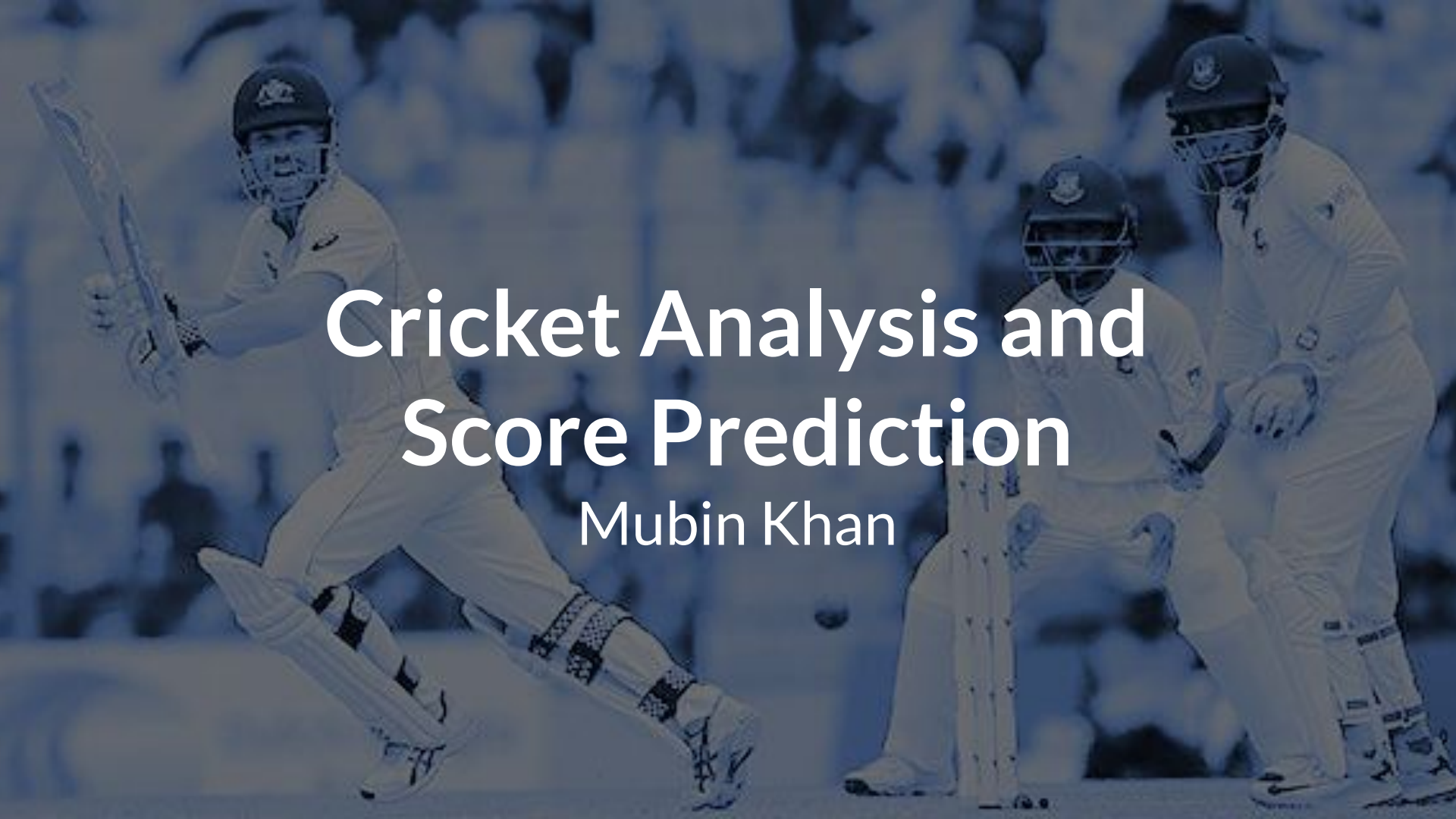


Cricket Analysis and Score Prediction

Mubin Khan

Understanding Cricket

What is it all about?

Cricket Players

Bowler - throw the ball

Batter - hits the ball

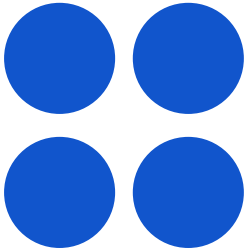
How to score in cricket?

