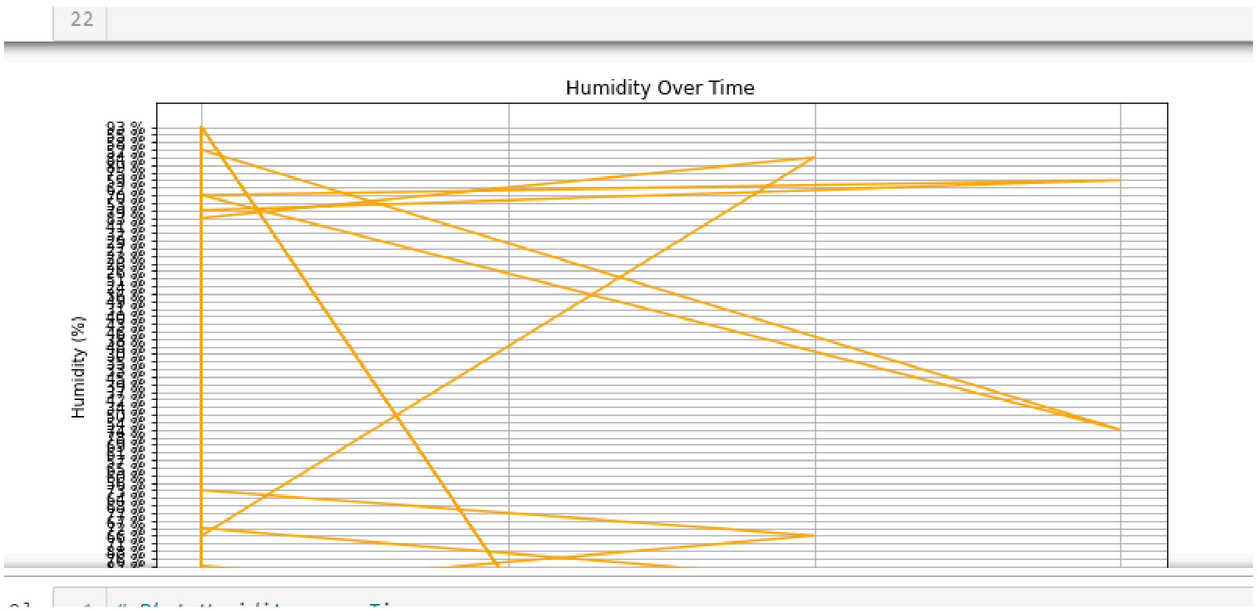
We can observe any sudden changes, daily cycles, or seasonal variations.

Further analysis could explore correlations with other factors (e.g., humidity, wind speed).

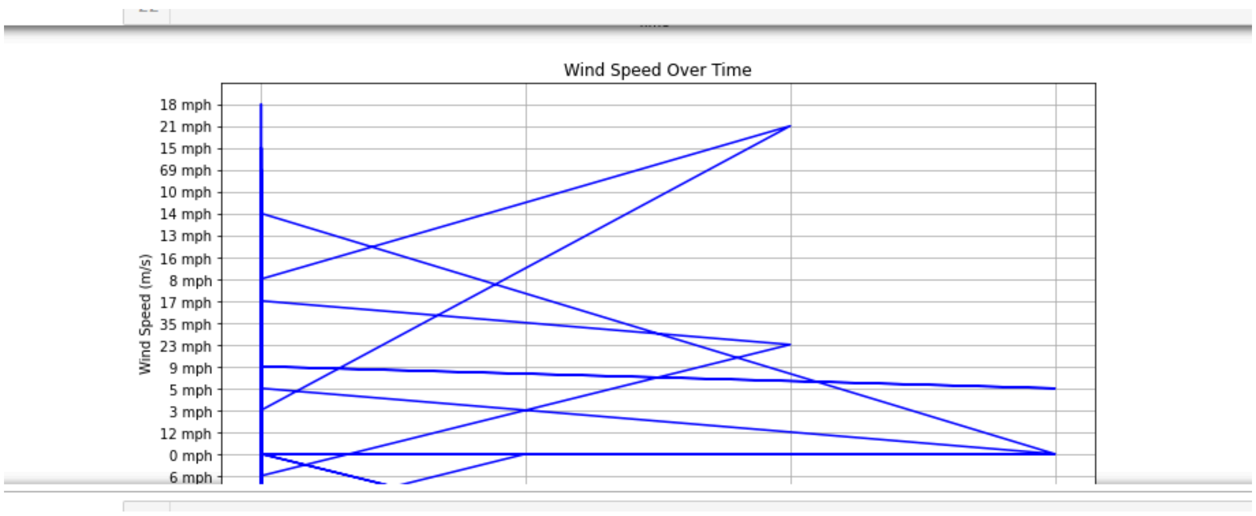
2)



