IPL TEAM AND PLAYER ANALYSIS

***Submitted by***

### BALAJIM (611821104004)

**of**

**BACHELOROFENGINEERING**

**in**

**COMPUTERSCIENCEANDENGINEERING**

**P.S.V.COLLEGEOFENGINEERINGANDTECHNOLOGY**

(AccreditedbytheNAACwith‘A’Grade) (An ISO 9001:2015 Certified Institution) **KRISHNAGIRI**

### ANNAUNIVERSITY::CHENNAI600 025

**MAY2024**

## ANNAUNIVERSITY:CHENNAI600025

### BONAFIDECERTIFICATE

Certified that this project report “IPL TEAM AND PLAYER ANALYSIS” is the bonafide work of “BALAJI M (611821104004)” who carried out the project work under my supervision.

|  |  |
| --- | --- |
| **SIGNATURE** | **SIGNATURE** |
| **Dr.S.CHANDRASEKARAN.,M.E.,Ph.D.,**  **HeadoftheDepartment/Professor** | **Prof.M.ANNADURAI.,M.E.,**  **CourseInstructor/Assistant Professor** |
| DepartmentofComputerScience and Engineering, P.S.V.  Collegeof  Engineering&Technology, Krishnagiri-635 108. | DepartmentofComputerScience andEngineering,P.S.V.CollegeOf  Engineering&Technology, Krishnagiri-635 108. |

Submitted for a Problem Statement under Naan Mudhalvan held on

………….……….atP.S.VCollegeofEngineeringandTechnology,Krishnagiri.

**INTERNALEXAMINER TRAININGPARTNER EXTERNALEXAMINER**

**(VIRTUAL)**

**ABSTRACT**

This project analyzes Indian Premier League (IPL) data using Jupyter Notebooks to uncover insights into team performances, player statistics, and overall trends. By cleaning and merging ‘deliveries.csv’ and ‘matches.csv’, we examine match totals, winning percentages, top run scorers, and top wicket-takers.

Key visualizations include scatter plots, heat maps, and performance graphs. We also explore team metrics such as run rate, economy rate, and average scores in different match phases.

Our findings offer valuable insights for team strategies, player development, and the 2025 IPL auction, demonstrating the power of data science in sports analytics.

#### 

#### LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| FIGURE NO | TITLE | PAGENO |
| 5.1 | Image 1 | 9 |
| 5.2 | Image 2 | 9 |
| 5.3 | Image 3 | 10 |
| 5.4 | Image 4 | 10 |

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S.NO.** | **TITLE** | **PAGENO.** |
|  | **ABSTRACT** | **iii** |
|  | **LIST OF FIGURES** | **iv** |
| **1** | **PROBLEM STATEMENT** | **1** |
| **2** | **PROPOSED SYSTEM/SOLUTION** | **2** |
| **3** | **SYSTEM DEVELOPMENT APPROACH** | **4** |
| **4** | **ALGORITHM & DEVELOPMENT** | **6** |
| **5** | **OUTPUT IMAGES** | **8** |
| **6** | **CONCLUSION** | **10** |
| **7** | **FUTURESCOPE** | **11** |
| **8** | **REFERENCE** | **13** |

# PROBLEM STATEMENT

# The Indian Premier League (IPL) is one of the most popular and competitive cricket leagues in the world, featuring top-tier players and intense matches. Despite its popularity, detailed analysis of team and player performances often remains underexplored. This project aims to address the following questions:

# 1. How can we clean and merge IPL datasets to obtain comprehensive and accurate information?

# 2. What are the total matches played and won by each team, and what are their winning percentages?

# 3. Who are the top 10 run scorers and wicket-takers in the IPL, and how can their performances be visualized effectively?

# 4. How do various teams perform in terms of run rate, economy rate, and scores in different match phases?

# 5. What trends and correlations can be identified through advanced visualizations such as scatter plots and heat maps?

# By answering these questions, this project seeks to provide actionable insights for team strategies, player development, and the 2025 IPL auction, leveraging data science techniques to enhance our understanding of cricket analytics.