1 Run

Both batsman run from respective ends of pitch once

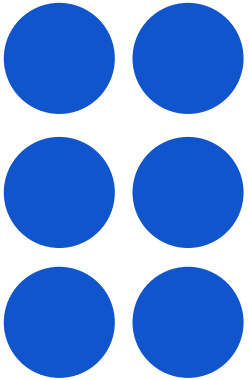
2 Runs

Both batsman run from respective ends of pitch twice



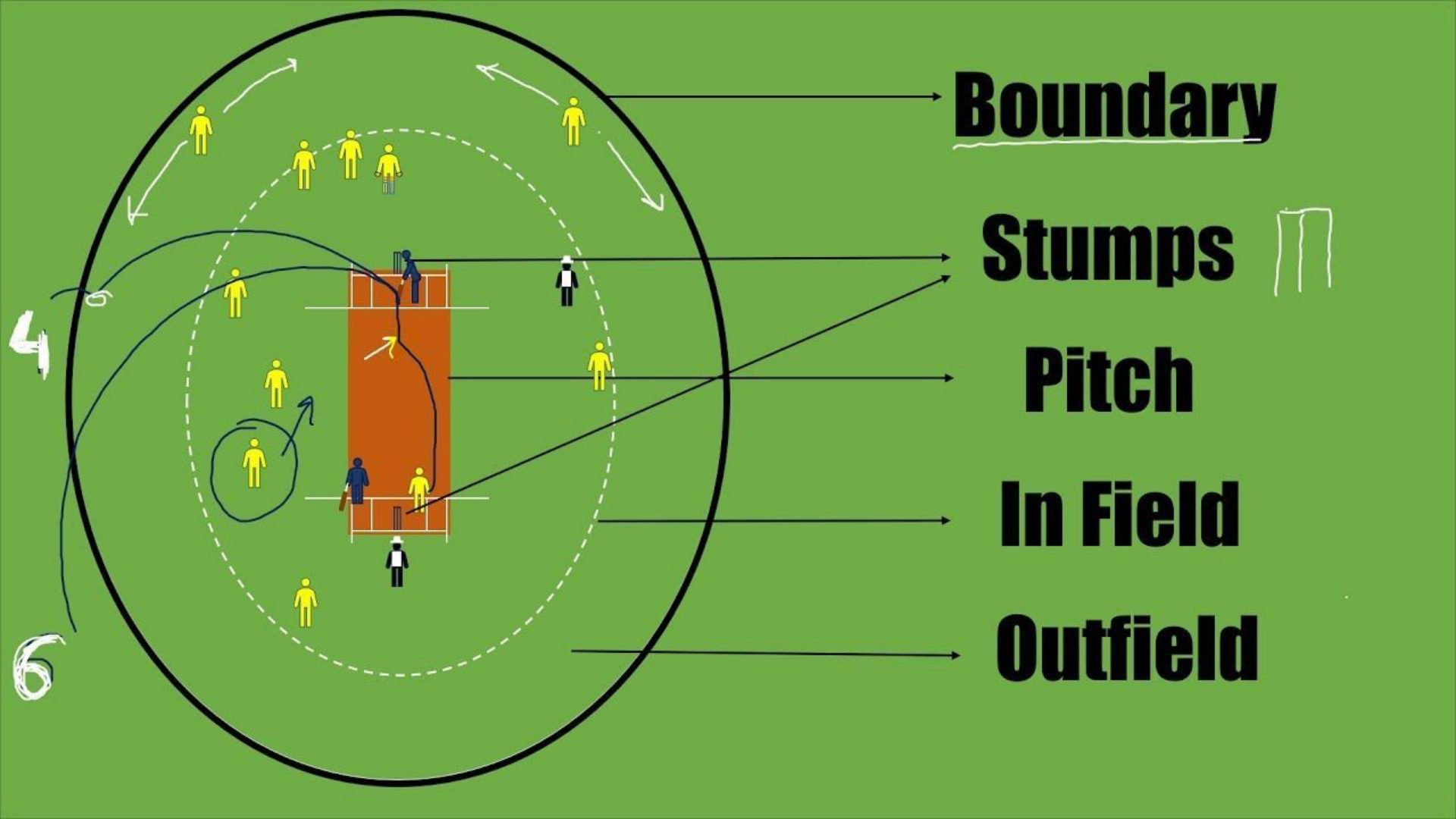
4 Runs

Ball hits the boundary line



6 Runs

Ball hits over the boundary line without bouncing


Boundary
Stumps
Pitch
In Field
Outfield
4
6

What am i trying to solve?

BUSINESS PROBLEM

Coaches and Captain of batting team needs to strategize the game plan based on current score statistics to increase winning chances.

Therefore, predicting score would help them optimise the performance of their team.

Predict the IPL score for the batting team

DATA SCIENCE PROBLEM

What are my goals?

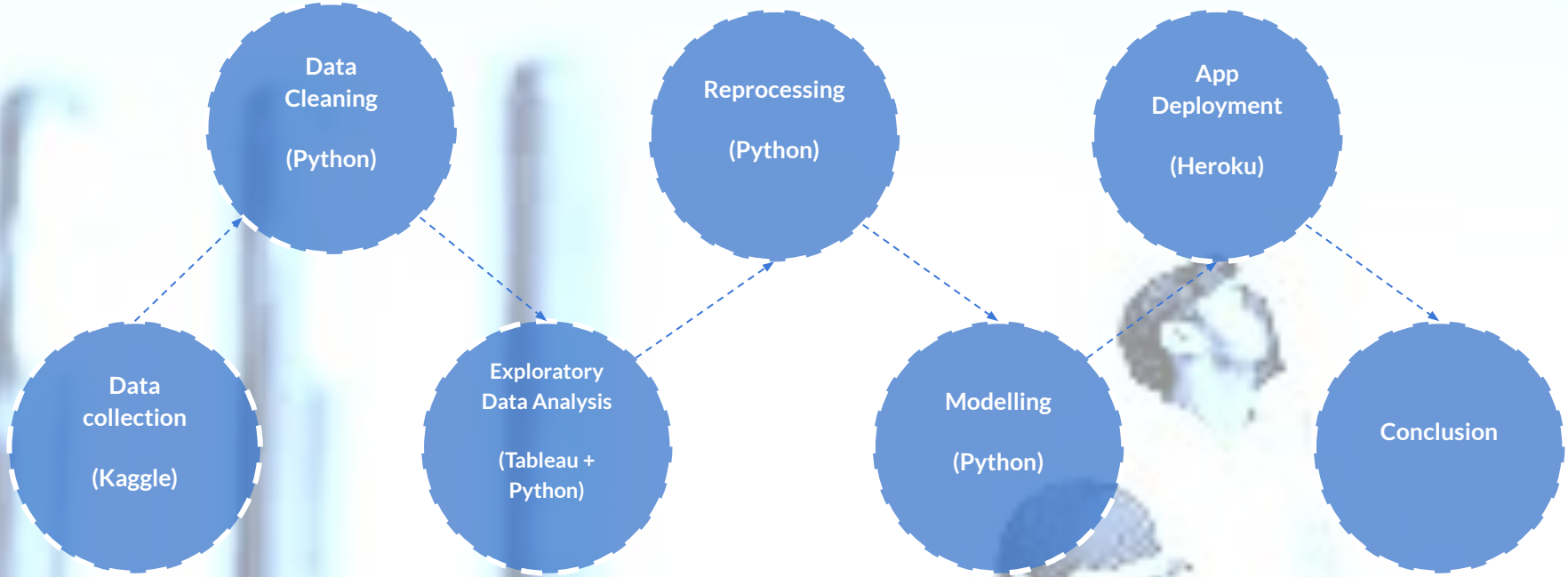
Gather dataset for IPL matches

Perform data analysis on the collected data to look for trends and select important variables

Use Linear Regression models to predict the score of IPL match for batting team.

