A Dashboard for Exploratory Climate Data Analysis and Visualization

Dristi Kafle Manasvi Sharma

***Abstract*—This project presents an interactive dashboard designed to explore and visualize meteorological datasets. Built using Python and Streamlit, the system enables users to upload structured CSV datasets containing region-wise time-series climate data. The dashboard supports time-series trends, feature distribution plots, clustering, anomaly detection and geospatial mapping. The dashboard incorporates customizable feature selection and is intended for researchers, students, and analysts working with environmental data across different geographic scopes.**

***Index Terms*—Streamlit, Climate Visualization, Meteorological Data Mining, Dashboard.**

I. INTRODUCTION

Understanding climate trends across regions is crucial in the face of increasing environmental uncertainty. Climate change research often involves interpreting large-scale, multi-dimensional datasets that span various regions and timeframes. To address such complex data analysis, we developed a web-based interactive dashboard using Streamlit, enabling exploratory data analysis (EDA) for meteorological datasets in a user-friendly way. Originally built around Nepal's district level data, the system was expanded to support datasets from various regions.

II. DATA SOURCES AND STRUCTURE

A. *Primary Dataset*

The initial dataset used for development was the publicly available "District-wise Monthly Climate Data for Nepal (1981-2019)", which has over 14,000 records from 62 districts. The key features include date, year, month, district, precipitation, specific and relative humidity at 2 meters, temperature at 2 meters (maximum, minimum, range), wind speed at 10 and 50 meters (maximum, minimum, range), earth skin temperature, surface pressure, wet bulb temperature at 2 meters. Among these, month, date and district are categorical attributes whereas the remaining are numerical attributes.

Since the original dataset lacked complete records for several districts, missing data was collected from NASA's Power API to ensure full nation-wide geographic coverage over 1981-2019.

B. *Generalization*

To extend the functionality beyond Nepal, the dashboard supports: dynamic file uploads for any region-level climate data, automatic column name preprocessing and validation, and flexible aggregation and visualization filters based on file contents.

III. SYSTEM DESIGN

A. *Technologies Used*

- Python: Core programming language
- Streamlit: Front-end interface for dashboard
- Pandas and NumPy: Data preprocessing and transformation
- Altair, Matplotlib, Seaborn,Plotly: Visualization libraries
- Scikit-learn: Clustering and anomaly detection algorithms

B. *Data Preprocessing*

1. Column Cleanup
   All column names are stripped of leading/ trailing whitespaces using:
   ```
   df.columns = df.columns.str.strip()
   ```
   This prevents mismatches due to inconsistent column labels.
2. Date Handling
   The system dynamically detects and parses the date column:
   ```
   date_col = next((col for col in df.columns
   if col.lower() == 'date'), None)
   df[date_col] = pd.to_datetime(df[date_col],
   errors='coerce')
   ```
   This ensures consistent date formatting across datasets.
3. Time Period Aggregation
   Climate data can be aggregated by daily, monthly or yearly frequency:
   ```
   filtered_df['Period'] =
   filtered_df[date_col].dt.to_period('M').ast
   ype(str)
   ```
   This enables grouped analysis across different timescales.This enables grouped analysis across different timescales.

C. *Page Architecture*

The app consists of four integrated modules:

1. Main Page (app.py)
   - File upload
   - Navigation via radio buttons

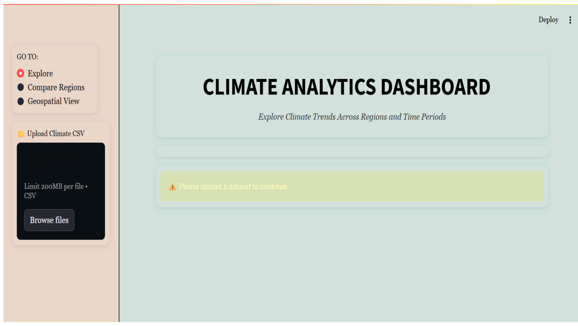


Fig. 1. Main page interface with navigation and file upload

2. Dashboard View-Single District (dashboard.py)
   - Time-series charts (line, area)
   - Box plots, histograms, KDE plots
   - Correlation matrix and pairplot
   - Clustering and anomaly detection

