ANALYZING RAINFALL PATTERNS OVER THE LAST DECADE IN A REGION

A PROJECT REPORT

Submitted by

SRIVATHSAN S (2303811724321108)

in partial fulfillment of requirements for the award of the course

AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

JUNE- 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report on **"ANALYZING RAINFALL PATTERNS OVER THE LAST DECADE IN A REGION"** is the bonafide work of **SRIVATHSAN S (2303811724321105)** who carried out the project work during the academic year 2024 - 2025 under my supervision.

SIGNATURE

Dr.T. AVUDAIAPPAN, M.E.,Ph.D.,

HEAD OF THE DEPARTMENT

PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

SIGNATURE

Ms.S.Murugavalli., M.E.,(Ph.D).,

SUPERVISOR

ASSISTANT PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

Submitted for the viva-voce examination held on **02.06.2025**.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION

I declare that the project report on **"ANALYZING RAINFALL PATTERNS OVER THE LAST DECADE IN A REGION"** is the result of original work done by me and best of my knowledge, similar work has not been submitted to **"ANNA UNIVERSITY CHENNAI"** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the completion of the course **AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R.**

Signature

SRIVATHSAN S

Place: Samayapuram

Date:02.06.2025

ACKNOWLEDGEMENT

It is with great pride that I express our gratitude and in-debt to our institution "**K.Ramakrishnan College of Technology (Autonomous)**", for providing us with the opportunity to do this project.

I glad to credit honourable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

I would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

I would like to thank **Dr. N. VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

I whole heartily thanks to **Dr. T. AVUDAIAPPAN, M.E.,Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

I express our deep expression and sincere gratitude to our project supervisor **Ms.S.Murugavalli., M.E.,(Ph.D).,** Department of **ARTIFICIAL INTELLIGENCE,** for her incalculable suggestions, creativity, assistance and patience which motivated us to carry out this project.

I render our sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

I wish to express our special thanks to the officials and Lab Technicians of our departments who rendered their help during the period of the work progress.

INSTITUTE

Vision:

- To serve the society by offering top-notch technical education on par with global standards.

Mission:

- Be a center of excellence for technical education in emerging technologies by exceeding the needs of industry and society.
- Be an institute with world class research facilities.
- Be an institute nurturing talent and enhancing competency of students to transform them as all – round personalities respecting moral and ethical values.

DEPARTMENT

Vision:

- To excel in education, innovation, and research in Artificial Intelligence and Data Science to fulfil industrial demands and societal expectations.

Mission

- To educate future engineers with solid fundamentals, continually improving teaching methods using modern tools.
- To collaborate with industry and offer top-notch facilities in a conducive learning environment.
- To foster skilled engineers and ethical innovation in AI and Data Science for global recognition and impactful research.
- To tackle the societal challenge of producing capable professionals by instilling employability skills and human values.

PROGRAM EDUCATIONAL OBJECTIVES (PEO)

- **PEO1:** Compete on a global scale for a professional career in Artificial Intelligence and Data Science.
- **PEO2:** Provide industry-specific solutions for the society with effective communication and ethics.
- **PEO3** Enhance their professional skills through research and lifelong learning initiatives.

PROGRAM SPECIFIC OUTCOMES (PSOs)

- **PSO1:** Capable of finding the important factors in large datasets, simplify the data, and improve predictive model accuracy.
- **PSO2:** Capable of analyzing and providing a solution to a given real-world problem by designing an effective program.

PROGRAM OUTCOMES (POs)

Engineering students will be able to:

1. **Engineering knowledge:** Apply knowledge of mathematics, natural science, computing, engineering fundamentals, and an engineering specialization to develop solutions to complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development.
3. **Design/development of solutions:** Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required.
4. **Conduct investigations of complex problems:** Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions.
5. **Engineering Tool Usage:** Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems.
6. **The Engineer and The World:** Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment.

7. **Ethics:** Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws.

8. **Individual and Collaborative Team work:** Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.

9. **Communication:** Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences.