Use flask to create web framework

The journey I took



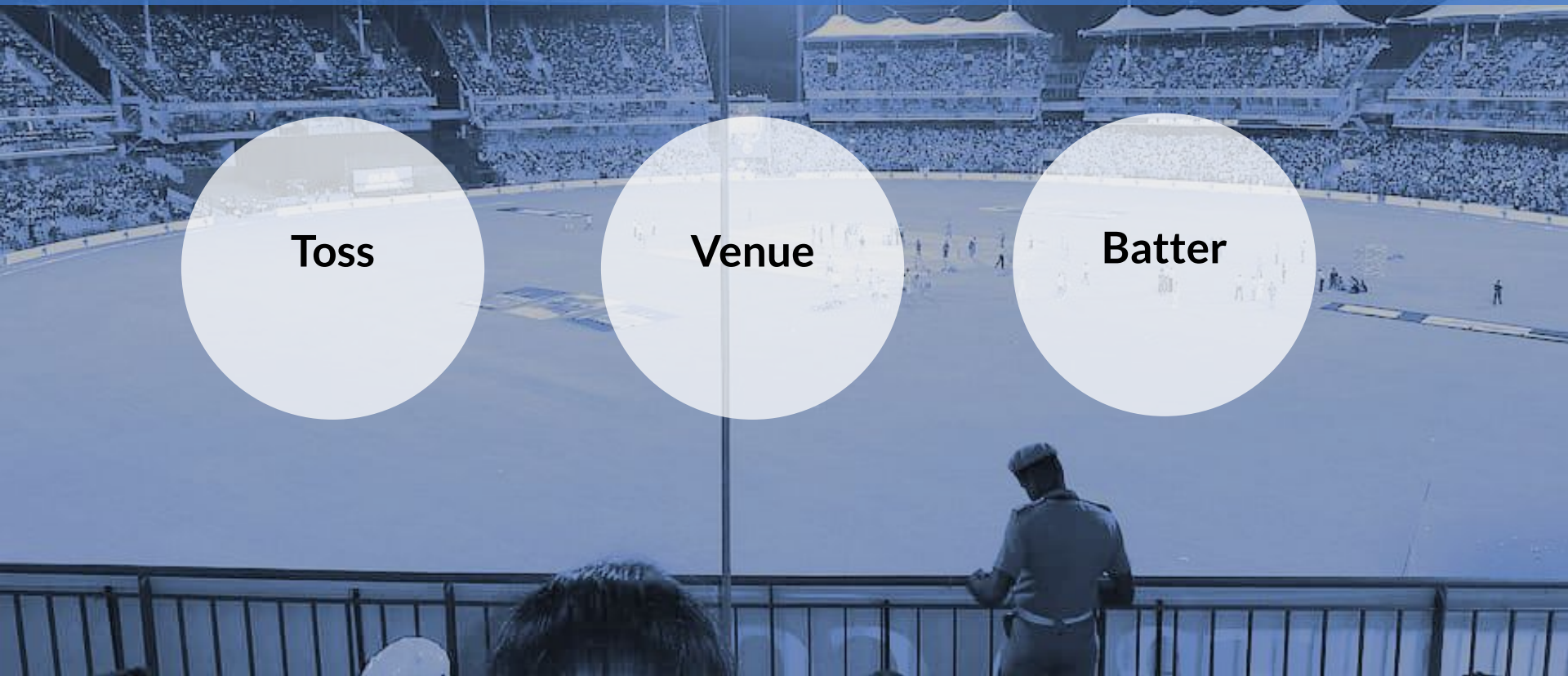
Data used to inform my decisions

	Macro Data (Overview)	Micro Data (Granular)
Data Collection	IPL matches dataset 2008-2020	IPL ball-by-ball dataset
Data Source	IPL [Kaggle Dataset 2008-2020] (https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020)	

