A Project Report

On

ClimateScope

Submitted by

Pujan Kumar Patel

In partial fulfillment of the requirements of the
Data Science Internship Program under Infosys Springboard

December 2025

List of Figures

List of Tables

Table of Content

Abstract

Climate variability and the increasing frequency of extreme weather events pose significant challenges to environmental monitoring and infrastructure resilience. The **ClimateScope** project presents a comprehensive data-driven web application designed to analyze, visualize, and interpret global climate data to support informed decision-making. The system integrates historical and near real-time weather data to examine key climatic parameters such as temperature, humidity, precipitation, and wind speed across multiple countries.

The project follows a milestone-based development approach encompassing data collection and preprocessing, exploratory data analysis, statistical evaluation, and interactive dashboard development using **Python, Pandas, Plotly, and Streamlit**. Advanced visualizations including choropleth maps, trend analyses, correlation heatmaps, and extreme event detection provide meaningful insights into climate patterns and anomalies.

A dedicated **Infrastructure Risk Management module** extends the analytical capabilities of the dashboard by translating climate indicators into a composite infrastructure risk index. This module dynamically evaluates risk levels across selected regions, identifies high-risk periods, and supports comparative analysis for resilience planning. The application is deployed as a scalable web platform using **Streamlit Community Cloud**, enabling real-time accessibility and user interaction.

Milestone 1: Data Collection and Preprocessing

Tasks Performed:

- Collected cleaned climate data containing temperature, humidity, precipitation, wind speed, country, and timestamp fields.
- Performed data cleaning: handling missing values, correcting data types, and standardizing column names.
- Converted date-time fields (last_updated) into pandas datetime format.
- Validated data consistency across countries and time ranges.

Insights Gained:

- Climate datasets contain temporal inconsistencies that must be handled before analysis.
- Data standardization is critical for multi-country comparison.
- Missing or malformed timestamps significantly affect time-series analysis.

