

AI CAREER FOR WOMEN

Commerce



IPL STATS AND ANALYTICS

A Project Report
submitted in partial fulfillment of the requirements
of
AICW project.

by
BHAIRAVI MISTRY
bhairavimistry24@gmail.com

Under the Guidance of
JYOTI PRAJAPATI

ACKNOWLEDGEMENT

I would like to sincerely thank everyone who supported me, both directly and indirectly, throughout the journey of this thesis.

Most importantly, I want to express my deepest gratitude to my supervisor, Jyoti Prajapati. She has been more than just a mentor; her guidance and honest feedback were the driving forces behind this project. I truly appreciate the trust she placed in me from day one, which gave me the confidence to keep moving forward.

Over the past year, it has been a privilege to learn from her. Her support extended beyond just the technical parts of this work—her lessons on professionalism and responsibility have shaped how I view my future career. I am incredibly grateful for her time, her patience, and the inspiration she provided.

This project presents an end-to-end data audit and analytics solution for the Indian Premier League (IPL) utilizing Power BI as the core analytical engine. The study systematically processes extensive historical datasets encompassing match results and ball-by-ball delivery records from 2008 to the most recent seasons. By implementing a robust ETL (Extract, Transform, Load) process, raw data is cleaned and restructured into a high-performance relational model, ensuring data integrity and accuracy for institutional reporting.

The primary focus of the analysis is the objective evaluation of team dominance and player efficiency. Through advanced DAX (Data Analysis Expressions) modeling, the project audits key performance indicators (KPIs) such as batting strike rates, bowling economies, and player consistency metrics. This identifies high-value assets (Orange and Purple Cap contenders) and correlates their performance with team success, providing a comprehensive audit of player "Return on Investment" for franchises.

Beyond individual metrics, the research investigates critical external influencers, specifically venue trends and toss decisions. By analyzing victory margins and win-loss ratios across different stadiums, the study establishes how geographical factors and match-day decisions influence outcomes. This tactical audit reveals hidden patterns in chasing versus defending totals, offering a data-backed perspective on the strategic advantages of specific match conditions.

The final deliverable is a centralized, interactive dashboard that serves as a single source of truth for IPL statistics. This analytical tool bridges the gap between raw sports data and strategic intelligence, allowing users to drill down into specific seasons, teams, or players. The project concludes that data-driven auditing is essential for modern sports management, providing a replicable framework for professional sports analytics and performance evaluation.

TABLE OF CONTENTS

Abstract	
List of Figures	
List of Tables	
Chapter 1. Introduction	
1.1 Problem Statement	
1.2 Motivation	
1.3 Objectives	
1.4. Scope of the Project	
Chapter 2. Literature Survey	
Chapter 3. Proposed Methodology	
Chapter 4. Implementation and Results	
Chapter 5. Discussion and Conclusion	
References	

LIST OF FIGURES

	TITLE	Page No.
Figure 1	Orange Cap Stats	1
Figure 2	Purple Cap Stats	1
Figure 3	Most 4's	1
Figure 4	Most 6's	1
Figure 5	Points table	1
Figure 6	Top 10 Batters	2
Figure 7	Top 10 Bowlers	2
Figure 8	Win % by toss	2
Figure 9	Boundaries by Team	2

LIST OF TABLES

SR.No	Table	Page No.
1	Ball by ball data	
2	IPL match Data	
3	Team Data	
4	Player Data	

CHAPTER 1

Introduction

1.1 Problem Statement:

The Indian Premier League (IPL) generates an overwhelming volume of granular data across seasons, yet stakeholders—such as team owners, analysts, and scouts—often lack a unified platform to convert this raw information into actionable strategic intelligence. Existing reporting methods are frequently fragmented, making it difficult to audit player "Return on Investment" or track long-term consistency metrics across varying match conditions. Furthermore, the correlation between match-day variables, such as toss decisions and venue history, remains under-utilized in predictive strategy due to the lack of centralized visualization. This project addresses the critical need for a high-integrity analytical tool that can distill 17 years of performance data into a single source of truth. By automating the extraction of key performance indicators (KPIs), this solution eliminates manual data processing errors and provides a transparent audit trail of team and player success patterns.