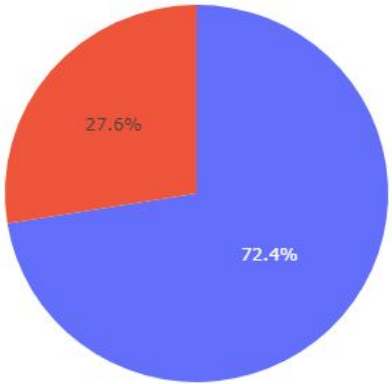
Uncovering trends and patterns

Key Findings from EDA

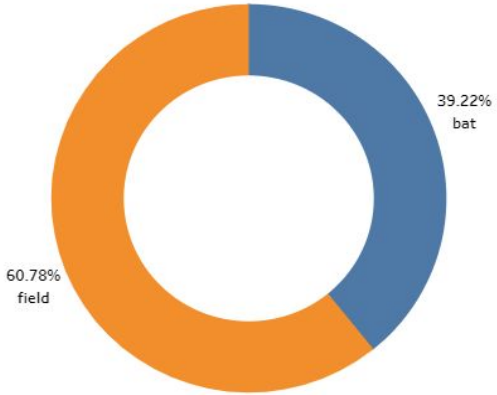
Key factors impacting the gameplay



How winning the toss impact the gameplay

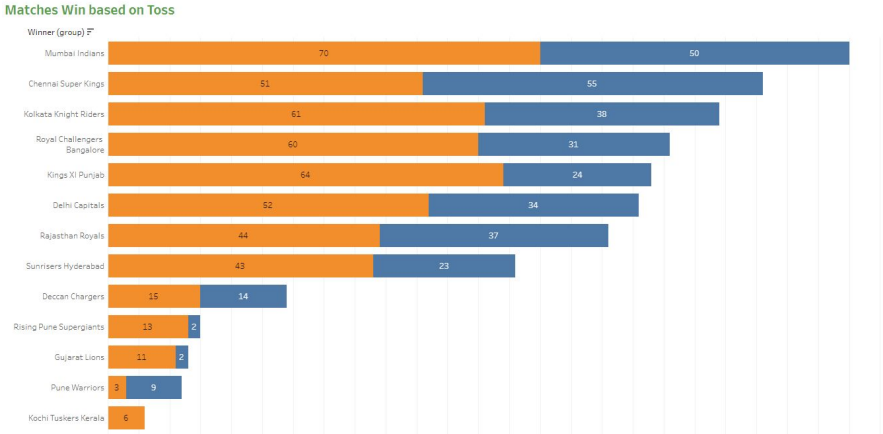


70% of the team who wins the toss Wins the game



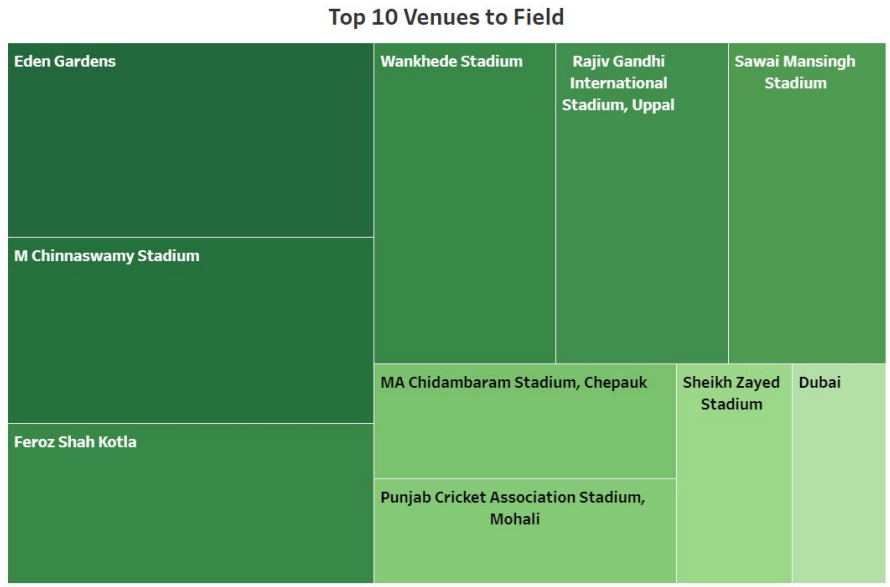
Those who win the toss would generally field

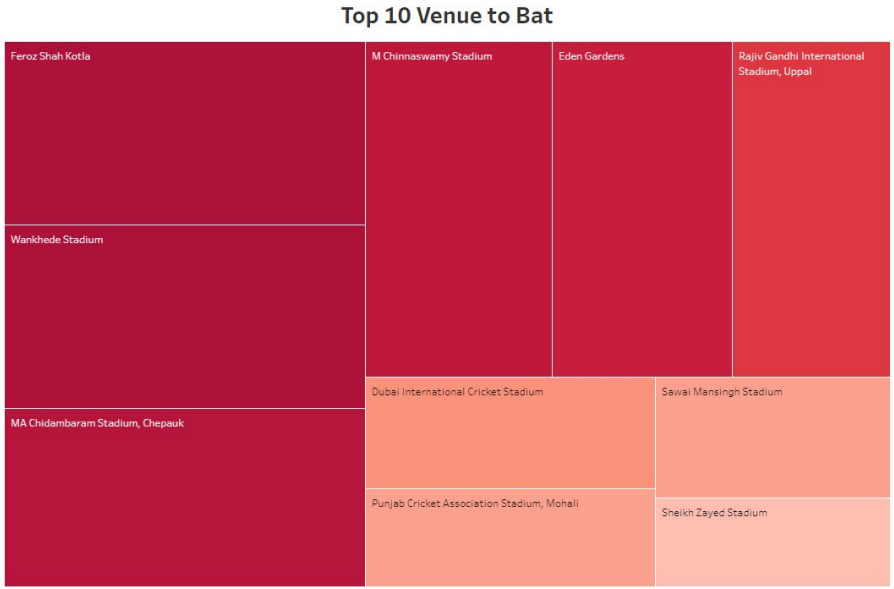
60% of those who field Wins the game



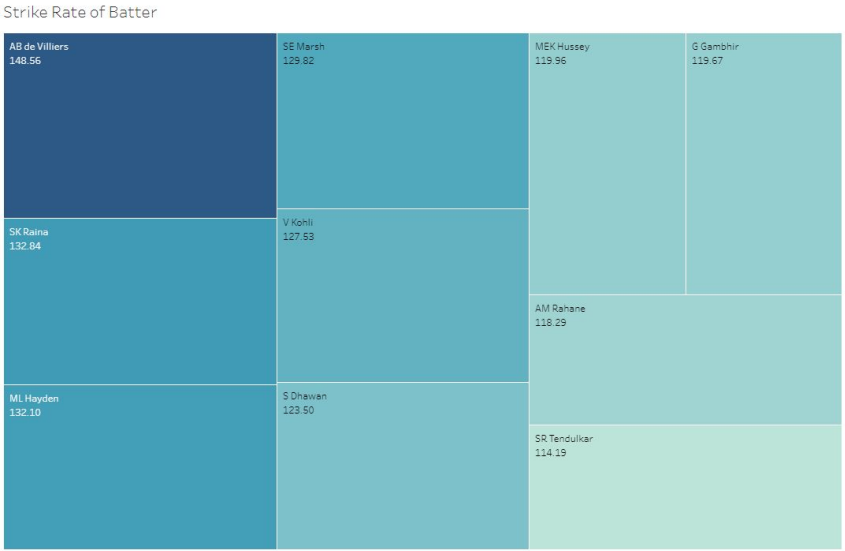
Details on the winning teams based on toss decision

