IPL 2008-2019 Dataset Analysis

Debmalya Biswas

2022-11-05

Introduction

The Indian Premier League (**IPL**) is a Men's T20 franchise cricket league of India that started in 2008. In 2010, the IPL became the first sporting event in the world to be broadcast live on YouTube. The brand value of the IPL in 2019 was 47,500 crore (US$5.9 billion). The IPL is the most-attended cricket league in the world and in 2014 was ranked sixth by average attendance among all sports leagues. In this visualization project, we are gonna perform Exploratory Data Analysis (**EDA**) on IPL 2008-2019 dataset by exploring through various statistical queries that might interest an inquisitive cricket mind.

Dataset & Variables

Source: https://www.kaggle.com/anuranroy/ipldatasets

The dataset comprises of 2 csv files:

- **deliveries2.csv**: This dataset consists of 179078 rows and 22 columns. It contains ball-by-ball delivery-wise statistics of every match played in the period 2008-2019. The variables are as follows:

Variable	Type	Description
match_id	Numeric (discrete)	match id number
inning	Binary	1 denotes 1st innings, 2 denotes 2nd innings
batting_team	Nominal	name of the batting team
bowling_team	Nominal	name of the bowling team
over	Numeric (discrete)	over number in a match
ball	Numeric (discrete)	ball number in an over
batsman	Nominal	name of the striking batsman in the delivery
non_striker	Nominal	name of the non-striking batsman in the delivery
bowler	Nominal	name of the bowler
is_super_over	Binary	1 if it is a super over delivery else 0
wide_runs	Numeric (discrete)	runs for the batting team due to a wide ball delivery
bye_runs	Numeric (discrete)	bye runs
legbye_runs	Numeric (discrete)	legbye runs

Variable	Type	Description
noball_runs	Numeric (discrete)	runs due to a no-ball delivery
penalty_runs	Numeric (discrete)	penalty runs
batsman_runs	Numeric (discrete)	runs scored by the batsman in the delivery
extra_runs	Numeric (discrete)	total extra runs scored in the delivery
total_runs	Numeric (discrete)	total runs scored in the delivery
player_dismissed	Nominal	player dismissed in the delivery
dismissal_kind	Nominal	type of dismissal
fielder	Nominal	fielder who contributed in the dismissal
winner	Nominal	winner of the match

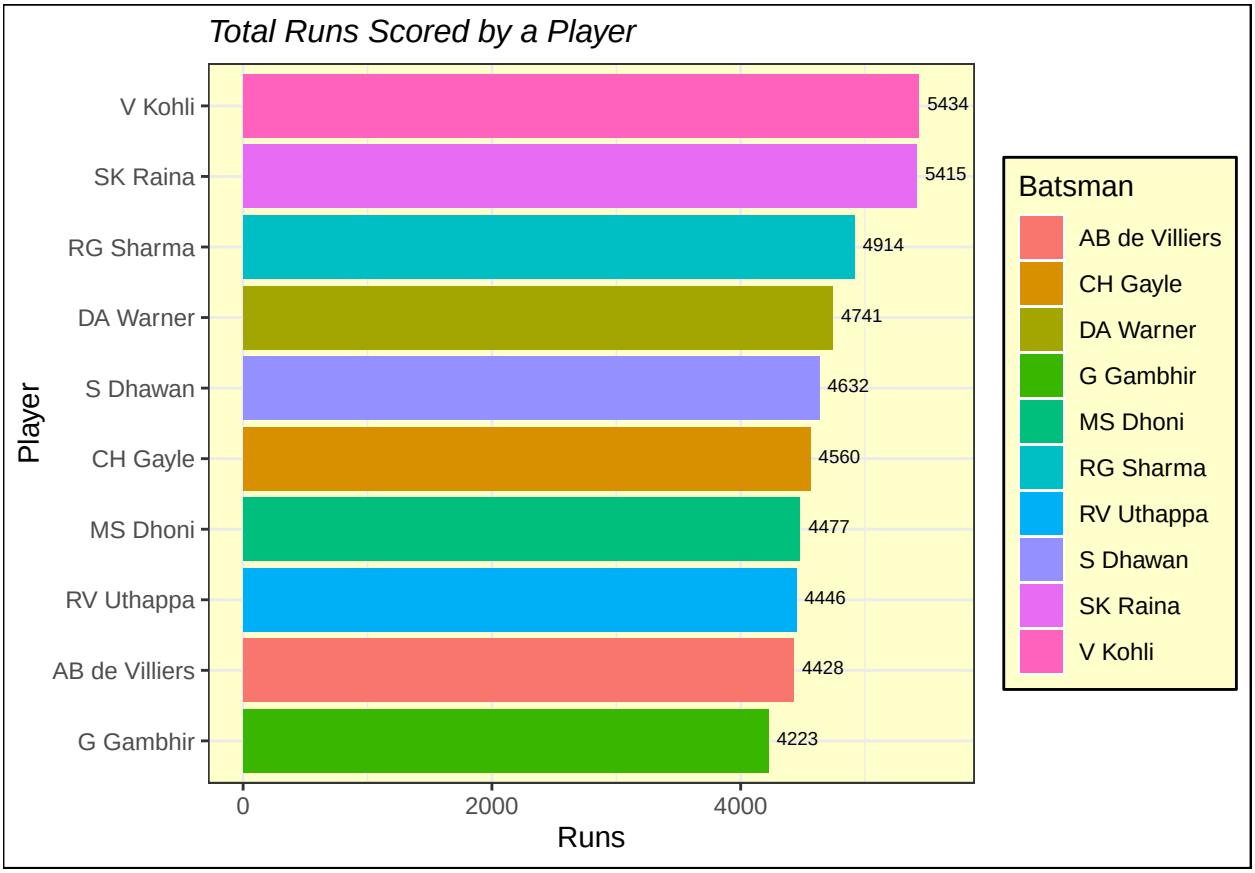
- **matches.csv**: This dataset consists of 756 rows and 18 columns. It contains match-wise statistics of every match played in the period 2008-2019. The variables are as follows:

Variable	Type	Description
id	Numeric (discrete)	match id number
season	Numeric (discrete)	IPL season
city	Nominal	city hosting the match
date	Ordinal	date of the match
team1	Nominal	first team
team2	Nominal	second team
toss_winner	Nominal	team winning the toss
toss_decision	Binary	toss winning team's decision to field or bat
result	Nominal	result type of the match is normal, tie or no result
dl_applied	Binary	1 if dls method applied else 0
winner	Nominal	winner of the match
win_by_runs	Numeric (discrete)	run difference if won by runs
win_by_wickets	Numeric (discrete)	wicket difference if won by wickets
player_of_match	Nominal	player of the match
venue	Nominal	venue where the match is played
umpire1	Nominal	1st umpire
umpire2	Nominal	2nd umpire
umpire3	Nominal	3rd umpire

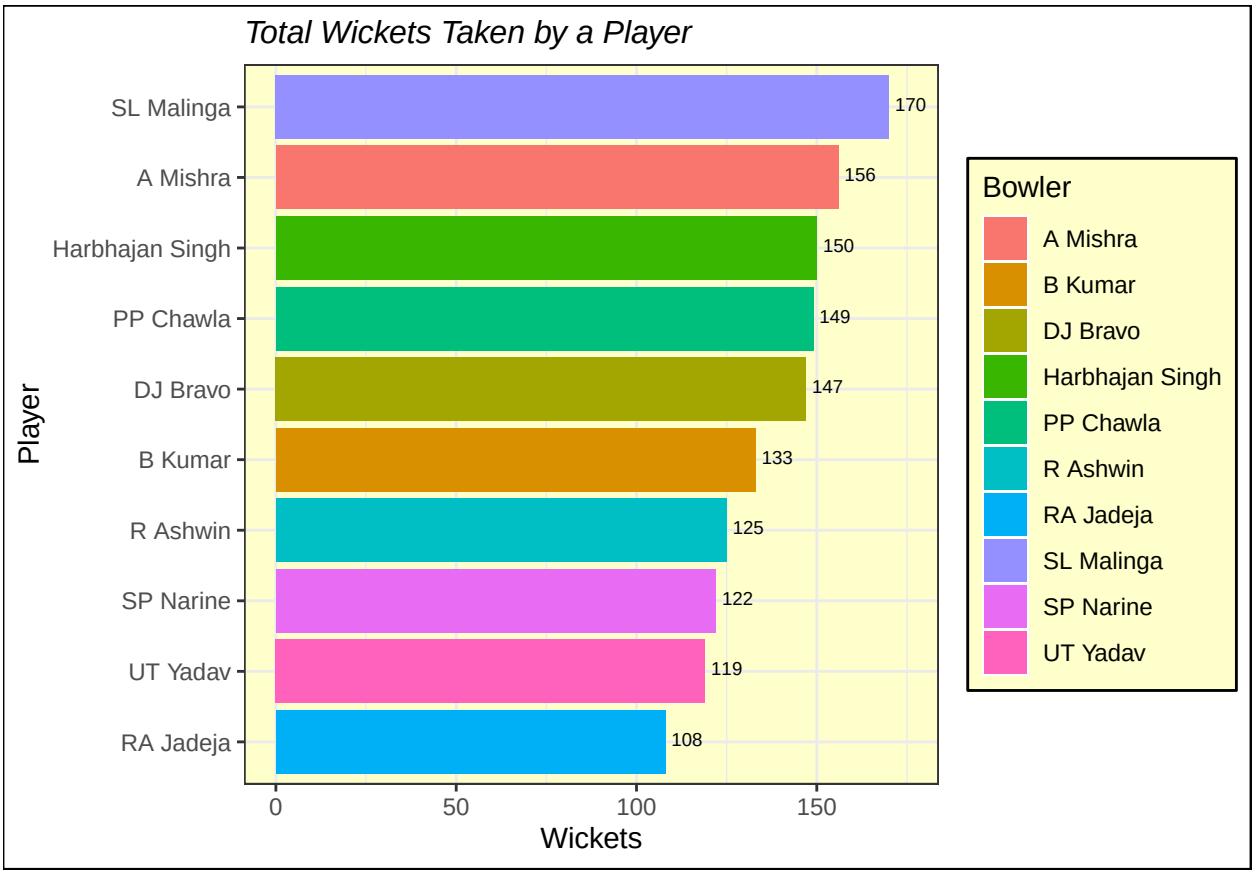
Data Analysis (DA)