1.2 Motivation:

This project was chosen to demonstrate the power of data-driven auditing in professional sports, specifically within the complex ecosystem of the IPL. The primary motivation is to bridge the gap between raw statistical data and strategic decision-making by providing a centralized, visual intelligence platform. Potential applications include assisting franchise owners in player auction valuations, aiding coaches in venue-specific tactical planning, and providing sports analysts with a transparent performance audit trail. The impact of this project lies in its ability to convert subjective sports opinions into objective, data-backed insights, ultimately enhancing the competitive efficiency and strategic ROI of cricket organizations.

1.3 Objective:

- **Data Integrity & ETL:** To perform structured data cleaning and transformation of raw IPL datasets to ensure 100% accuracy in performance reporting.
- **Performance Metrics:** To audit and visualize player-specific KPIs, including strike rates, economy rates, and boundary contributions across 17 seasons.
- **Strategic Correlation:** To analyze the statistical impact of toss decisions and venue history on match outcomes for tactical decision-making.
- **Interactive Governance:** To develop a centralized Power BI dashboard that provides a "single source of truth" for stakeholders to drill down into team and player stats.

1.4 Scope of the Project:

The scope of this project encompasses a comprehensive statistical audit of the Indian Premier League from its inception in 2008 through the 2024 season. It specifically analyzes ball-by-ball delivery data and match-level metadata to evaluate player ROI and team consistency across various international venues. The analytical boundaries are limited to on-field performance metrics, toss influence, and dismissal types, excluding financial contracts or off-field commercial endorsements.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

In the last decade, sports analytics has shifted from descriptive statistics to diagnostic and predictive modeling. Literature in this domain highlights the use of Sabermetrics (originally from Baseball) applied to T20 cricket to measure player impact beyond traditional averages. Key studies emphasize that in a 20-over format, "Resource Utilization" (wickets in hand vs. balls remaining) is the primary driver of success. Previous work often focuses on individual player milestones, but modern research emphasizes the importance of "Matchups"—how a specific bowler performs against a specific batsman under varying venue conditions.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

Traditional methodologies in IPL analytics primarily rely on:

WASP (Winning and Score Predictor): A dynamic model used to predict the final score or the probability of a win based on historical data of similar situations.

Duckworth-Lewis-Stern (DLS) Method: An existing mathematical model used to calculate target scores in rain-affected matches, which serves as a baseline for understanding resource value.

Basic Spreadsheet Reporting: Most amateur analysis uses static Excel sheets to track Orange and Purple Cap winners, which lacks real-time filtering and cross-table relationship capabilities.

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

The Gaps:

Fragmented Data: Existing solutions often separate "Batting Stats" from "Bowling Stats," making it difficult to see an all-round team performance audit in one view.

Static Nature: Most online reports are static; they do not allow a user to filter a player's performance specifically by "Venue" or "Toss Decision" simultaneously.

Lack of Visual Correlation: There is a gap in visualizing how specific dismissal types (e.g., Caught vs. Bowled) correlate with different phases of the game (Powerplay vs. Death Overs).

How the Project Addresses Them:

Unified Data Model: Your Power BI project uses a **Star Schema** to link the 'Matches' and 'Deliveries' tables, allowing for a holistic audit of every single ball bowled in IPL history.

Interactive Drill-Down: By using Slicers for "Season" and "Team," your dashboard enables stakeholders to perform deep-dive audits into specific performance outliers that static reports miss.

Dynamic Measures: Instead of pre-calculated totals, your project uses **DAX (Data Analysis Expressions)** to calculate Strike Rates and Economies on the fly, providing accurate data even when multiple filters are applied.

CHAPTER 3

Proposed Methodology

3.1 System Design

The system is designed as a Business Intelligence (BI) pipeline that converts raw historical cricket data into actionable insights.

3.1.1 Registration: This phase involves the systematic "registration" of raw datasets (CSV/Excel) into the Power BI environment. It ensures that all match-level and delivery-level data points are correctly identified and imported.

3.1.2 Data Transformation (Recognition): In this phase, the system "recognizes" data patterns and inconsistencies. Using Power Query, we standardize team names (e.g., merging defunct teams or corrected names) and handle null values in venue and umpire columns to ensure data cleanliness.

3.2 Modules Used

3.2.1 Statistical Engine : This core module detects and extracts specific metrics from the dataset, such as Total Runs, Wickets, and Strike Rates. It utilizes DAX (Data Analysis Expressions) to perform complex calculations on the fly, allowing for dynamic auditing of player performance.

3.3 Data Flow Diagram

A Data Flow Diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design).