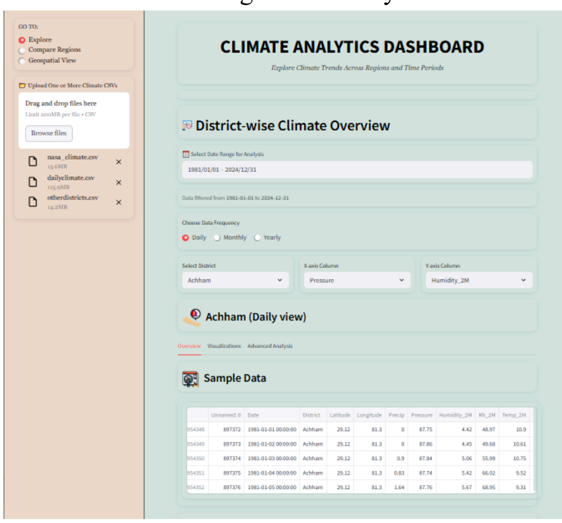


Fig. 2. District-wise dashboard with filters and charts

3. Comparison View-Two Districts (compare.py)
   - Side-by-side visualization of two districts
   - Area/line/bar chart comparisons
   - KMeans clustering and silhouette scoring

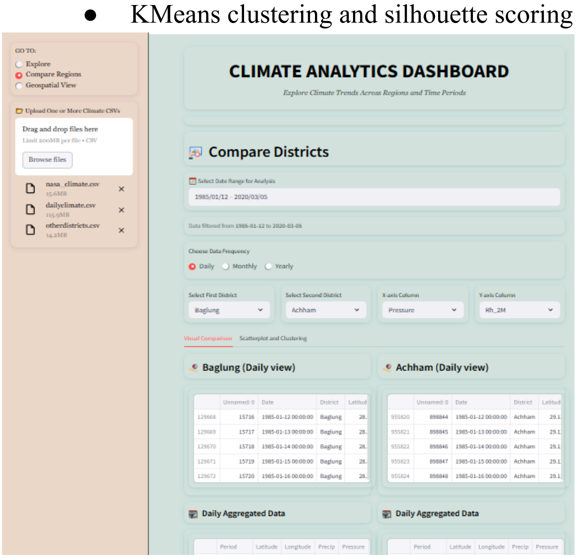


Fig. 3. Districts comparison page

4. Geospatial View (map.py)
   - Choropleth map of Nepal using GeoJSON
   - Filter by a specific year, any numeric climate feature, color scale
   - Highlight maximum and minimum values