Player-wise

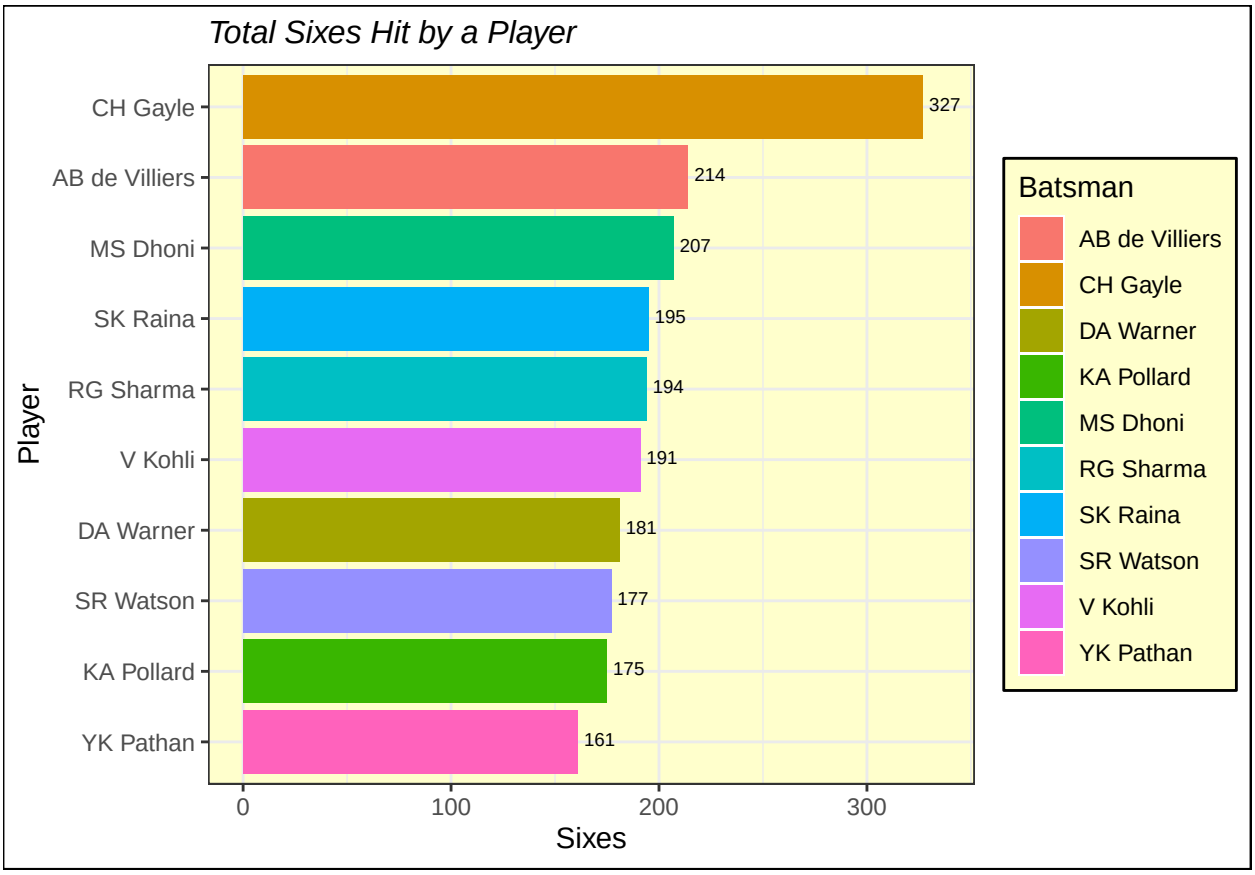
- Top 10 Run Scorers in IPL



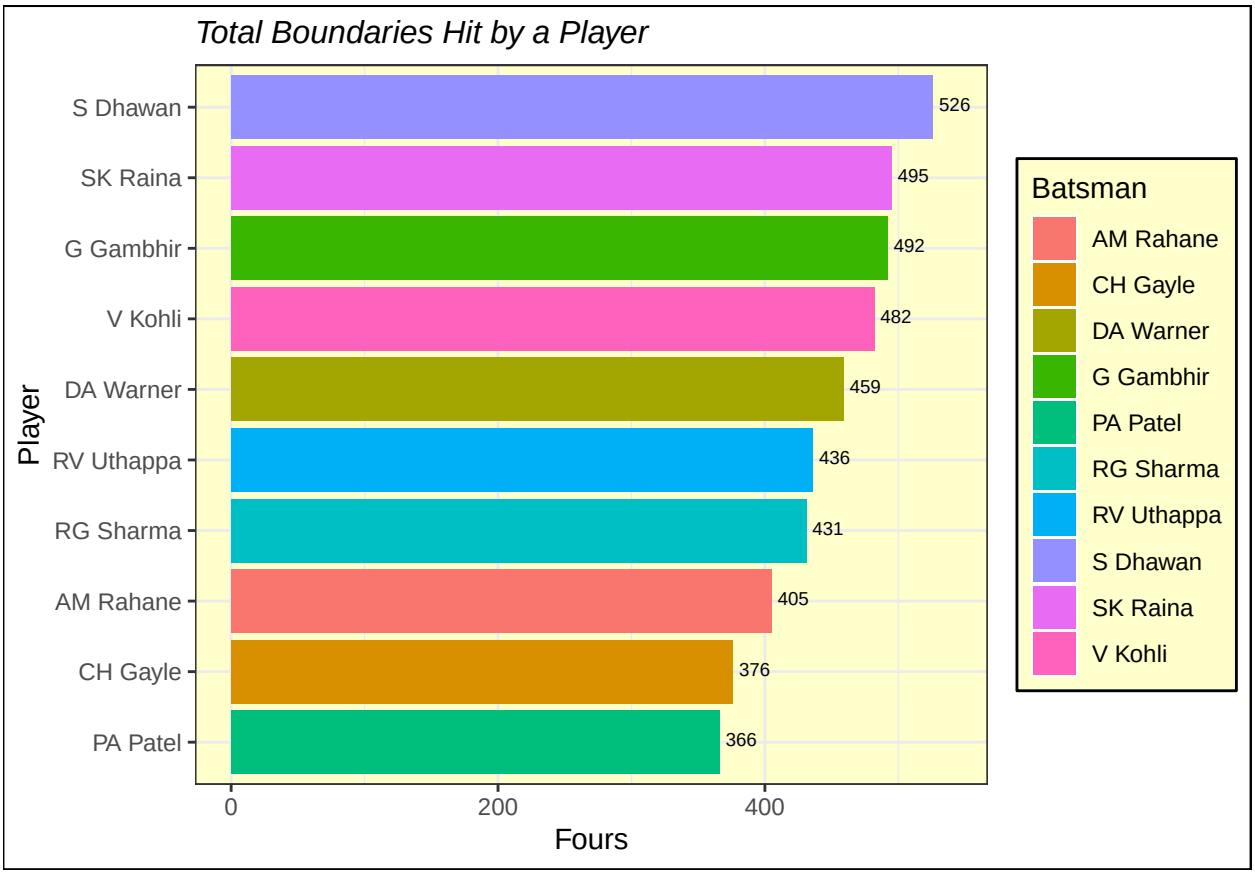
- Top 10 Wicket Takers in IPL



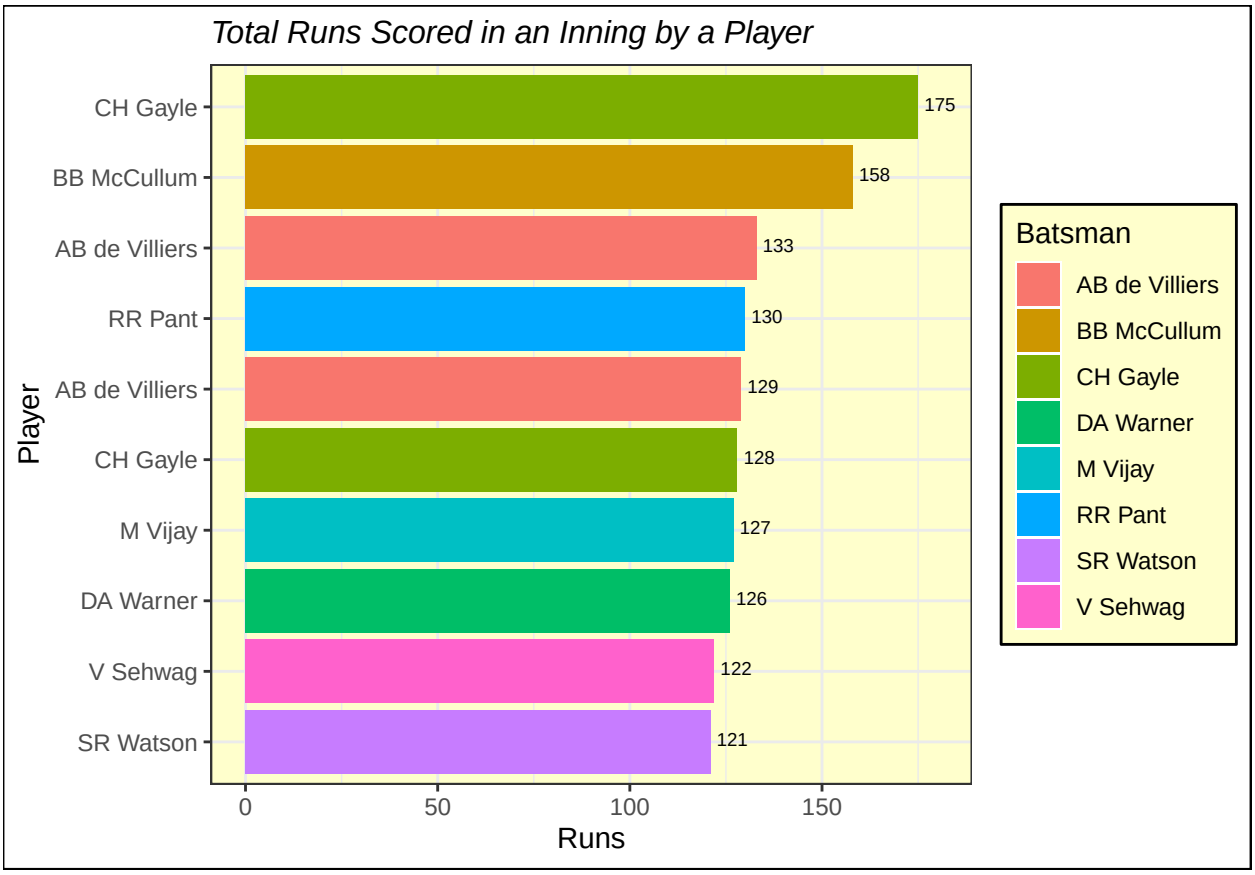