10. **Project management and finance:** Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

11. **Life-long learning:** Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change.

ABSTRACT

This project presents an interactive R Shiny web application developed for analyzing rainfall patterns over the last ten years in a selected geographical region. The tool allows users to upload rainfall datasets in CSV format and interactively explore trends, seasonality, and statistical summaries using real-time filters and visualizations. Users can select a specific region and time window, and the system dynamically renders daily time-series plots, seasonal distribution via boxplots, and descriptive summaries.

Built using core R packages including shiny, dplyr, ggplot2, lubridate, and DT, the application ensures a smooth and responsive user experience. The dashboard transforms raw meteorological data into insightful visualizations and statistical indicators, enabling better understanding of rainfall distribution and variability. This tool is particularly useful for environmental researchers, agricultural planners, and climate analysts aiming to monitor regional rainfall patterns over a decadal timeline.

ABSTRACT WITH POs AND PSOs MAPPING

CO 5 : BUILD DATA SCIENCE USING R PROGRAMMING FOR SOLVING REAL-TIME PROBLEMS.

ABSTRACT	POs MAPPED	PSOs MAPPED
This project presents an interactive R Shiny web application developed for analyzing rainfall patterns over the last ten years in a selected geographical region. The tool allows users to upload rainfall datasets in CSV format and interactively explore trends, seasonality, and statistical summaries using real-time filters and visualizations. Users can select a specific region and time window, and the system dynamically renders daily time-series plots, seasonal distribution via boxplots, and descriptive summaries.	**PO1 -3** **PO2 -3** **PO3 -3** **PO5 -3** **PO9 -2**	**PSO1 -3** **PSO2 -3**

Note: 1- Low, 2-Medium, 3- High

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

11.1 Objective

Understanding long-term rainfall patterns is crucial for sustainable water resource management, agricultural planning, and climate monitoring. This project aims to analyze and visualize rainfall data across a ten-year period for user-selected regions. The primary goal is to build an R Shiny dashboard that enables users to upload region-specific rainfall datasets, filter them by custom date ranges, and generate dynamic, interpretable visual outputs.

11.2 Overview

Rainfall data often contains temporal and seasonal patterns that require visual exploration to reveal insights. This project uses daily rainfall data with attributes such as date, region, and rainfall amount (in millimeters). Through filtering options, users can narrow down the dataset and examine fluctuations in rainfall over time or across seasons. The output includes time-series plots for daily rainfall, monthly boxplots to capture seasonal variability, and a detailed summary of the selected dataset.

11.3 R Programming Concepts Used

The interactive nature of the dashboard supports real-time feedback and easy interpretation of data without the need for deep statistical expertise. With the integration of modern R packages, this application is a practical demonstration of how R can be used for environmental data science and decision-making.