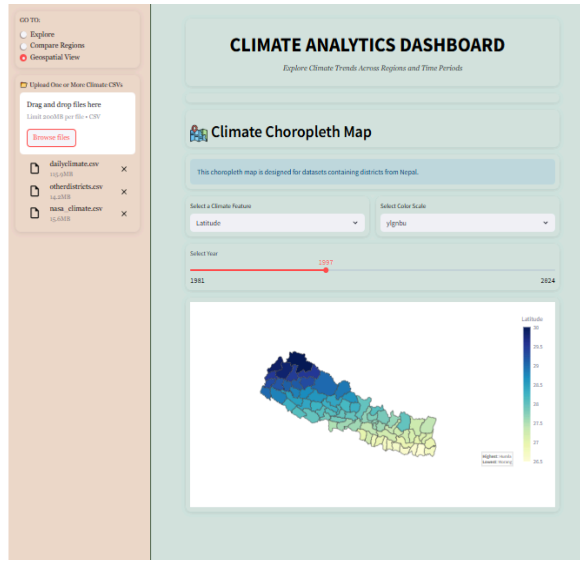


Fig. 4. Choropleth map view showing spatial climate variations across Nepal

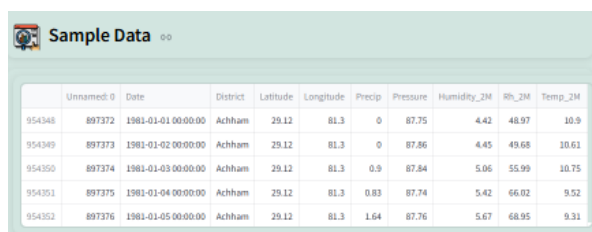
IV. VISUALIZATION AND ANALYSIS

A. *Single District Analysis*

This module enables users to explore the climate behavior of a selected district over time using multiple visual tools.

1. Sample Data Table

A preview of the dataset is shown after filtering for the selected district. This provides immediate verification of the dataset's structure and contents.

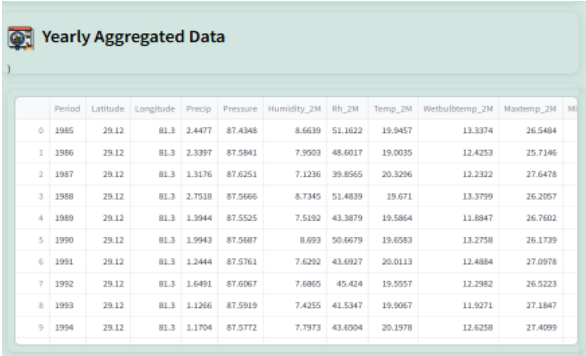

Sample Data

	Unnamed: 0	Date	District	Latitude	Longitude	Precip	Pressure	Humidity_2M	Rh_2M	Temp_2M
954348	897372	1981-01-01 00:00:00	Achham	29.12	81.3	0	87.75	4.42	48.97	10.9
954349	897373	1981-01-02 00:00:00	Achham	29.12	81.3	0	87.86	4.45	49.68	10.61
954350	897374	1981-01-03 00:00:00	Achham	29.12	81.3	0.9	87.84	5.06	55.99	10.75
954351	897375	1981-01-04 00:00:00	Achham	29.12	81.3	0.83	87.74	5.42	66.02	9.52
954352	897376	1981-01-05 00:00:00	Achham	29.12	81.3	1.64	87.76	5.67	68.95	9.31

Fig. 5. Sample Data Table for Achham District

2. Aggregated Data Table

This table presents daily, monthly or yearly aggregated climate statistics for the selected district.


Yearly Aggregated Data

	Period	Latitude	Longitude	Precip	Pressure	Humidity_2M	Rh_2M	Temp_2M	Wetbulbtemp_2M	Maxtemp_2M
0	1985	29.12	81.3	2.4477	87.4348	8.6639	51.1622	19.9457	13.3374	26.5484
1	1986	29.12	81.3	2.3397	87.5841	7.9503	48.6017	19.0035	12.4253	25.7146
2	1987	29.12	81.3	1.3176	87.6251	7.1236	39.8565	20.3296	12.2322	27.6478
3	1988	29.12	81.3	2.7518	87.5666	8.7345	51.4839	19.671	13.3799	26.2057
4	1989	29.12	81.3	1.3944	87.5525	7.5192	43.3879	19.5864	11.8847	26.7602
5	1990	29.12	81.3	1.9943	87.5687	8.693	50.6679	19.6583	13.2758	26.1739
6	1991	29.12	81.3	1.2444	87.5761	7.6292	43.6927	20.0113	12.4884	27.0978
7	1992	29.12	81.3	1.6491	87.6067	7.6865	45.424	19.5557	12.2982	26.5223
8	1993	29.12	81.3	1.1266	87.5919	7.4255	41.5347	19.9067	11.9271	27.1847
9	1994	29.12	81.3	1.1704	87.5772	7.7973	43.6504	20.1978	12.6258	27.4099

Fig. 6. Yearly Aggregated Data Table

3. Feature Statistics

Descriptive statistics such as mean, standard deviation, min/max values, and quartiles are displayed for all numeric weather parameters.


