Kerala Rainfall Analysis (2009–2024): Trends, Anomalies & Climate Vulnerability

Author: Muhammad Rayif
Course: Data Analytics Final Project
Mentor: Krishna Priya **Submission Date:** 13/09/2025 **Repository:**
https://github.com/rayif007/Kerala-Rainfall-Analysis

Abstract

This study analyzes Kerala's daily rainfall (2009–2024) using IMD state-level daily rainfall data merged with state vulnerability indicators from the India Data Portal. The goal is to quantify rainfall variability, identify anomalies and extreme rainfall days, and assess implications for society and business. Preprocessing, feature engineering, and exploratory visualizations were performed in Python (Pandas, Matplotlib, Seaborn). Key findings show strong seasonal dependence (monsoon), years with extreme excess rainfall days, and clear linkages between anomaly magnitude and extreme-event frequency.

---

1. Introduction

Kerala's economy and livelihoods strongly depend on monsoon rainfall. Climate variability has led to notable flood and drought events in the past decades. This project aims to produce data-driven insights into how Kerala's rainfall has changed over 2009–2024, how extremes are distributed, and what the societal and business implications are.

Objectives

- Clean and merge daily rainfall and vulnerability datasets.
- Engineer temporal and climate features (year, month, season, anomaly, excess flag).
- Perform EDA to reveal trends, seasonality, anomalies, and extremes.
- Document findings and recommend policy/business actions.

---

2. Data Description

Sources

- Daily rainfall: India Meteorological Department (IMD), via India Data Portal (IDP).
- Climate vulnerability indicators: India Data Portal (state-level vulnerability table).

Coverage & Granularity

- Period: 2009–2024
- Spatial: Kerala (state-level)
- Granularity: Daily rainfall (actual, normal), yearly vulnerability indicators

Columns used (final)

- `date`, `year`, `month`, `season`, `actual` (mm), `normal` (mm), `deviation` (%), `rainfall_anomaly` (mm), `excess_rainfall` (flag)

> *Note:* Many state-level socio-economic vulnerability columns were empty/constant for Kerala in the provided vulnerability dataset; correlation heatmap with vulnerability indicators was therefore omitted and documented as a limitation.

---

3. Methodology

3.1 Preprocessing

- Loaded raw CSVs and filtered for Kerala (state_name == "Kerala").
- Converted `date` to `datetime` (dayfirst=True) and dropped invalid dates.
- Extracted `year`, `month`, and assigned `season` (Monsoon: Jun–Sep; Post-Monsoon: Oct–Nov; Winter: Dec–Feb; Summer: Mar–May).
- Merged the daily rainfall table (daily) with yearly vulnerability features on `year`.
- Handled missing values:
  - For numeric columns: mean imputation (only where sensible).
  - For categorical columns: mode imputation.
  - Dropped columns that were all-NaN or constant for Kerala (e.g., `rainfed_agriculture`, `climate_vul_in`).
- Feature engineering:
  - `rainfall_anomaly = actual - normal`
  - `excess_rainfall = 1 if deviation > 20 else 0`
- Final dataset saved as: `kerala_rainfall_vulnerability_cleaned.csv`.

3.2 Exploratory Data Analysis (plots & steps)

- Daily trends: raw daily plot, monthly averages, 30-day rolling average.
- Yearly totals: sum of daily rainfall per year.
- Anomaly distribution: histogram of `rainfall_anomaly`.

- Seasonal analysis: seasonal bar plot and boxplot.
- Extremes: percent of excess rainfall days per year (excess >20%).
- Relationship: scatter (mean yearly anomaly vs percent excess days) with point size reflecting excess share.

---

4. Results & Key Insights

1. **Seasonality confirmed:** Monsoon months (Jun–Sep) dominate rainfall.
2. **Annual variability:** Some years are clearly flood-prone (large annual totals), others show deficit.
3. **Anomalies are common:** Distribution shows substantial numbers of both deficit and excess days.
4. **Extremes cluster in certain years:** Percentage of excess rainfall days spiked in known flood years.
5. **Anomaly–extreme link:** Years with larger mean anomaly have higher share of excess days (scatter and regression support).

---

5. Limitations

- State-level vulnerability indicators were missing/constant for Kerala in the provided file, hence correlation analysis with socio-economic variables was not possible.
- Analysis is state-level (Kerala); district-level analysis could provide finer policy insights.
- Feature imputation (mean/mode) was used where necessary and documented.

---

6. Societal & Business Implications

Society: Increased flood risk causes displacement, infrastructure damage, and health hazards. Insights can guide disaster-preparedness planning and targeted support to vulnerable groups.
Business: Agriculture, insurance and infrastructure sectors should incorporate observed extreme-frequency metrics into planning, pricing models, and building codes.

---

7. Conclusions

This project documents that Kerala's rainfall (2009–2024) has significant variability and recurring anomalies that translate into extreme rainfall days. Even without full socio-economic data, the rainfall signals themselves are actionable and recommend targeted adaptation measures.

---

8. Future Work

- Integrate district-level rainfall and crop yield or flood-damage datasets.
- Add socio-economic time-series from alternate sources (state statistical reports, census).
- Build predictive models (time-series forecasting) for anomaly / extreme-day probabilities.

---

9. Reproducibility & Files

- Notebook: `notebooks/Kerala_Rainfall_Analysis.ipynb`
- Cleaned data: `data/kerala_rainfall_vulnerability_cleaned.csv`
- Output plots: `outputs/*.png`
- Final report: `report/Kerala_Rainfall_Report.pdf`
- GitHub repo: https://github.com/your-github/YourRepoName

---

10. References

- India Data Portal (IDP): https://indiadataportal.com
- India Meteorological Department (IMD): https://mausam.imd.gov.in


```
Start coding or generate with AI.
```