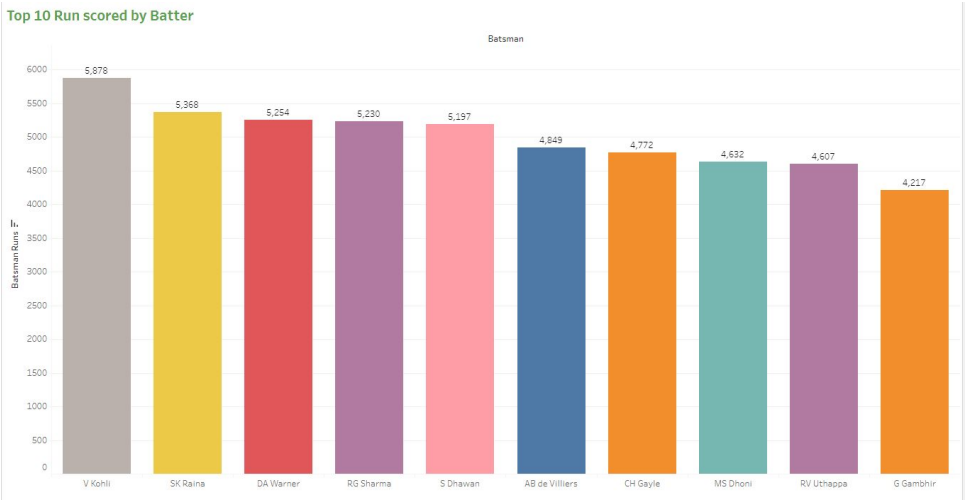
How the venues impact the gameplay





How the efficiency rate of batter impacts the gameplay



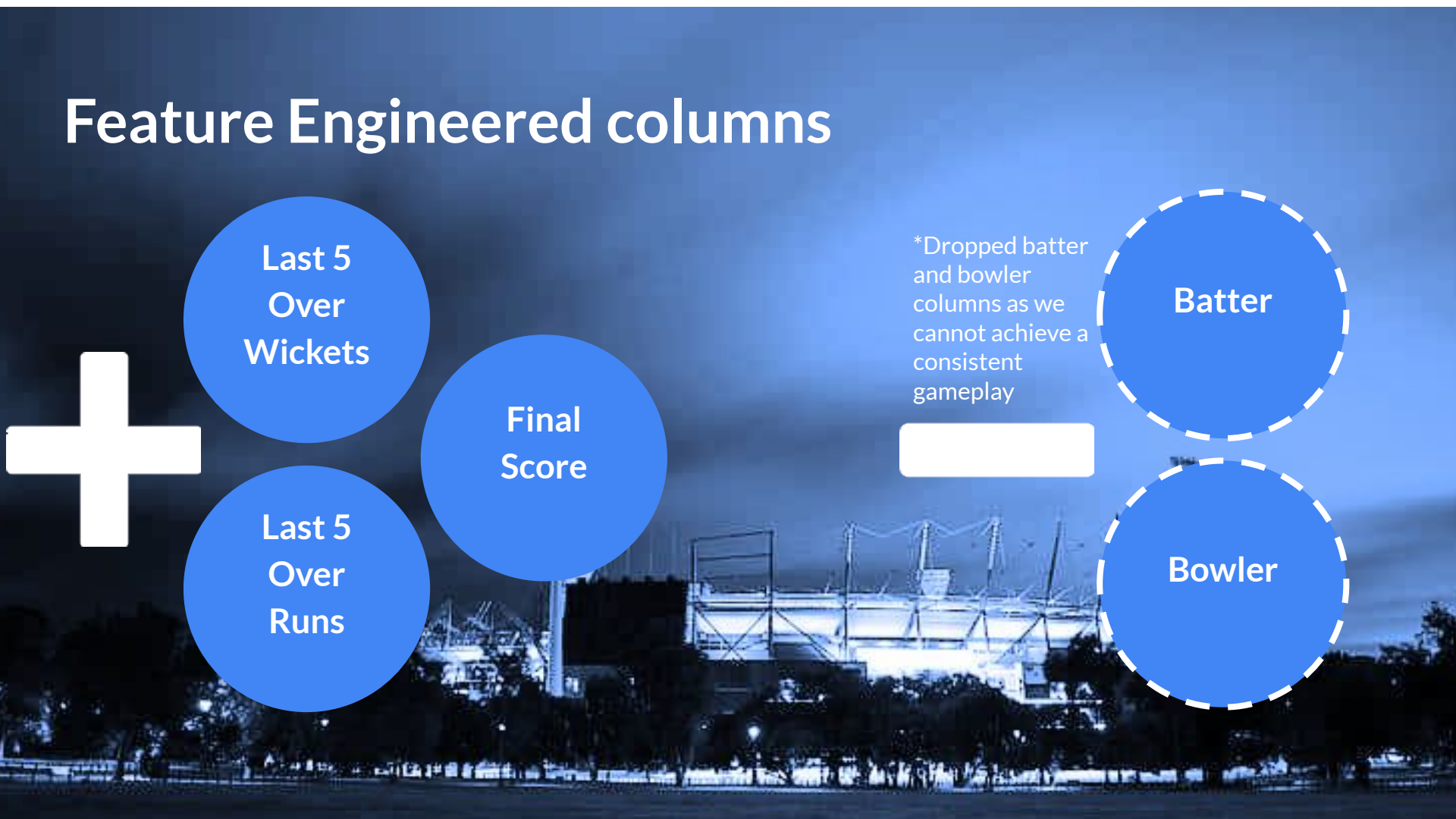


This data would be useful in selecting the right players during the game

Variable selection for Feature Engineering

Selected the 8 Most Consistent Teams

Feature Engineered columns



Data Modeling

TRAIN DATA				
Model Name	**Train Score**	**Mean Absolute Error**	**Mean Squared Error**	**Root Mean Squared Error**
Linear Regression	0.52503	14.93999	409.602129	20.23863
Ridge	0.52503	14.940166	409.603182	20.238656
Lasso	0.497843	15.527418	433.341759	20.816862
Decision Tree	0.496933	15.501726	422.262332	20.549023
Random Forest	**0.661075**	**11.419714**	**242.500067**	**15.572414**
Ada Boost	0.402423	17.61751	514.087121	22.673489
Gradient Boost	**0.832477**	**1.603841**	**7.332912**	**2.707935**