Feature Statistics

	count	mean	std	min	25%	50%	75%	max
Humidity_2M	11326	8.5619	5.3822	0.98	4.02	6.42	14.24	20.67
Rh_2M	11326	50.119	22.9876	5.57	31.07	45.835	71.57	97.28
Temp_2M	11326	19.6637	5.8452	3.4	14.43	20.96	24.25	32.3
Wetbulbtemp_2M	11326	13.2041	7.0281	-3.55	7.0525	13.155	20.47	24.56
Maxtemp_2M	11326	26.2639	5.5898	5.8	22.04	26.39	30.2275	40.25
Mintemp_2M	11326	14.104	6.0389	-3.87	8.6	14.58	19.91	25.27
Temprange_2M	11326	12.1599	3.813	2.28	9.23	12.43	14.9675	23.95
Earthskintemp	11326	18.8534	7.148	-0.03	12.1625	20.485	24.6775	33.39
Windspeed_10M	11326	2.205	0.424	0.79	1.9	2.16	2.5	4.68
Maxwindspeed_10M	11326	4.2473	0.9016	1.31	3.63	4.2	4.84	8.71

Fig. 7. Feature Statistics Table

4. Line and Area Chart

The dashboard allows the user to switch between a line chart and an area chart. Area charts help highlight cumulative patterns, while line charts emphasize fluctuations.

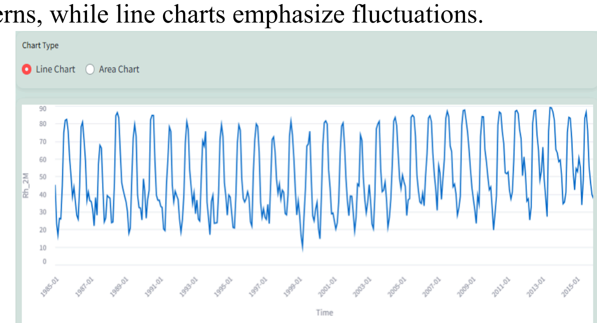


Fig. 8. Line chart for Relative Humidity at 2 meters vs Time

5. Scatter Plot

A scatter plot is used to examine the relationship between any two numeric parameters.

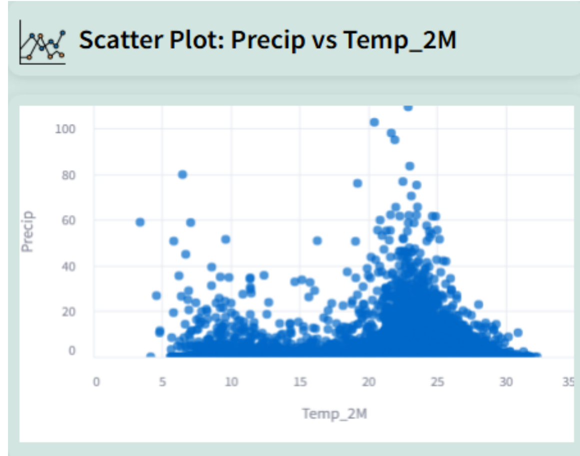


Fig. 9. Scatter Plot for Precipitation vs Temperature at 2 meters

6. Box Plot

A box plot displays distribution of a selected feature and highlights median, interquartile range, and potential outliers at a glance.

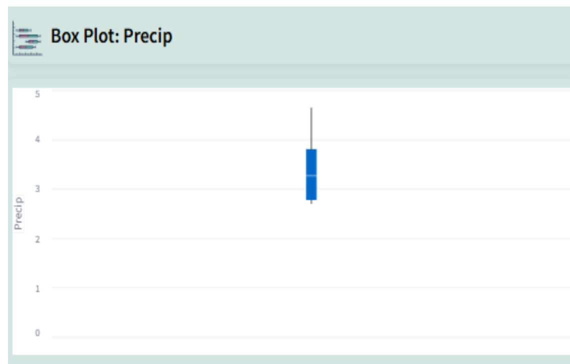


Fig. 10. Box Plot showing distribution of Precipitation

7. Distribution and KDE Plot

To better understand the frequency distribution of selected climate features, histograms and kernel density estimation (KDE) plots are presented side-by-side.

