Exploratory Data Analysis of Indian Premier League Data

Srinidhi Kalgundi Srinivas
Ritika Kishore Kumar
Sepehr Aiden Bostan
LingYi Li
Jonathan Cayaban

INDIAN PREMIER LEAGUE ®

Cricket - Overview

- Cricket is a bat-and-ball game played between two teams of eleven players on a field
- **The batting side** - scores runs by striking the ball bowled (pitched) by the opposite side and then running between the wickets
- **The bowling/fielding side** - tries to prevent this by preventing the ball from leaving the field and getting the ball to either wicket
- 2 innings of 20 overs; the toss-winner decides if they want to bat or field in the first inning and then the teams switch; a bowler bowls 6 times in one over
- A team wins by scoring more runs or getting all of the opposing team's batters out

Indian Premier League (IPL):

- IPL is a men's cricket league of India (started in 2008)
- Players from all of the international teams play in this franchise (Soccer fans: think La Liga or Premier League)

Goal of the EDA

- See how toss-winner/toss-decision affects game outcome
- Compare statistics between different players
- See how player statistics can be used by teams to strategize games
- See how management decisions for auctioning players affects team success

Datasets

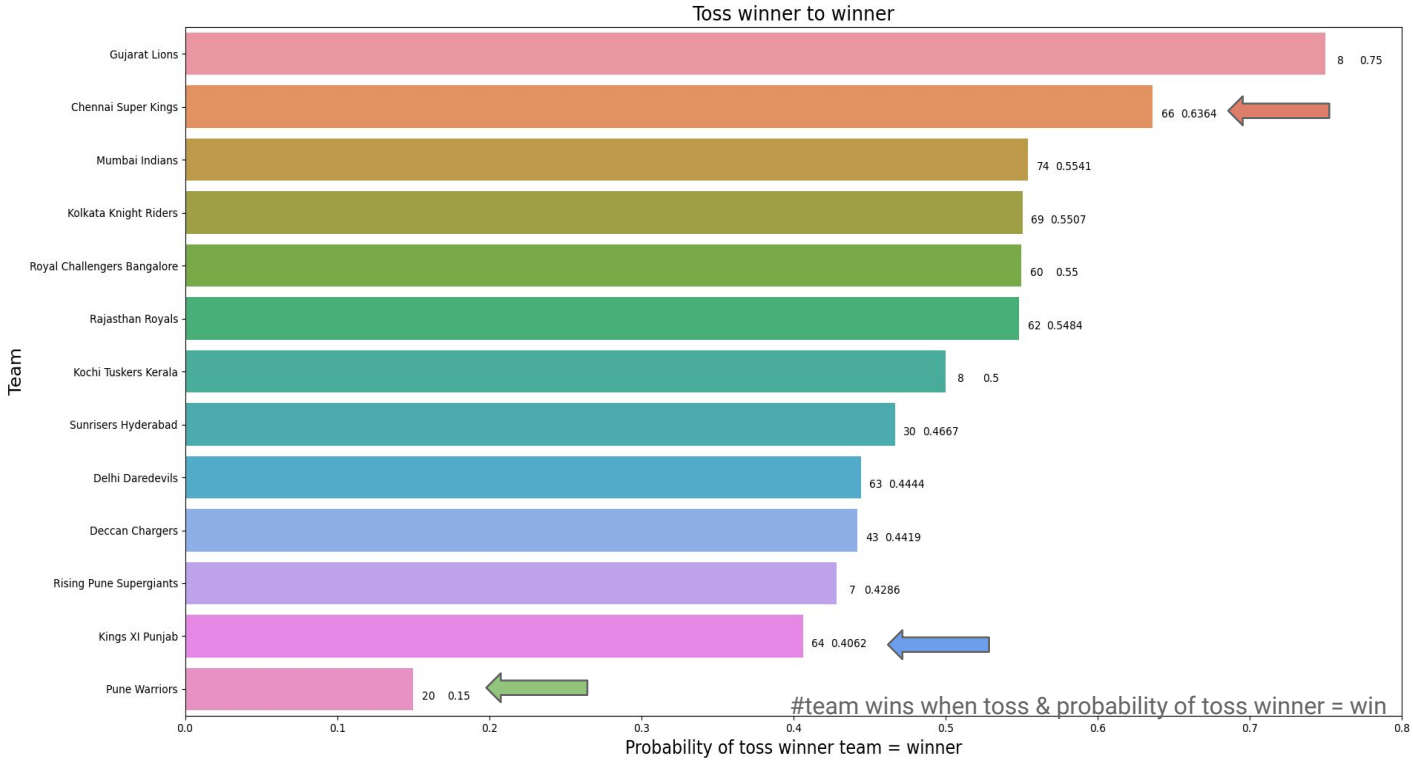
1. Exploratory Data Analysis for IPL (2008-2022)[1]
    - Matches
    - Deliveries
2. Kaggle: IPL Player Auction Dataset - From Start to Now (2013-2022)[3]
3. IPL Auction dataset scraping Wikipedia[2] (2008 - 2012)
    - Data was scraped using **BeautifulSoup** module
    - Some data cleaning and editing needed to be done for efficiency

