

PROJECT WRITEUP

IPL Data Analysis

Group no. 3

Prabhdeep Singh (2024010078)

Shamandeep Singh (2024010095)

Dataset Link - <https://www.kaggle.com/datasets/chaitu20/ipl-dataset2008-2025>

IPL Data Visualisation Project Report

1. Introduction- Project Synopsis: This study outlines a data visualisation effort centred on the Indian Premier League (IPL). The IPL is a well acclaimed and internationally recognised professional T20 cricket league in India, offering a substantial dataset for analysis and visualisation. The research intends to utilise this dataset to reveal significant trends, player performance insights, and strategy patterns that have influenced the league from its start to the latest season.

Dataset: The project employs the extensive IPL dataset from Kaggle, encompassing all seasons from 2008 to 2025. This dataset is a significant resource with over 60 attributes, offering detailed, ball-by-ball information for each match. It encompasses vital information such match metadata (date, venue, toss outcomes), player-specific contributions (runs, wickets, balls faced), and comprehensive match events (extras, types of dismissals).

Aim: The principal objective of this research is to convert unprocessed IPL data into significant visual storytelling. The analysis concentrates on several critical domains, including:

- Evaluating the supremacy of particular teams and players over time.
- Assessing the influence of strategic choices, such as the toss.
- Identifying performance patterns across various settings.
- Emphasising notable player accomplishments and milestones.

Instruments and Technologies: The investigation utilised Python for data processing with prominent libraries for data visualisation, like Pandas for data manipulation and Matplotlib or Seaborn for generating graphic elements.

2. Methodology Data Preparation: The initial phase encompassed comprehensive data cleansing and preparation. This involved addressing discrepancies in team names, stadium locations, or player identifiers. New elements were developed from the raw data to enable more intricate analysis, including the computation of batting strike rates, bowling economy rates, and win percentages.

Exploratory Data Analysis (EDA): An first exploration was conducted to comprehend the dataset's structure, identify essential variables, and assess data quality concerns. This method entailed examining the distribution of matches per season, the number of competing teams, and the comprehensive statistical summaries of runs and wickets.

3. Principal Insights and Visual Representations:

This section delineates the conclusions from the investigation, with each point substantiated by an appropriate data visualisation.

Venue-Dependent Performance: An important finding from the analysis is that specific locations have a distinct link with high-scoring games. Visualisations comparing the average first-innings scores at various stadiums indicate that venues such as Wankhede Stadium in Mumbai and M. Chinnaswamy Stadium in Bangalore routinely yield greater totals, presumably due to favourable pitch conditions and shorter boundaries.

Toss and Match Outcome: An examination of toss data indicates a notable tendency. A visualisation comparing the percentage of matches won by the side that won the toss and chose to field first against those that batted first illustrates that pursuing a goal has a superior success record in the IPL. This observation underscores the strategic significance of the toss decision and its possible influence on the ultimate outcome.

Preeminent Teams and Athletes: The data unequivocally illustrates the historical preeminence of select franchises. A bar chart illustrating the overall victories of each club verifies that the Mumbai Indians and Chennai Super Kings are the most successful in the league's history. From a player's standpoint, visual representations showcasing leading run-scorers and wicket-takers, such as a bar chart depicting the top 10 batsmen with the highest runs and the top 10 bowlers with the most wickets, underscore the sustained excellence of athletes like Virat Kohli, Rohit Sharma, and Lasith Malinga.

Performance Trends Across Seasons: A line chart depicting the average total runs scored per season illustrates the evolution of batting in the IPL. The visualisation illustrates a consistent increasing trajectory in scoring across the years, signifying a transition towards more aggressive batting techniques and elevated run rates.

4. Conclusion Synopsis of Results: This research effectively utilised the extensive IPL dataset to deliver a data-driven account of the league. The visualisations provided persuasive evidence of strategic tendencies, including the benefits of pursuing, and validated the consistent superiority of elite teams and players.

Constraints and Prospective Research: This project establishes a robust foundation, although there are aspects warranting more investigation. Future endeavours may involve integrating advanced analytics to develop a match prediction model, evaluating player performance throughout several game phases (e.g., powerplay, death overs), or designing an interactive dashboard for enhanced user discovery. These extensions would enhance the insights acquired and offer a more thorough comprehension of the IPL.