- Top 10 Six Hitters in IPL



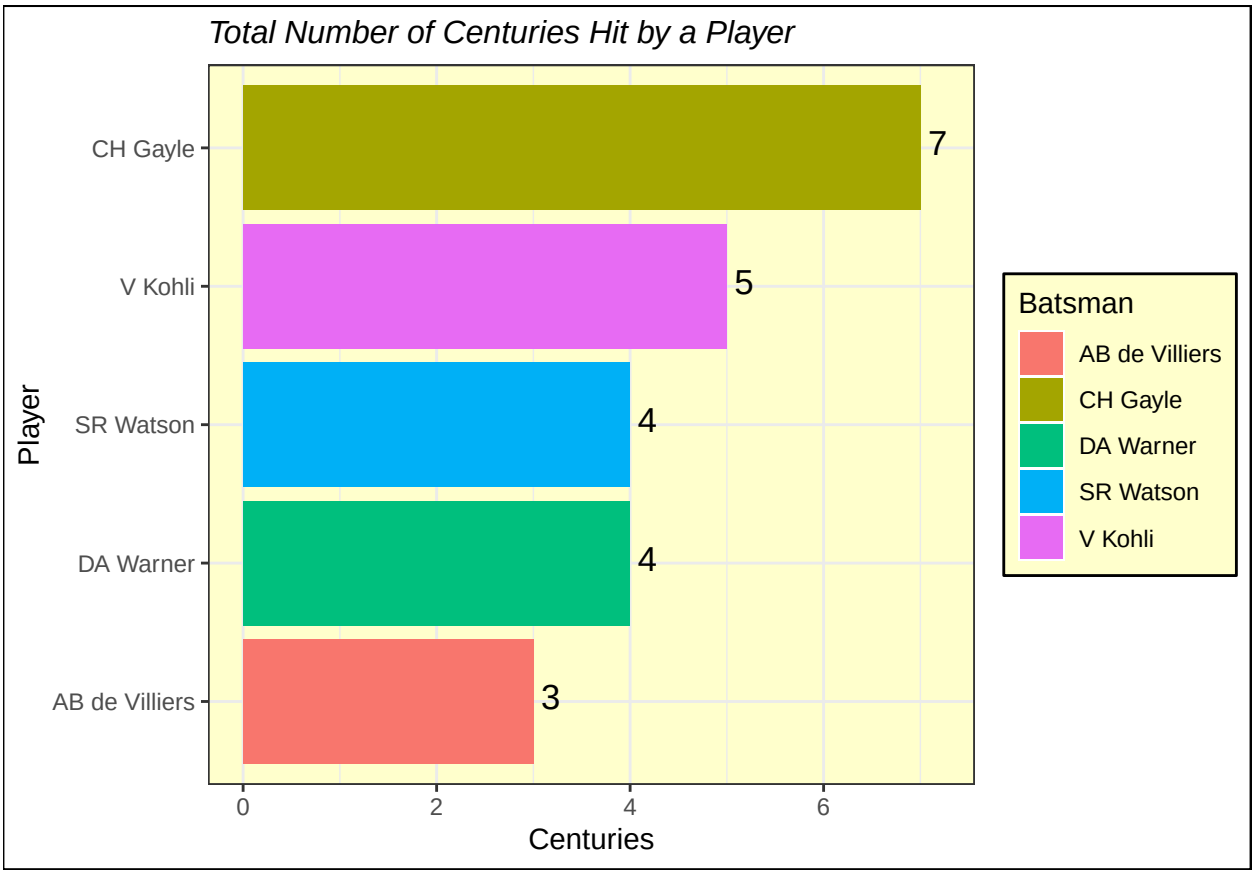
- Top 10 Boundary Hitters in IPL



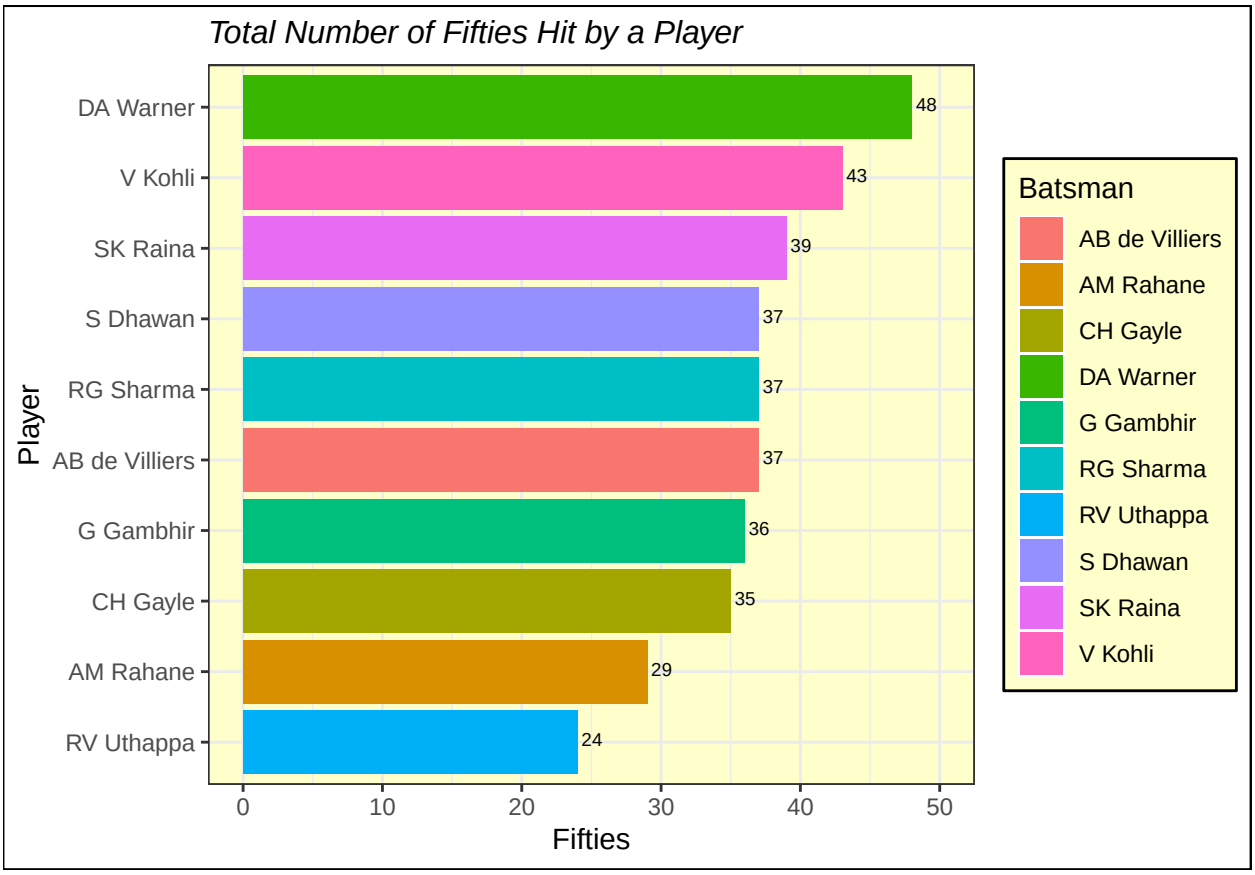
- Top 10 Individual Run Scorers in a Match