[1] - https://www.kaggle.com/code/ambarish/exploratory-data-analysis-ipl/data
[2] - https://www.kaggle.com/datasets/kalilurrahman/ipl-player-auction-dataset-from-start-to-now
[3] - https://en.wikipedia.org/wiki/List_of_2008_Indian_Premier_League_auctions_and_personnel_signings

Methodology

- Data gathered from Kaggle
- Created classes to read from the data file and to clean the data
  - IPL auction information was scraped from the web to find the correlation between money spent and the number of wins
- Created functions for analysis and plotting:
  - Match statistics
  - Player statistics
  - Team spending

Team performance as toss winner



- Plot showing probability a team wins when it wins the toss

Interpretation:

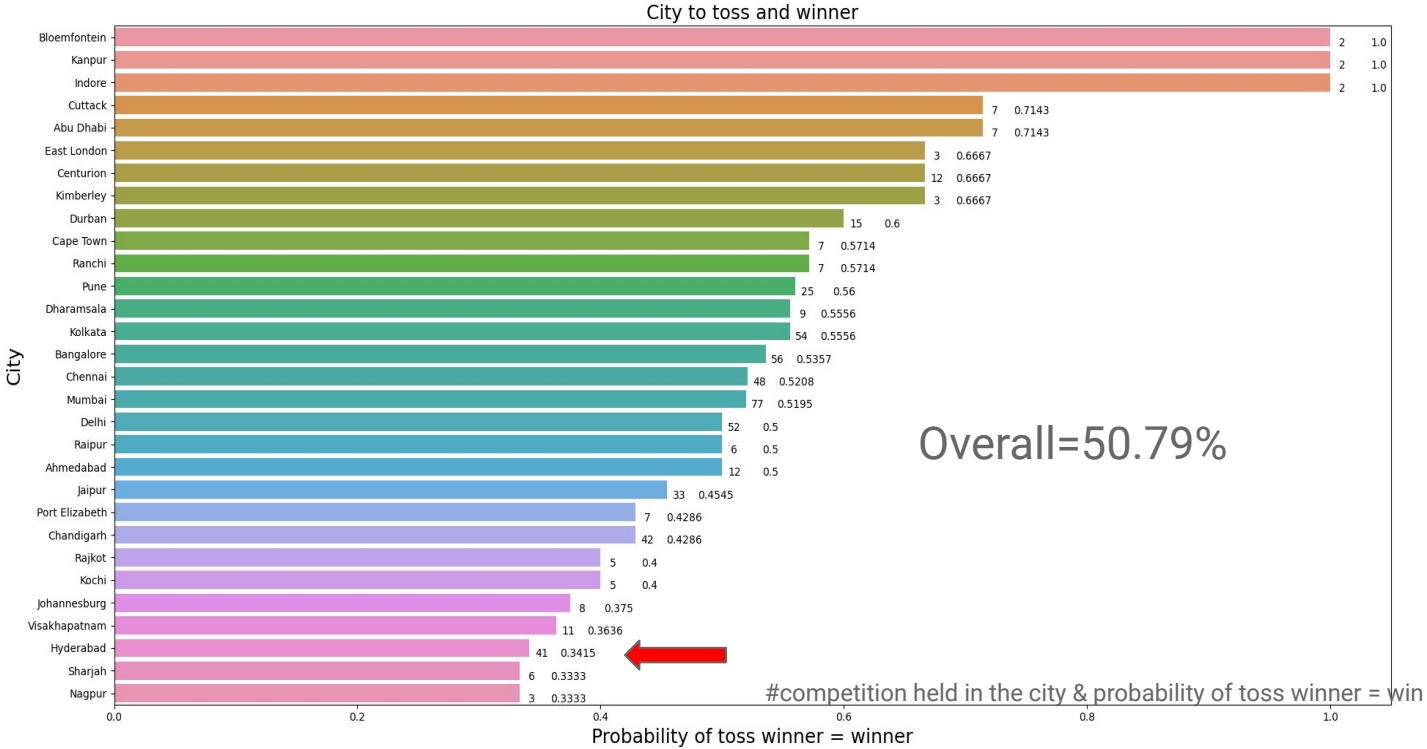
Various between teams

Chennai Super Kings wins most of time when it is toss winner

Kings XI Punjab is relatively weak

Pune Warriors should change strategy to take advantage of toss winner

Venue affects the outcome



- Plot shows the chances of a team winning the game given they won the toss in a particular ground