CHAPTER 2
PROJECT METHODOLOGY

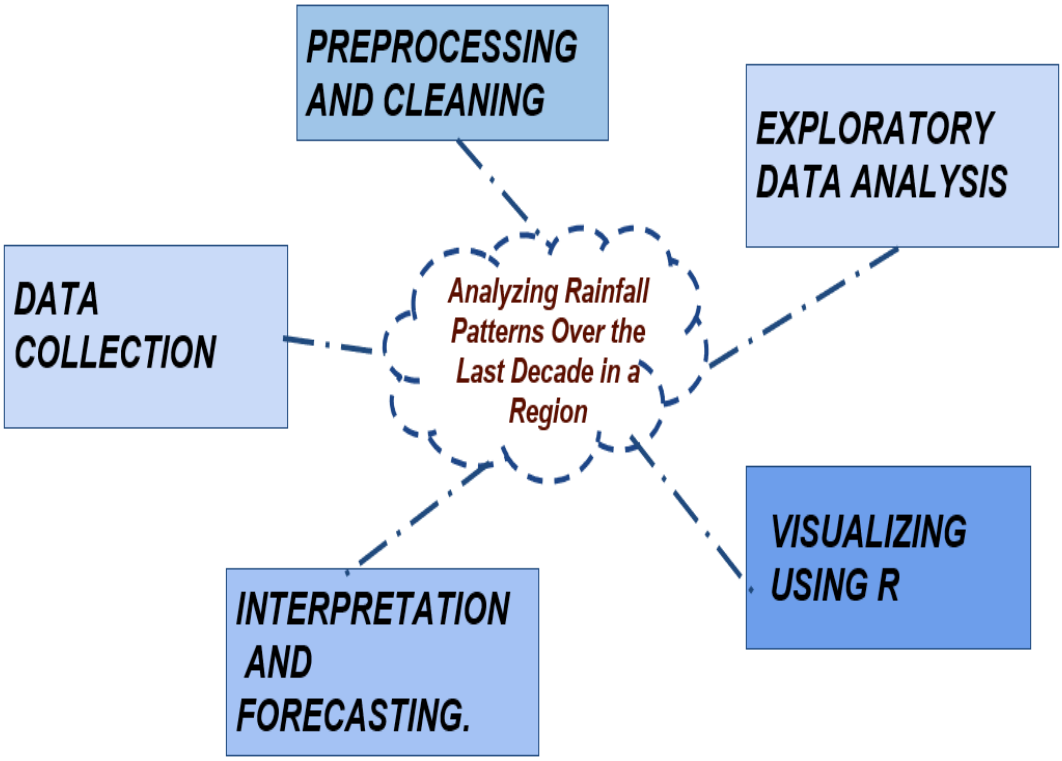
2.1 Proposed Work

The project methodology begins with dataset acquisition, wherein the user is expected to upload a CSV file containing rainfall data with at least three columns: Region, Date, and Rainfall_mm. Once the file is uploaded, the system reads and parses it, converting the date column into a proper Date format using the lubridate package. This preprocessing ensures compatibility with date range filters and visual plotting functions.

Upon successful data loading, the unique regions in the dataset are extracted and presented as options in a dropdown menu. The user then selects a specific region of interest and defines a date range (defaulted to the last ten years). On triggering the “Analyze” button, the system filters the dataset according to the specified parameters.

The application calculates and renders a time-series plot of daily rainfall values using ggplot2, summarizing rainfall trends over time. A seasonal analysis is provided through a monthly boxplot, which aggregates rainfall distributions by month across years. Additionally, a summary tab generates statistical details such as the total rainfall, average daily rainfall, wettest day, and total number of records. These outputs are dynamically updated based on user input, making the dashboard both flexible and responsive.

2.2 Block Diagram



CHAPTER 3
MODULE DESCRIPTION

3.1 Data Upload Module

The **Data Upload Module** provides users with a file input field to upload their rainfall dataset. This module accepts .csv files containing daily rainfall observations and parses them into a structured data frame. The column “Date” is converted to the Date data type, ensuring compatibility with time-based filters.

3.2 Filtering Module

The **Filtering Module** enables users to select a specific region and a date range using dropdown and calendar controls. Upon selection, the dataset is filtered in real-time using dplyr, isolating records that match the specified region and fall within the selected date window.

3.3 Time Series Visualization Module

The **Time Series Visualization Module** processes the filtered dataset and groups rainfall data by individual date. A daily rainfall line chart is generated to visualize how rainfall has fluctuated over the chosen period. This module is particularly useful in detecting trends, rainy spells, or dry phases.

3.4 Seasonal Boxplot Module

The **Seasonal Boxplot Module** analyzes monthly rainfall variation by aggregating the data across all years and plotting the distribution of rainfall values by month. This boxplot visualization helps users understand the seasonal nature of rainfall, detect outliers, and identify peak monsoon months.

3.5 Summary Reporting Module

The **Summary Reporting Module** provides key statistics about the filtered dataset. It displays the selected region and period, total number of entries, cumulative rainfall, average daily rainfall, and the date and value of the wettest day recorded. This module enhances the dashboard's interpretability and supports evidence-based conclusions.