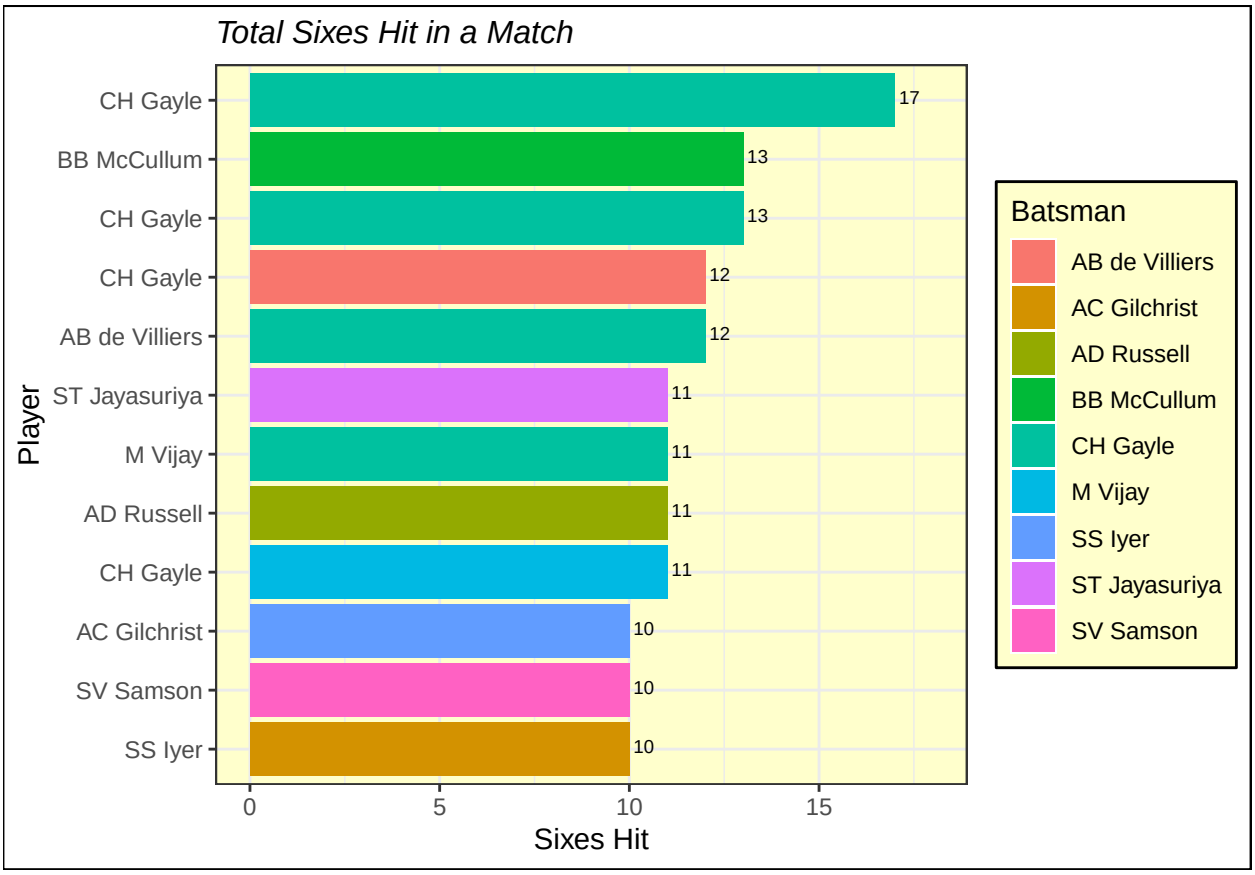
- Top 5 Players with Most IPL Centuries



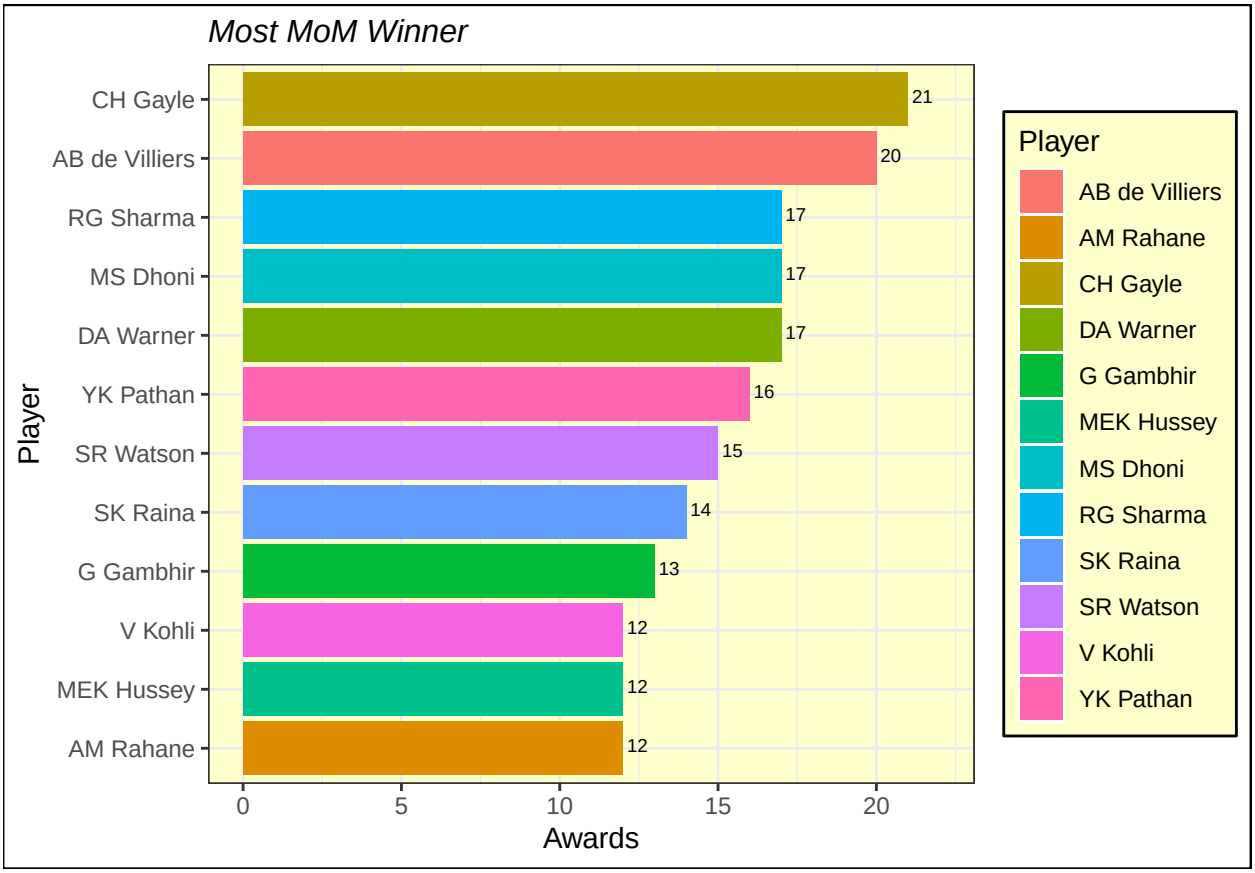
- Top 10 Players with Most IPL Fifties



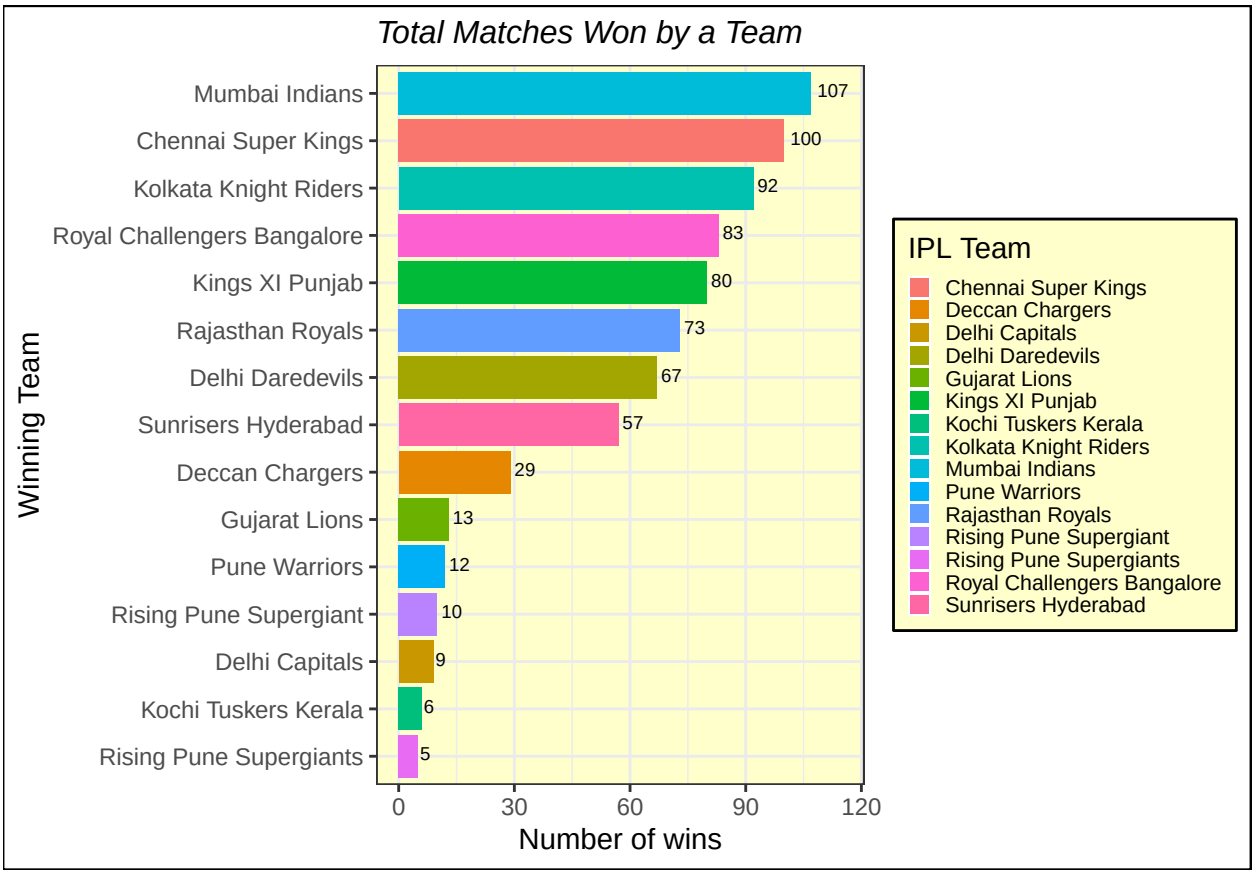
- IPL Players with Most Sixes Hit in a Match



