Weather Patterns Analysis and Prediction

A presentation by Aditya L.N.

OBJECTIVES:

- Perform exploratory data analysis (EDA).
- Implement K-Nearest Neighbors (KNN) classification for weather prediction.
- Apply K-Means clustering to identify natural weather pattern groupings.

DATABASE OVERVIEW:

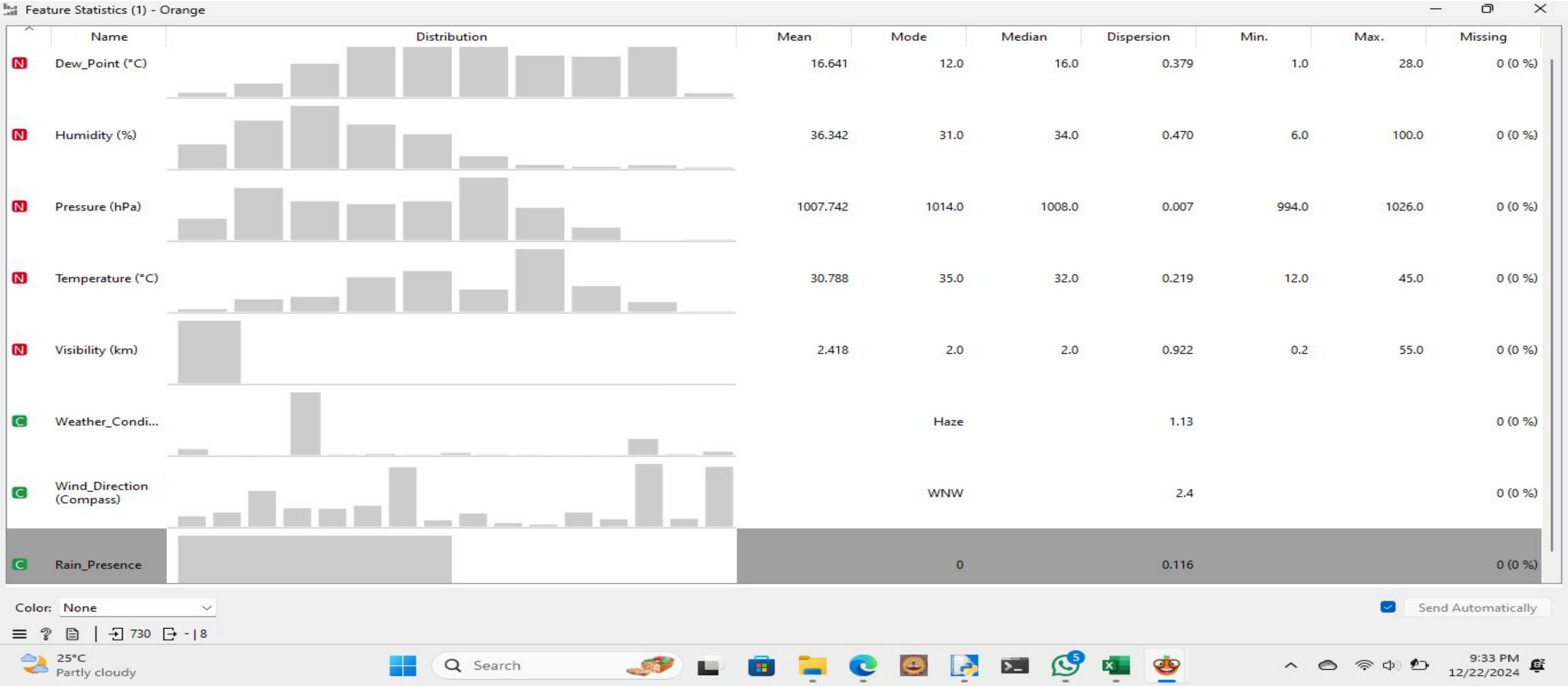
The dataset consists of 731 records of Bangalore weather with the following features included:

- **Date:** Date of the record.
- **Weather Condition:** Weather condition on that day (e.g., Smoke, Light Rain, Mist, Haze).
- **Dew Point (°C):** Dew point temperature.
- **Humidity (%):** Atmospheric humidity percentage.
- **Pressure (hPa):** Atmospheric pressure.
- **Temperature (°C):** Recorded temperature.
- **Visibility (km):** Visibility in kilometers.
- **Wind Direction:** Compass direction of wind (e.g., NNW, NE).
- **Rain Presence:** Binary value indicating rain presence (1 = Yes, 0 = No).

