
✅ IPL Cricket Data Analysis Project — Final Report

🔗 Project Objective:

The primary goal of this project was to perform comprehensive Exploratory Data Analysis (EDA) on IPL cricket data from **2007 to 2024**, providing in-depth insights into team and player performances, match outcomes, venue impacts, and overall trends using data-driven visualizations.

📁 Datasets Used:

1. **Raw (Uncleaned) Datasets:**
 - matches.csv — Match-level data
 - deliveries.csv — Ball-by-ball delivery data
 2. **Cleaned Datasets** (created through preprocessing):
 - Updated_Matches.csv
 - Updated_Deliveries.csv
 - No_result_matches.csv
 - Duckworth_lewis_matches.csv
 - Super_over_matches.csv
-

⚙️ Data Preprocessing & Cleaning Steps:

- **Missing values** were identified and either removed or flagged for special handling (example: no-result matches).
- Created **separate datasets** for:
 - Matches with Duckworth-Lewis method applied
 - Matches ending in Super Overs
 - No-result matches
- Renamed columns and standardized team/player names where necessary for consistency.

- Converted columns to appropriate data types (e.g., dates, integers, categories).
 - Merged matches and deliveries datasets to enable venue-based and player-season-based analyses.
-

✓ Analysis Workflow Followed:

1 Team-Level Analysis:

- Counted team-wise total wins and identified the **most successful team**.
 - Found **season-wise winning teams** with the most victories.
 - Analyzed **highest and lowest team scores** in IPL history.
 - Conducted **venue-wise team performance analysis** (both in terms of runs and wickets).
 - Visualized this data with horizontal bar plots and combined bar charts.
-

2 Batting Performance Analysis:

- Identified top players with:
 - **Most runs**
 - **Most fours**
 - **Most sixes**
 - **Highest individual scores per season**
 - **Total balls faced**
 - Created player-wise **heatmaps** to show performance across different venues.
 - Tracked batting **performance over seasons** using line plots for total runs and strike rate evolution.
-

3 Bowling Performance Analysis:

- Identified top bowlers with:
 - **Most wickets**
 - Best **economy rates** (overall and by season)
- Special analysis for **powerplay** and **death overs** bowling performance.

- Created **venue-wise wicket heatmaps** for top-performing bowlers.
 - Analyzed bowling consistency over seasons.
-

4 Match Outcome & Toss Decision Analysis:

- Investigated how often the toss winner was also the match winner (calculated win percentage).
 - Analyzed the effect of toss decisions (bat or field first) on match outcomes.
 - Created bar charts and percentages for toss impact.
-

5 Venue & Umpire Statistics:

- Analyzed:
 - The venues with **highest team totals** and performance consistency.
 - **Umpires** with the most matches officiated.
 - Umpires involved in **no-result** matches.
 - Visualized umpire statistics and venue impact graphs.
-

Visualization Techniques Used:

- Horizontal & vertical bar charts with annotations for clarity.
 - Heatmaps for player performance (runs and wickets) by venue.
 - Dual bar charts for highest and lowest team scores.
 - Combined plots for team runs and wickets across venues.
 - Seasonal line plots for player runs, strike rate, wickets, and economy rates.
-

⚙️ Tools and Libraries Used:

- **Python**
- **Jupyter Notebook**
- **Libraries:**
 - pandas for data manipulation

- numpy for calculations
 - matplotlib and seaborn for data visualization
 - Git & GitHub for version control and project hosting
-



Key Insights Discovered:

- **Mumbai Indians** emerged as the most successful IPL team.
 - Eden Gardens, Wankhede Stadium, and M. Chinnaswamy Stadium are historically **high-scoring venues**.
 - **Virat Kohli**, **AB de Villiers**, and **Chris Gayle** top the batting charts in runs and strike rates.
 - **Lasith Malinga** and **Dwayne Bravo** dominate the wicket-taking charts.
 - Toss-winning teams have a **50.9% probability** of also winning the match, showing a slight advantage.
 - Batting-friendly venues show clear favoritism toward certain players' peak performances.
-



Project Structure Created:

- **Notebooks:** Single end-to-end notebook (Cricket_Project.ipynb) with structured sections and markdown explanations.
 - **Cleaned Datasets:** Separate folder for post-cleaning datasets.
 - **Uncleaned Datasets:** Separate folder for raw datasets.
 - **README.md:** Detailed documentation with setup instructions, project overview, and visual summaries.
 - **.gitignore:** To exclude unnecessary files from version control.
 - **requirements.txt:** For environment setup and reproducibility.
-



Data Handling Techniques Used:

- Groupby aggregations, pivot tables, merges for linking deliveries and match data.
- Outlier detection in scores and wickets to highlight peak performance matches.
- String manipulation for standardizing team names and venues.

- Conditional filtering to identify matches affected by Duckworth-Lewis, super overs, and no-results.
-

🔧 Version Control & Hosting:

- Entire project is version-controlled using Git.
 - Hosted on GitHub with structured folders, markdown documentation, and dataset separation.
-

🚀 Possible Future Extensions:

- Predictive modeling for match outcomes.
 - Player form tracking using rolling averages.
 - Interactive dashboards in Power BI or Tableau.
 - Inclusion of Women's IPL data for comparative analysis.
-

★ Conclusion:

This project successfully provides a full-fledged IPL cricket data analysis with well-structured datasets, insightful visualizations, and player/team-centric findings. The data cleaning, handling, and presentation methodology are robust, making the notebook easily understandable and reusable for presentations, portfolio building, and future model development.
