**DATA SCIENCE MINOR PROJECT REPORT**

**CONTENTS OF THE REPORT**

- Cover page

- Declaration

- Certificate

- Acknowledgement

- Table of Content

1. Introduction

2. Source of dataset

3. EDA process

4. Analysis on dataset (for each analysis)

1. Introduction
2. General Description
3. Specific Requirements, functions and formulas
4. Analysis results
5. Visualization

5. Conclusion

6. Future scope

7. References

**<DATA TOOLBOX: PYTHON PROGRAMMING>**

**PROJECT REPORT**

(Project Semester January-April 2025)

***(Analysis of India’s Rainfall Patterm using Python and IMD Data)***

Submitted by

(Priteek Singh Manhas)

Registration No 12326131

Programme and Section Btech CSE KM005

Course Code INT375

Under the Guidance of

**(Mrs Maneet Kaur UID 15709)**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Priteek Singh Manhas (student’s name) bearing Registration no. 12326131 has completed INT 375 <Course Code> project titled, **“*Analysis of India’s Rainfall Patterm using Python and IMD Data*”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Maneet Kaur**

**Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 2025-04-10

**DECLARATION**

I, Priteek Singh Manhas, student of Data Science Toolbox:Python Programming under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 2025-04-10 Signature PriteekSingh

Registration No. 12326131 Name of the student: Priteek Singh Manhas

# TABLE OF CONTENTS

1. Introduction  
2. Source of Dataset  
3. EDA Process  
4. Analysis on Dataset  
 i. Introduction  
 ii. General Description  
 iii. Specific Requirements, Functions and Formulas  
 iv. Analysis Results  
 v. Visualization  
5. Conclusion  
6. Future Scope  
7. References

# 1. INTRODUCTION

Rainfall plays a crucial role in maintaining agricultural productivity, water resources, and ecological balance in India. Analyzing historical rainfall data helps identify climatic trends and anomalies that can affect these systems. This project leverages rainfall records from the Indian Meteorological Department (IMD) to analyze changes in rainfall patterns across regions and time. The analysis includes trend detection, correlation analysis, and predictive modeling to support decision-making in climate-sensitive sectors.

# 2. SOURCE OF DATASET

# The dataset used in this project was sourced from the Indian Meteorological Department (IMD) and accessed via the Government of India’s open data portal (<https://data.gov.in>). It contains monthly rainfall data (in mm) for various meteorological subdivisions in India over several years.

# Key features include:

# SUBDIVISION: Regional area name

# YEAR: Observation year

# JAN to DEC: Monthly rainfall totals

# JJAS: Rainfall during the monsoon months (June–September)

# ANNUAL: Total annual rainfall

# This dataset is suitable for time series analysis, trend detection, and regional comparison.

# 3. EDA PROCESS

A thorough Exploratory Data Analysis (EDA) was conducted to clean and prepare the dataset for insights.

**Steps involved:**

1. **Missing Data Handling**: Missing numeric values were filled using median imputation, which is robust to outliers.
2. **Zero Value Treatment**: Zeroes in numeric columns were replaced with mean values to correct anomalies without dropping rows.
3. **Data Type Conversion**: The YEAR column was cast to integer for modeling and plotting.
4. **Outlier & Inconsistency Removal**: Rows missing crucial values like ANNUAL were excluded.

This process ensured the dataset was clean, consistent, and suitable for further analysis like regression and visualization.

# 4. ANALYSIS ON DATASET

## i. Introduction

## This section involves multiple analyses on rainfall data to extract trends, detect regional extremes, and build predictive models. It focuses on time-series behavior, seasonal trends, correlation between features, and subdivision-specific insights using both descriptive and inferential statistics.

## ii. General Description

## Rainfall data in India varies considerably across subdivisions and seasons. The dataset captures monthly, annual, and monsoon rainfall amounts over several decades. The analyses are designed to explore:

## Temporal trends (annual, decadal)

## Spatial patterns (wettest/driest regions)

## Correlations (JJAS vs Annual)

## Predictive modeling using linear regression

## iii. Specific Requirements, Functions and Formulas

## The analysis was performed using Python with libraries such as:

## Pandas: for data wrangling and processing

## Seaborn & Matplotlib: for visualization

## Scikit-learn: for linear regression modeling

## SciPy: for calculating linear trends using linregress

## Key Formulas and Functions:

## df.fillna(df.median()): Median imputation

## linregress(x, y): Calculates slope (trend) and correlation

## LinearRegression().fit(X, y): Fits a regression model

## corr(): Pearson correlation coefficient

## .groupby(): For aggregating values by year/decade

## iv. Analysis Results

## Trend Analysis: Annual rainfall in Jammu & Kashmir showed year-to-year variation, with some declining trends.

## Extreme Values: The subdivision with the highest rainfall in 2015 was identified, aiding in detecting flood-prone zones.