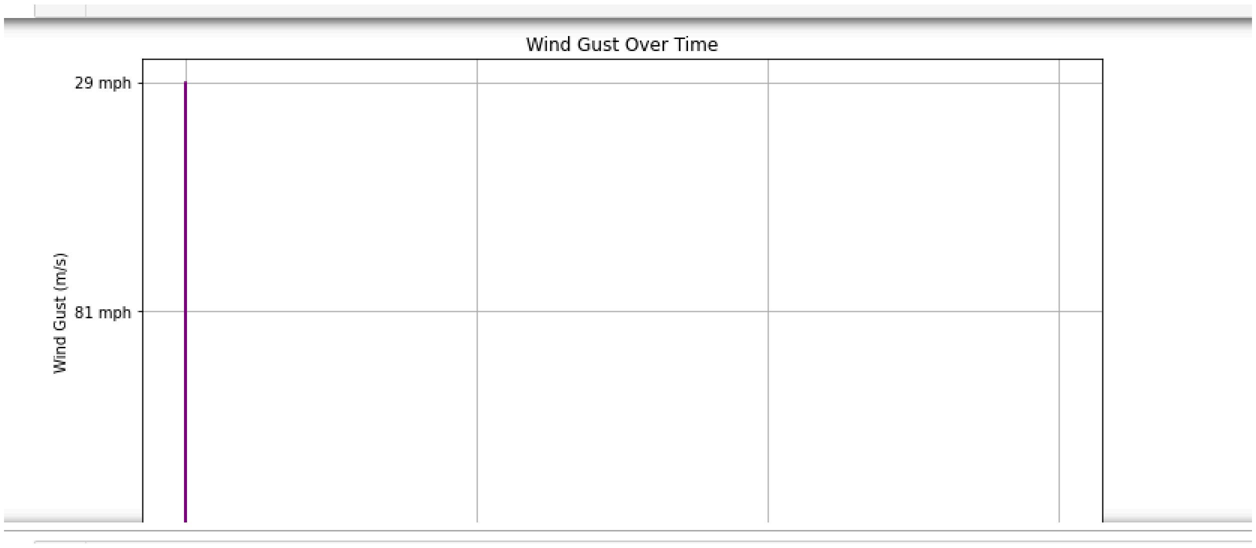
3)


Humidity Over Time
Humidity (%)
93 %

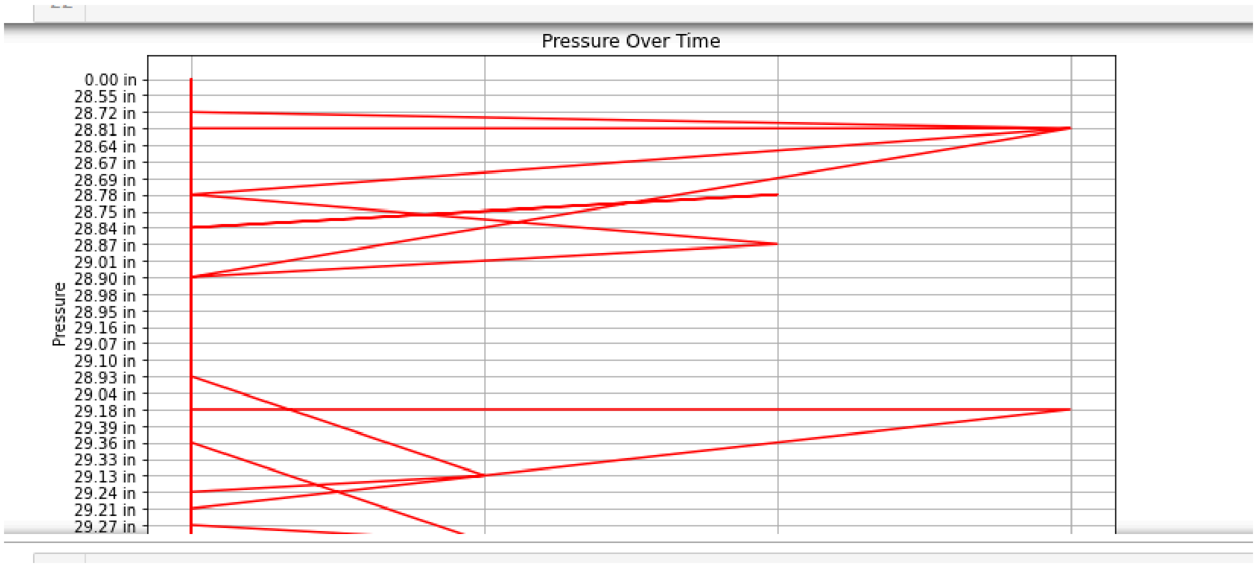
4)


Wind Speed Over Time
Wind Speed (m/s)
18 mph
21 mph
15 mph
69 mph
10 mph
14 mph
13 mph
16 mph
8 mph
17 mph
35 mph
23 mph
9 mph
5 mph
3 mph
12 mph
0 mph
6 mph

5)



6)