CHAPTER 4
CONCLUSION & FUTURE SCOPE

Conclusion

This project successfully demonstrates how R and Shiny can be used to build an interactive, real-time dashboard for visualizing and analyzing rainfall data. By transforming raw meteorological records into structured and insightful graphics, the dashboard aids in understanding regional rainfall patterns and variability over a decade. Users can interactively explore temporal and seasonal patterns, derive key statistics, and support climate-related decision-making.

Future Scope

In the future, this project can be extended in several ways:

- Integration of geospatial visualizations using maps to compare rainfall patterns across multiple regions.
- Addition of trend analysis and decomposition techniques (e.g., moving averages, seasonal adjustment).
- Use of machine learning models to forecast future rainfall patterns based on historical trends.
- Real-time data streaming support from online meteorological services or government APIs.
- Enhanced UI elements such as interactive dashboards with drill-down capabilities and region-wise comparison.

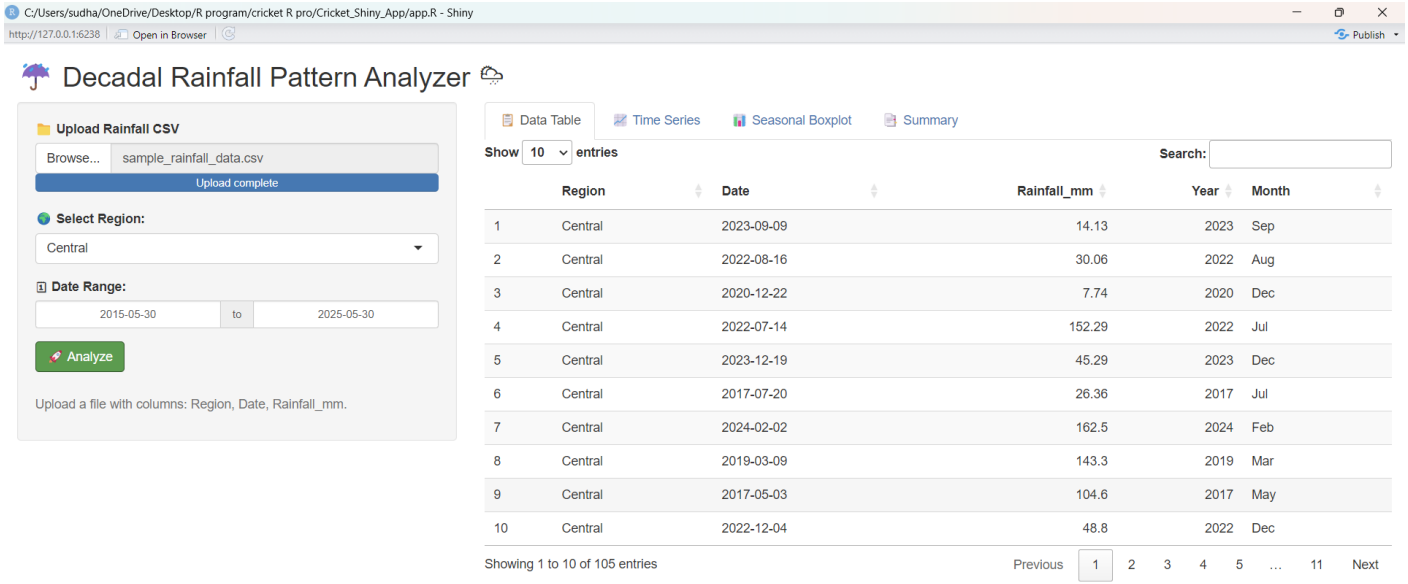