TOOLS AND TECHNIQUES:

- **Tools :**
  - Orange data management software
  - Google Sheets
  - Google Slides
  - Python (IDLE)
- **Algorithms :**
  - K-Nearest Neighbors (KNN)
  - K-Means Clustering

Exploratory Data Analysis (EDA) (using orange):



Name	Distribution	Mean	Mode	Median	Dispersion	Min.	Max.	Missing
Dew_Point (°C)		16.641	12.0	16.0	0.379	1.0	28.0	0 (0 %)
Humidity (%)		36.342	31.0	34.0	0.470	6.0	100.0	0 (0 %)
Pressure (hPa)		1007.742	1014.0	1008.0	0.007	994.0	1026.0	0 (0 %)
Temperature (°C)		30.788	35.0	32.0	0.219	12.0	45.0	0 (0 %)
Visibility (km)		2.418	2.0	2.0	0.922	0.2	55.0	0 (0 %)
Weather_Condi...			Haze		1.13			0 (0 %)
Wind_Direction (Compass)			WNW		2.4			0 (0 %)
Rain_Presence			0		0.116			0 (0 %)

Methodology (KNN):

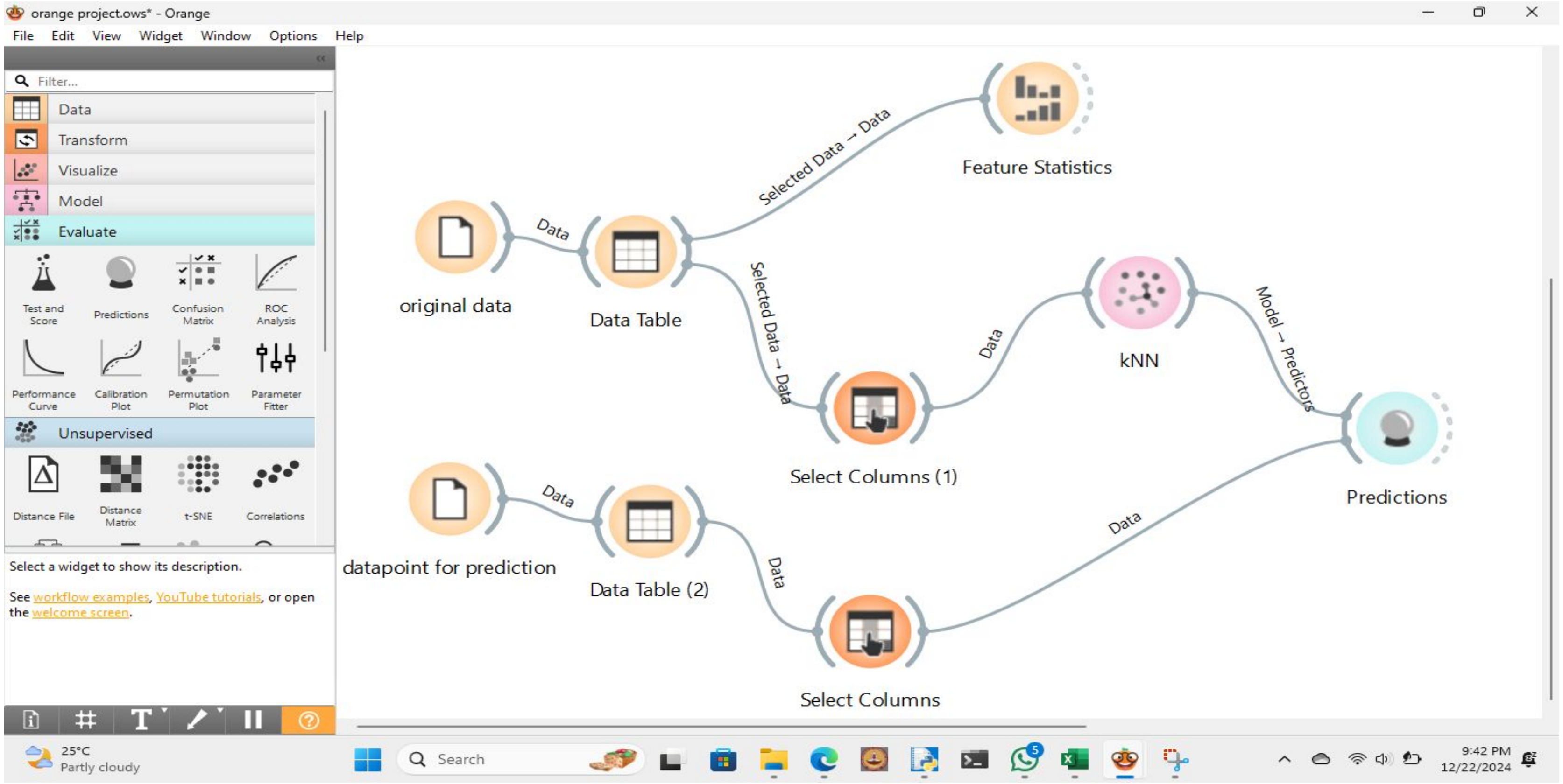
- **Tool Used:** Orange data mining software.
- **Workflow:**
    - Set the value of K = 2.
    - Implemented the KNN workflow in Orange to classify weather conditions.