Outcome :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 106208 entries, 0 to 106207
Data columns (total 41 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   country                106208 non-null  object
 1   location_name          106208 non-null  object
 2   latitude               106208 non-null  float64
 3   longitude              106208 non-null  float64
 4   timezone               106208 non-null  object
 5   last_updated_epoch     106208 non-null  int64
 6   last_updated           106208 non-null  object
 7   temperature_celsius    106208 non-null  float64
 8   temperature_fahrenheit 106208 non-null  float64
 9   condition_text         106208 non-null  object
 10  wind_mph               106208 non-null  float64
 11  wind_kph               106208 non-null  float64
 12  wind_degree            106208 non-null  int64
 13  wind_direction         106208 non-null  object
 14  pressure_mb            106208 non-null  float64
 15  pressure_in            106208 non-null  float64
 16  precip_mm              106208 non-null  float64
 17  precip_in              106208 non-null  float64
 18  humidity               106208 non-null  int64
 19  cloud                  106208 non-null  int64
...
 39  moon_phase             106208 non-null  object
 40  moon_illumination      106208 non-null  int64
dtypes: float64(23), int64(7), object(11)
memory usage: 33.2+ MB
```

	latitude	longitude	last_updated_epoch	temperature_celsius	temperature_fahrenheit	wind_mph	wind_kph	wind_degree	pressure_mb	pressure_in	...	gust_kph
date												
2024-05-31	19.211157	21.571142	1.716479e+09	25.146254	77.262991	8.896344	14.320574	169.975227	1013.077644	29.915350	...	22.333444
2024-06-30	19.183897	21.743390	1.718499e+09	26.458191	79.626374	9.127356	14.694160	178.469251	1011.956417	29.882174	...	22.217657
2024-07-31	19.104503	22.120577	1.721100e+09	26.801201	80.244012	8.693869	13.995249	183.542488	1011.525457	29.869846	...	20.843492
2024-08-31	19.107315	22.088584	1.723811e+09	26.791679	80.226683	8.641522	13.912407	180.549545	1011.988586	29.883474	...	20.411166
2024-09-30	19.103744	22.084018	1.726442e+09	25.124671	77.226911	8.215216	13.224226	169.930415	1013.015216	29.913498	...	18.951906
2024-10-31	19.143867	22.503692	1.729068e+09	22.335382	72.204086	7.602126	12.237375	168.554651	1014.646512	29.962058	...	16.539236
2024-11-30	19.152160	22.633795	1.731712e+09	19.424566	66.966495	7.545711	12.148272	159.128589	1015.511604	29.987767	...	16.986729
2024-12-31	19.107269	22.088991	1.734345e+09	17.751894	63.953780	8.116890	13.065691	163.333333	1016.078246	30.004304	...	18.372787
2025-01-31	19.180727	22.675545	1.737028e+09	17.392724	63.309158	8.078022	13.004829	164.437063	1015.991841	30.001558	...	18.340826
2025-02-28	19.115330	22.085891	1.739571e+09	17.424684	63.366942	8.283013	13.333773	160.083746	1017.145135	30.035684	...	18.434451
2025-03-31	19.110843	22.102421	1.742118e+09	20.095299	68.173432	8.242013	13.267721	161.620427	1014.315180	29.951998	...	18.006042

Name : Pujan Kumar Patel

Milestone 2: Exploratory Data and Statistical Analysis

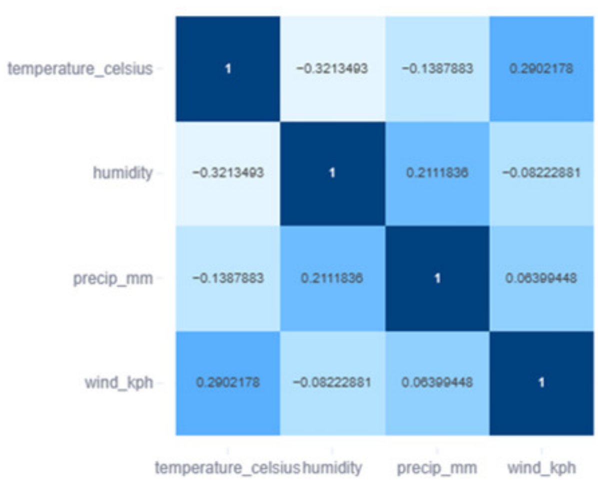
Tasks Performed:

- Conducted descriptive statistical analysis for temperature, humidity, precipitation, and wind.