CHAPTER 5
Appendix A – Source Code

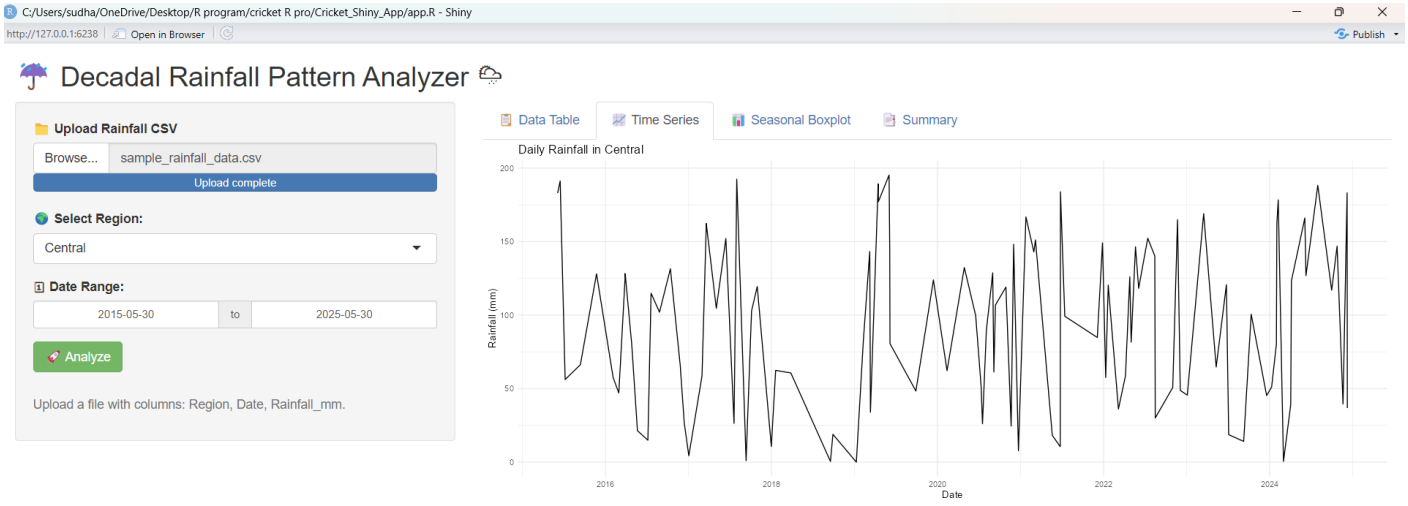
```
# app.R: Shiny App for Analyzing Rainfall Patterns Over the Last Decade in a Region
# Place this file and your rainfall_data.csv in the same directory, then run shiny::runApp(".")

library(shiny)
library(dplyr)
library(ggplot2)
library(lubridate)
library(DT)

ui <- fluidPage(
  titlePanel("☔ Decadal Rainfall Pattern Analyzer 🌦"),
  sidebarLayout(
    sidebarPanel(
      fileInput("file", "📁 Upload Rainfall CSV", accept = c(".csv")),
      selectInput("region", "🌎 Select Region:", choices = NULL),
      dateRangeInput("dates", "🗓 Date Range:",
                     start = today() - years(10), end = today()),
      actionButton("run", "🚀 Analyze", class = "btn-success"),
      br(), br(),
      helpText("Upload a file with columns: Region, Date, Rainfall_mm.")
    ),
    mainPanel(
      tabsetPanel(
        tabPanel("📋 Data Table", DTOutput("table")),
        tabPanel("📈 Time Series", plotOutput("tsPlot")),
        tabPanel("📊 Seasonal Boxplot", plotOutput("seasonPlot")),
        tabPanel("📑 Summary", verbatimTextOutput("summaryText"))
      )
    )
```