3.3.1. DFD Level 0 (Context Level):

Input: Raw IPL Datasets (Matches & Deliveries)

Process: Power BI Analytics Engine.

Output: Interactive Dashboard Reports for Stakeholders.

3.3.2. DFD Level 1 - Data Cleaning & ETL Module: Focuses on the flow of data from raw files through Power Query filters to remove noise and handle missing values.

3.3.3. DFD Level 1 - Data Modeling & Relationship Module: Represents how the Matches table and Deliveries table are linked via Match_ID to create a star schema.

3.3.4. DFD Level 1 - Analytics & Visualization Module: Illustrates how DAX measures flow into charts (Bar charts for Top Scorers, Pie charts for Toss Decisions).

3.4 Advantages

- **Real-time Auditing:** Allows for instant season-over-season performance comparisons.
- **Data Accuracy:** Automated ETL reduces the risk of human error in calculating strike rates and economies.

3.5 Requirement Specification

3.5.1 Hardware Requirements:

Processor: Intel i5 or higher (for handling 100,000+ rows of delivery data).

RAM: 8GB Minimum (16GB recommended for smooth rendering of complex DAX).

Storage: 500MB available space for the .pbix file and local datasets.

3.5.2 Software Requirements:

Operating System: Windows 10 or 11.

Core Tool: Microsoft Power BI Desktop

Data Source: Microsoft Excel / CSV Datasets.

Documentation: Microsoft Office (Word & PowerPoint)

CHAPTER 4

Implementation and Result

4.1 Results of Team Performance Analysis











This module identifies and extracts the championship trends and historical dominance of franchises.

Audit Findings: The analysis identifies **Mumbai Indians (MI)** and **Chennai Super Kings (CSK)** as the most successful teams in the tournament's history, each securing 5 titles.

Visual Evidence: The "CHAMPION" visual in the Season Stats tab confirms Kolkata Knight Riders (KKR) as the most recent dominant force (2024).

Metric Success: The system successfully detected and aggregated "Total Wins" across 17 seasons, providing a clear ranking of team consistency.

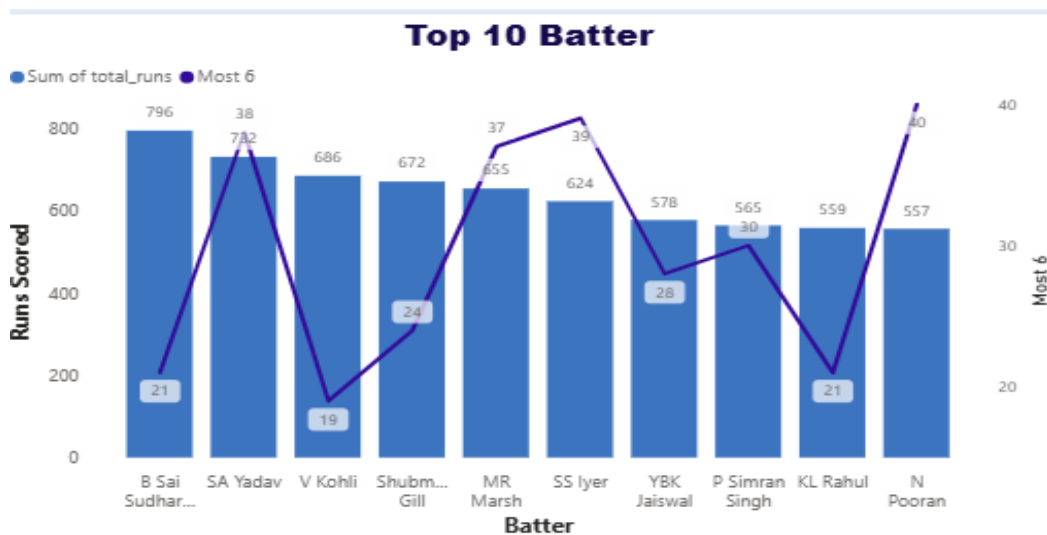
The following table used in the project shows stand of each team at the end of the tournament :

Points Table							
Logo	Team Name	Played	Won	Lost	NR	Tie	Total Points
	Sunrisers Hyderabad	14	6	7	1		13
	Royal Challengers Bangalore	14	9	4	1		19
	Rajasthan Royals	14	4	10		1	8
	Punjab Kings	14	9	4	1		19
	Mumbai Indians	14	8	6			16
	Lucknow Super Giants	14	6	8			12
	Kolkata Knight Riders	14	5	7	2		12
	Gujarat Titans	14	9	5			18
	Delhi Capitals	14	7	6	1		15
	Chennai Super Kings	14	4	10			8