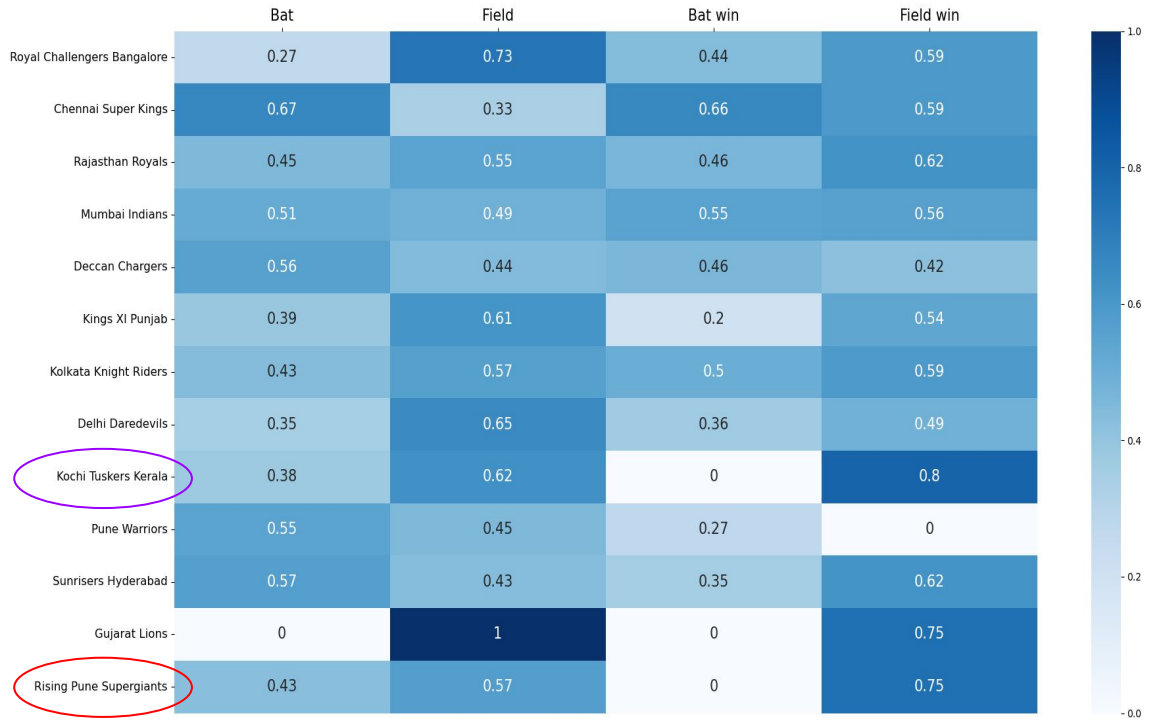
Interpretation:

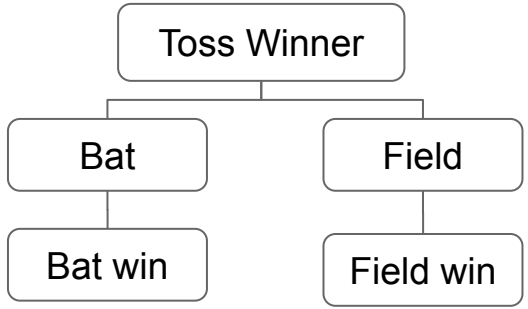
Overall prob slightly > 50%

If a city holds more competition, the probability will converge to 50%

Hyderabad is special which held relatively **more** competitions but with relatively **lower** probability for toss winner=winner