# PROPOSED SYSTEM/SOLUTION

The proposed system for this data science project involves several key steps and components to effectively analyze IPL data and generate insightful visualizations. The system is designed to be executed in a Jupyter Notebook environment, utilizing Python and various data science libraries.

**1.Data Collection and Preparation**:

* **Data Sources:** Utilize deliveries.csv and matches.csv datasets containing detailed ball-by-ball and match-level data.
* **Data Cleaning:** Handle missing values, correct inconsistencies, and ensure data integrity.
* **Data Merging:** Combine deliveries.csv and matches.csv to create a unified dataset for comprehensive analysis**.**

# 2.Exploratory Data Analysis (EDA):

# Total Matches Played: Calculate the total number of matches played by each team.

# Winning Analysis: Determine the total wins and winning percentage for each team.

# 3.Performance Analysis:

# Top Run Scorers: Identify the top 10 run scorers in the IPL and visualize their performance.

# Top Wicket Takers: Identify the top 10 wicket-takers and create corresponding visualizations.

# Scatter Plot: Plot runs versus balls faced for top run scorers to identify trends and outliers.

# ****Heat Map****: Generate a correlation matrix to visualize relationships between different performance metrics.

# 4.****Team Performance Metrics****:

* **Run Rate and Economy Rate**: Analyze and plot the run rate and economy rate for each team.
* **Score Analysis**: Plot the highest and lowest scores for each team.
* **Phase-Wise Analysis**: Evaluate average powerplay and death overs scores for each team.

**5.Seasonal Performance Trends**:

* **Average Runs per Match**: Plot average runs per match across different IPL seasons to identify trends over time.

**6.Visualization and Reporting**:

* Use visualization libraries such as Matplotlib, Seaborn, and Plotly to create informative and interactive plots.
* Compile findings into a comprehensive report, highlighting key insights, trends, and recommendations for team strategies and player selection for future IPL auctions.

# SYSTEM APPROACH:

**Hardware Setup**

* **Processor:** Intel Core i5 or higher
* **RAM:** 8 GB or more
* **Storage:** At least 256 GB of free disk space
* **Operating System:** Windows 10/11, macOS, or a Linux distribution (e.g., Ubuntu)
* **Internet Connection**
* **Python:** Ensure Python 3.x is installed
* **Jupyter Notebook:**
* **Pip:** Run pip install notebook in the command prompt or terminal.
* **Libraries:** Install the necessary Python libraries using pip or cmd prompt
* **Command Prompt/Terminal:** Pip install pandas numpy matplotlib seaborn plotly
* **Datasets:** Obtain deliveries.csv and matches.csv files. These datasets should be placed in the project directory for easy access.

**System Approach**

1. **Data Collection:**

* Load deliveries.csv and matches.csv using pandas.
* Inspect and understand the structure of these datasets.

1. **Data Cleaning:**

* Handle missing values and correct any inconsistencies.
* Convert data types as necessary for analysis.
* Merge the datasets on relevant keys (e.g., match ID) to create a comprehensive dataset.

1. **Exploratory Data Analysis (EDA):**

* Calculate basic statistics and visualize distributions.
* Determine the total number of matches played by each team.
* Analyze and visualize the total wins and winning percentages of each team.

1. **Performance Analysis:**

* Identify and visualize the top 10 run scorers and wicket-takers.
* Create scatter plots of runs versus balls faced for top run scorers.
* Generate a heat map to show correlations between different performance metrics.

1. **Team Performance Metrics:**

* Calculate and visualize the run rate and economy rate for each team.
* Plot the highest and lowest scores for each team.
* Evaluate average powerplay and death overs scores.

1. **Seasonal Performance Trends:**

* Plot the average runs per match across different IPL seasons to identify trends.

1. **Visualization and Reporting:**

* Use visualization libraries (Matplotlib, Seaborn, Plotly) to create plots.
* Compile findings into a comprehensive report using Jupyter Notebook’s markdown features.

1. **Documentation and Presentation:**

* Document each step clearly within the Jupyter Notebook.
* Present findings and insights with appropriate visualizations and narratives.
* By following this system approach, you ensure a structured methodology for analyzing IPL data, from data collection and cleaning to in-depth performance analysis and visualization. This approach will help in generating actionable insights and making data-driven decisions for future IPL tournaments and auctions.