From 2 to 6 pictures

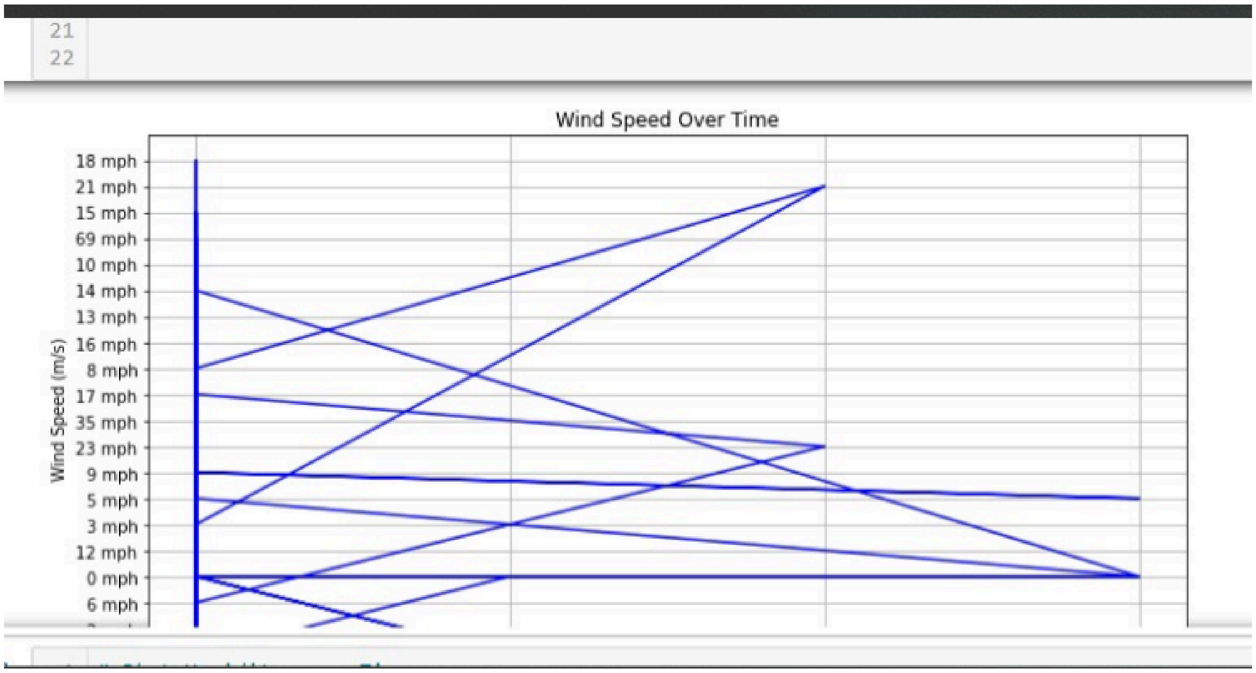
This Python script segment visualizes the variation of dew point over time using a line

plot. It operates on a DataFrame named `dataframe_2022`, which should contain at least two columns: 'Time' representing time points and 'Dew Point' representing corresponding dew point values.

The script first converts the 'Time' column to string type to ensure proper formatting for plotting. Then, it creates a line plot with dew point values plotted against time values. The plot is customized with labels for the x-axis ('Time') and y-axis ('Dew Point'), a title ('Dew Point Over Time'), rotated x-axis tick labels for better readability, grid lines for clarity, and an optimized layout.

Finally, the plot is displayed using `plt.show()`.

This script segment provides a straightforward approach to visualize the trend of dew point over time, aiding in the analysis and interpretation of weather data.



Convert the ‘Time’ column to a string representation to ensure proper formatting for plotting.

3. Plotting Wind Speed

The code creates a line plot with the x-axis representing time and the y-axis representing wind speed values.

The wind speed data from the DataFrame (dataframe_2022['Wind Speed']) is plotted in blue.

X-axis labels are rotated for better readability.

Grid lines are added for visual clarity.

Cleaning and Preprocessing:

- Remove text within cells: Use text processing techniques like regular expressions to remove unwanted text from cells.
- Convert necessary categorical columns to numeric: Utilize techniques like one-hot encoding or label encoding to convert categorical variables into numerical format.
- Remove outliers: Identify and remove outliers using statistical methods such as z-score, IQR, or domain-specific knowledge.
- Deal with missing values: Impute missing values using techniques such as mean imputation, median imputation, or predictive imputation.
- Deal with duplicate values: Remove duplicate rows to ensure data integrity.
- Ensure proper formatting & consistency in date and time columns: Standardize date and time formats across different dataframes.

Temporal Features Extraction:

- Extract day of the week, month, and year from the date: Utilize datetime functions to extract temporal features from date columns.
- Combine date and time into a single datetime feature: Merge date and time columns into a single datetime column for better analysis.

Data Visualization:

- Visualize data using libraries like Matplotlib, Seaborn, or Plotly to gain insights into relationships between variables, trends over time, etc.
- Analyze relationships between temporal features (date, time) and other variables like temperature.

Machine Learning Model:

- Split the combined dataset into training and testing sets using techniques like train-test split or cross-validation.

- Choose an appropriate machine learning model based on the nature of your data (e.g., linear regression, random forest, etc.).
- Train the model on the training set and evaluate its performance on the testing set using appropriate evaluation metrics.
- Tune hyperparameters of the model using techniques like grid search or random search to improve performance if necessary.