Choice & Result

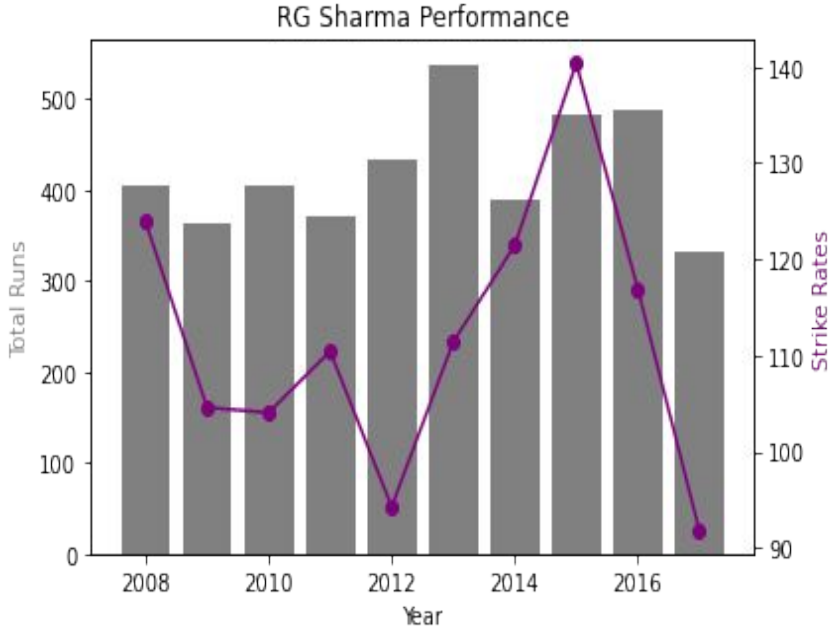




Most of teams do right choice

Rising Pune Supergiants
Kochi Tuskers Kerala
should always choose “Field” as toss winner

Batsman Performance



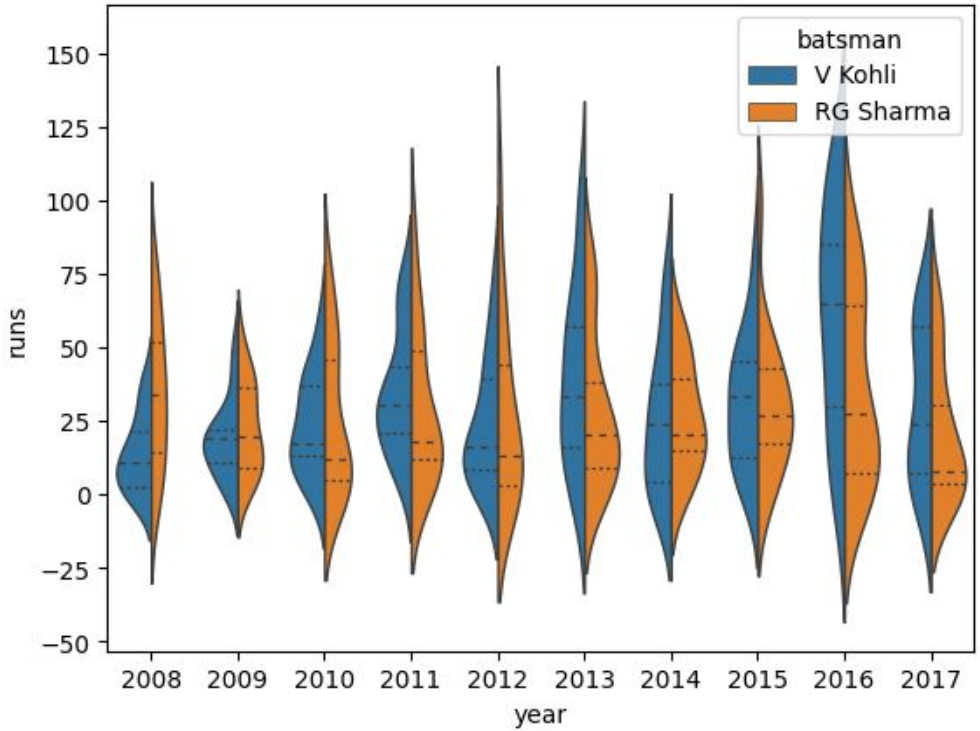
This graph plots the total runs scored by a batsman each season and the line plot gives an intuition of their strike rate in that season.