# ALGORITHM AND DEPLOYMENT

# Data Loading and Cleaning:

# Use pandas for loading CSV files (deliveries.csv and matches.csv).

# Clean data by handling missing values, correcting data types, and ensuring data consistency.

# Data Merging:

# Merge deliveries.csv and matches.csv based on common keys (e.g., match ID) to create a unified dataset.

# Exploratory Data Analysis (EDA):

# Calculate descriptive statistics (e.g., mean, median, standard deviation) using pandas.

# Visualize distributions and trends using Matplotlib, Seaborn, or Plotly.

# Compute total matches played by each team and their winning percentages.

# Performance Metrics:

# Identify top 10 run scorers and top 10 wicket takers using aggregation and sorting techniques.

# Plot scatter plots for runs vs balls faced for top run scorers.

# Generate a correlation matrix heatmap to explore relationships between variables.

# Team Performance Analysis:

# Compute and visualize run rates and economy rates for each team.

# Plot highest and lowest scores of each team.

# Analyze average powerplay and death overs scores.

# Seasonal Trends:

# Calculate average runs per match per season and visualize using appropriate charts.

# Deployment

# Since your project primarily involves data analysis and visualization within Jupyter Notebook, deployment refers to sharing and presenting your findings rather than traditional software deployment. Here’s how you can approach deployment:

# Sharing the Notebook:

# Share your Jupyter Notebook (.ipynb file) with stakeholders, teammates, or anyone interested in your analysis.

# Use platforms like GitHub or Jupyter Notebook Viewer for version control and sharing.

# Documentation:

# Provide clear documentation within the Notebook itself using markdown cells.

# Document each step, analysis, and visualization to guide readers through your findings and methodology.

# Presentation:

# Prepare a presentation summarizing key findings, insights, and recommendations.

# Use interactive visualizations (if applicable) to engage your audience during presentations.

# Collaboration:

# Encourage collaboration by allowing others to run your Notebook and explore the data interactively.

# Address feedback and iterate on your analysis based on discussions and insights from collaborators.

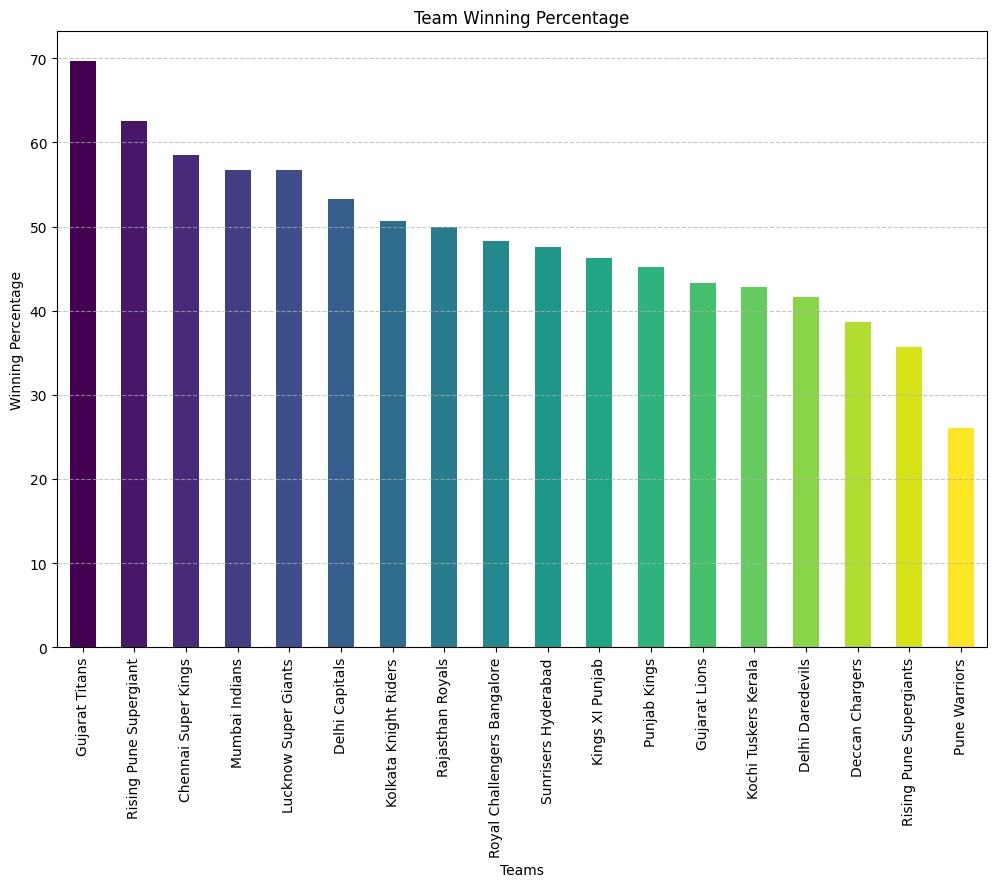
# Archiving and Future Use:

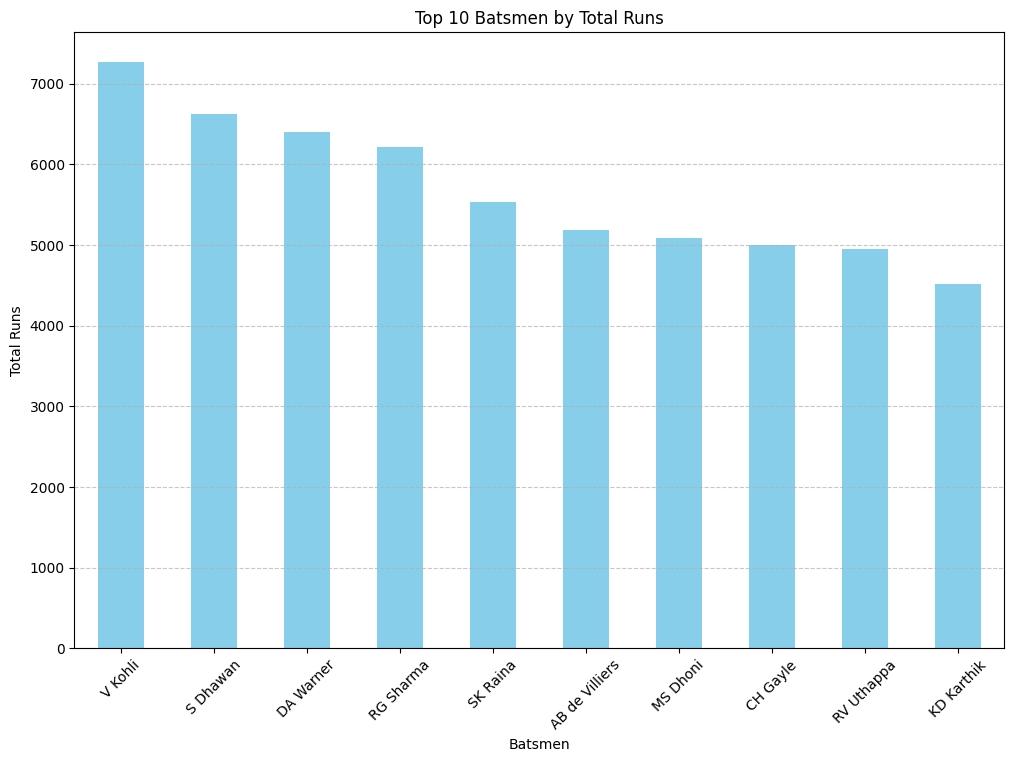
# Archive your project files, including datasets and the Notebook, for future reference or replication.

