VISUALIZING SPORTS DATA: SCORED BY PLAYERS IN CRICKET

A PROJECT REPORT

Submitted by

SIBI CHAKKARAVARTHI S (2303811724321105)

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 **"VISUALIZING SPORTS DATA: RUNS SCORED BY PLAYERS IN CRICKET"** is the bonafide work of **SIBI CHAKKARAVARTHI 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 **"VISUALIZING SPORTS DATA: RUNS SCORED BY PLAYERS IN CRICKET"** 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

SIBI CHAKKARAVARTHI 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 a dynamic and interactive Shiny web application developed using the R programming language. The purpose of this application is to analyze and visualize sports performance data, specifically the runs scored by cricket players in the Indian Premier League (IPL). Cricket is one of the most data-rich sports, and this project harnesses IPL data to offer insights into individual and team performances through an intuitive interface. The application allows users to explore and compare player statistics, including total runs, seasonal performances, strike rates, and team-wise contributions. By leveraging tools such as ggplot2, dplyr, plotly, and Shiny, the project enables users to interactively filter, analyze, and visualize data. This dashboard highlights the practical utility of R programming in transforming raw datasets into meaningful analytics and provides a valuable platform for sports analysts, fans, and data science learners.

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 showcases an interactive Shiny web application built with R to analyze and visualize IPL cricket performance data, focusing on runs scored by players. It enables users to explore statistics like total runs, seasonal performance, strike rate, and team contributions. Using tools like ggplot2, dplyr, and plotly, the app provides a user-friendly platform for data analysis, offering valuable insights for sports analysts, fans, and learners in data science.	**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

The aim of this project is to develop an interactive dashboard that visualizes the batting performance of players in IPL matches. The system focuses primarily on runs scored by players and uses IPL datasets available in the public domain. With the increasing popularity of cricket and the availability of detailed match data, sports analytics has become an area of interest not only for professional analysts but also for aspiring data scientists. This project emphasizes how data science techniques can be used to enhance the understanding of sports performance through visual and interactive means.

11.2 Overview

The developed dashboard allows users to filter data by IPL season, team, and player name, and to view insights such as top scorers, team totals, and strike rate comparisons. The application is built using R's Shiny framework, which supports dynamic and user-friendly web interfaces. Through this platform, users can explore trends in batting statistics across seasons and among different teams, helping to identify consistent performers, emerging players, and team strategies.

11.3 R Programming Concepts Used

From a technical perspective, the project integrates several R packages. The data is manipulated using dplyr and tidyr, visualized through ggplot2 and plotly, and displayed in tables with the DT package. Interactive programming concepts like reactive expressions, dynamic UI rendering, and data-driven event handling are employed using Shiny functions. Overall, the project demonstrates the practical application of R in building real-time data dashboards for sports analytics.