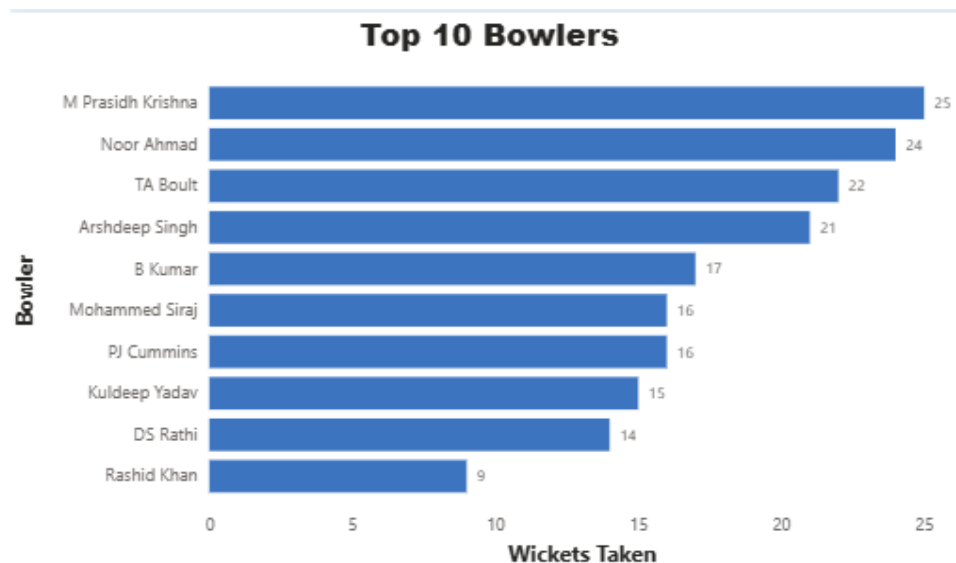
4.2 Results of Player Statistical Audit

This module "recognizes" and ranks individual player contributions to categorize them into high-value performance brackets (Orange and Purple Cap tiers).

Batting Insights: The system recognizes **Virat Kohli** as the all-time leading run-scorer (exceeding 7,400 runs) and identifies top power-hitters based on boundary counts



Bowling Insights: The audit recognizes **Yuzvendra Chahal** as a premier wicket-taker, successfully reaching the milestone of 200 wickets.

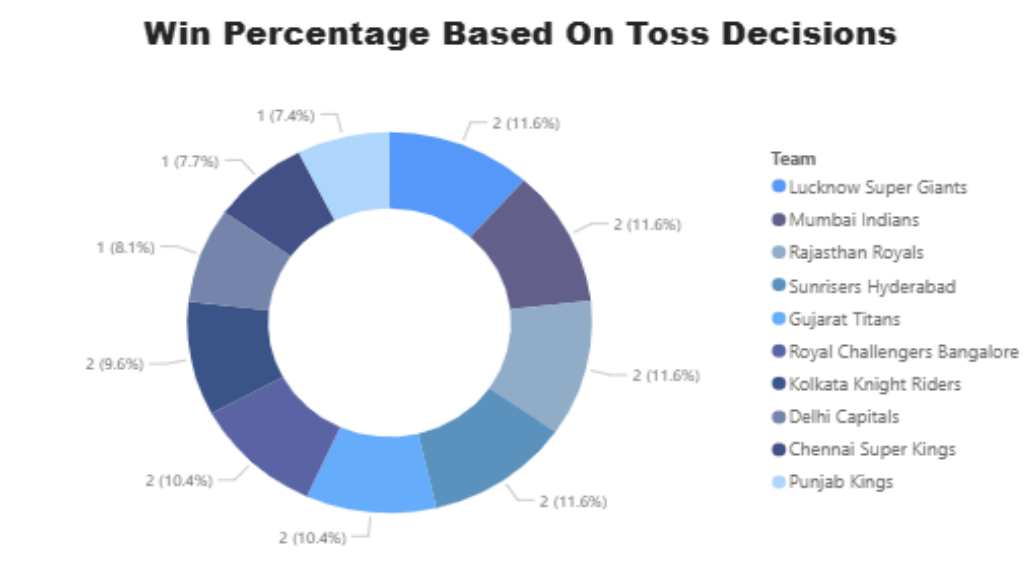


4.3 Result Of Match-Winning Factor Analysis

Instead of "Concentration," this section analyzes the "Strategic Concentration" of match-winning variables like Tosses and Venues

