Exploratory Analysis of Rainfall Data in India for Agriculture

Abstract

This project focuses on performing Exploratory Data Analysis (EDA) on real Indian rainfall data sourced from the Indian Meteorological Department (IMD) via the Open Government Data Platform. The objective is to analyze rainfall trends, seasonal monsoon behavior (Southwest and Northeast Monsoon), state-wise and district-wise rainfall distribution, and statistical characteristics such as mean, variance, correlation, and regression trends. The insights derived aim to support agricultural planning, irrigation management, and crop selection strategies across India.

Phase 1: Problem Definition

• Understand how rainfall patterns influence Indian agriculture. • Identify variability across states and districts. • Analyze monsoon-specific rainfall behavior. • Detect long-term rainfall trends using regression analysis.

Phase 2: Planning

Data Source: IMD District-wise and State-wise rainfall dataset. Time Period: Multi-year analysis (e.g., 2010–2024). Tools Used: Python, Pandas, NumPy, Matplotlib, Scikit-learn. Environment: Jupyter Notebook / Python Script. Output: Statistical report with visualizations and insights.

Phase 3: Data Collection

• Download district-wise rainfall dataset from data.gov.in (IMD source). • Verify dataset completeness and missing values. • Convert raw CSV data into structured Pandas DataFrame. • Standardize date formats and region names.

Phase 4: Data Preprocessing

• Handle missing rainfall values. • Aggregate monthly rainfall into seasonal rainfall. • Define Southwest Monsoon (June–September). • Define Northeast Monsoon (October–December). • Compute annual rainfall totals.

Phase 5: Statistical Analysis

1. Mean Rainfall: Average rainfall per state/district. 2. Variance & Standard Deviation: Measure rainfall variability. 3. Correlation Analysis: Relationship between seasonal rainfall and annual rainfall. 4. Regression Analysis: Detect long-term rainfall trends over years. 5. Outlier Detection: Identify extreme rainfall events.

Phase 6: Monsoon-Specific Analysis

• Southwest Monsoon Contribution Percentage. • Northeast Monsoon Contribution Percentage. • Comparison of monsoon variability across regions. • Impact of monsoon deviation on agriculture.

Phase 7: State-wise & District-wise Analysis

• Rank states based on annual rainfall. • Identify drought-prone districts. • Identify high rainfall zones. • Compare agricultural dependency on rainfall across regions.

Phase 8: Implementation (Coding Steps)

Step 1: Import Libraries (Pandas, NumPy, Matplotlib, Sklearn). Step 2: Load Dataset using pd.read_csv(). Step 3: Perform data cleaning and aggregation. Step 4: Compute statistical metrics using .mean(), .var(), .corr(). Step 5: Apply Linear Regression using sklearn.linear_model. Step 6: Visualize trends using Matplotlib. Step 7: Interpret results and document insights.

Conclusion

The exploratory analysis of real Indian rainfall data provides actionable insights for agricultural planning. Seasonal and regional rainfall variability significantly affects crop yield and irrigation demand. Regression analysis helps in understanding long-term climate trends. Data-driven agricultural strategies can enhance sustainability and productivity in India.