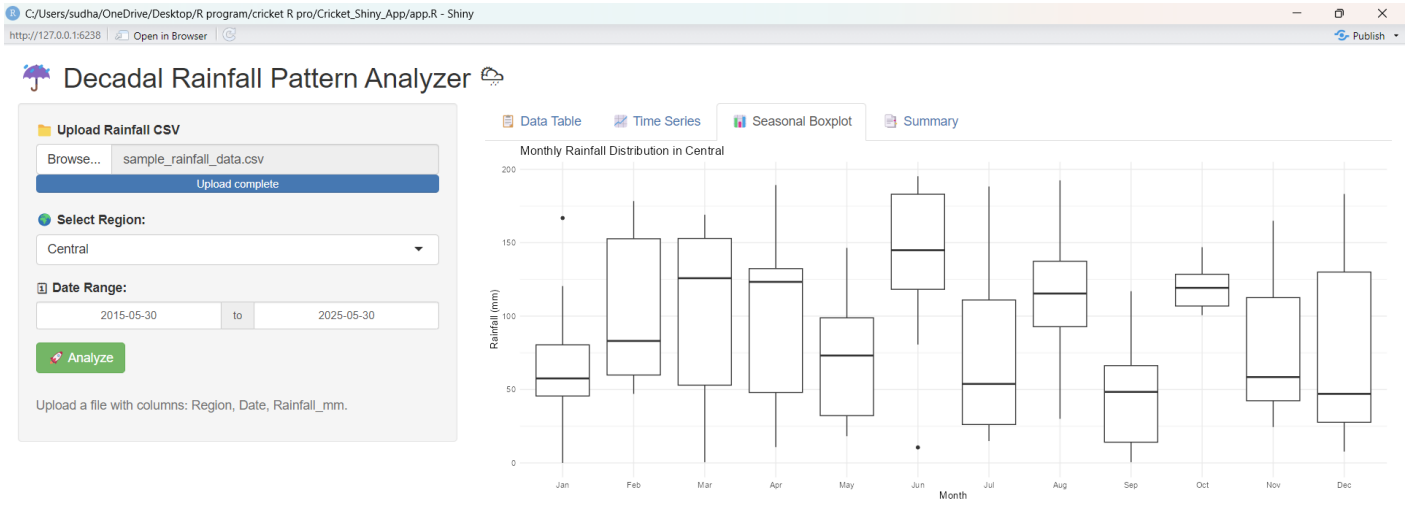
```
  )
)
server <- function(input, output, session) {
  rawData <- reactive({
    req(input$file)
    df <- read.csv(input$file$datapath, stringsAsFactors = FALSE)
    df$Date <- as.Date(df$Date)
    df
  })
  observeEvent(rawData(), {
    regions <- unique(rawData()$Region)
    updateSelectInput(session, "region", choices = regions)
  })
  filtered <- eventReactive(input$run, {
    df <- rawData() %>%
      filter(Region == input$region,
             Date >= input$dates[1], Date <= input$dates[2])
    df$Year <- year(df$Date)
    df$Month <- month(df$Date, label = TRUE)
    df
  })
  output$table <- renderDT({
    datatable(filtered(), options = list(pageLength = 10, scrollX = TRUE))
  })

  output$tsPlot <- renderPlot({
    df <- filtered() %>%
      group_by(Date) %>%
      summarize(Rainfall = sum(Rainfall_mm))
    ggplot(df, aes(x = Date, y = Rainfall)) +
      geom_line() +
      labs(
        title = paste("Daily Rainfall in", input$region),
        x = "Date", y = "Rainfall (mm)"
      ) +
```

