Clustering and Fitting

Name: Muhammad Umer Asif
Student Id: 23098286
GitHub link: https://github.com/UmerAsif435/Umerasif.git

Introduction

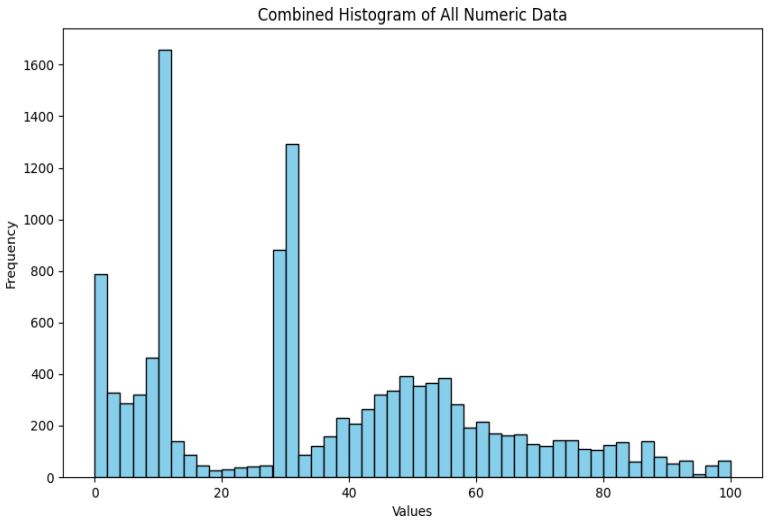
In this project, we studied a weather dataset with important details like temperature, humidity, pressure, visibility, and wind speed. The goal was to find patterns and connections in the data using simple tools and methods. We used charts like histograms, scatter plots, and heatmaps to look at the data visually. We also grouped similar weather conditions using k-means clustering and found how weather features like humidity, pressure, and temperature are related using linear regression. These methods helped us understand how different weather factors interact with each other.

Data Visualization

1. Histogram for All Numeric Columns

The first chart is a combined histogram showing how the numbers in the dataset are spread out.

- **Purpose**: It shows how often different values appear for features like temperature, pressure, and visibility.
- **What we noticed**:
    - Most weather features have a normal or concentrated spread.
    - For example, temperature has a clear peak, showing common temperature patterns, while visibility and pressure vary more.
    - This chart helps us spot unusual values, uneven spreads, and overall differences in the data.

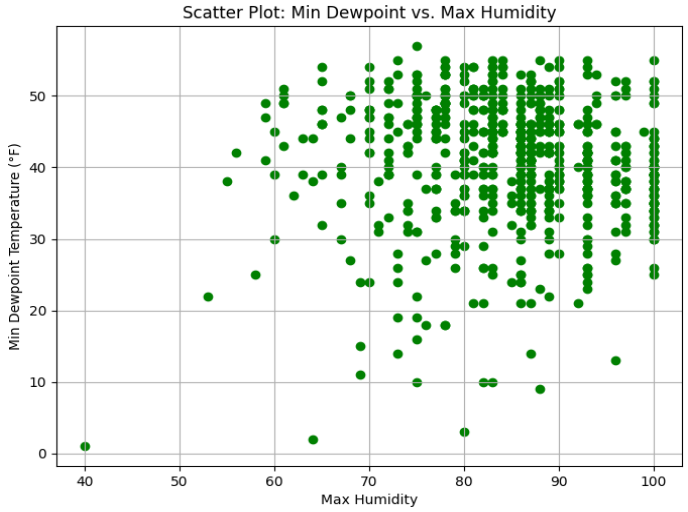


Graph 1: The histogram provides a clear view of the data spread and density across different weather-related numeric features.

2. Scatter Plot: Min Dewpoint vs. Max Humidity

We used a scatter plot to look at the connection between Min Dewpoint Temperature and Max Humidity.

- **Purpose**: To see how humidity affects dewpoint temperature.
- **What we noticed**:
    - There is a strong link: higher humidity usually means a higher dewpoint temperature.
    - This shows that more moisture in the air raises the dewpoint.
    - Understanding this helps predict weather changes and how moisture behaves in the atmosphere.