CHAPTER 2
PROJECT METHODOLOGY

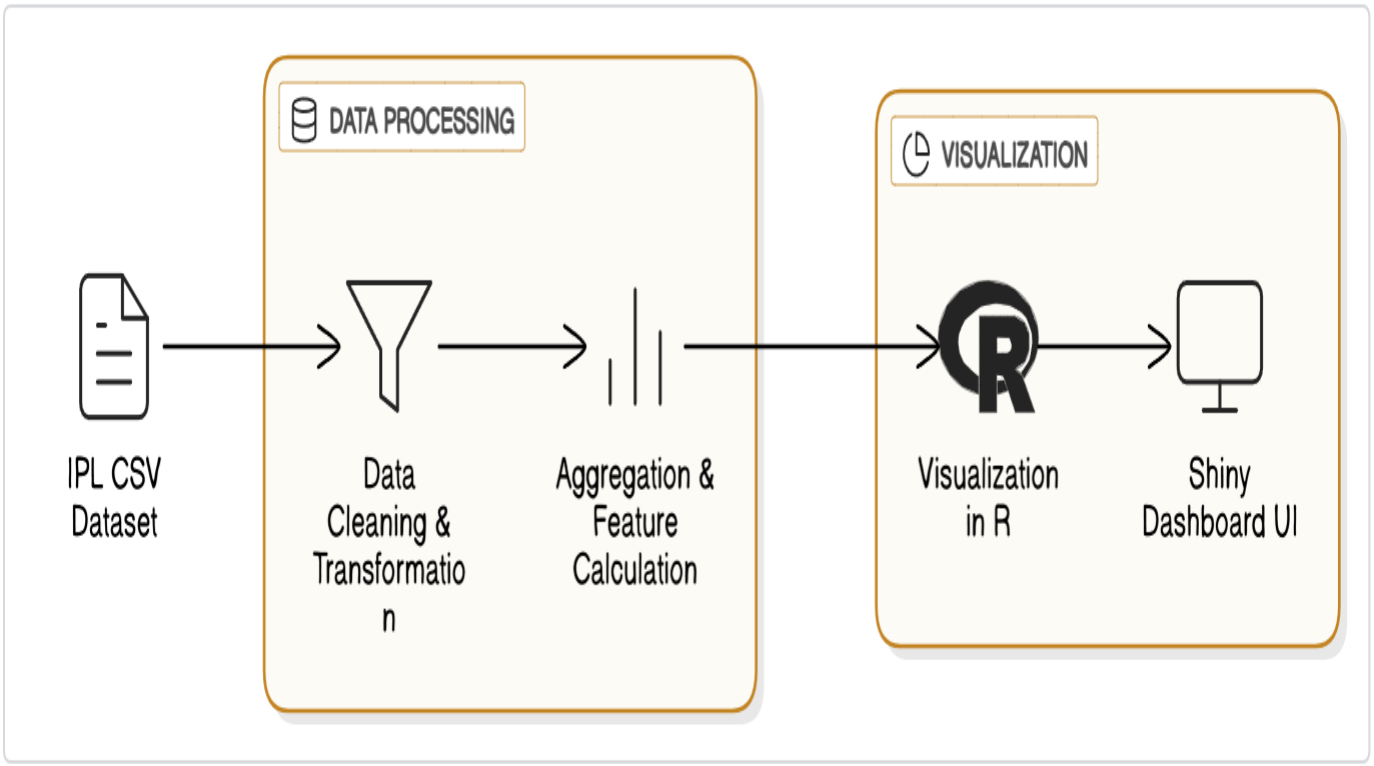
2.1 Proposed Work

The project methodology is structured around the core principles of data acquisition, cleaning, transformation, and visualization. Initially, data related to IPL matches and player statistics was sourced from Kaggle, a reliable open-source platform for datasets. The dataset includes ball-by-ball data with details such as match ID, player name, team, runs scored, and season.

The first step was to clean and preprocess the dataset. This involved removing null values, renaming inconsistent columns, converting data types (especially date and numeric fields), and preparing subsets of the data suitable for visual analysis. Once cleaned, the data was aggregated based on the analysis requirements. For instance, total runs were calculated per player per season, while strike rates were computed based on balls faced and runs scored.

The second phase involved designing visualizations using ggplot2 and plotly. Bar graphs were used to display top scorers, while line charts illustrated player trends across seasons. Interactive elements like dropdowns and sliders were incorporated using Shiny, allowing users to customize the data views. Finally, all the visual components were integrated into a Shiny UI, organized using layout elements such as fluidPage, sidebarLayout, and tabsetPanel. The result is a cohesive and visually appealing dashboard that updates outputs in real time based on user selections.

2.2 Block Diagram



CHAPTER 3
MODULE DESCRIPTION

3.1 Data Input Module

The Data Input Module, allows users to load IPL match datasets into the application. Once the data is uploaded, it undergoes preprocessing where missing values are handled, and fields are formatted appropriately. This ensures data consistency across all visualization components.

3.2 Filter and Aggregation Module

The Filter and Aggregation Module, which processes the dataset based on user input. Users can filter data by year, team, or player. Aggregations such as total runs, number of matches, and strike rates are computed using group_by() and summarise() functions from the dplyr package.

3.3 Top Scorer Detection Module

The Top Scorer Visualization Module, identifies and displays the top-performing players in terms of runs scored. The data is plotted in interactive bar charts using plotly, with the ability to hover over each bar to view specific details such as the player’s name and total runs.

3.4 Strike Rate & Team-wise Analysis Module

The Strike Rate and Team Analysis Module calculates the batting strike rate of players and presents team-wise performance comparisons. These visualizations help users understand not only the volume of runs scored but also the efficiency and consistency of players.