## Monthly Patterns: Bar plots showed July and August are the peak rainfall months across India.

## Regional Extremes: Top 5 wettest and driest years for Andaman & Nicobar Islands revealed changing climate behavior.

## Decadal Shifts: Monsoon rainfall analyzed by decade showed trends in intensity.

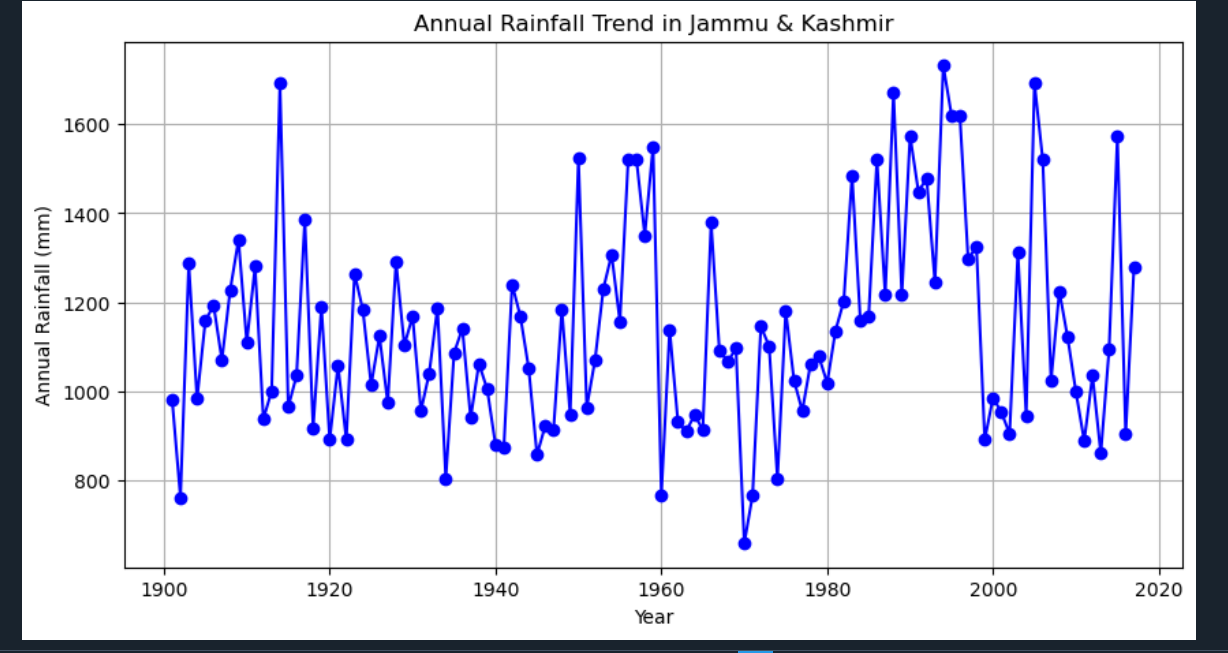
## Decreasing Trend Detection: Linear regression slopes highlighted subdivisions with negative rainfall trends.