Methodology (KMeans)

- **One** Python and one Orange version was made.
- **Process:**
  1. Selected initial means using records from 01-Jan-2015 and 02-Jan-2015.
  2. Generated a distance matrix to compute distances from each mean.
  3. Created an assignment list to group data points.
  4. Recalculated means based on new groupings.
  5. Repeated the process using a custom Python function `kmeans()`.

Orange Kmeans workflow:

Orange K-NN workflow:



Results:

KNN Results:

- Successfully classified a test data points with 83% accuracy.

K-Means Results:

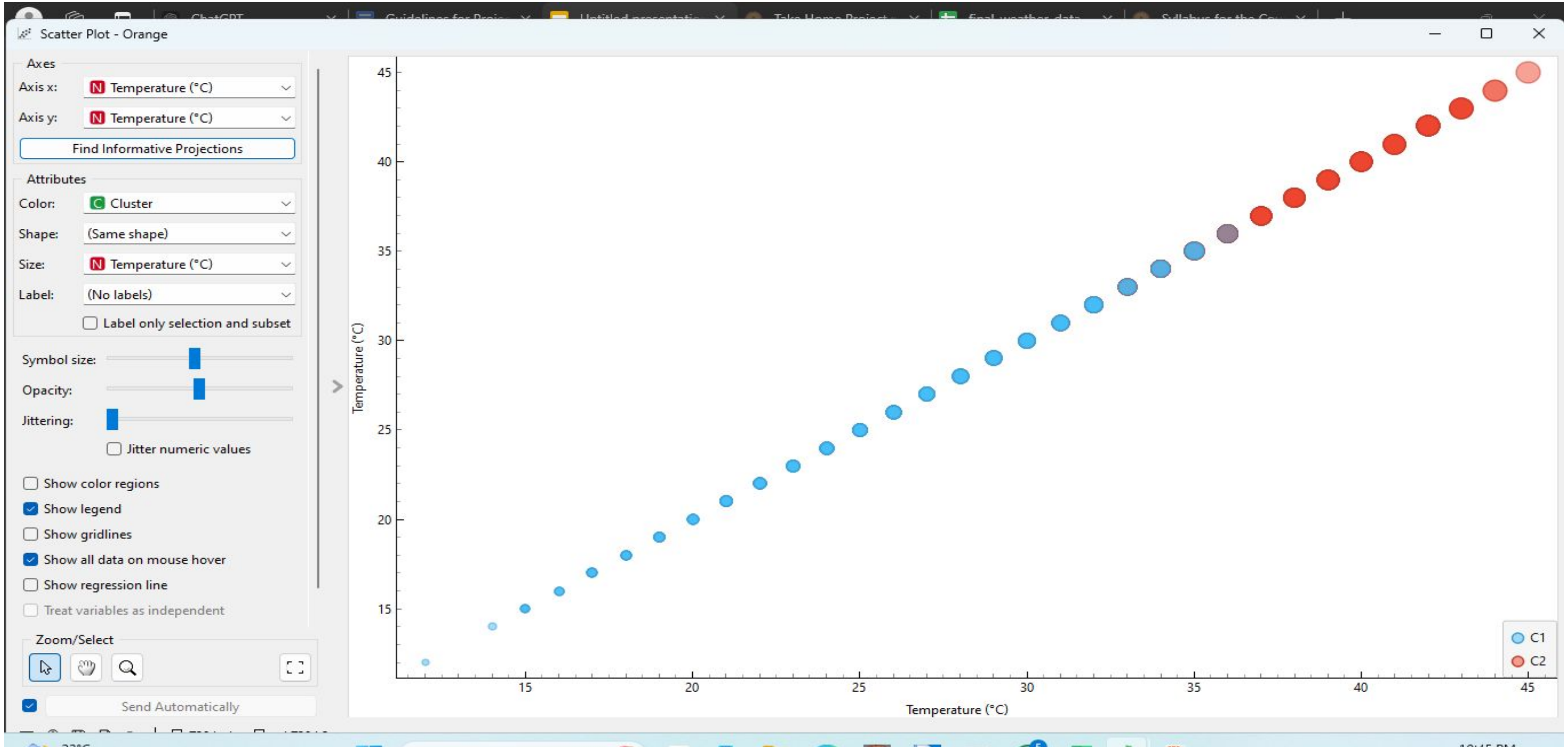
- Output highlighted natural groupings:
    - **Cluster 1:** Colder days with higher atmospheric pressure.
    - **Cluster 2:** Hotter days with lower atmospheric pressure.