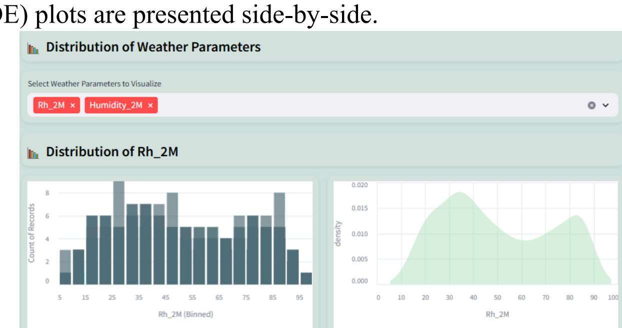


Fig. 11. Distribution Plot for Relative Humidity at 2 meters

8. Correlation Matrix

A color-coded correlation matrix is generated to measure the pairwise correlation between numerical variables. This can help identify variables that may be redundant or strongly related.

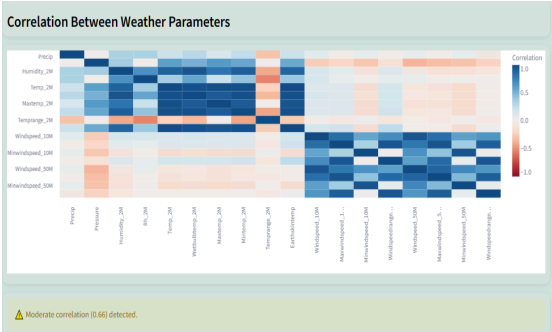


Fig. 12. Correlation Matrix

9. Pairplot

Pairplots are generated for illustrating pairwise relationships and clustering potential between multiple features.

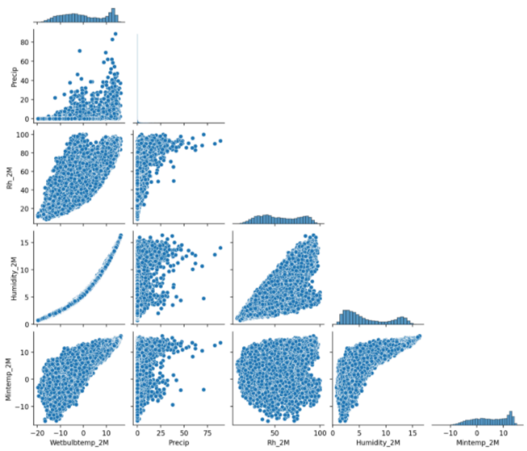


Fig. 13. Pairplot for features like Wet Bulb Temperature at 2 meters, Precipitation, Relative Humidity at 2 meters, Specific Humidity at 2 meters, Minimum Temperature at 2 meters

10. Clustering Analysis

KMeans clustering is applied on selected numeric features to group similar data points within the district. Users can control the number of clusters and visualize the results in an interactive scatter plot, selecting any two features as axes.

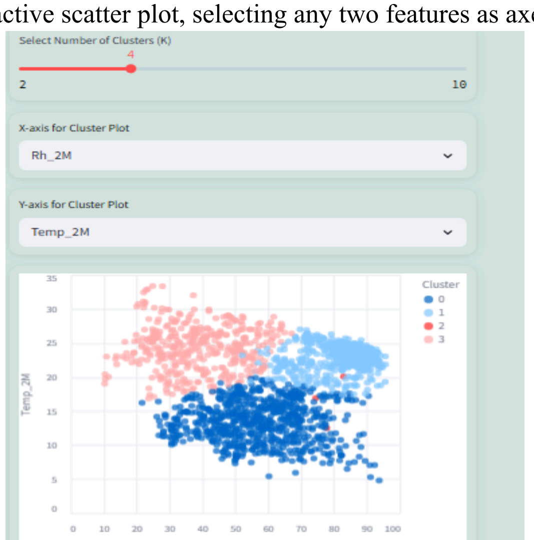


Fig. 14.Clustering Analysis for Relative humidity at 2 meters vs Temperature at 2 meters

11. Anomaly Detection

Anomaly detection uses Z-score statistics to highlight unusual observations. Detected outliers are marked in red over the original time-series plot.