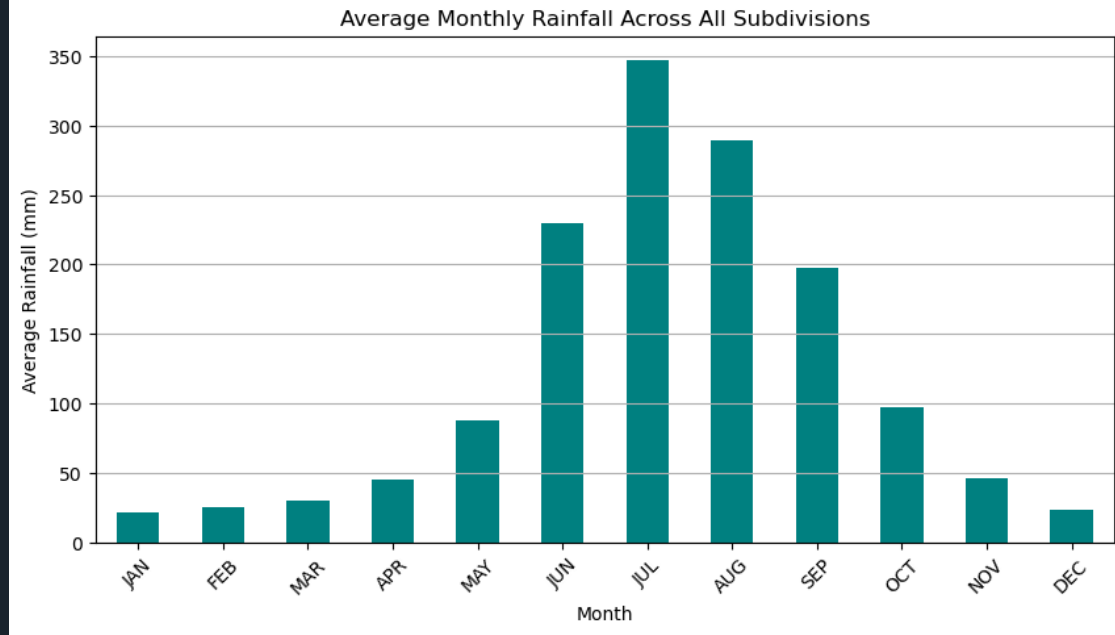
## JJAS vs Annual Correlation: A strong correlation confirmed monsoon rainfall's significant contribution to annual totals.

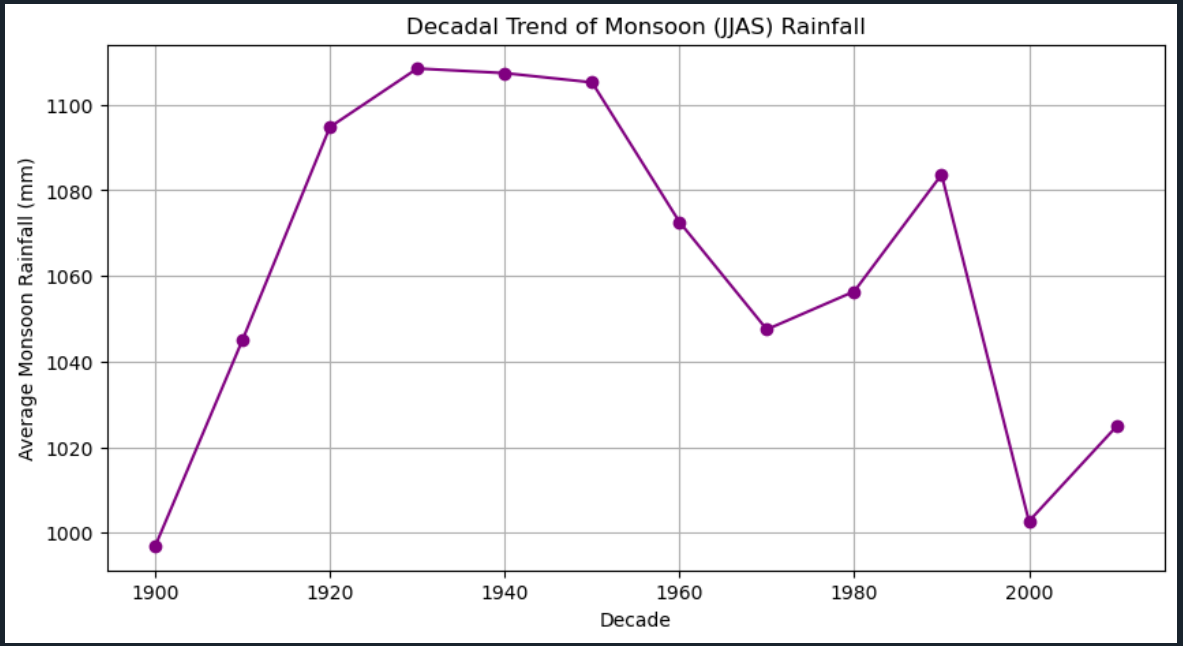
## Prediction: A regression model predicted the JJAS rainfall for 2025 based on past trends.