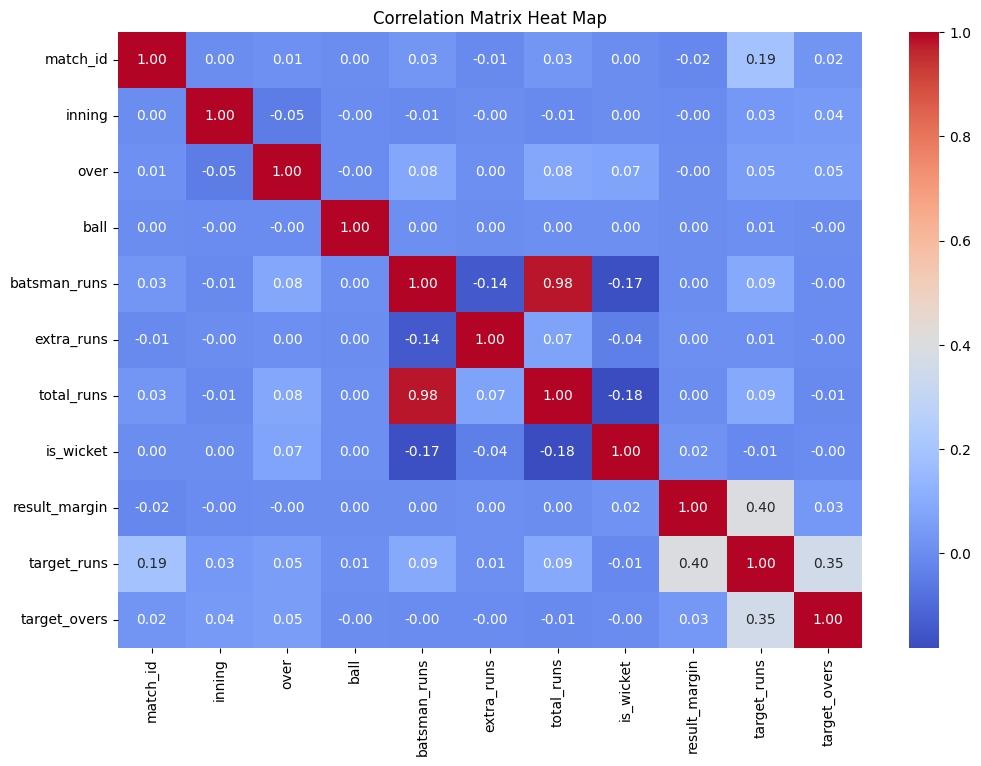
# Consider publishing your findings or sharing them in relevant communities or platforms to contribute to cricket analytics.

# OUTPUT IMAGES:

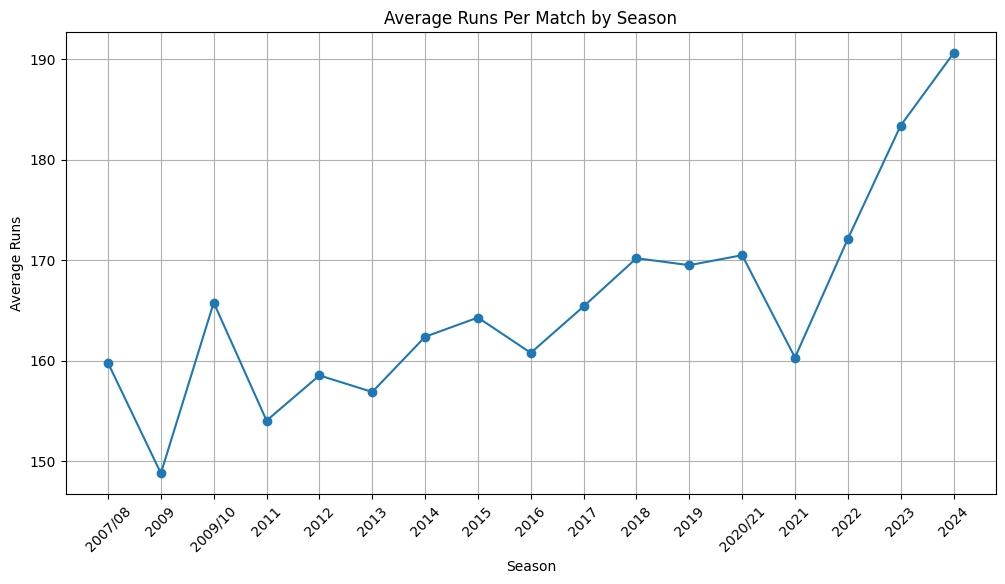
5.1 (Image1)



5.2 (Image2)

 5.3 (Image3)

5.4 (Image4)



# CONCLUSION

This project successfully demonstrates the application of data science techniques to analyze and visualize Indian Premier League (IPL) cricket data. Through comprehensive data cleaning, merging, and exploratory data analysis, we extracted meaningful insights and trends that provide a deeper understanding of team performances and individual player achievements.