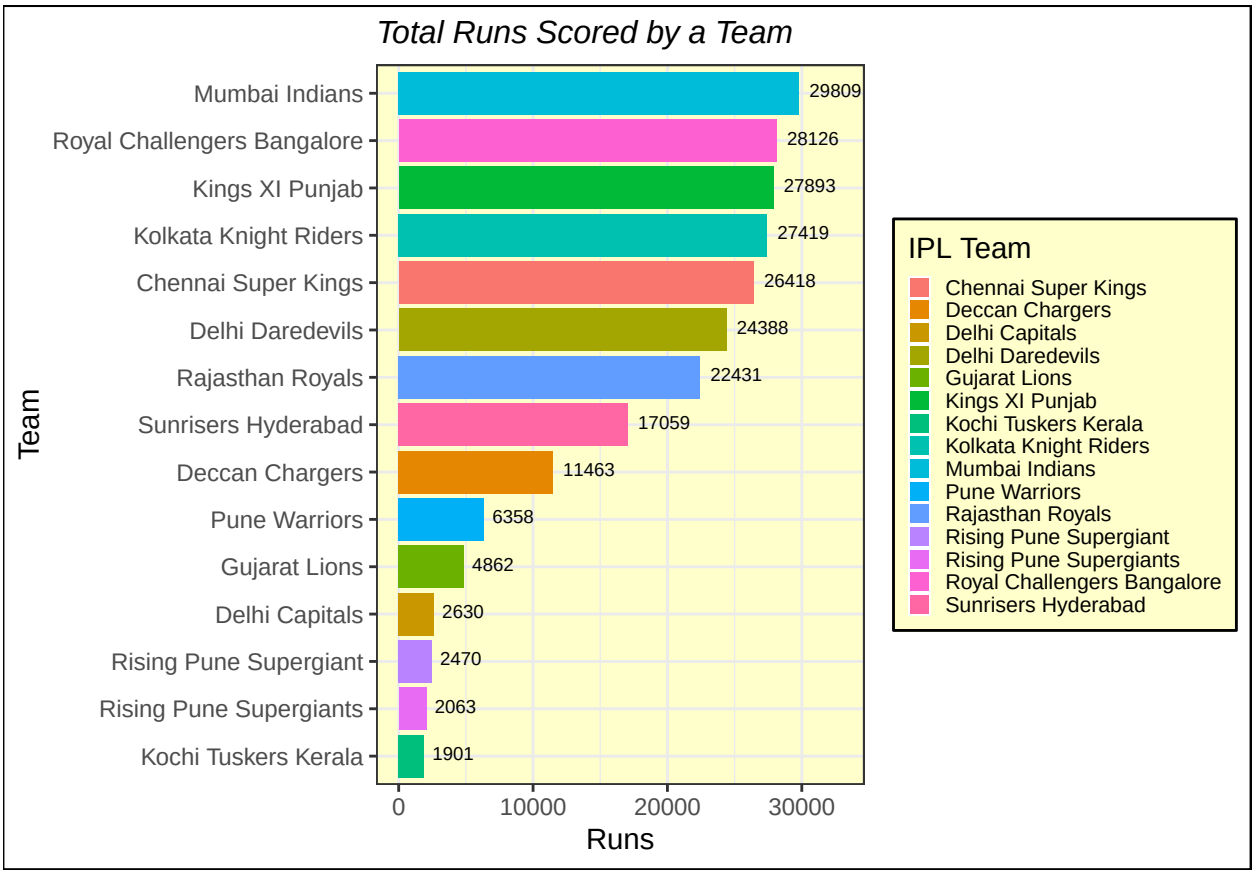
- Most Man of The Match Award Winners in IPL



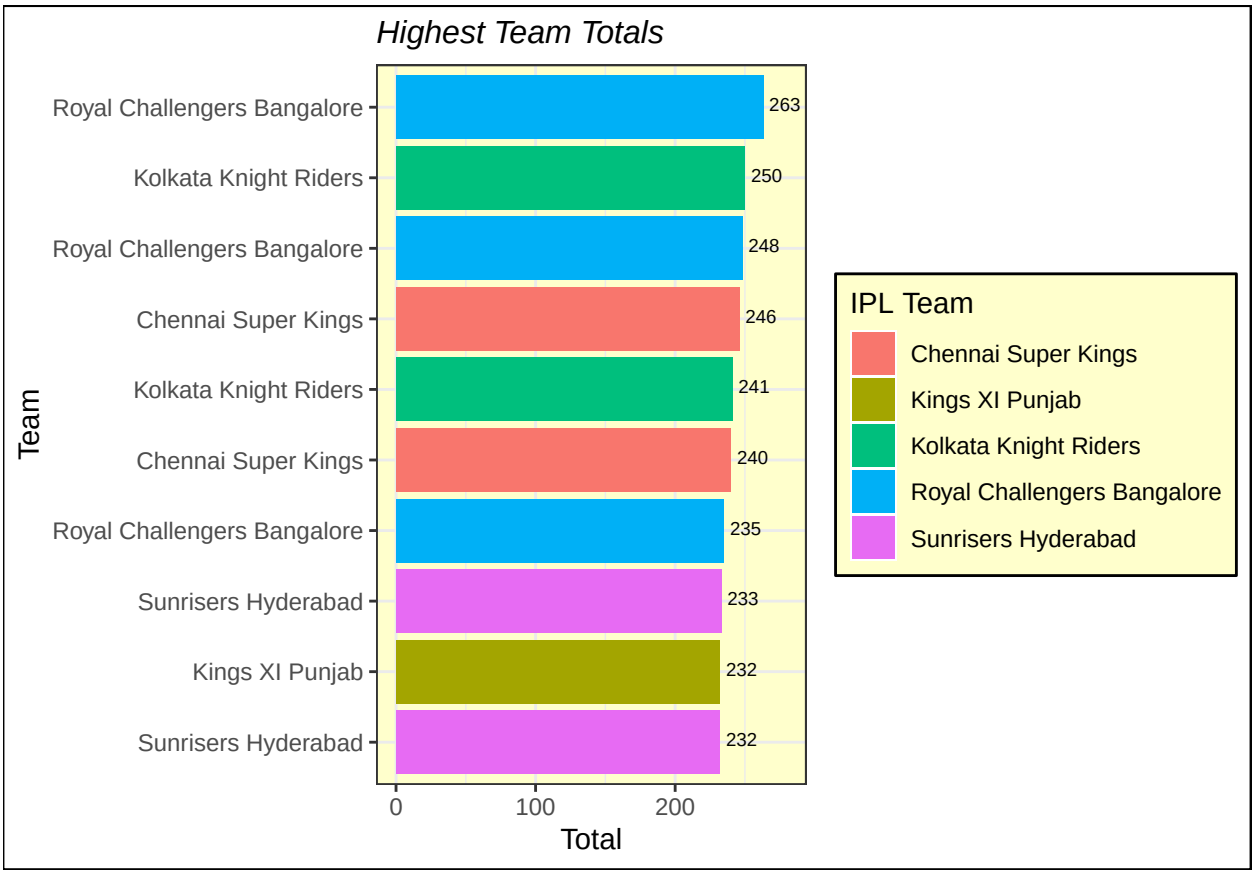
- Top 10 ### Team-wise
- Total Wins by Teams in IPL