Insights and learnings:

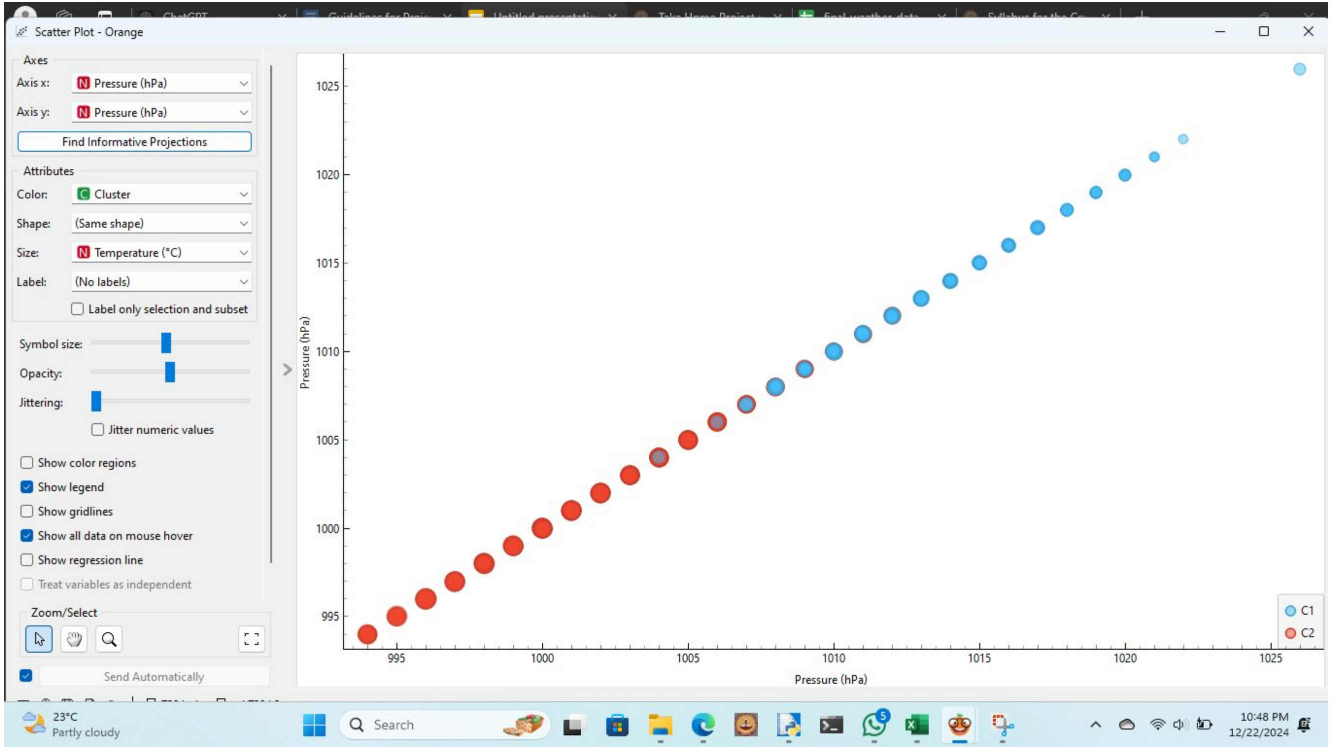
- EDA revealed significant relationships between weather variables.
- KNN effectively classified unknown data points but provided limited insights beyond classification.
- K-Means clustering demonstrated:
    - Clear separation of hot and cold days.
    - Inverse correlation between temperature and pressure.

Following are the relevant graphs

It appears our most natural grouping has resulted in a split of hotter and colder days into 2 clusters



We notice days with higher pressure are more in the low temp group



Challenges and recommendations for future projects

- Orange software was unreliable due to glitches and inflexibility.
- Python provided a more robust and versatile solution for data processing and algorithm implementation.

Recommendation: Future projects should prioritize Python over Orange for greater control and reliability.

conclusion

1. This project successfully applied KNN and K-Means algorithms to analyze and classify weather data.
2. Key findings:
   a. Clear clustering of hot and cold days.
   b. Observed inverse relation between atmospheric pressure and temperature.

Implications:

1. Perhaps the natural division between hot and cold days in our source suggests a KNN algorithm to classify days as hot and cold will be quite accurate. Perhaps BBMP may utilize such a program for sending heat warnings on hot days or deploying cover on any dams to save water on hot days
2. The inverse relation between pressure and temperature is of note, perhaps a physicist may be inclined to investigate this violation of Gay-Lussac's law

Thank you!