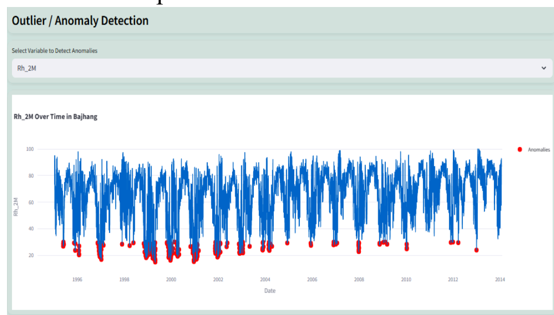
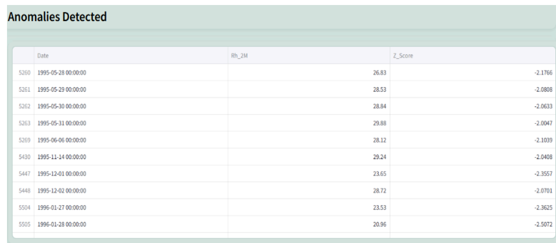


Fig. 15. Anomaly Detection for Relative Humidity at 2 meters in Bajhang district



Anomalies Detected

	Date	Rh_2M	Z_Score
5260	1995-05-28 00:00:00	26.83	-2.1766
5261	1995-05-29 00:00:00	28.53	-2.0808
5262	1995-05-30 00:00:00	28.84	-2.0633
5263	1995-05-31 00:00:00	29.88	-2.0047
5269	1995-06-06 00:00:00	28.12	-2.1039
5430	1995-11-14 00:00:00	29.24	-2.0408
5447	1995-12-01 00:00:00	23.65	-2.3557
5448	1995-12-02 00:00:00	28.72	-2.0701
5504	1996-01-27 00:00:00	23.53	-2.3625
5505	1996-01-28 00:00:00	20.96	-2.5072

Fig. 16. Detected Anomalies (Z-score Method)

12. Bar Chart

The bar chart displays the average value of the selected climate feature across aggregated time periods. It provides a clear comparative view of trends within the selected district.

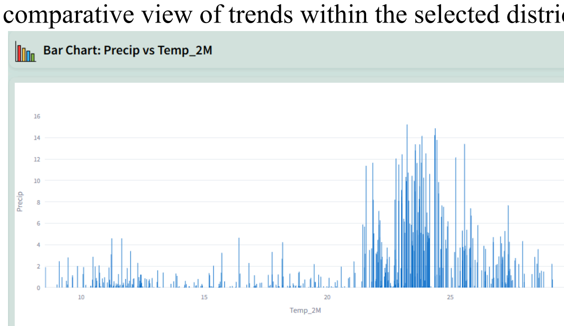
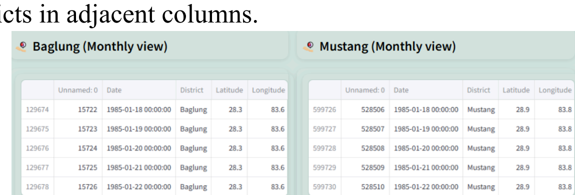


Fig. 17. Bar Chart showing Precipitation levels with respect to Temperature at 2 meters

B. District Comparison

1. Sample Data Tables

The module displays filtered sample data from both selected districts in adjacent columns.