- Total Runs Scored by Teams in IPL

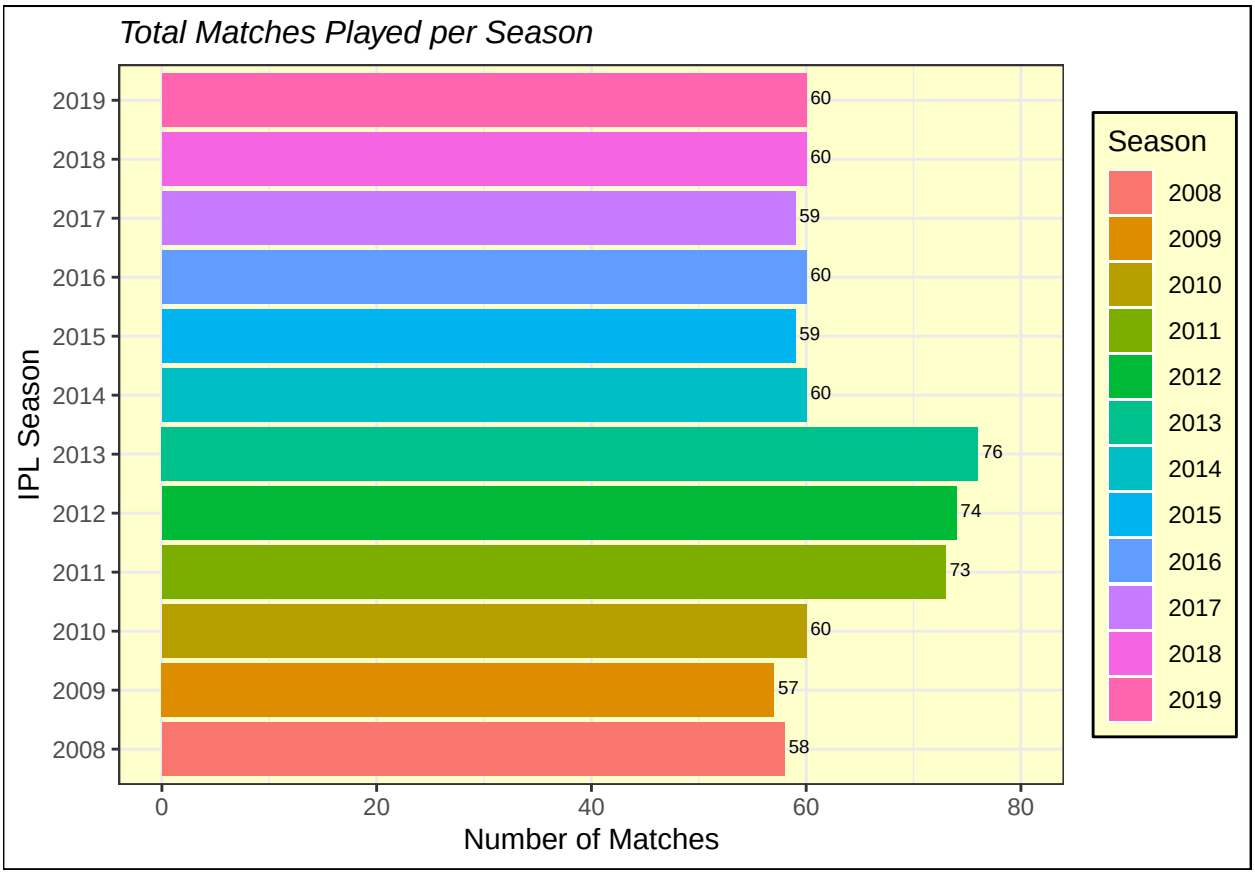


- Highest Total by Teams in IPL

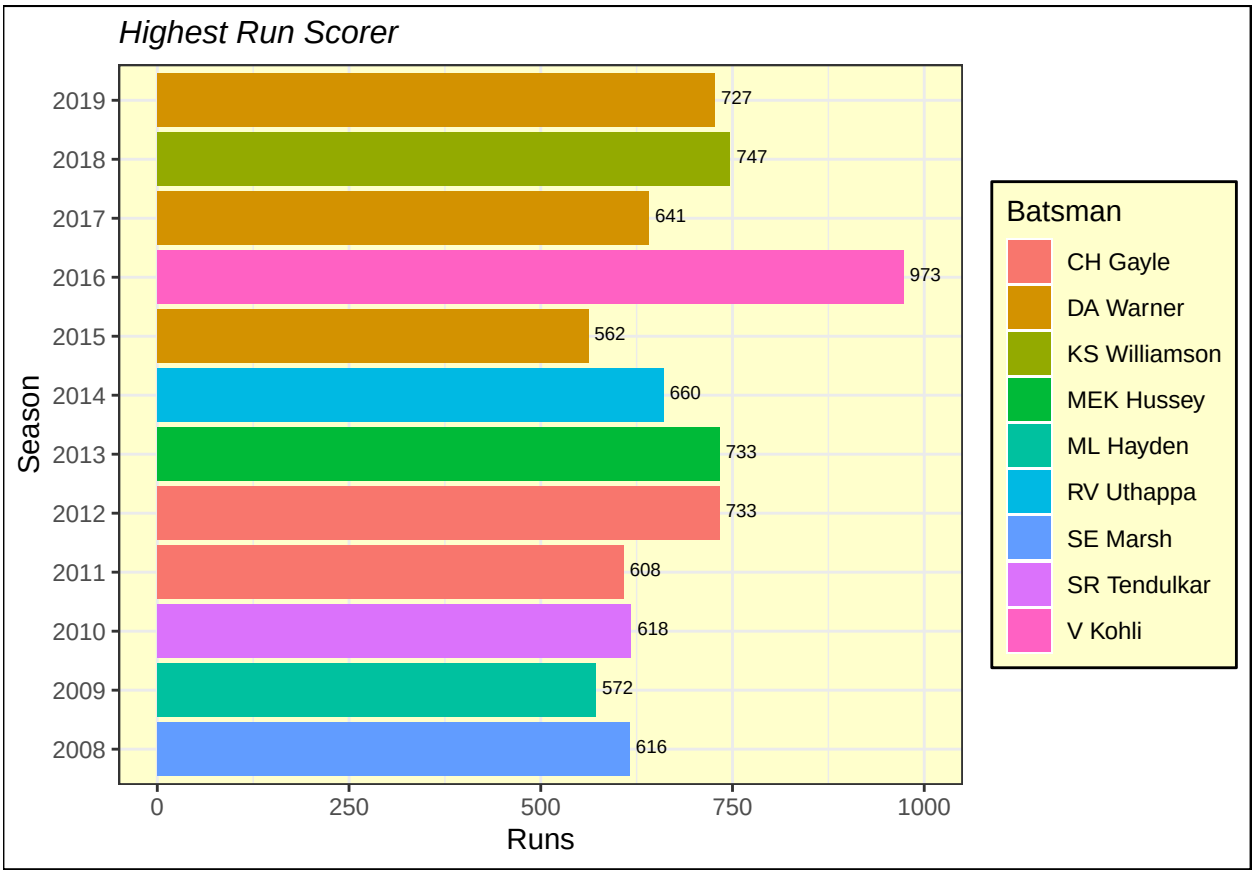


Season-wise

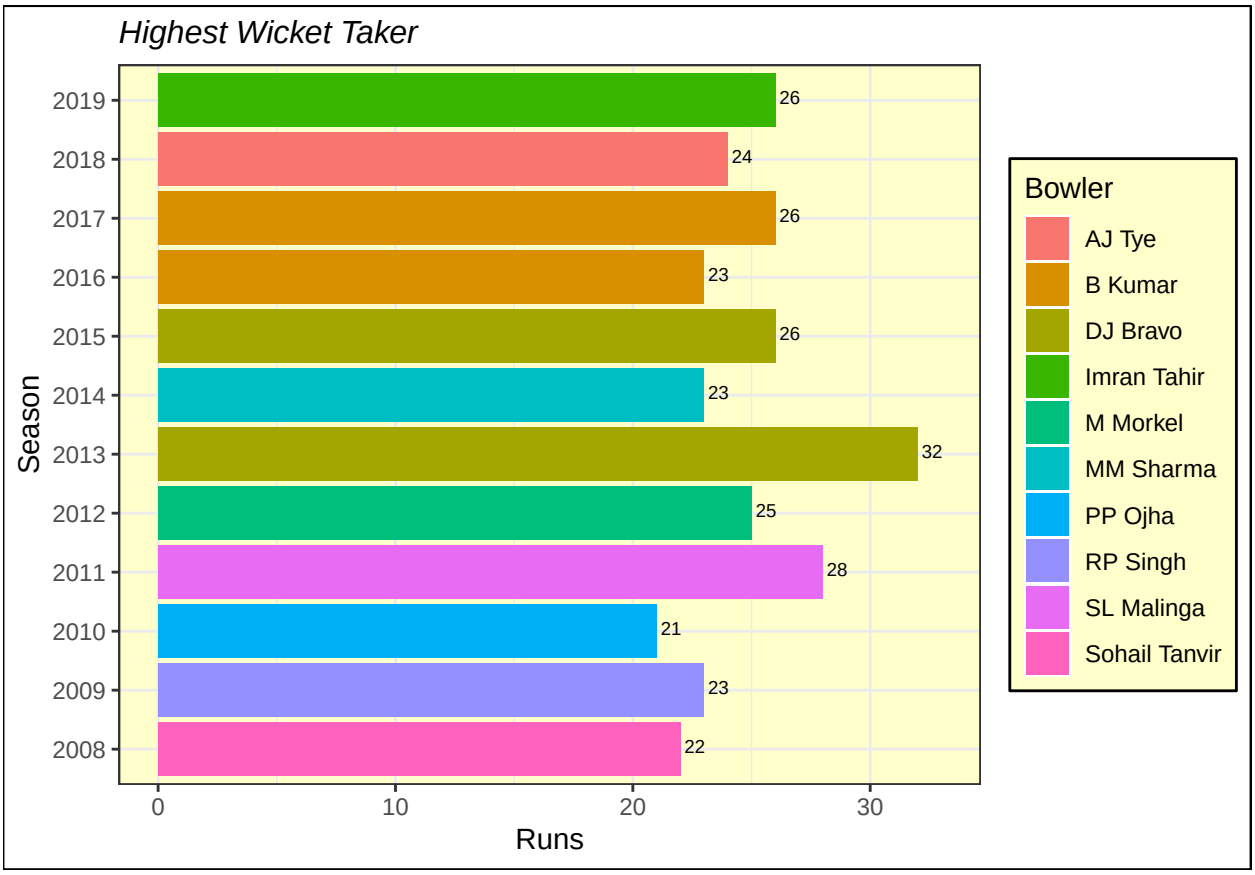
- Number of Matches Played Each Season



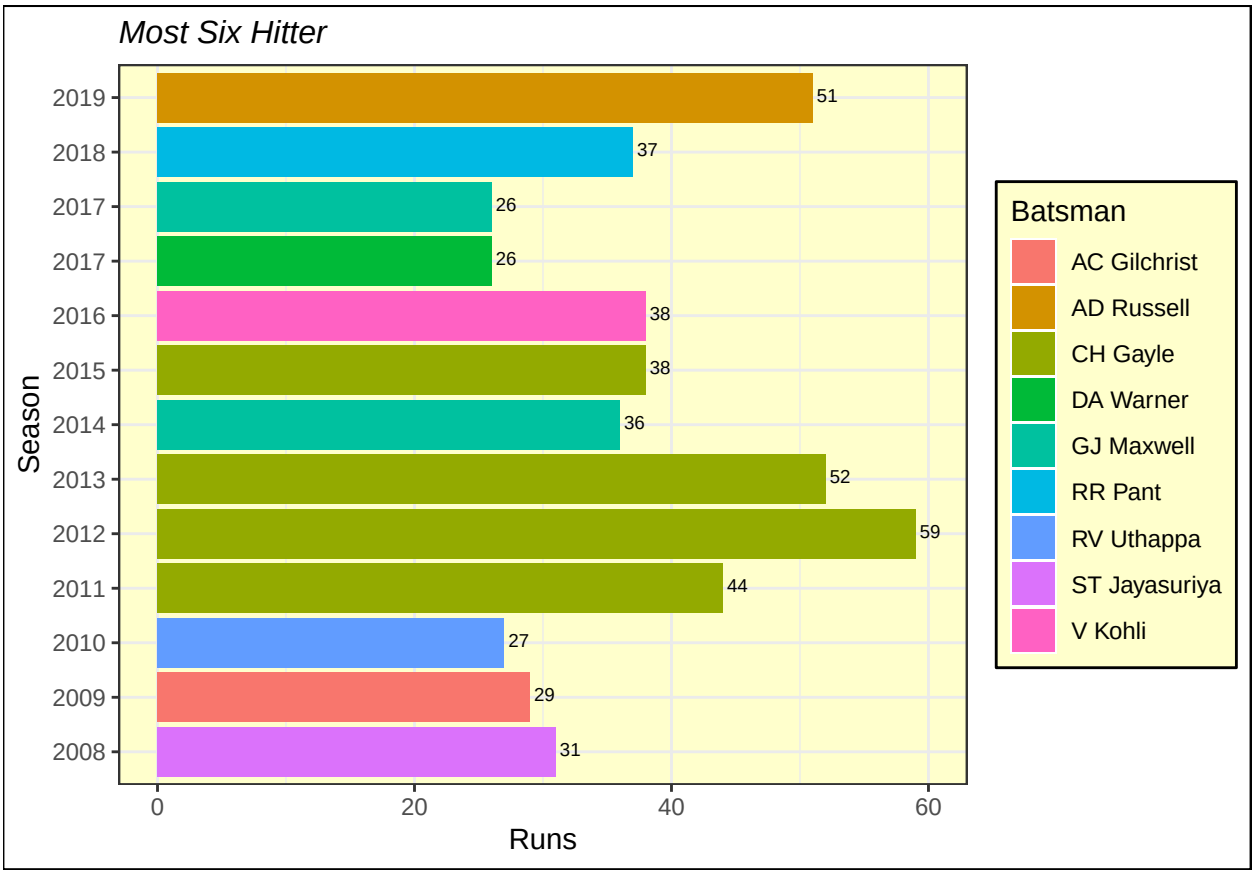
- Highest Run Scorer by Season in IPL