- Built correlation heatmaps to identify relationships between climate variables.
- Analyzed metric distributions using histograms, box plots, and violin plots.
- Implemented country-wise and temporal aggregations (monthly and seasonal).

Insights Gained:

- Temperature shows moderate correlation with humidity in several regions.
- Seasonal aggregation reveals clear climatic cycles across countries.
- Extreme values significantly influence averages, justifying separate extreme event analysis.

Outcome:



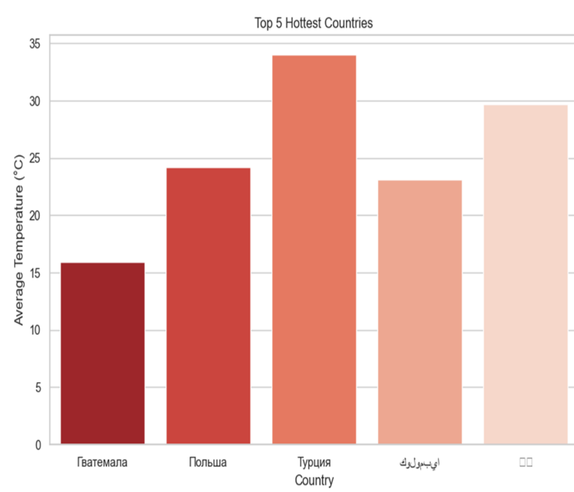


Fig 1.1 Heat Map **Fig 1.2 Top 5 Hot Countries**

Name : Pujan Kumar Patel

Milestone 3: Interactive Dashboard Development

Tasks Performed:

- Developed a Streamlit-based interactive dashboard.
- Implemented global filters for:
- Country selection
- Date range
- Climate metric selection
- Time aggregation (monthly / seasonal)
- Multiple interactive dashboard sections were developed, including the Executive Dashboard, Statistical Analysis, Climate Trends, and Extreme Events, to support comprehensive climate data exploration.

Insights Gained:

- Interactive filtering enables better comparative analysis across regions.
- Choropleth maps effectively highlight spatial climate variation.
- Executives benefit from KPI-style summaries before deep analysis.

Outcome:

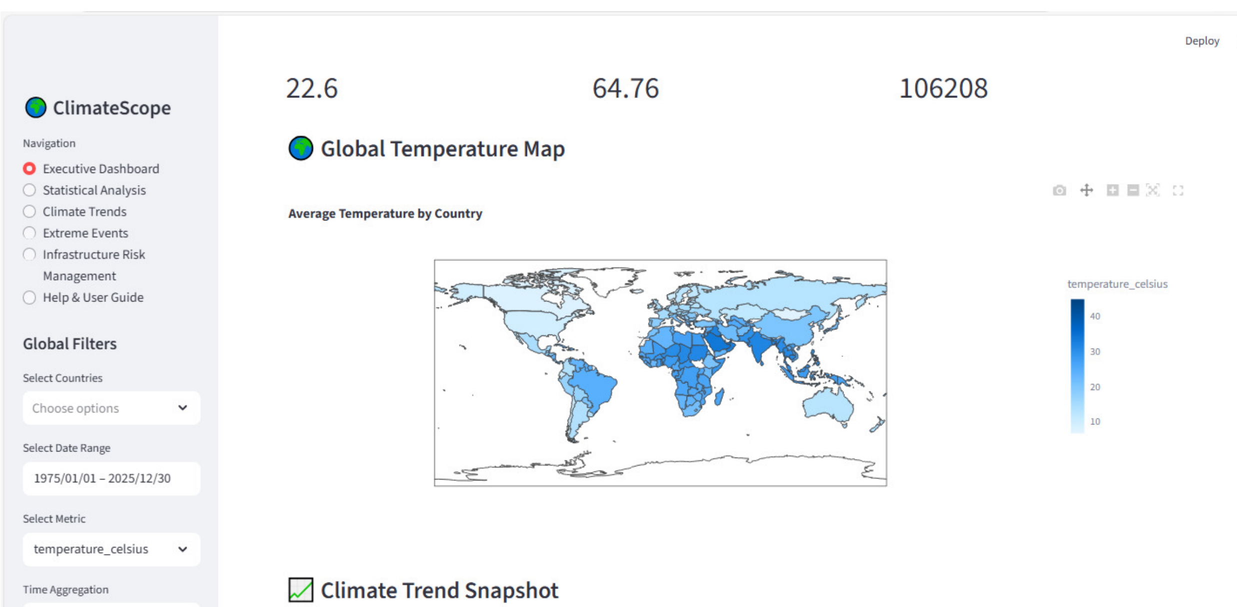


Fig 2.1 Dashboard

Milestone 4: Infrastructure Risk Management Application

Tasks Performed:

- Designed a derived Infrastructure Risk Index using weighted climate parameters.
- Formula used: - Risk Index = (0.4 × Temperature) + (0.35 × Precipitation) + (0.25 × Wind Speed)
- Classified risk levels into Low, Moderate, and High categories.
- Developed a dedicated dashboard section for Infrastructure Risk Management.
- Enabled dynamic updates of risk metrics based on selected countries and date range.

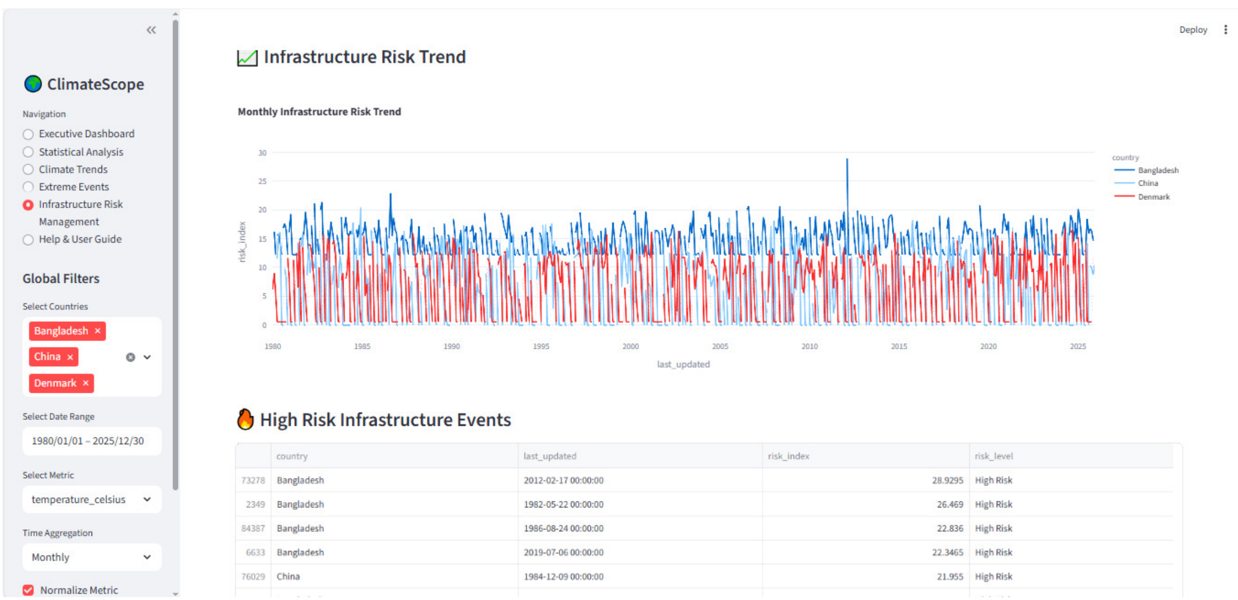
Insights Gained:

- Combined climate parameters provide a more actionable risk indicator.
- Certain countries consistently exhibit higher infrastructure risk trends.
- Risk-based dashboards bridge the gap between analysis and policy planning.

Name : Pujan Kumar Patel

Outcome:

Fig 3.1 Infrastructure Risk Management App



Test Cases

Test Case ID	Scenario	Input	Expected Output	Result
TC01	No country selected	Empty selection	All countries selected by default	Pass
TC02	Date filter	Valid range	Filtered records shown	Pass
TC03	Metric change	Temperature → Wind	Charts update correctly	Pass
TC04	Risk calculation	Multiple countries	Risk index computed correctly	Pass
TC05	Extreme threshold	High threshold	Only extreme events shown	Pass

Errors Encountered and Resolutions

- Error 1: No countries seleceted

➢ When no country is selected in the global filter, the system automatically defaults to all available countries, ensuring continuous data visualization without missing or null values.

- Error 2: Date filter:

➢ Applying a valid date range correctly filters the dataset, and only records within the selected period are displayed across all dashboard sections.

- Error 3: Risk calculation

➢ For multiple selected countries, the infrastructure risk index is computed accurately using weighted climate parameters and reflected consistently in all risk visualizations.