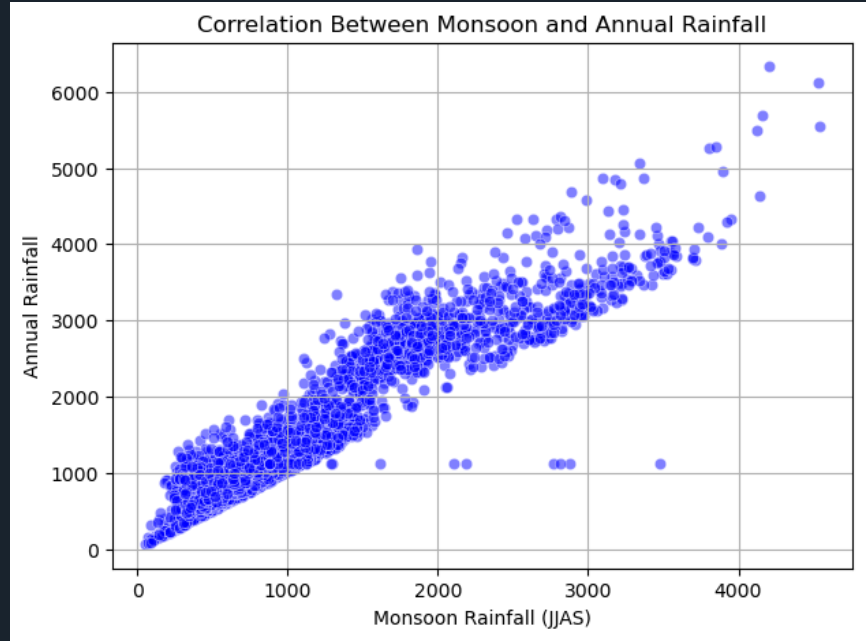
## v. Visualization

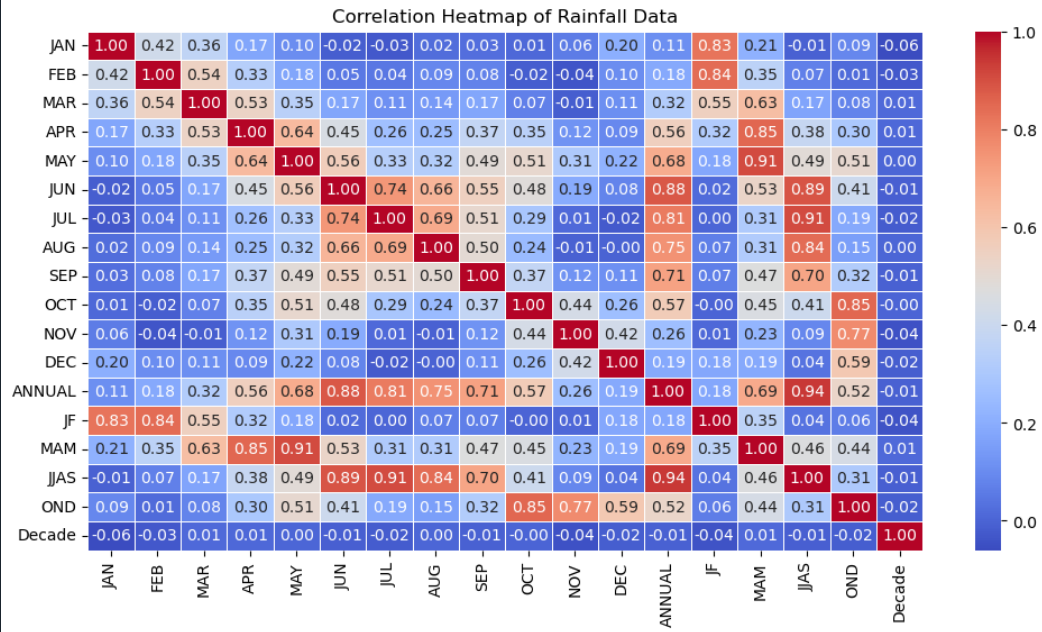
Figure 1: Annual Rainfall Trend in Jammu & Kashmir