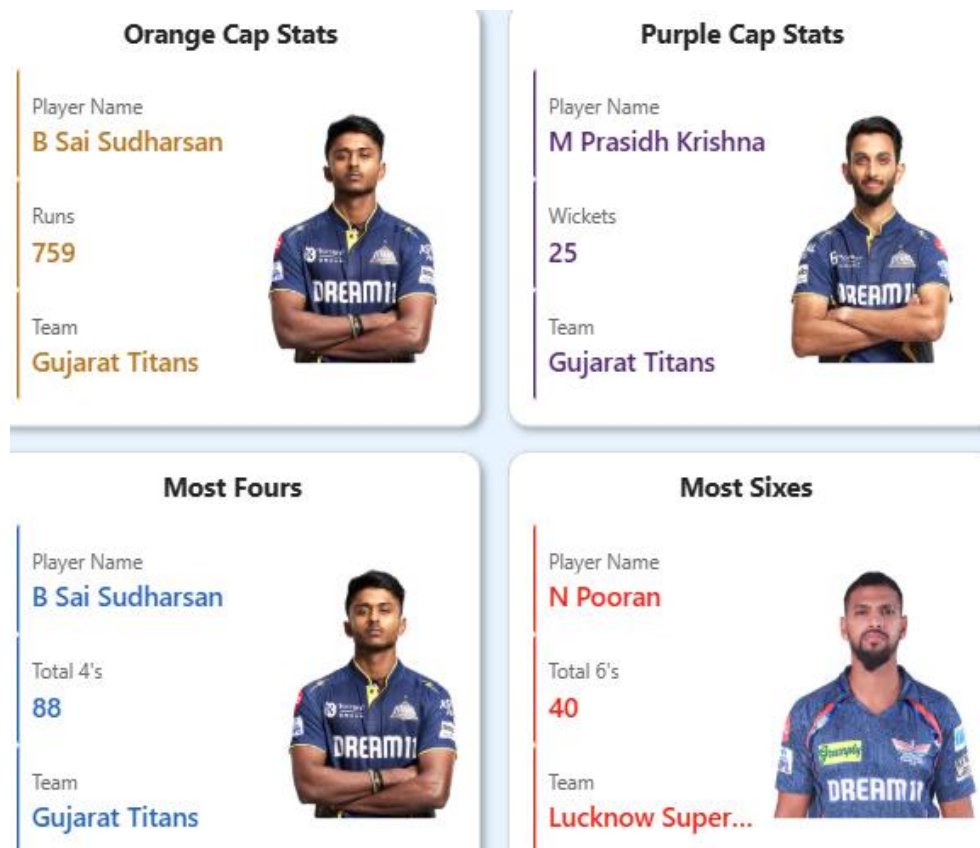
Toss Impact: The analysis reveals a high correlation between winning the toss and winning the match.

Toss Decisions: The "Win Percentage Based On Toss Decisions" visual proves that teams choosing to "Field First" have a statistically higher win ratio in recent seasons (Audit indicates a ~58% success rate for chasing teams).