Interpretation -

Line plot: Higher the better

Bar plot: Higher the better

Batsman Stats

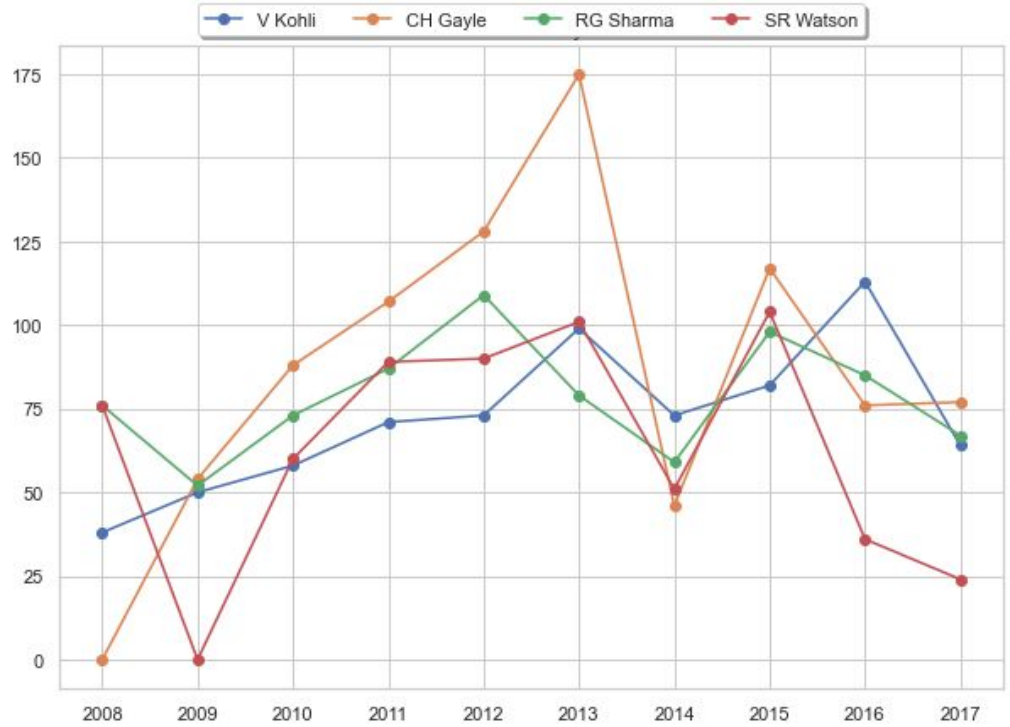


- This graph talks about a batsman consistency (or form) over a whole season.

  This can be used by auctioneers to see how consistent a player is through a season. Also compare the same stats with another batsman.

- Blue is the bell curve of runs over all matches by V Kohli in the given year. Orange is the same but for RG Sharma

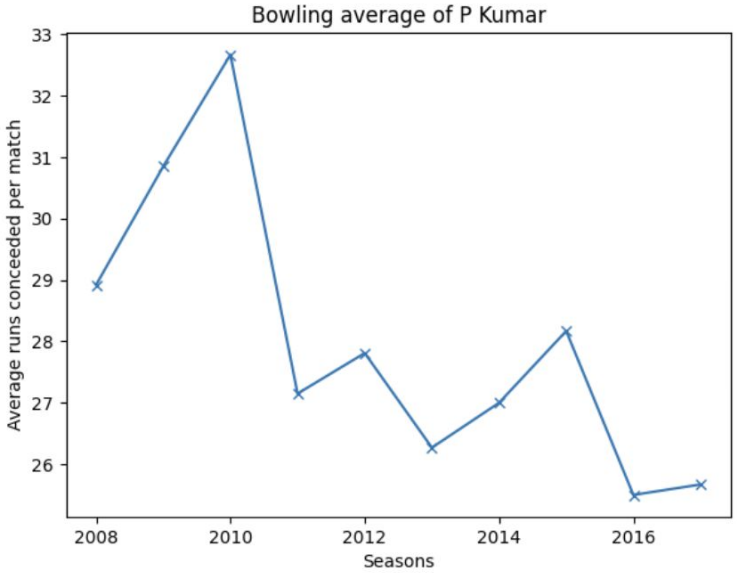
Batsman Comparison

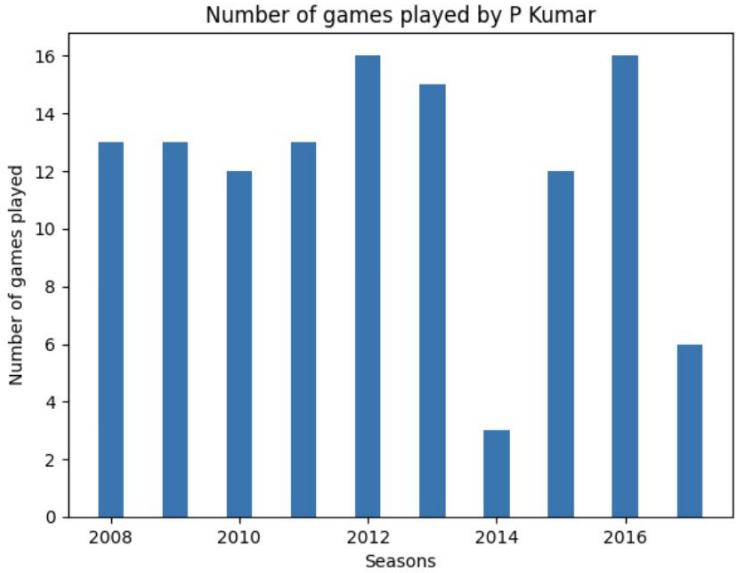


- Plot showing comparison of top 4 batsmen performance over the course of multiple seasons
- Commentators use this data to predict each season who would be the top player and so.

