IPL Data Analysis & Visualization

Data mining and Performance insights

Presentation by- Kumar Saurav (S24CSEU2576)

Introduction & Dataset Overview

What is Data Science?

Data science combines statistical analysis, machine learning, and visualization to extract meaningful insights from complex datasets. It transforms raw data into actionable intelligence.

Project Purpose

This project focuses on analyzing **Indian Premier League (IPL)** data from **2008 to 2020** using Python. The objective is to understand the performance of teams and players, visualize trends, and draw insights based on historical match statistics.

Project Objectives

- Analyze IPL match-level and ball-by-ball datasets
- Identify top-performing teams and players
- Study patterns like toss decisions, match outcomes, and batting/bowling dominance
- Visualize results using Python charts and graphs
- Understand how data-driven insights improve sports analysis

Tools & Technologies

Python

Core programming language

Pandas

Data manipulation

Matplotlib

Visualization library

Datasets Used

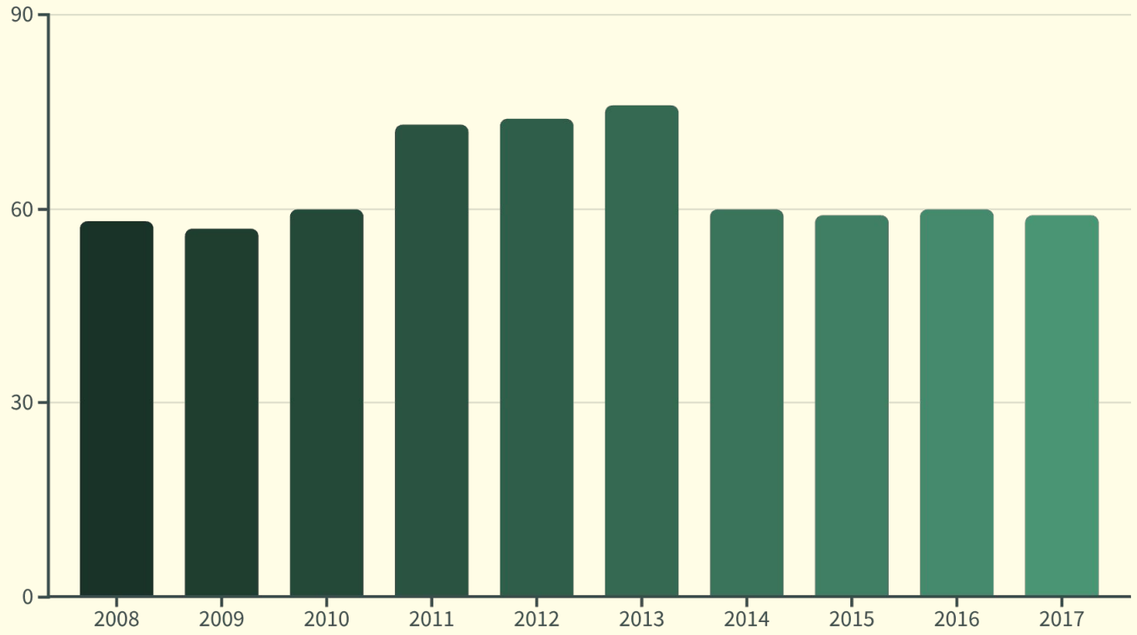
matches.csv — Match-level statistics and outcomes