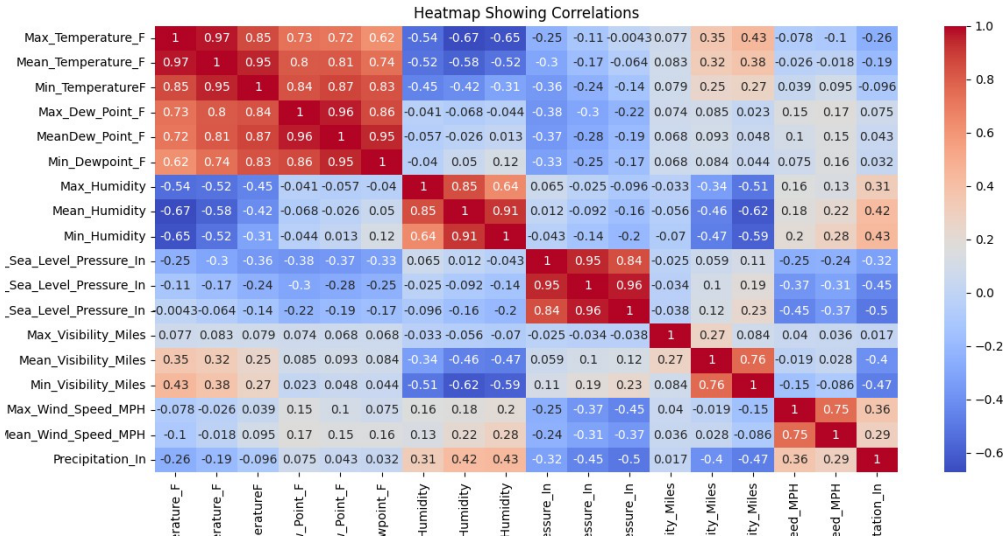


Graph 2: The scatter plot shows the correlation between **humidity and dewpoint temperature**, emphasizing their dependency.

3. Heatmap Showing Correlations

The third chart is a heatmap showing how weather features like temperature, pressure, visibility, and wind speed are connected.

- **Purpose**: To find out how strongly these features are related to each other.
- **What we noticed**:
    - The colors show positive or negative connections, with cool and warm colors representing the strength of the relationship.
    - For example, Max Temperature and Sea-Level Pressure are strongly linked, showing a connection between temperature and air pressure.
    - These insights help us understand how different weather conditions affect each other.



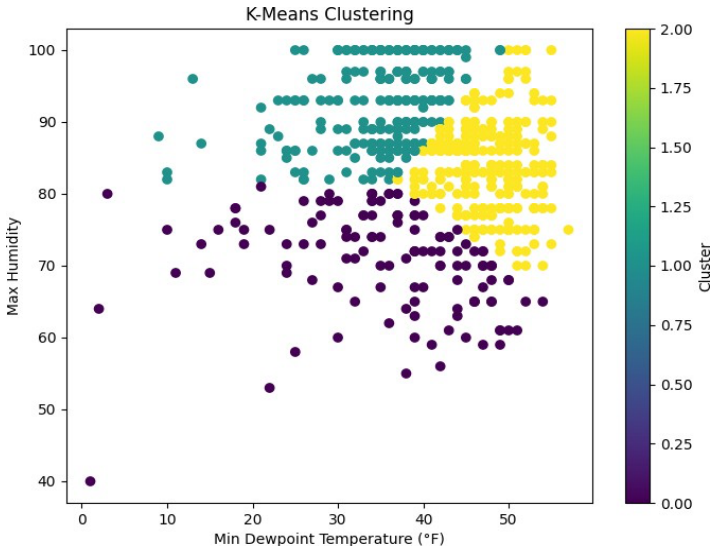
Graph 3: The heatmap effectively highlights correlations among various atmospheric parameters, aiding meteorological analysis.

Clustering and Regression Techniques

1. K-Means Clustering

We used k-means clustering to group the data based on Min Dewpoint, Max Humidity, and Sea-Level Pressure.

- **Purpose**: To find groups with similar weather patterns.
- **What we noticed**:
  - The k-means method created three groups, each representing areas with different weather conditions.
  - Each group shows a unique mix of temperature, humidity, and pressure.
  - This helps us understand how weather varies by region and find similar weather patterns.