Interpretation: **Higher** the better, 0 indicates that the player has not participated in that season

Bowler Statistics

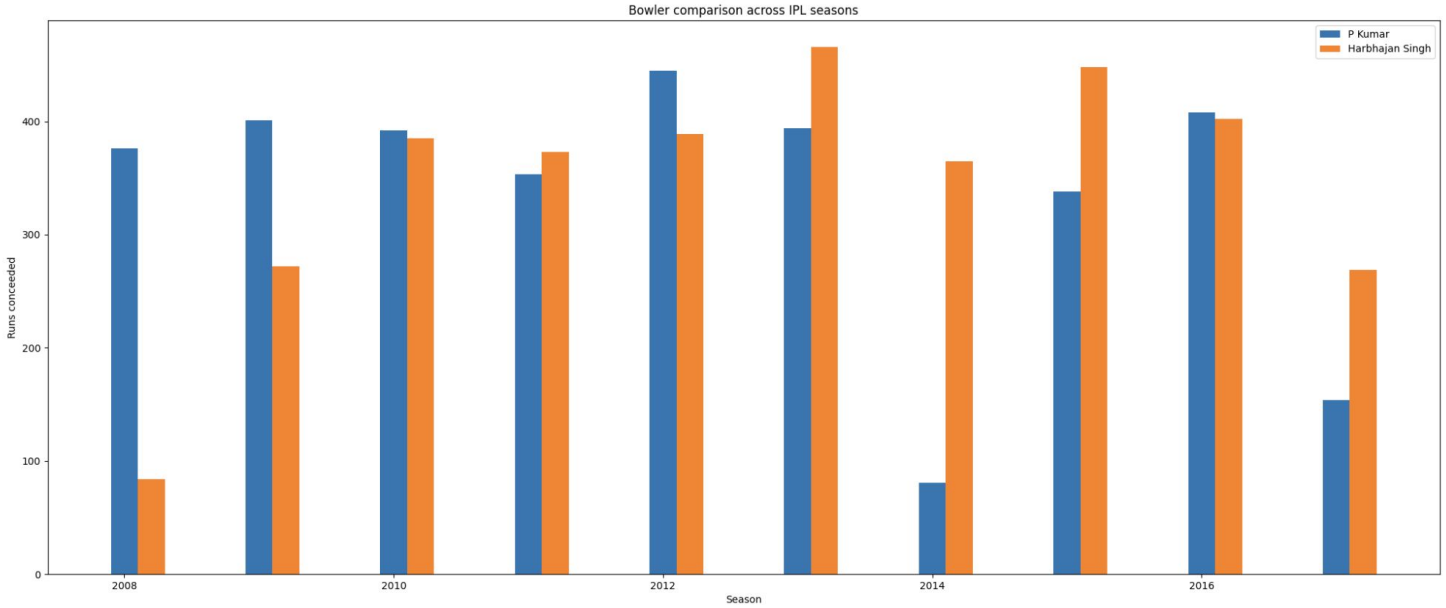




Interpretation **Lower** the better

- Useful information on how costly the bowler is
- Useful for the teams to decide on the bowling strategy