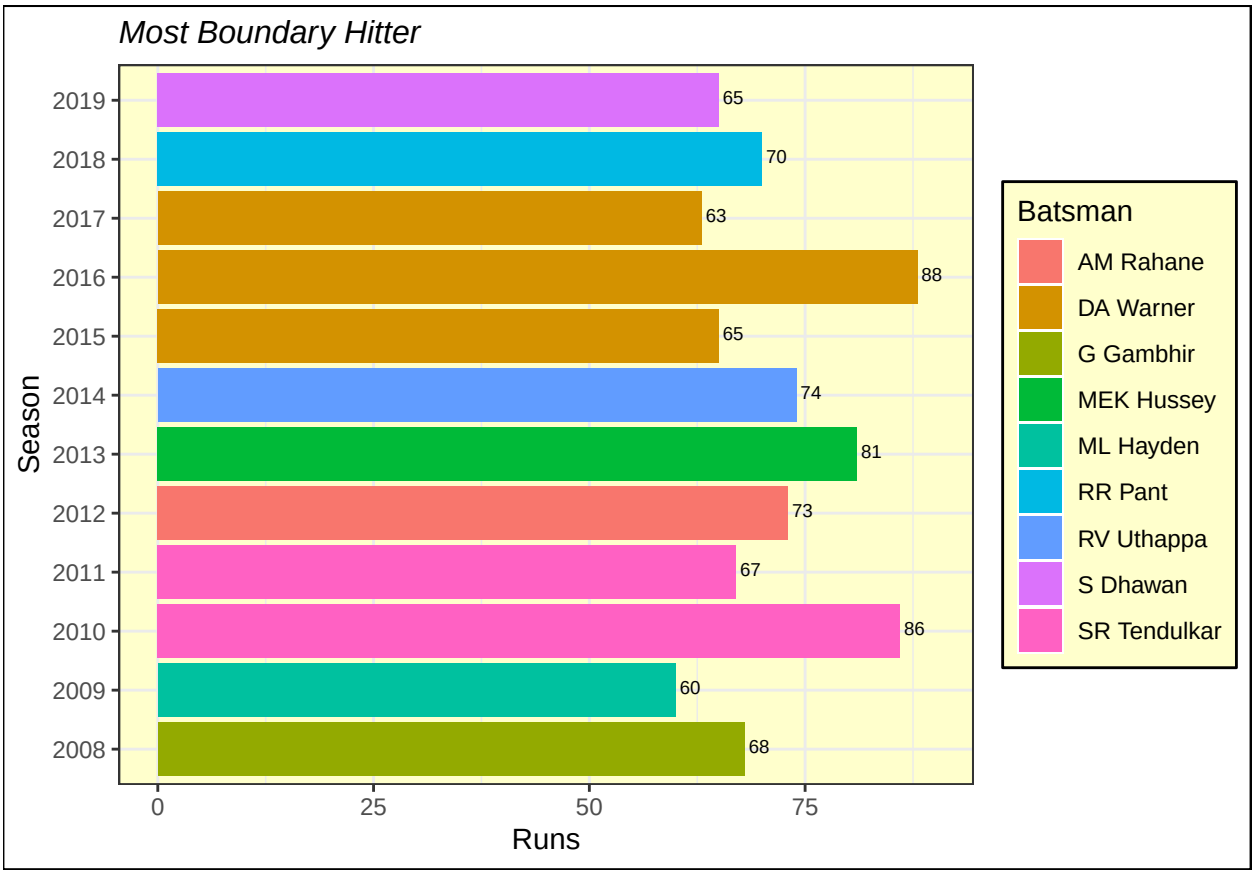
- Highest Wicket Taker by Season in IPL



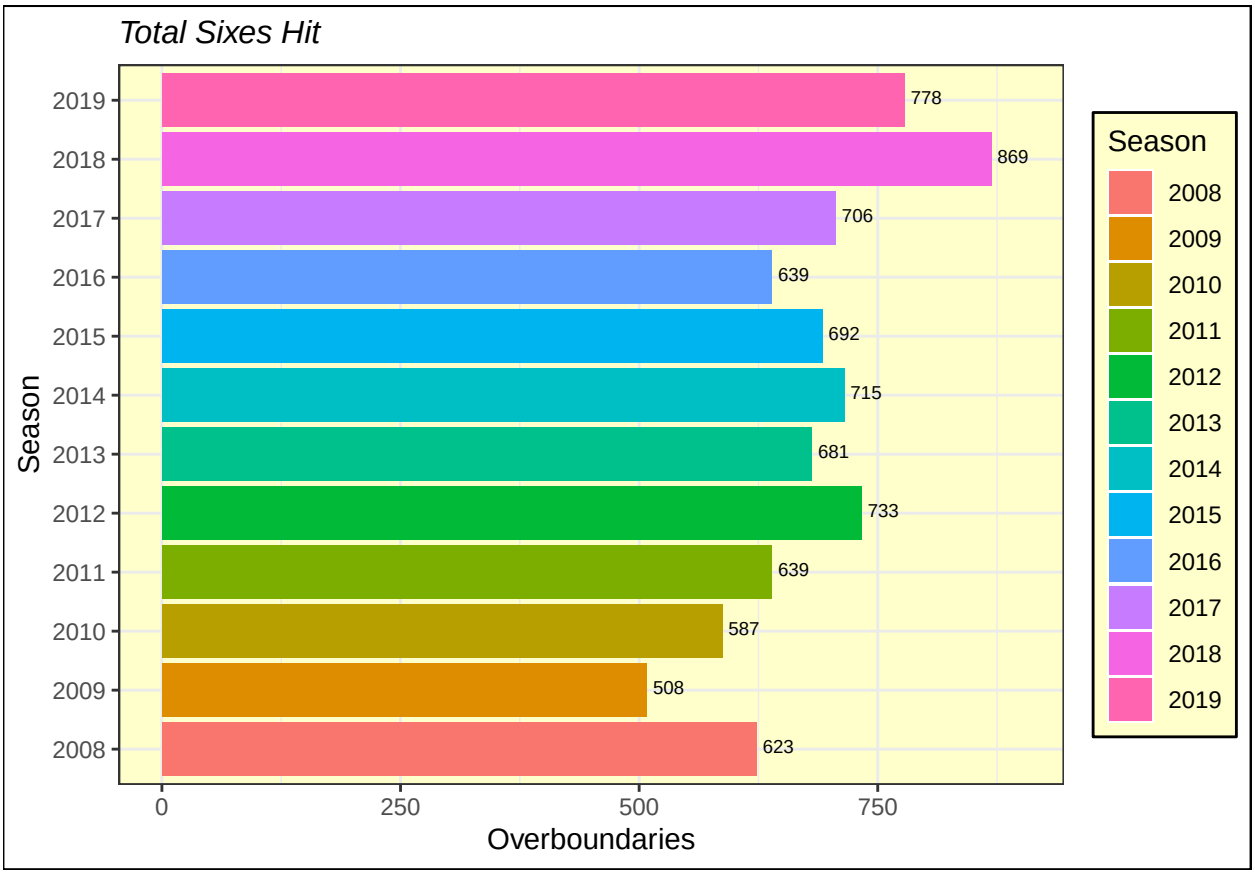
- Most Six Hitters by Season



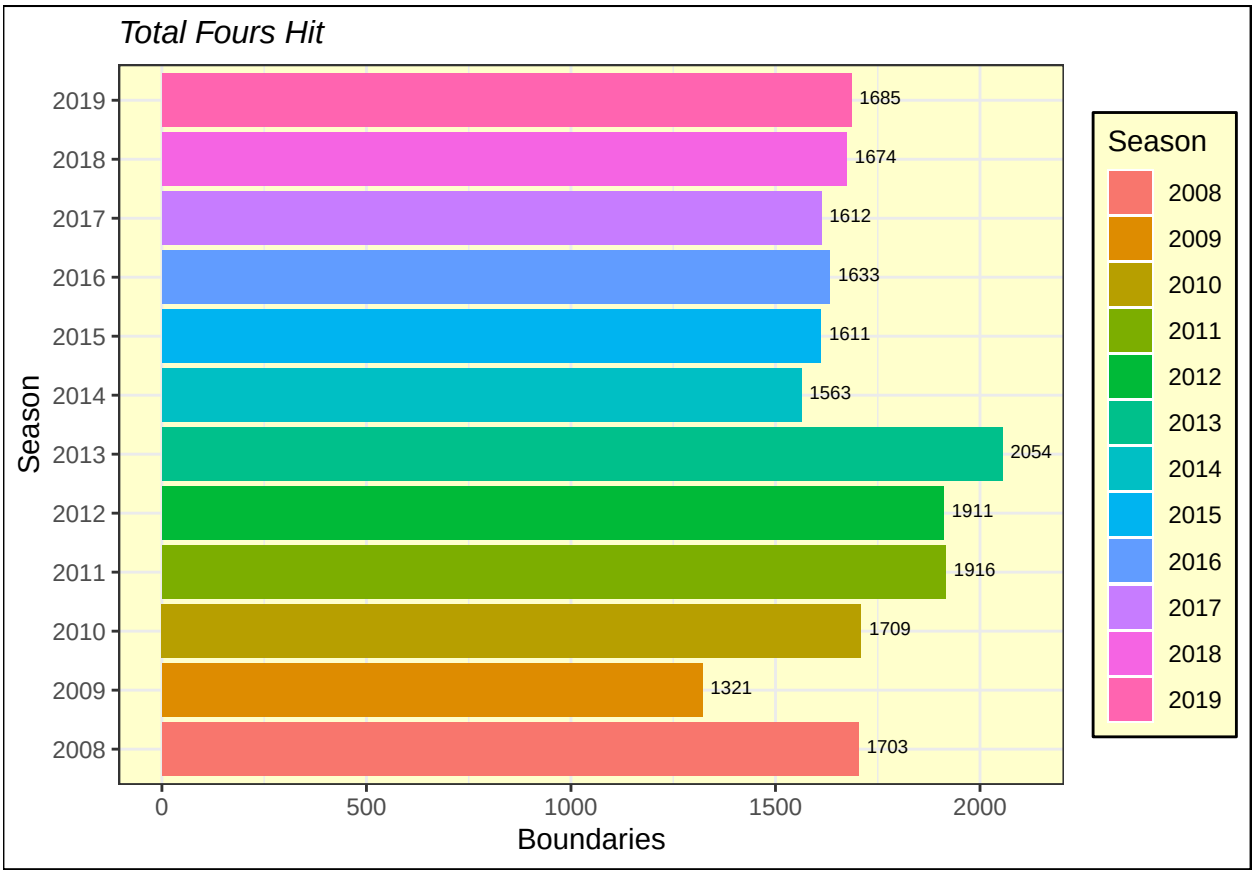
- Most Boundary Hitters by Season



- Total Sixes Hit by Season

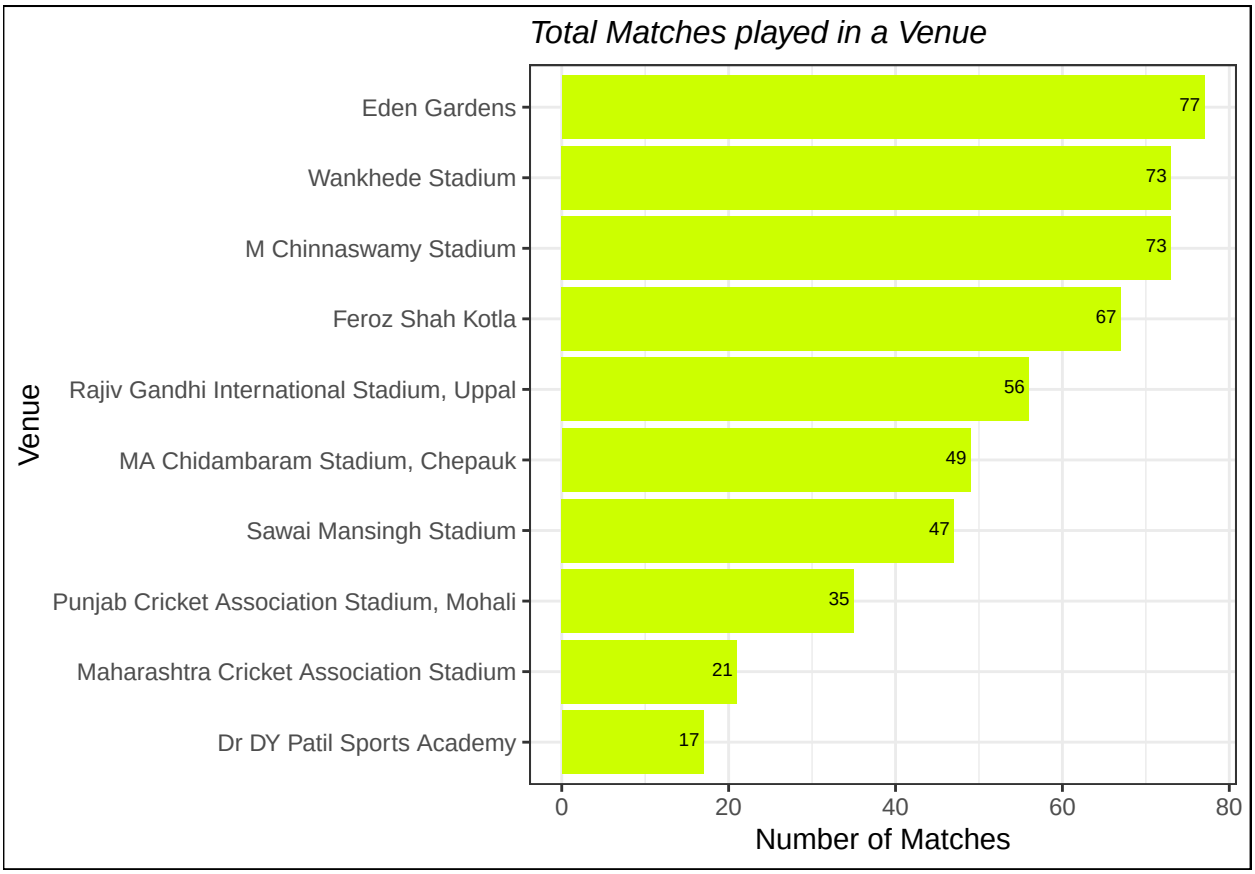


- Total Boundaries Hit by Season



Venue-wise

- Top 10 Venues with most IPL Matches Played



Conclusion

We have successfully performed an EDA by exploring thorugh various general, batting and bowling queries grouping them by player, team, season and venue, and answering them in a visual way for ease of capturing the data analysis and statistics.