3.5 Visualization Module

The Shiny Dashboard Interface, which combines all visual outputs and filters into a single user interface. It uses tab-based navigation to separate different views such as raw data tables, graphs, and summary statistics. Interactive elements are linked to the server logic through reactive programming, ensuring that all plots and tables update based on user input.

CHAPTER 4
CONCLUSION & FUTURE SCOPE

Conclusion

This project demonstrates the power and flexibility of R and Shiny for building interactive data applications. The dashboard effectively translates raw IPL match data into insightful visualizations that allow users to explore and analyze player performance. It provides a platform for discovering trends, comparing players, and evaluating team strategies based on runs scored and strike rate metrics.

Future Scope

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

- Integration of additional batting metrics such as boundaries (fours and sixes), batting average, and dismissal types to enrich the player profile analysis.
- Expansion of the dashboard to include bowling statistics, such as wickets taken, economy rate, and bowling strike rate, offering a complete player performance overview.
- Incorporation of real-time data streaming from live IPL APIs to analyze ongoing matches dynamically.
- Enhancement of the user interface with advanced visualizations, including heatmaps, radar charts, and performance comparisons across seasons.
- Implementation of predictive modeling techniques to forecast player performance based on historical trends.

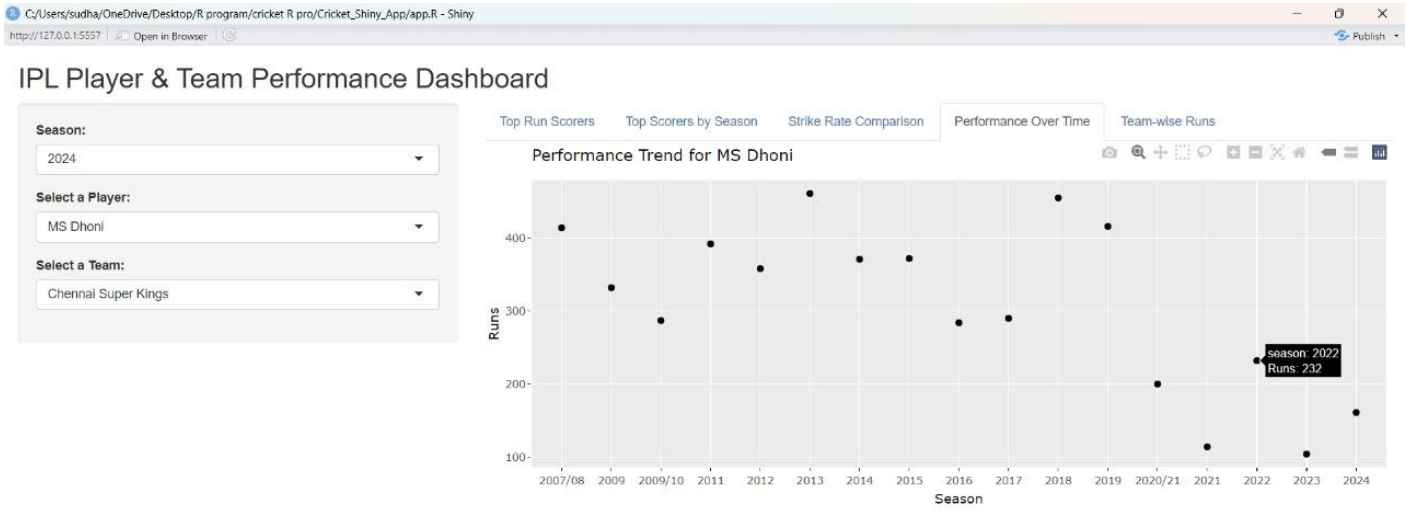
CHAPTER 5
Appendix A – Source Code

```
library(shiny)
library(dplyr)`
library(ggplot2)
library(plotly)
library(readr)

matches <- read_csv("matches.csv")
deliveries <- read_csv("deliveries.csv")

ui <- fluidPage(
  titlePanel("IPL Player & Team Performance Dashboard"),
  sidebarLayout(
    sidebarPanel(
      selectInput("season", "Season:", choices = unique(matches$season), selected =
max(matches$season)),
      selectInput("player", "Select a Player:", choices = unique(deliveries$batter)),
      selectInput("team", "Select a Team:", choices = unique(deliveries$batting_team))
    ),
    mainPanel(
      tabsetPanel(
        tabPanel("Top Run Scorers", plotlyOutput("topScorers")),
        tabPanel("Top Scorers by Season", plotlyOutput("scorersBySeason")),
        tabPanel("Strike Rate Comparison", plotlyOutput("strikeRate")),
        tabPanel("Performance Over Time", plotlyOutput("trendPlot")),
        tabPanel("Team-wise Runs", plotlyOutput("teamRuns"))
      )
    )
  )
)

server <- function(input, output) {

  output$topScorers <- renderPlotly({
    top_runs <- deliveries %>%
      group_by(batter) %>%
      summarise(Runs = sum(batsman_runs)) %>%
      arrange(desc(Runs)) %>%
      head(10)
```