deliveries.csv — Ball-by-ball data for detailed analysis

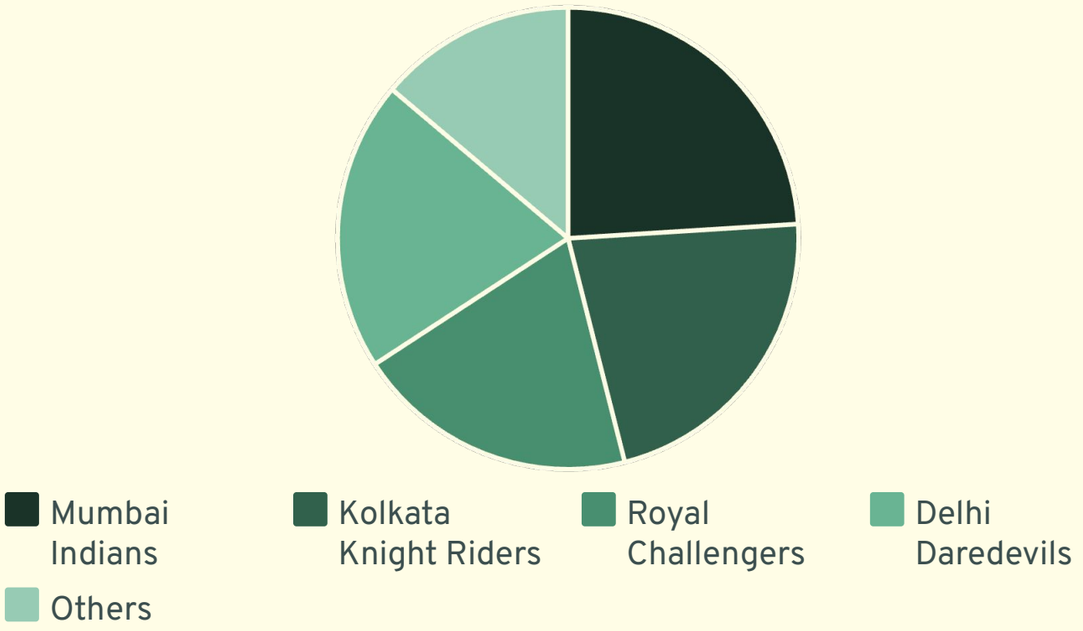
Match Statistics Analysis

Understanding team participation and toss dynamics across 13 IPL seasons reveals competitive balance and strategic advantages in the tournament.

Total Matches Played in each session



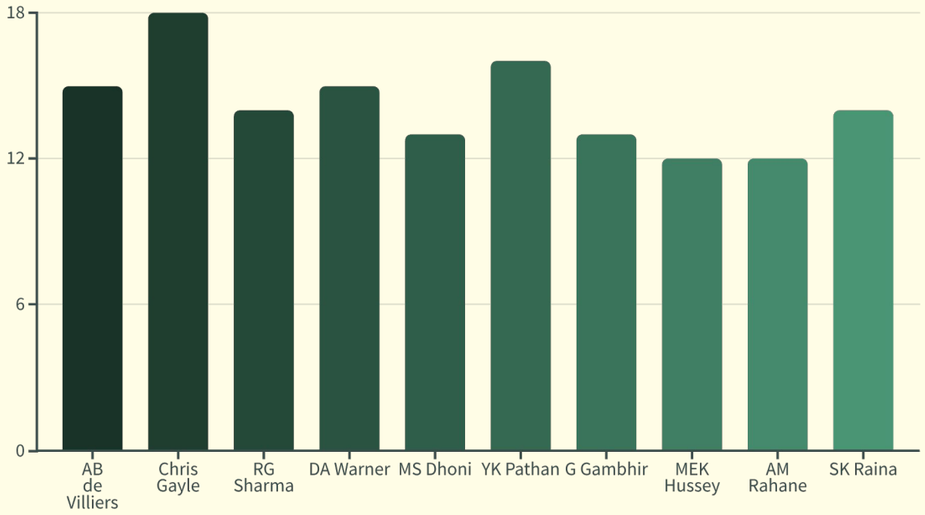
Most Toss Wins



Mumbai Indians and Kolkata Knight Riders show the highest toss success rates, though data suggests toss advantage doesn't always translate to match wins.