Baglung (Monthly view)

	Unnamed: 0	Date	District	Latitude	Longitude
129674	15722	1985-01-18 00:00:00	Baglung	28.3	83.6
129675	15723	1985-01-19 00:00:00	Baglung	28.3	83.6
129676	15724	1985-01-20 00:00:00	Baglung	28.3	83.6
129677	15725	1985-01-21 00:00:00	Baglung	28.3	83.6
129678	15726	1985-01-22 00:00:00	Baglung	28.3	83.6

Mustang (Monthly view)

	Unnamed: 0	Date	District	Latitude	Longitude
599726	528506	1985-01-18 00:00:00	Mustang	28.9	83.8
599727	528507	1985-01-19 00:00:00	Mustang	28.9	83.8
599728	528508	1985-01-20 00:00:00	Mustang	28.9	83.8
599729	528509	1985-01-21 00:00:00	Mustang	28.9	83.8
599730	528510	1985-01-22 00:00:00	Mustang	28.9	83.8

Fig. 18. Side-by-side Sample Table for Baglung and Mustang Districts

2. Aggregated Comparison Table

It shows yearly or monthly mean values for a chosen variable in both districts side by side.

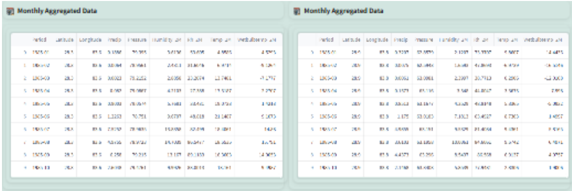


Fig. 19. Aggregated Comparison Tables for Baglung and Mustang Districts

3. Scatter Plot Comparison

A combined scatter plot overlays climate values from both districts on the same time axis. Each district is color-coded, enabling visual comparison of trends, peaks, and anomalies between the regions.

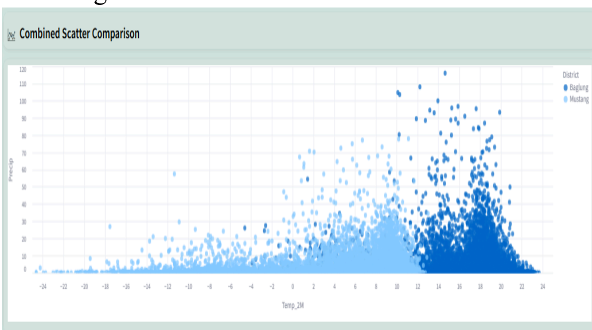


Fig. 20. Scatter Plot Comparison for Precipitation and Temperature at 2 meters of Baglung & Mustang

4. Clustering Analysis with Silhouette Score

The module supports clustering in two modes:

- Per-District Time-Series Clustering: Clusters weather patterns within each district based on selected features. Visualized using colored scatter plots for each region.
- All-District Aggregated Clustering: Clusters districts using average climate statistics, and evaluates cluster validity using silhouette score and silhouette plot. Higher scores indicate well-separated, dense clusters.

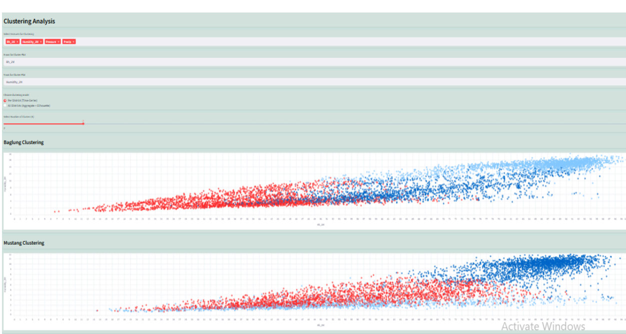


Fig. 21. Clustering Comparison for Baglung & Mustang

5. Bar Chart Comparison

This bar chart compares aggregated averages of the selected variable between the two districts.