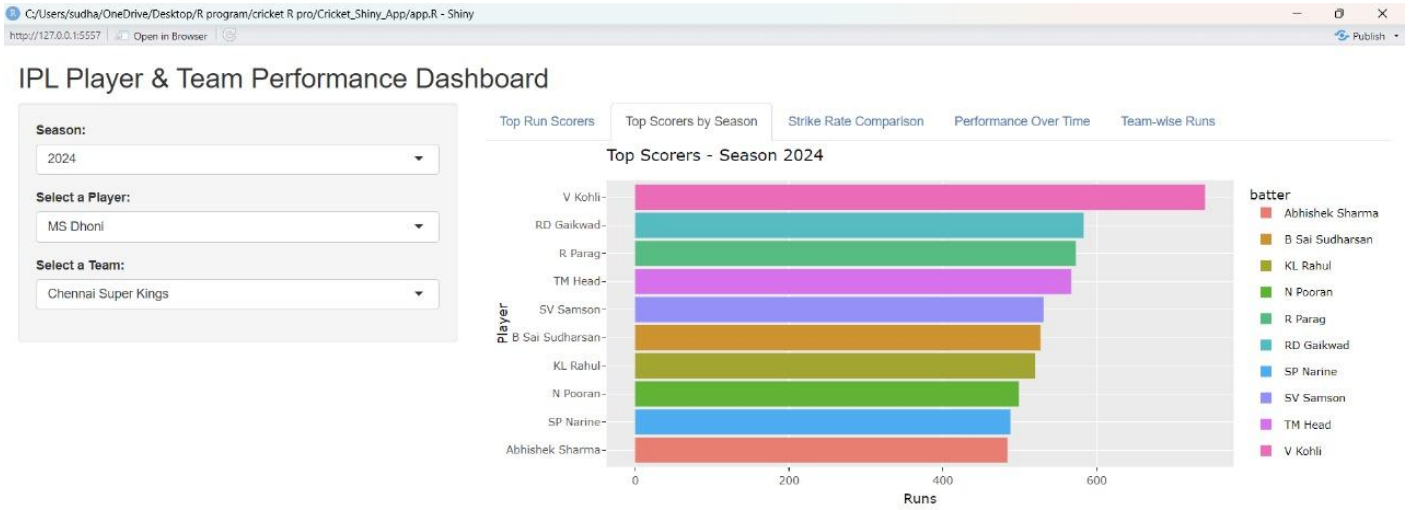
```
    gg <- ggplot(top_runs, aes(x = reorder(batter, Runs), y = Runs, fill = batter)) +
      geom_col() +
      coord_flip() +
      labs(title = "Top 10 Run Scorers", x = "Player", y = "Total Runs")

    ggplotly(gg)
  })

  output$scorersBySeason <- renderPlotly({
    top_season <- matches %>%
      filter(season == input$season) %>%
      select(id) %>%
      inner_join(deliveries, by = c("id" = "match_id")) %>%
      group_by(batter) %>%
      summarise(Runs = sum(batsman_runs)) %>%
      arrange(desc(Runs)) %>%
      head(10)

    gg <- ggplot(top_season, aes(x = reorder(batter, Runs), y = Runs, fill = batter)) +
      geom_col() +
      coord_flip() +
      labs(title = paste("Top Scorers - Season", input$season), x = "Player", y = "Runs")

    ggplotly(gg)
  })

  output$strikeRate <- renderPlotly({
    sr_data <- deliveries %>%
      group_by(batter) %>%
      summarise(Runs = sum(batsman_runs), Balls = n()) %>%
      mutate(StrikeRate = round((Runs / Balls) * 100, 2)) %>%
      arrange(desc(StrikeRate)) %>%
      filter(Runs > 200) %>%
      head(10)

    gg <- ggplot(sr_data, aes(x = reorder(batter, StrikeRate), y = StrikeRate, fill =
batter)) +
      geom_col() +
      coord_flip() +
      labs(title = "Top Strike Rates (Min 200 Runs)", x = "Player", y = "Strike Rate")

    ggplotly(gg)
  })
```

```
  output$trendPlot <- renderPlotly({
    trends <- matches %>%
      inner_join(deliveries, by = c("id" = "match_id")) %>%
      filter(batter == input$player) %>%
      group_by(season) %>%
      summarise(Runs = sum(batsman_runs))

    gg <- ggplot(trends, aes(x = season, y = Runs)) +
      geom_line(color = "blue") +
      geom_point() +
      labs(title = paste("Performance Trend for", input$player), x = "Season", y =
"Runs")

    ggplotly(gg)
  })

  output$teamRuns <- renderPlotly({
    team_data <- deliveries %>%
      filter(batting_team == input$team) %>%
      group_by(batter) %>%
      summarise(Runs = sum(batsman_runs)) %>%
      arrange(desc(Runs)) %>%
      head(10)

    gg <- ggplot(team_data, aes(x = reorder(batter, Runs), y = Runs, fill = batter)) +
      geom_col() +
      coord_flip() +