- Error 4: Extreme threshold:

➢ When a high extreme-event threshold is set, the dashboard correctly filters and displays only those climate records that exceed the defined limit.

Conclusion

The ClimateScope project effectively demonstrates how raw and heterogeneous climate data can be transformed into meaningful and actionable insights using data science techniques and interactive web technologies.

Through a structured, milestone-driven approach, the project progressed from data collection and preprocessing to exploratory data analysis, statistical evaluation, and the development of an interactive, web-based dashboard.

The integration of advanced visualizations enabled intuitive exploration of climate patterns, trends, and extreme events across multiple regions.

Furthermore, the addition of the **Infrastructure Risk Management module** significantly enhances the practical relevance of the system by translating climatic indicators into a composite risk index.

This enables stakeholders such as planners, policymakers, and researchers to assess potential climate-induced risks and support informed decision-making.

The successful deployment of the application as a web platform demonstrates its scalability, usability, and real-world applicability.

Future Enhancements

- While ClimateScope provides a strong foundation for climate analysis and risk assessment, several enhancements can further extend its capabilities.
- Future work may include the integration of real-time climate data through external APIs to enable live monitoring and up-to-date analysis.
- Machine learning models can be incorporated to predict future climate trends and infrastructure risks based on historical patterns.
- Additionally, the system can be enhanced to support city-level analysis and infrastructure-type-specific risk modeling, allowing more granular and targeted assessments. Features such as automated alerts for high-risk events and exportable analytical reports in PDF or CSV formats can further improve usability and decision support, making ClimateScope a more comprehensive climate intelligence platform.

References

- IPCC, *Climate Change 2023: Synthesis Report*, Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2023.

- Pandas Development Team, *pandas: Python Data Analysis Library*, 2024. Available: https://pandas.pydata.org/

- Plotly Technologies Inc., *Plotly Python Graphing Library*, 2024. Available: https://plotly.com/python/

- Streamlit Inc., *Streamlit: A Framework for Building Data Apps*, 2024. Available: https://streamlit.io/

- McKinney, W., "Data Structures for Statistical Computing in Python," *Proceedings of the 9th Python in Science Conference*, pp. 51–56, 2010.

- Hunter, J. D., "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

- NOAA, *National Centers for Environmental Information – Climate Data*, National Oceanic and Atmospheric Administration. Available: https://www.ncei.noaa.gov/

- World Meteorological Organization (WMO), *Guide to Climatological Practices*, WMO-No. 100, Geneva, Switzerland.

- Pedregosa, F. et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

- Streamlit Community Cloud, *Deploying Streamlit Applications*, 2024. Available: https://docs.streamlit.io/streamlit-community-cloud