Exploratory Data Analysis (EDA) of Rainfall Patterns for Indian Agriculture

Final Project Report

1. Project Overview

India is an agrarian economy where nearly 54% of the arable land is rain-fed. The Southwest Monsoon (June–September) accounts for nearly 75% of India’s annual rainfall. Recent climate change impacts have increased rainfall variability, making agricultural planning uncertain. This project performs Exploratory Data Analysis (EDA) on long-term rainfall data to study trends, anomalies, and spatial shifts affecting crop productivity.

2. Problem Statement

Indian agriculture is highly sensitive to rainfall variability. The gambling nature of monsoons, spatial inconsistency across regions, and increasing climate-induced dry spells and extreme rainfall events pose significant risks to crop yield and farmer livelihoods.

3. Objectives

The objectives of this study are:
- To identify long-term temporal rainfall trends
- To analyze spatial rainfall variability across India
- To assess rainfall reliability using statistical measures
- To study correlation between rainfall deviation and crop yield

4. Dataset Description

The dataset is sourced from the India Meteorological Department (IMD) and World Bank Open Data.
Features include Year, Month, State, District, and Monthly Rainfall (mm).
The data spans from 1901 to 2024 with state and subdivision-level aggregation.

5. Methodology

Data preprocessing involved handling missing values, temporal aggregation, and normalization. Feature engineering included rainfall anomaly detection and Coefficient of Variation (CV) calculation. EDA techniques such as trend analysis, heatmaps, and correlation analysis were applied.

6. Exploratory Data Analysis

Temporal analysis shows stable long-term rainfall with increasing variability. Seasonal analysis highlights the dominance of the Southwest Monsoon. Spatial analysis reveals high rainfall in Northeast India and Western Ghats, and high variability in arid and semi-arid regions. CV analysis identifies rainfall reliability across regions.

7. Rainfall and Crop Yield Relationship

Rice shows strong dependence on monsoon rainfall, wheat shows weaker correlation due to irrigation support, and maize is sensitive to both deficit and excess rainfall.

8. Tools and Technologies

Python libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Statsmodels were used. GeoPandas supported spatial analysis, and Jupyter Notebook ensured reproducibility.

9. Key Findings

Rainfall variability has increased over time. Spatial inequality in rainfall is widening. Rain-fed regions are at higher agricultural risk, making irrigation and climate-resilient practices essential.

10. Conclusion and Future Scope

Rainfall variability poses a greater threat than total rainfall reduction. Future work can integrate soil moisture, ENSO indices, and machine learning models for yield prediction and drought early-warning systems.