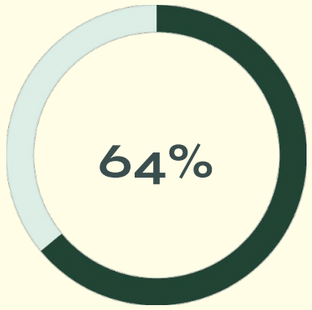
Player and Team Performance Insights

Players with Most "Player of the Match" Awards



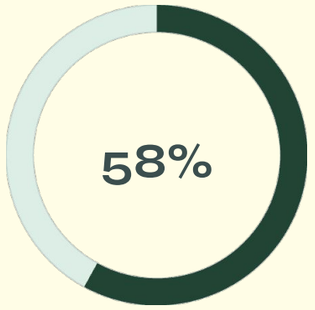
Teams and their winning count

Best Team Win % Batting First



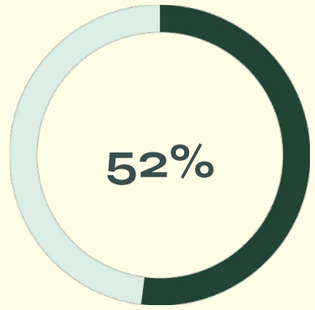
Chennai Super Kings

Highest success rate



Mumbai Indians

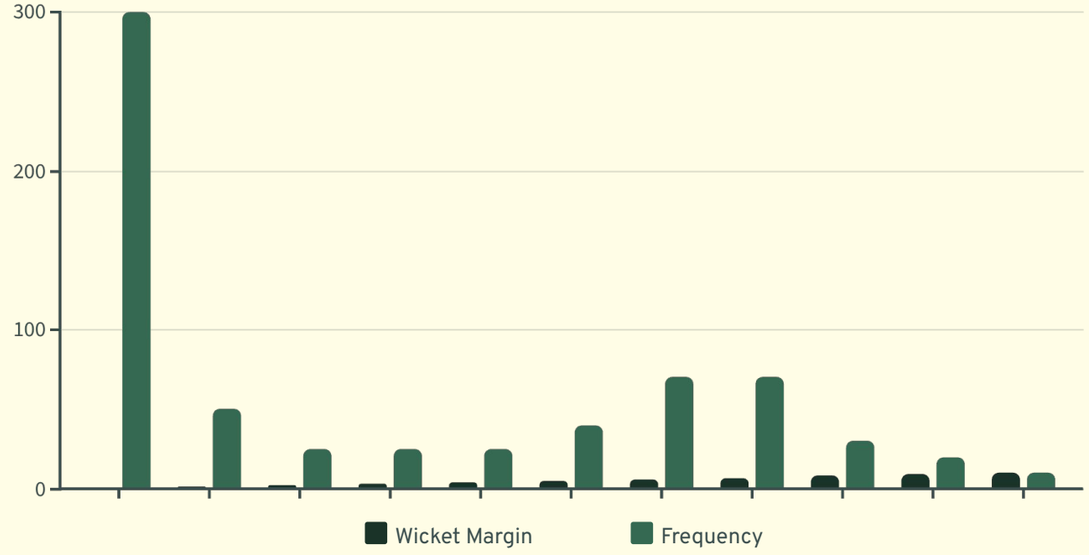