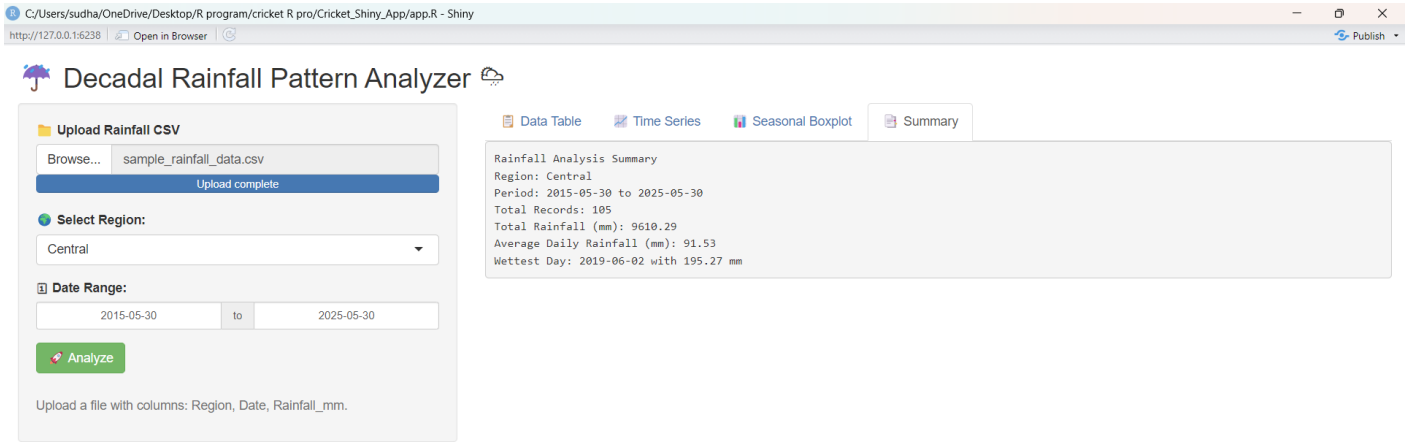
```
    theme_minimal()
  })

  output$seasonPlot <- renderPlot({
    df <- filtered()
    ggplot(df, aes(x = Month, y = Rainfall_mm)) +
      geom_boxplot() +
      labs(
        title = paste("Monthly Rainfall Distribution in", input$region),
        x = "Month", y = "Rainfall (mm)"
      ) +
      theme_minimal()
  })

  output$summaryText <- renderPrint({
    df <- filtered()
    cat("Rainfall Analysis Summary\n")
    cat("Region:", input$region, "\n")
    cat("Period:", format(input$dates[1]), "to", format(input$dates[2]), "\n")
    cat("Total Records:", nrow(df), "\n")
    cat("Total Rainfall (mm):", sum(df$Rainfall_mm), "\n")
    cat("Average Daily Rainfall (mm):", round(mean(df$Rainfall_mm), 2), "\n")
    cat("Wettest Day:", as.character(df$Date[which.max(df$Rainfall_mm)]), "with",
        max(df$Rainfall_mm), "mm\n")
  })
}

shinyApp(ui = ui, server = server)
```