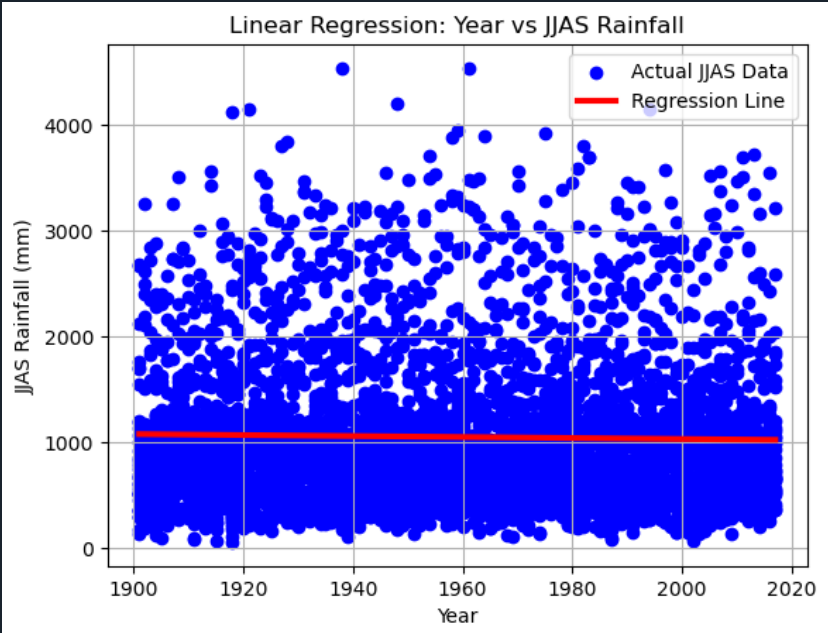
  
Figure 2: Average Monthly Rainfall

  
Figure 3: Decadal Monsoon Rainfall Trend

  
Figure 4: Correlation Between JJAS and Annual Rainfall

  
Figure 5: Correlation Heatmap

  
Figure 6: Linear Regression: Year vs JJAS Rainfall



# 5. CONCLUSION

This data-driven analysis provides key insights into rainfall trends across Indian subdivisions, focusing on both temporal and spatial patterns.

Rainfall Trend Over Time:

The trend analysis for Jammu & Kashmir reveals how rainfall has fluctuated over the years. Visual inspection of the plot can help detect whether there’s a rising, falling, or stable trend, useful for local planning and climate adaptation strategies.

Extremes in Annual Rainfall:

The analysis highlights the wettest subdivision in 2015, identifying areas prone to extreme rainfall, which can inform disaster preparedness and resource allocation.

Monthly Rainfall Patterns:

By calculating the average monthly rainfall, it is evident which months contribute most to the annual total—typically the monsoon months (June to September), which is crucial for agriculture and water management.

Wettest and Driest Years:

Identifying the top 5 wettest and driest years in Andaman & Nicobar Islands helps understand climatic extremes, which can support environmental and economic risk assessment in island regions.

Decadal Rainfall Trends:

Analysis of JJAS (monsoon) rainfall across decades indicates how climate patterns are shifting, providing evidence for or against climate change hypotheses.

Subdivisions with Decreasing Rainfall:

Using linear regression, subdivisions showing declining rainfall trends were identified, suggesting potential vulnerability to droughts and calling for targeted water conservation efforts.

Correlation Insights:

A strong positive correlation between JJAS and Annual rainfall (around 0.97) confirms that monsoon rainfall is a significant contributor to the annual totals—vital for forecasting models and planning agricultural cycles.

Predictive Modeling:

A linear regression model predicts a decline or change in rainfall for future years, with an estimate for 2025 JJAS rainfall aiding future preparedness.

Heatmap Correlation Analysis:

The correlation matrix helps understand interdependencies between different months, supporting feature selection for more advanced predictive modeling.'''

# 6. FUTURE SCOPE

Future work can explore:

* Use of **time-series forecasting models** like ARIMA, Prophet, or LSTM for better long-term predictions.
* Integration of **geospatial analysis** to assess local rainfall variations using latitude/longitude coordinates.
* Inclusion of **satellite rainfall data** and real-time feeds for dynamic modeling.
* Use of rainfall data alongside **crop yield** or **hydrological datasets** for more holistic environmental analysis.

# 7. REFERENCES

 [1] IMD Rainfall Data – <https://data.gov.in>

 [2] Scikit-learn Documentation – <https://scikit-learn.org>

 [3] Seaborn Documentation – https://seaborn.pydata.org

 [4] Pandas Documentation – https://pandas.pydata.org

 [5] SciPy Documentation – https://docs.scipy.org

 [6] ARIMA Forecasting in Python – <https://www.statsmodels.org>