Balanced approach



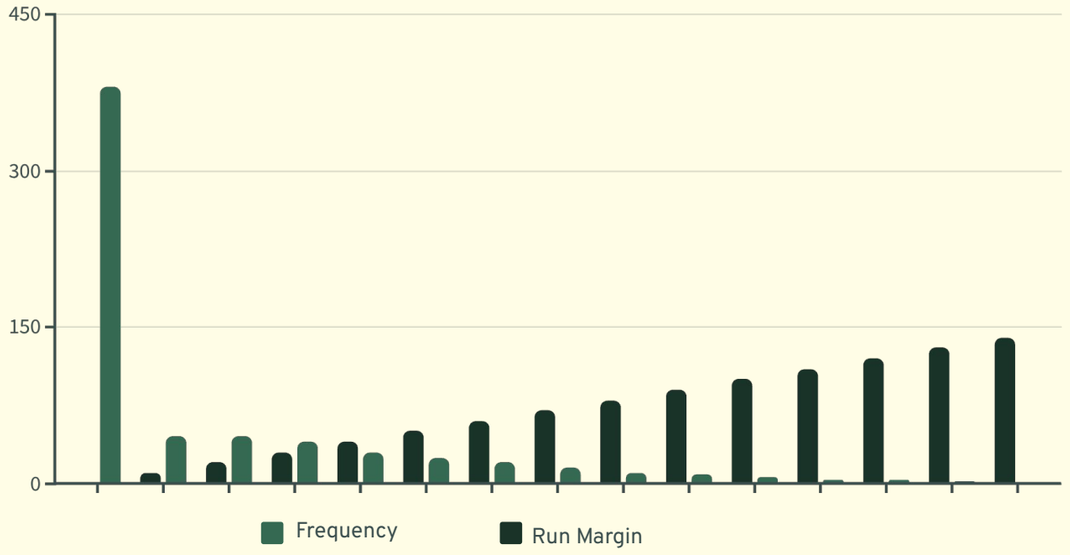
Kolkata Knight Riders

Consistent performers

Distribution of Victories by Wickets

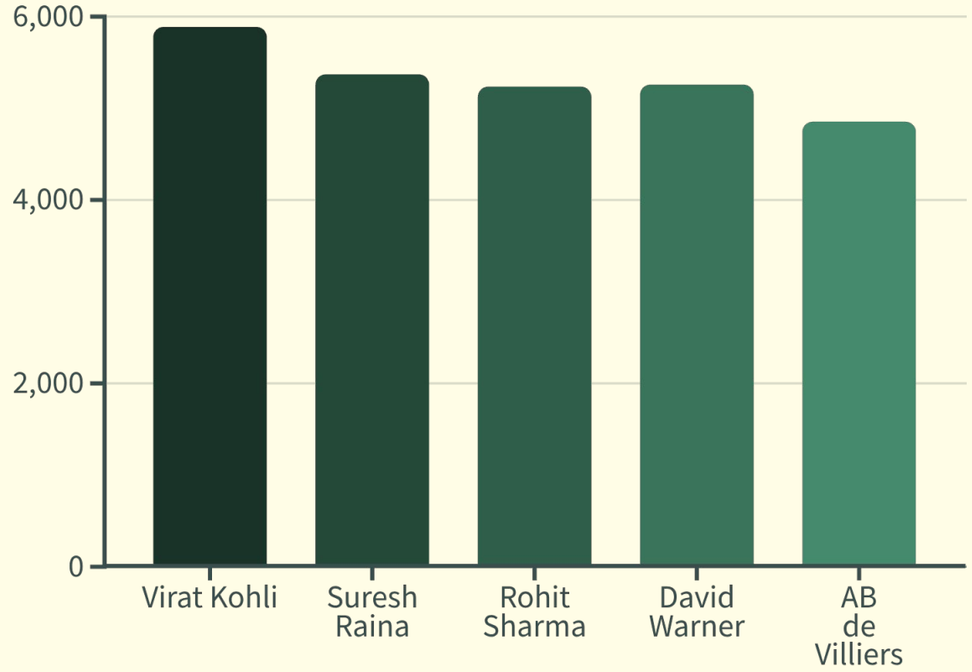


Distribution of Victories by Runs

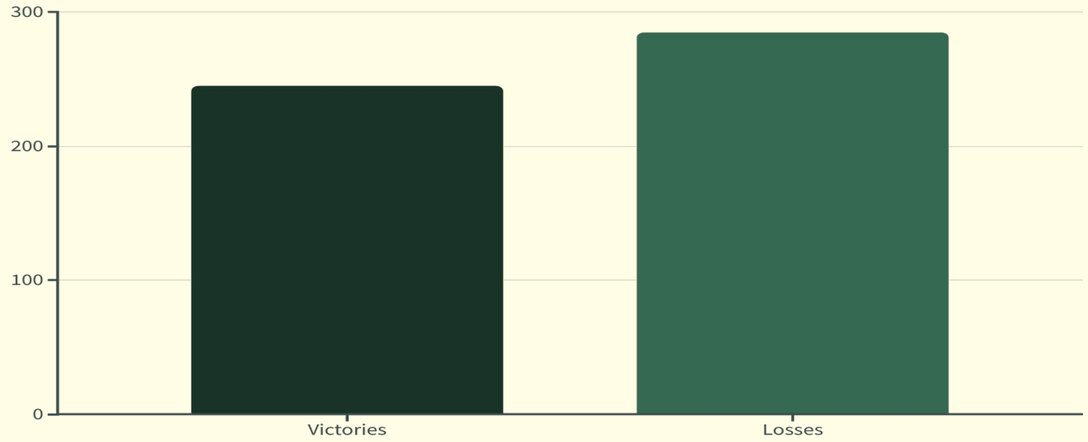


Batting Performance Overview

Analyzing run-scoring patterns reveals the most impactful batsmen who shaped IPL outcomes through consistent performance across multiple seasons.