      labs(title = paste("Top Scorers for", input$team), x = "Player", y = "Runs")

    ggplotly(gg)
  })
}

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://127.0.0.1:5557
Open in Browser
Publish
IPL Player & Team Performance Dashboard
Season:
2024
Select a Player:
MS Dhoni
Select a Team:
Chennai Super Kings
Top Run Scorers
Top Scorers by Season
Strike Rate Comparison
Performance Over Time
Team-wise Runs
Top Strike Rates (Min 200 Runs)
J Fraser-McGurk
WG Jacks
PD Salt
T Stubbs
TM Head
AD Russell
BCJ Cutting
H Klaasen
TH David
SP Narine
Player
0
50
100
150
200
Strike Rate
batter
AD Russell
BCJ Cutting
H Klaasen
J Fraser-McGurk
PD Salt
SP Narine
T Stubbs
TH David
TM Head
WG Jacks


C:/Users/sudha/OneDrive/Desktop/R program/cricket R pro/Cricket_Shiny_App/app.R - Shiny
http://127.0.0.1:5557
Open in Browser
Publish
IPL Player & Team Performance Dashboard
Season:
2024
Select a Player:
MS Dhoni
Select a Team:
Chennai Super Kings
Top Run Scorers
Top Scorers by Season
Strike Rate Comparison
Performance Over Time
Team-wise Runs
Top Scorers - Season 2024
V Kohli
RD Gaikwad
R Parag
TM Head
SV Samson
B Sai Sudharsan
KL Rahul
N Pooran
SP Narine
Abhishek Sharma
Player
0
200
400
600
Runs
batter
Abhishek Sharma
B Sai Sudharsan
KL Rahul
N Pooran
R Parag
RD Gaikwad
SP Narine
SV Samson
TM Head
V Kohli


C:/Users/sudha/OneDrive/Desktop/R program/cricket R pro/Cricket_Shiny_App/app.R - Shiny
http://127.0.0.1:5557
Open in Browser
Publish
IPL Player & Team Performance Dashboard
Season:
2024
Select a Player:
SC Ganguly
Select a Team:
Kolkata Knight Riders
Top Run Scorers
Top Scorers by Season
Strike Rate Comparison
Performance Over Time
Team-wise Runs
Top 10 Run Scorers
V Kohli
S Dhawan
RG Sharma
DA Warner
SK Raina
MS Dhoni
AB de Villiers
CH Gayle
RV Uthappa
KD Karthik
Player
0
2000
4000
6000
8000
Total Runs
batter
reorder(batter, Runs): S Dhawan
Runs: 6769
batter: S Dhawan
AB de Villiers
CH Gayle
DA Warner
KD Karthik
MS Dhoni
RG Sharma
RV Uthappa
S Dhawan
SK Raina
V Kohli

REFERENCES

1. Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2021). *Shiny: Web Application Framework for R*. RStudio.
2. Kaggle IPL Dataset. Retrieved from https://www.kaggle.com
3. Wickham, H., & Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.