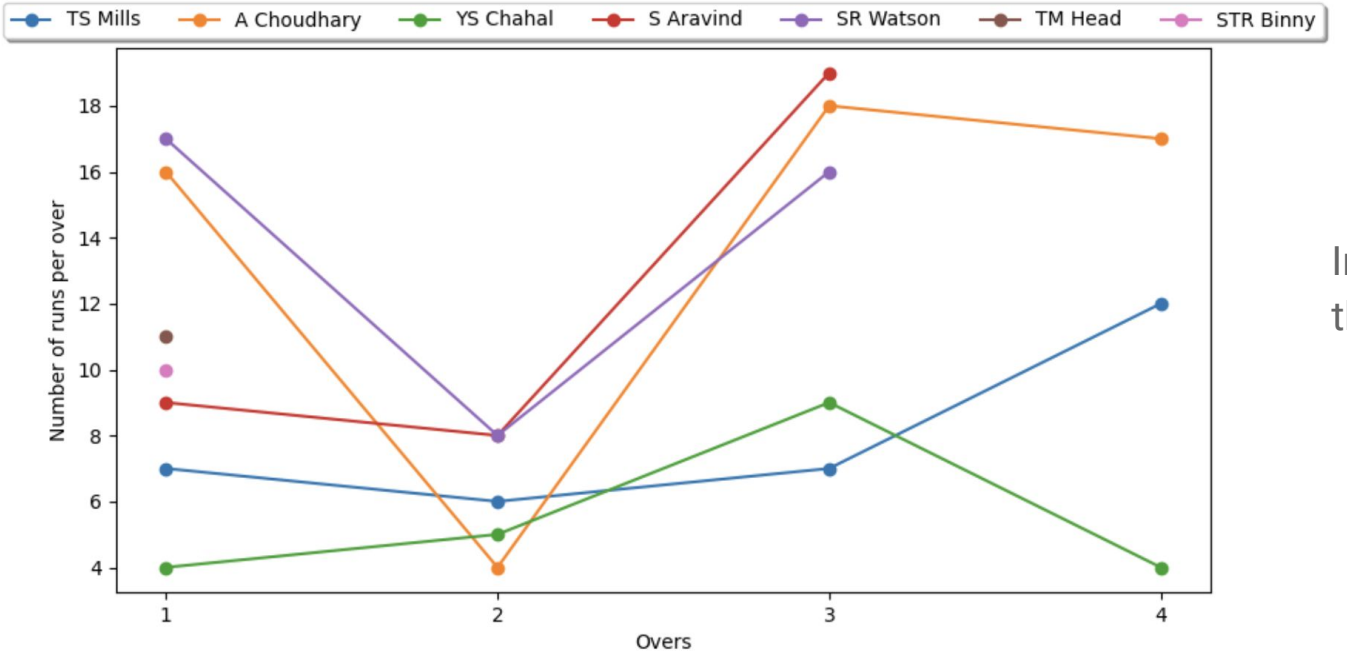
Bowler Statistics



Interpretation: **Lower** the better

- Provides neat comparison between two bowlers across their IPL career
- Aids teams to make the right decision during auction

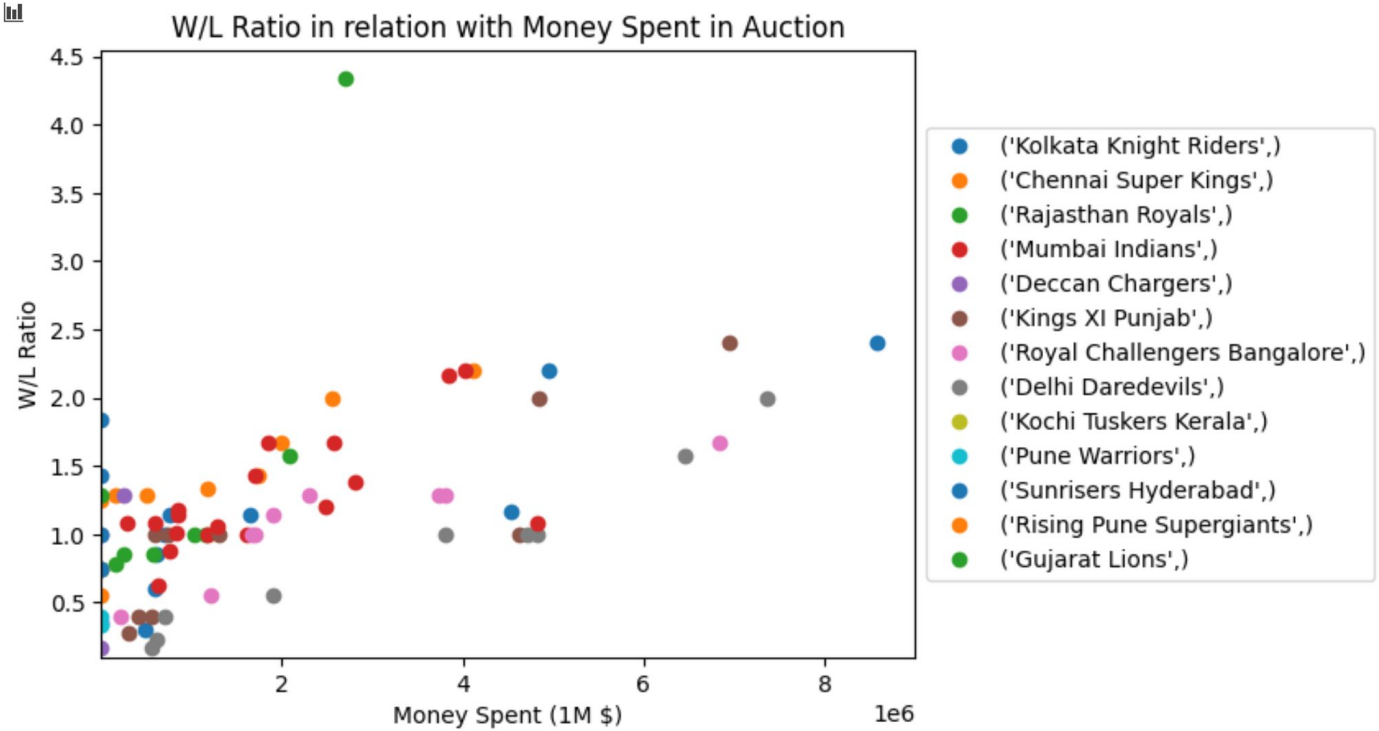
Bowling statistics (per inning)