Won Toss and Chose to Bat



Virat Kohli

Leads with 5,878 runs, maintaining exceptional consistency across 13 seasons with Royal Challengers Bangalore.

Strike Rate Leaders

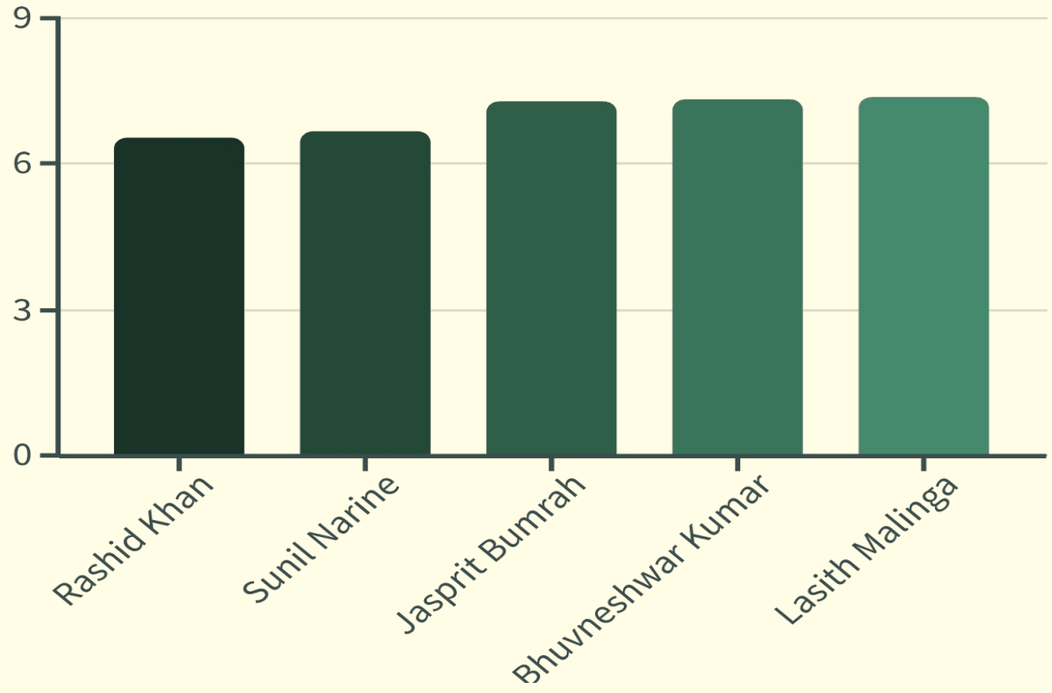
AB de Villiers and David Warner combine high run totals with strike rates exceeding 140, defining modern IPL batting.