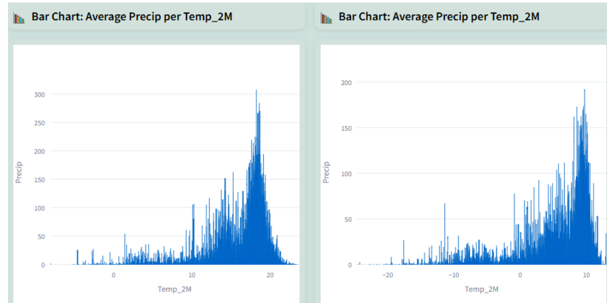


Fig. 22. Bar Chart Comparison for Precipitation and Temperature at 2 meters of Baglung & Mustang

6. Line and Area Chart Comparison

This dual-line or area chart compares the same climate variable across two selected districts over time, allowing users to identify similar or divergent temporal patterns.

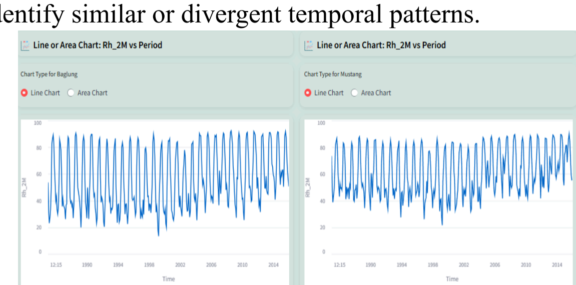


Fig. 23. Line Chart Comparison of Relative Humidity at 2 meters vs Time for Baglung & Mustang

C. Geospatial View

This module provides a choropleth map-based visualization of climate features across Nepal's districts. Key aspects include:

- Choropleth Mapping: Displays districts color-coded based on the average value of a selected climate feature (e.g., temperature, humidity). Interactive tooltips allow users to explore district-wise data.
- Year Filtering: Users can select any year in the dataset using a slider.
- Feature and Color Scale Selection: Users can dynamically choose any numeric variable and its color gradient from Plotly named color scales.
- Highlighting Extremes: Annotates the district with the highest and lowest values of the selected feature.
- District Name Validation: Built-in logic automatically corrects or excludes district names that don't match the GeoJSON boundary file, ensuring accurate visualization.
- Region Scope: The map is currently tailored for Nepal. Future improvements may allow dynamic loading of other countries' boundaries through user-uploaded GeoJSON files.

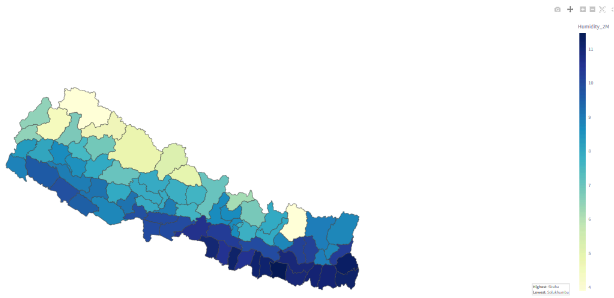


Fig. 24. Choropleth Map showing district-wise variation in Specific Humidity at 2 meters

V. CONCLUSION

The climate dashboard successfully delivers an interactive platform for performing exploratory climate data analysis. It supports both researchers and students by simplifying data visualization and analysis tasks using widely adopted Python libraries and frameworks. Its modular architecture ensures maintainability, while its dynamic data handling allows adaptation to various datasets and regions.

Future improvements may include dynamic map loading via user-uploaded GeoJSON, automated interpretation of detected anomalies, multi-format data upload, UI-based preprocessing tools for enhanced user experience, etc.

References

[1] *District wise climate data for Nepal - Open Data Nepal*. (n.d.). District Wise Climate Data for Nepal - Dataset

[2] *Mesaugat. (n.d.). GitHub - mesaugat/geoJSON-Nepal: GeoJSON for Nepal. 🇳🇵. GitHub.* https://github.com/mesaugat/geojson-nepal

[3] *NASA POWER Data Access Viewer (DAV). (n.d.). Prediction of Worldwide Energy Resources (POWER).* https://power.larc.nasa.gov/data-access-viewer/

[4] *Streamlit Docs. (n.d.).* https://docs.streamlit.io/