Interpretation: **Lower** the better

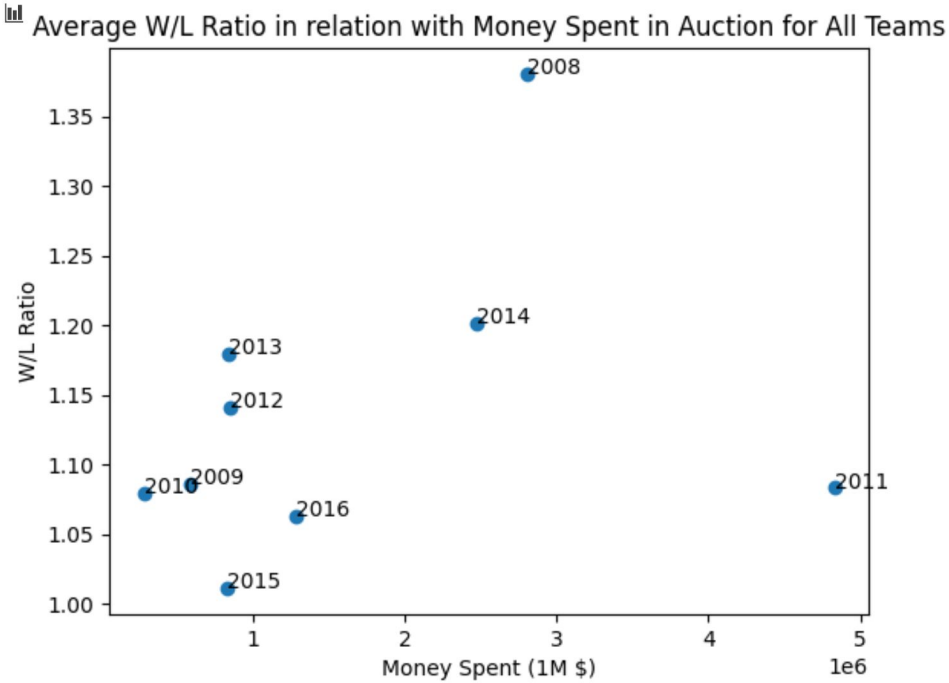
- Provides useful information about bowler's ability to bowl final overs

Win/Loss Ratio V.S. Auction Spendings



- Win:Loss to money spent per year by each team data points
- Positive correlation b/w them
- More money spent indicates higher chance of success

Average W:L vs Auction Spending



- Average W:L to average of total spending by all teams per year
- Positive but rather weak correlation
  - Money matters but there are other more important factors such as:
    - Familiarity with team members
    - Strategy
    - Etc

Summary and Applications

- Summary of results:
  - Team success based on in-game strategy:
    - Able to find useful information for teams to strategize based on toss-decision and player rotation
    - Other data to consider: field positions, specific player matchups (bowler vs. batter)
  - Team success based on management decisions:
    - Positive but weak correlation between win-rate and money spent on players
    - Other data to consider: team synergy (which players to put on the same team), money spent on training facilities, money spent on coaches
- Applications:
  - Team management can base their investment decisions on these statistics
  - Individual players can compare their own stats against other players to determine areas where they should improve
  - Sports-bettors can use these statistics to make better-educated bets
  - The methodology in this EDA can be applied to any sport