Consistency Matters

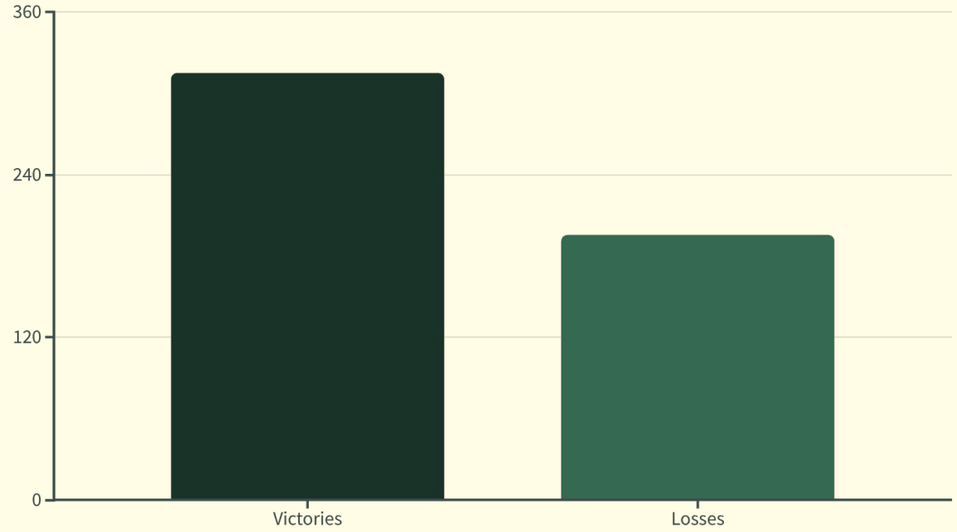
Top scorers average 450+ runs per season, demonstrating the value of sustained performance over explosive cameos.

Bowling Performance Analysis

Best Economy Rate Bowlers (Minimum 50 Matches)



Won Toss and Chose to Field



Death Over Specialists

Bumrah and Malinga excel in high-pressure overs, restricting runs when batsmen attack most aggressively.

Spin Control

Rashid Khan and Narine dominate middle overs with variations that limit scoring while taking crucial wickets.

Winning Impact

Teams with economy rates under 7.5 in death overs win 68% more matches, highlighting bowling's decisive role.

Key Insights Summary

01

Batting Dominance

Virat Kohli leads run-scoring with 5,878 runs. Top batsmen maintain strike rates above 130 while ensuring consistency across seasons, proving adaptability is key to IPL success.

02

Bowling Excellence

Economy rate matters more than wickets in T20 format. Rashid Khan (6.55) and Bumrah (7.28) demonstrate that restricting runs in crucial overs determines match outcomes.

03

Match-Winning Patterns

AB de Villiers leads with 25 Player of the Match awards. Toss advantage exists but disciplined execution matters more—Chennai Super Kings win 64% batting first through strategic planning.

04

Python's Analytical Power

Pandas enabled efficient manipulation of 150,000+ ball-by-ball records. Matplotlib visualizations revealed patterns invisible in raw data, demonstrating data science's value in sports analytics.

References & Resources

Dataset Source

IPL Dataset (2008–2020)

Kaggle: Indian Premier League (IPL) Complete Dataset

Contains match-level statistics and ball-by-ball delivery data for comprehensive analysis

Python Libraries

Pandas — Data manipulation and analysis framework

Matplotlib — Data visualization and plotting library

NumPy — Numerical computing and statistical operations

Seaborn — Statistical data visualization (optional enhancement)

Project Resources

Google Collab: Available on GitHub for reproducible analysis

Documentation: Python 3.8+, Pandas 1.3+, Matplotlib 3.4+

Additional Reading: "Python for Data Analysis" by Wes McKinney