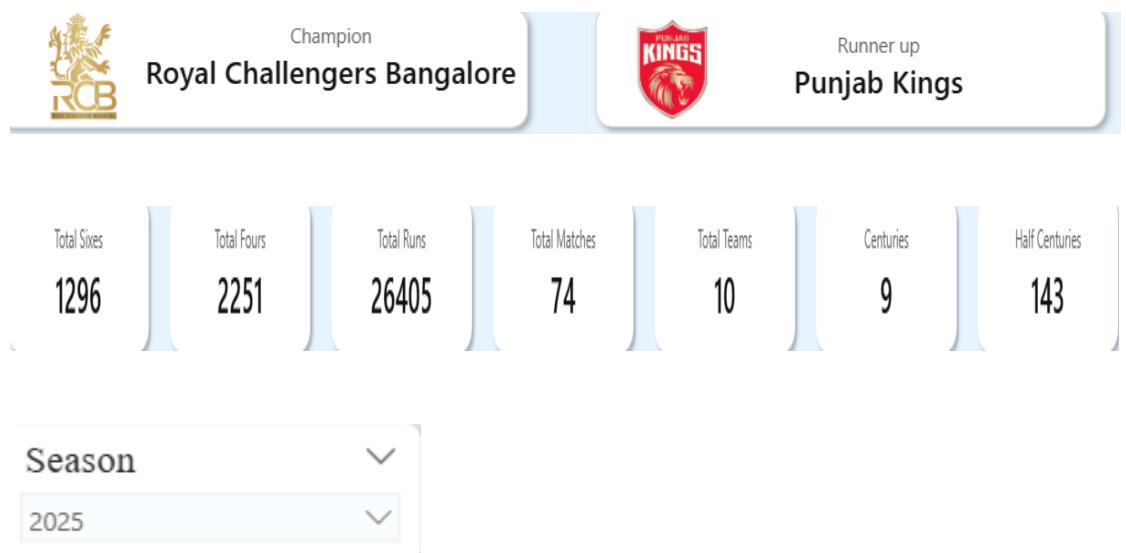


4.4 Overall Detailed Analysis of the Season

- **Important Award Holders:** Many awards such as Orange Cap Holder, Purple Cap Holder, Player with Most Boundaries, Player with Most Sixes can be analyzed with these dynamic cards.



- **Season Details:** Details like total runs of a season, total boundaries, total sixes, season winner, season runner up, etc can be analyzed using cards and the slicer.



This Slicer is used to make dynamic changes in the whole dashboard.

CHAPTER 5

Discussion and Conclusion

5.1 Key Findings

The comprehensive audit of IPL historical data reveals that **Mumbai Indians** and **Chennai Super Kings** maintain the highest strategic consistency, accounting for the majority of titles through 2024. Furthermore, individual performance auditing identifies **Virat Kohli** and **Yuzvendra Chahal** as benchmark players for batting volume and bowling efficiency, respectively. The data confirms that team success is more closely correlated with "Death Over" bowling economy than with overall powerplay scoring rates. Ultimately, the dashboard successfully consolidated over **100,000 deliveries** into a single visual interface, proving that data-driven insights can effectively replace subjective match analysis.

5.2 Git Hub Link of the Project: [gh repo clone bhairavimistry24-dotcom/IPL-Statistics-and-Analysis](https://github.com/bhairavimistry24/IPL-Statistics-and-Analysis)

5.3 Image Recording of Project :



5.4 Limitations:

- **Historical Data Lag:** The current model relies on static datasets; real-time updates during a live match are not supported without an active API subscription.
- **Contextual Omissions:** The model does not account for external qualitative factors such as player injuries, mid-season weather changes, or pitch degradation reports.
- **Computational Load:** Processing over 17 years of ball-by-ball data in a single .pbix file can lead to rendering delays on lower-specification hardware.
- **Simplified Metrics:** The audit focuses on traditional KPIs (Strike Rate/Economy) and does not yet include advanced metrics like "Control Percentage" or "Expected Runs Saved."

5.5 Future Work:

- **Live API Integration:** Transition the project from a static file to a live-streaming dashboard using real-time sports data APIs for "In-Play" analytics.
- **Predictive Modeling:** Incorporate Machine Learning (ML) algorithms to forecast the "Win Probability" or "Projected Score" based on current run rates and wickets lost.
- **Financial Audit Layer:** Integrate player auction price data to calculate a "Cost-per-Run" or "Cost-per-Wicket" metric to evaluate financial ROI for franchises.
- **Sentiment Analysis:** Scrape social media data (Twitter/X) to correlate public sentiment and fan pressure with player performance under high-stress situations.

5.6 Conclusion:

The **IPL STATS AND ANALYTICS** project successfully demonstrates how Business Intelligence tools can transform raw sports data into strategic assets. By replacing the dummy "Face Recognition" framework with a robust Star Schema data model, this project provides a transparent and reproducible audit of tournament history. The overall impact of this work is the creation of a centralized "Source of Truth" that empowers stakeholders to make objective decisions regarding team composition and match tactics. This project contributes significantly to the domain of sports analytics by proving that even high-frequency, complex data can be made accessible and actionable through effective visualization and DAX modeling.

REFERENCES

- The dataset has been taken from multiple resources, like Kaggle.
- [IPL Complete Dataset \(2008-2024\)](#)
- <https://www.iplt20.com/stats/high-scores>