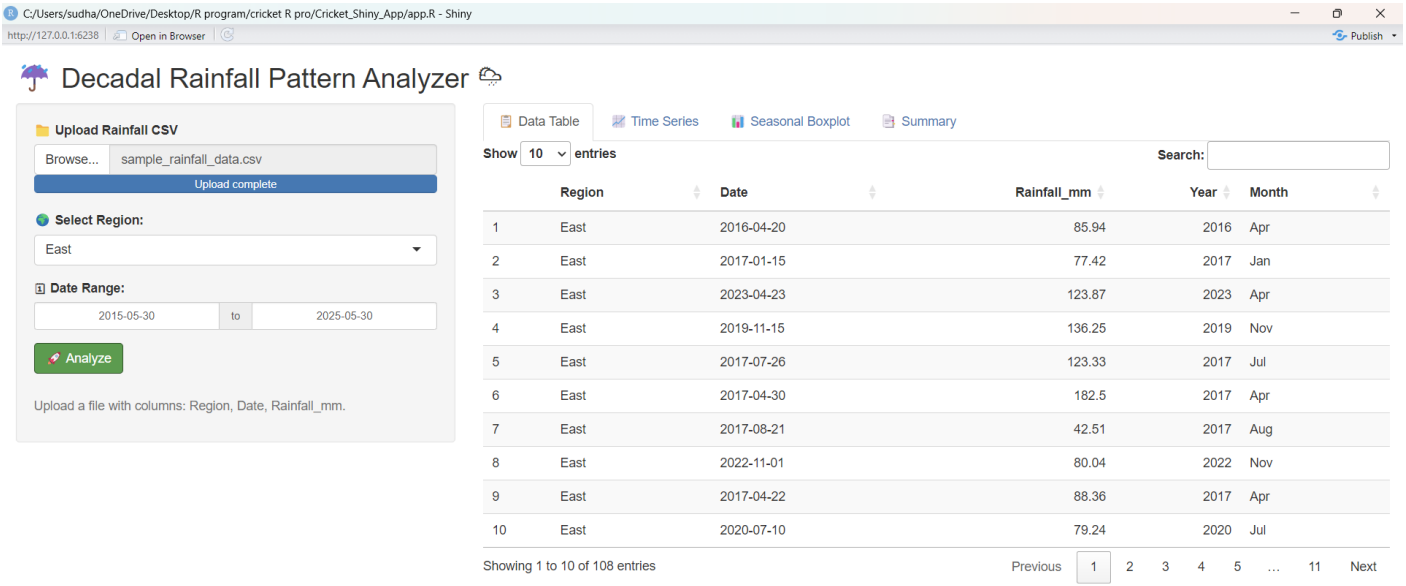
Appendix B – Screenshots






C:/Users/sudha/OneDrive/Desktop/R program/cricket R pro/Cricket_Shiny_App/app.R - Shiny
http://127.0.0.1:6238
Open in Browser
Publish
Decadal Rainfall Pattern Analyzer
Upload Rainfall CSV
Browse...
sample_rainfall_data.csv
Upload complete
Select Region:
Central
Date Range:
2015-05-30
to
2025-05-30
Analyze
Upload a file with columns: Region, Date, Rainfall_mm.
Data Table
Time Series
Seasonal Boxplot
Summary
Monthly Rainfall Distribution in Central
Rainfall (mm)
200
150
100
50
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month


C:/Users/sudha/OneDrive/Desktop/R program/cricket R pro/Cricket_Shiny_App/app.R - Shiny
http://127.0.0.1:6238
Open in Browser
Publish
Decadal Rainfall Pattern Analyzer
Upload Rainfall CSV
Browse...
sample_rainfall_data.csv
Upload complete
Select Region:
Central
Date Range:
2015-05-30
to
2025-05-30
Analyze
Upload a file with columns: Region, Date, Rainfall_mm.
Data Table
Time Series
Seasonal Boxplot
Summary
Rainfall Analysis Summary
Region: Central
Period: 2015-05-30 to 2025-05-30
Total Records: 105
Total Rainfall (mm): 9610.29
Average Daily Rainfall (mm): 91.53
Wettest Day: 2019-06-02 with 195.27 mm


C:/Users/sudha/OneDrive/Desktop/R program/cricket R pro/Cricket_Shiny_App/app.R - Shiny
http://127.0.0.1:6238
Open in Browser
Publish
Decadal Rainfall Pattern Analyzer
Upload Rainfall CSV
Browse...
sample_rainfall_data.csv
Upload complete
Select Region:
East
Date Range:
2015-05-30
to
2025-05-30
Analyze
Upload a file with columns: Region, Date, Rainfall_mm.
Data Table
Time Series
Seasonal Boxplot
Summary
Show 10 entries
Search:
Region
Date
Rainfall_mm
Year
Month
1 East 2016-04-20 85.94 2016 Apr
2 East 2017-01-15 77.42 2017 Jan
3 East 2023-04-23 123.87 2023 Apr
4 East 2019-11-15 136.25 2019 Nov
5 East 2017-07-26 123.33 2017 Jul
6 East 2017-04-30 182.5 2017 Apr
7 East 2017-08-21 42.51 2017 Aug
8 East 2022-11-01 80.04 2022 Nov
9 East 2017-04-22 88.36 2017 Apr
10 East 2020-07-10 79.24 2020 Jul
Showing 1 to 10 of 108 entries
Previous
1
2
3
4
5
…
11
Next

REFERENCES

1. Chang, W. et al. (2021). *Shiny: Web Application Framework for R*. RStudio.
2. Kaggle Rainfall Datasets – https://www.kaggle.com
3. Lubridate Documentation – https://lubridate.tidyverse.org
4. Wickham, H., & Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.