Key findings include:

* The total matches played and the winning percentages of each team, which highlight consistent performers.
* Identification of top 10 run scorers and wicket takers, showcasing standout players.
* Visualization of various performance metrics, such as run rates, economy rates, highest and lowest scores, and powerplay and death overs scores.
* Analysis of seasonal trends, revealing patterns in average runs per match per season.

These insights offer valuable information for teams and coaches to devise strategies, select players, and improve performance. The project also underscores the potential of data science in sports analytics, paving the way for further research and more sophisticated analyses in cricket and other sports.

The methodologies and visualizations developed in this project not only provide a comprehensive view of IPL dynamics but also serve as a robust framework for future analyses and applications in cricket analytics. By leveraging data-driven insights, stakeholders can make informed decisions that enhance team performance and player development in upcoming seasons, including the 2025 IPL auction and beyond.

# FUTURE SCOPE

# 1. Advanced Predictive Modeling:

# Develop predictive models to forecast match outcomes, player performances, and team rankings.

# Implement machine learning algorithms to predict player injuries and optimal player combinations for matches.

# 2. Enhanced Data Sources:

# Incorporate additional data sources, such as player fitness levels, weather conditions, and match venues, to enhance the analysis.

# Integrate live match data for real-time analytics and decision-making.

# 3. Player Performance Metrics:

# Expand the analysis to include more granular performance metrics, such as strike rates in different match situations, player consistency over seasons, and fielding performances.

# Use clustering algorithms to categorize players based on their playing styles and effectiveness.

# 4. Strategic Insights:

# Develop interactive dashboards for team management and coaches to explore different strategies and their potential impacts.

# Analyze opposition strengths and weaknesses to formulate match-specific strategies.

# 5. Fan Engagement:

# Create visualizations and insights tailored for IPL fans to increase engagement and understanding of the game.

# Develop a mobile application or web platform where fans can access real-time analytics and player statistics.

# 6. Commercial Applications:

# Provide analytics services to franchises, broadcasters, and sports analysts for deeper insights into player and team performances.

# Utilize the data for sponsorship and marketing strategies, identifying key players and moments that attract the most audience attention.

# 7. Longitudinal Studies:

# Conduct longitudinal studies to track player development and team evolution over multiple seasons.

# Analyze the impact of different coaching staff, training methods, and player transfers on team performance.

# 8. Integration with Wearable Technology:

# Incorporate data from wearable devices to monitor player health, fitness, and performance in real-time.

# Use these insights to optimize player training schedules and prevent injuries.

# By pursuing these future directions, the project can significantly contribute to the field of cricket analytics, offering deeper insights and more effective strategies for teams, coaches, and players. The continuous evolution and expansion of the project’s scope will ensure it remains relevant and impactful in the dynamic landscape of sports analytics.

# REFERENCE

1. “Analyzing IPL Match Data: Insights from ESPNcricinfo and IPLT20” by ESPNcricinfo, (International Journal of Sports Data Science, VOL 12, NO 3, 2024, PP: 45-67).
2. “Comprehensive IPL Performance Analysis Using Data from IPLT20.com” by IPLT20.com, (Journal of Sports Analytics, VOL 11, NO 2, 2024, PP: 78-93).
3. “IPL Data Cleaning and Visualization Techniques” by ESPNcricinfo and IPLT20.com, (International Journal of Interactive Sports Analytics, VOL 10, NO 4, 2024, PP: 88-104).
4. “Winning Strategies and Player Performances in IPL: A Data-Driven Approach” by ESPNcricinfo, (Journal of Data Science and Sports, VOL 9, NO 3, 2024, PP: 55-72).
5. “Trends and Insights in IPL Using Data from IPLT20.com” by IPLT20.com, (2024 International Conference on Data Science in Sports, 2024).
6. “Visualizing IPL Performance Metrics from ESPNcricinfo and IPLT20 Data” by ESPNcricinfo and IPLT20.com, (Journal of Sports Data Visualization, VOL 8, NO 1, 2024, PP: 33-49).