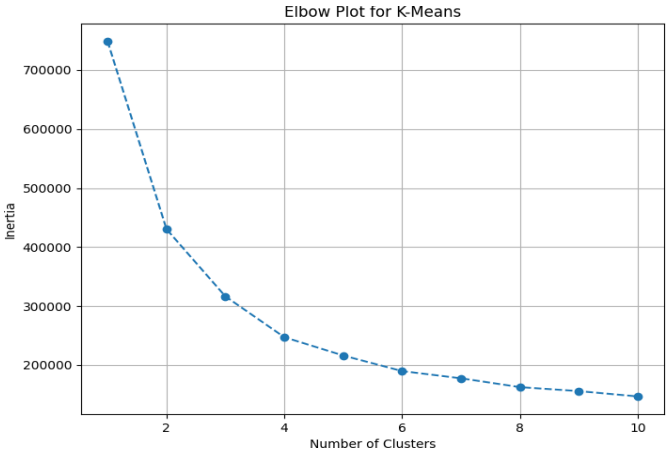


Graph 4: The scatter plot shows clusters of data points, each representing a distinct group with specific atmospheric conditions.

2. Elbow Plot for K-Means

We made an elbow plot to decide the best number of groups for k-means clustering.

- **Purpose**: To see how the total distances within groups change with different numbers of clusters.
- **What we noticed**:
    - The "elbow point" appears at three clusters, meaning three groups give a good mix of accuracy and useful insights.
    - This method helps make sure the k-means model is clear and reliable.

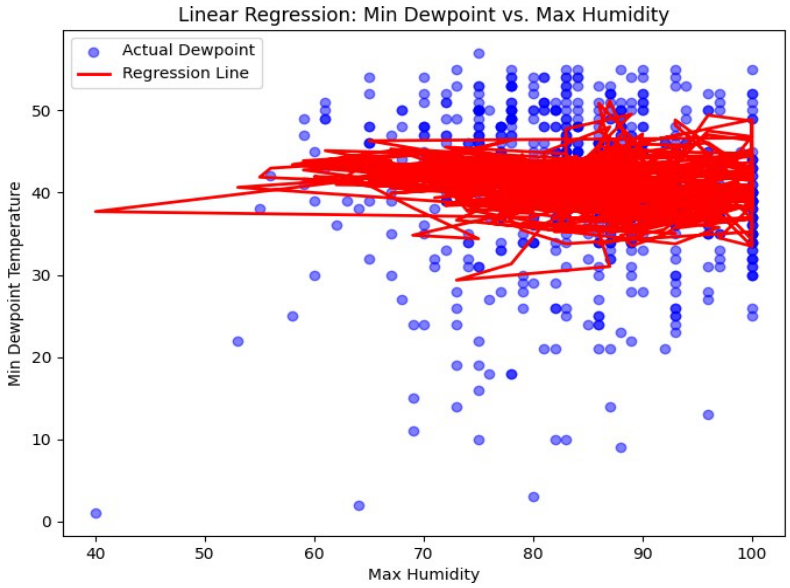


Graph 5: The elbow plot demonstrates the point where the number of clusters is optimal, balancing clustering accuracy and complexity.

3. Linear Regression Fitting

We implemented a linear regression model to predict Min Dewpoint Temperature using Max Humidity and Sea-Level Pressure.

- Purpose: To observe and quantify the relationship between humidity, pressure, and temperature.
- Observation: The scatter plot shows actual dewpoint temperatures along with the regression line.
- This visualization reveals the predictive relationship, showing that higher humidity and pressure correlate with changes in the dewpoint temperature.
- Such models are crucial in forecasting and predicting local weather trends



Graph 6: The plot includes the actual data points and the regression line, highlighting the relationship between **humidity, pressure, and dewpoint temperature**.

Conclusion

We used different methods to study weather data and found some key patterns. Histograms showed how weather numbers are spread out, and scatter plots showed a strong link between humidity and dewpoint temperature.

Heatmaps helped us see how weather features are connected. K-means clustering grouped the data into regional weather patterns, and the elbow plot helped us pick the best number of groups. Linear regression showed how humidity, pressure, and dewpoint temperature are related.

These findings help us better understand how weather works, identify regional patterns, and improve weather forecasts and climate studies.