TEST DATA				
Model Name	**Test Score**	**Mean Absolute Error**	**Mean Squared Error**	**Root Mean Squared Error**
Linear Regression	0.529161	14.806188	402.628817	20.065613
Ridge	0.52917	14.806251	402.621463	20.06543
Lasso	0.5033	15.370086	424.743787	20.609313
Decision Tree	0.505106	15.507938	423.199466	20.571812
Random Forest	**0.683246**	**12.099161**	**270.866555**	**16.458024**
Ada Boost	0.410373	17.496106	504.208386	22.454585
Gradient Boost	**0.861848**	**6.731762**	**118.137958**	**10.869129**

Value of Score Prediction

- Increase chances of winning
- Strategize players to onboard during game
- Leverage on data-driven decision making

IPL Score Prediction

Venue: M Chinnaswamy Stadium

Batting Team: Delhi Capitals

Bowling Team: Mumbai Indians

Overs: 10

Runs: 60

Wickets: 3

Runs in last 5 overs: 30

Wickets in last 5 overs: 2

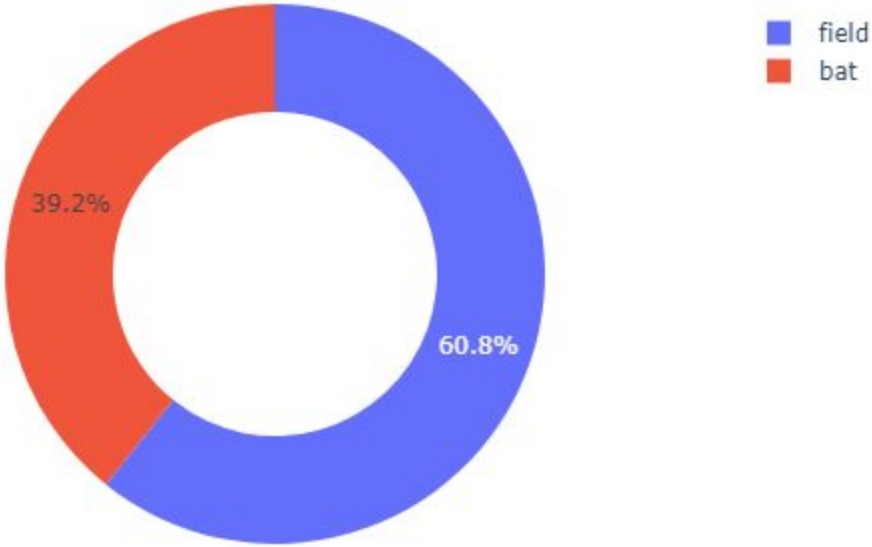
Estimate Score

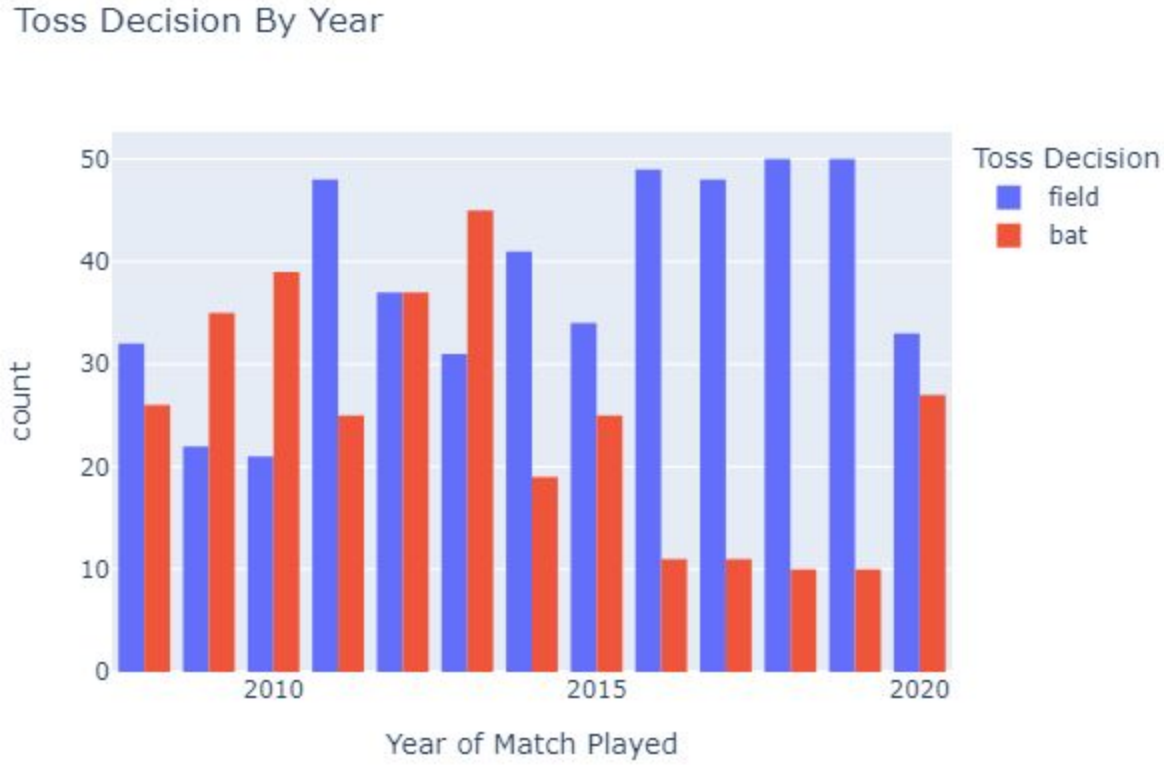
developed by mk

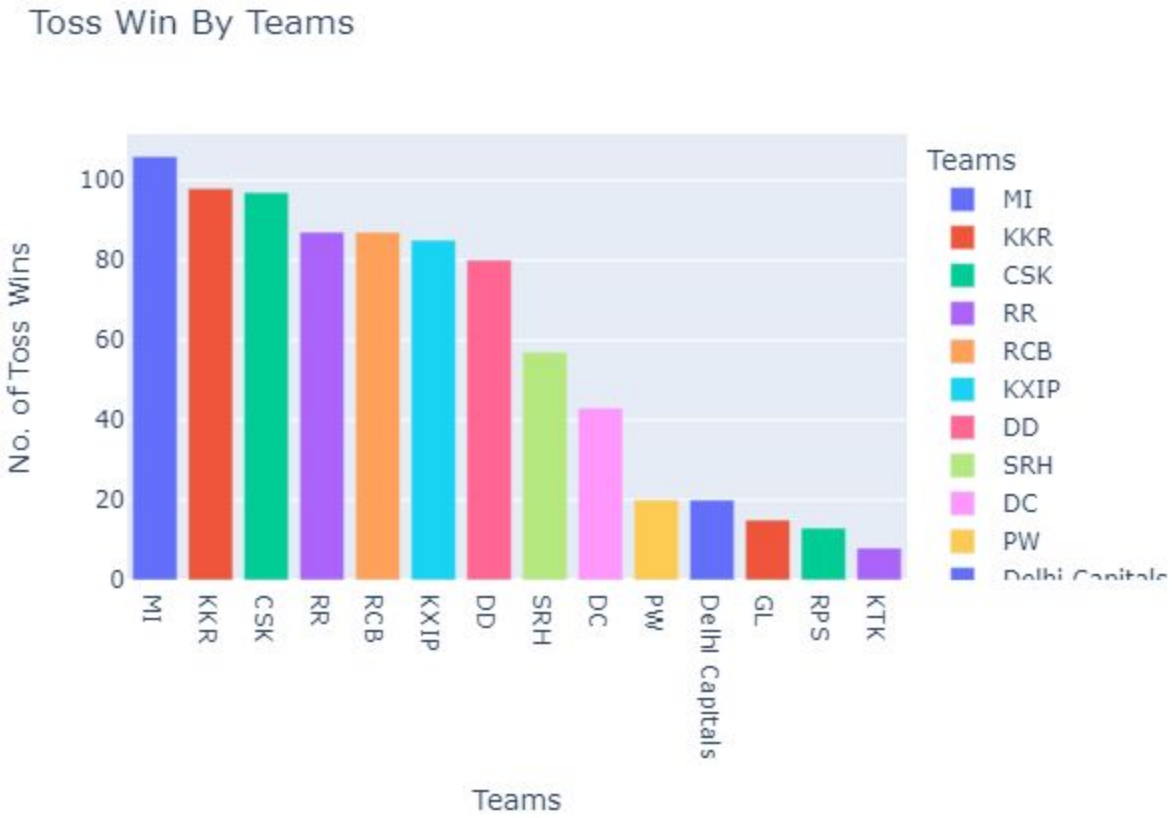
App QR Code

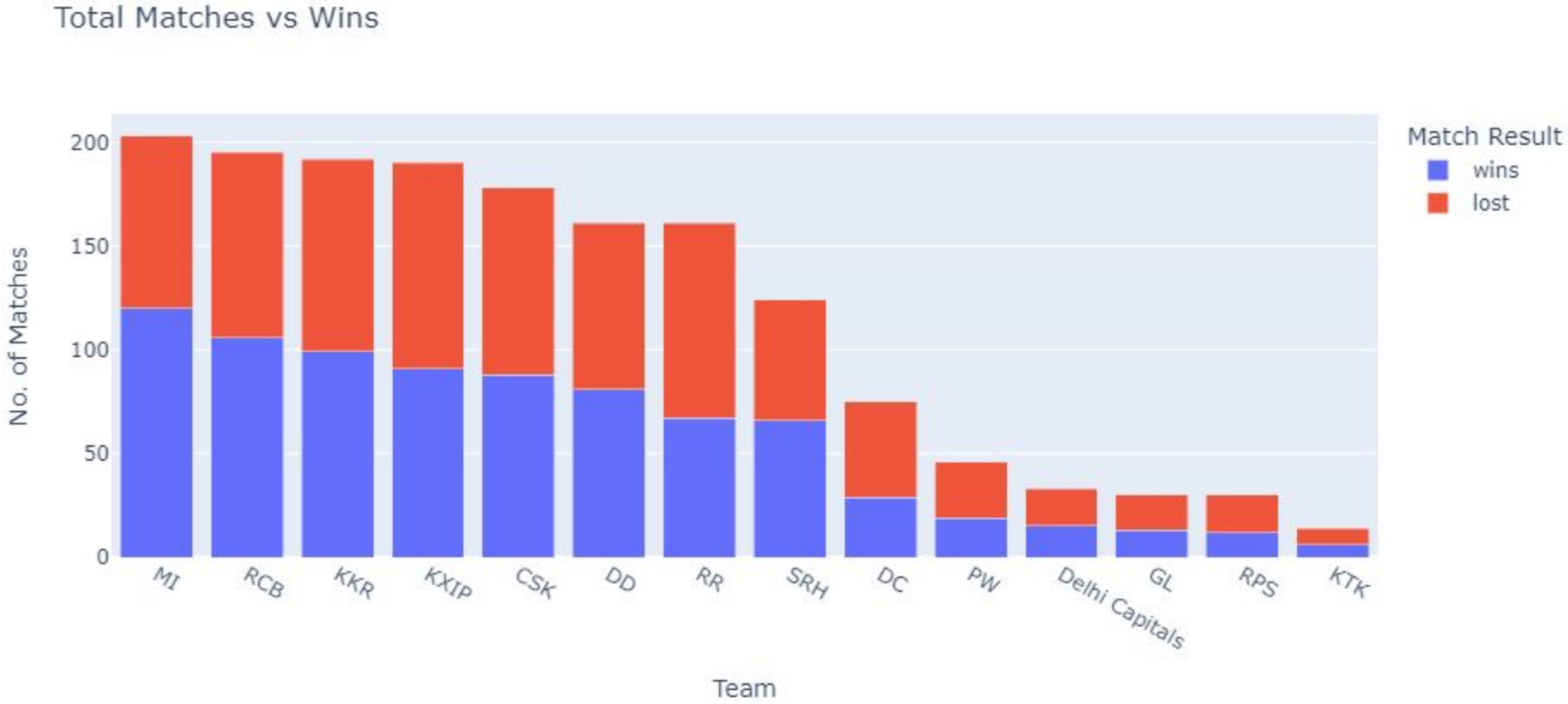
Annex

Percentage of Toss Decision

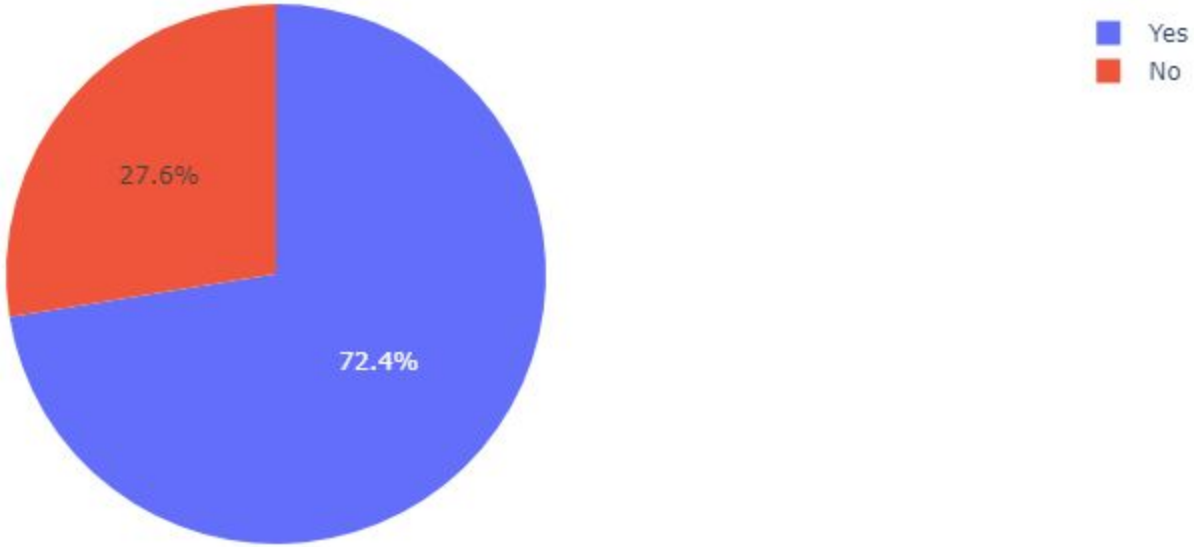


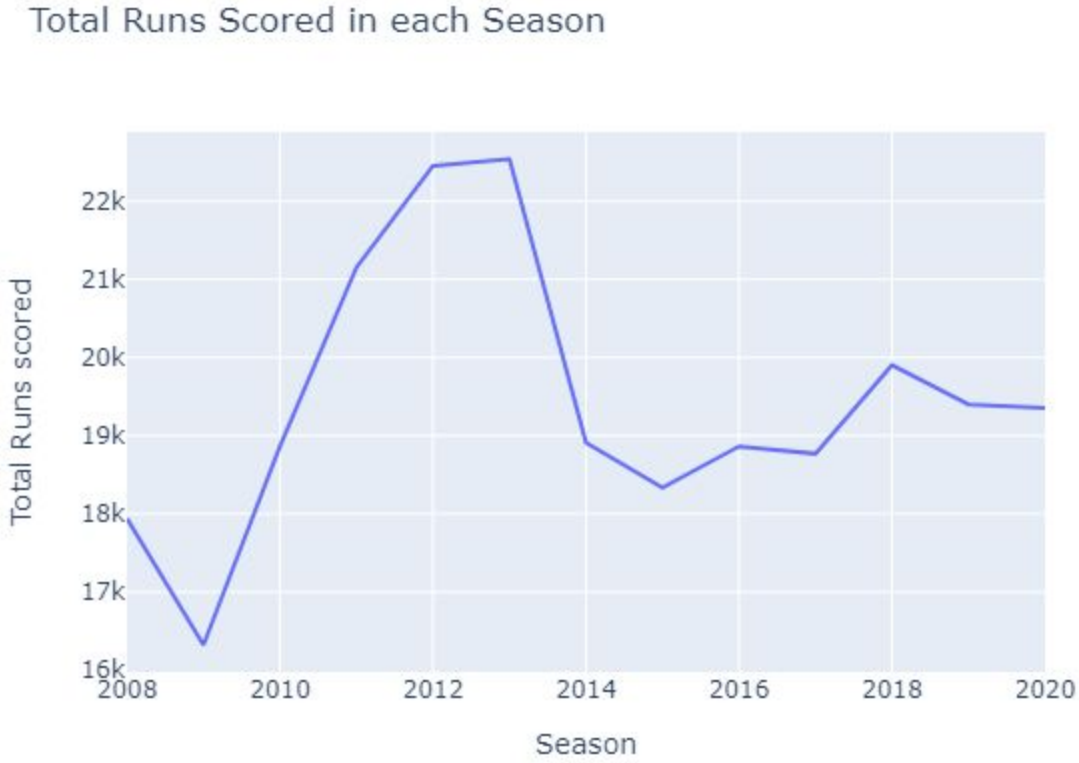

Toss Decision By Year
Toss Decision
field
bat
count
0
10
20
30
40
50
2010
2015
2020
Year of Match Played

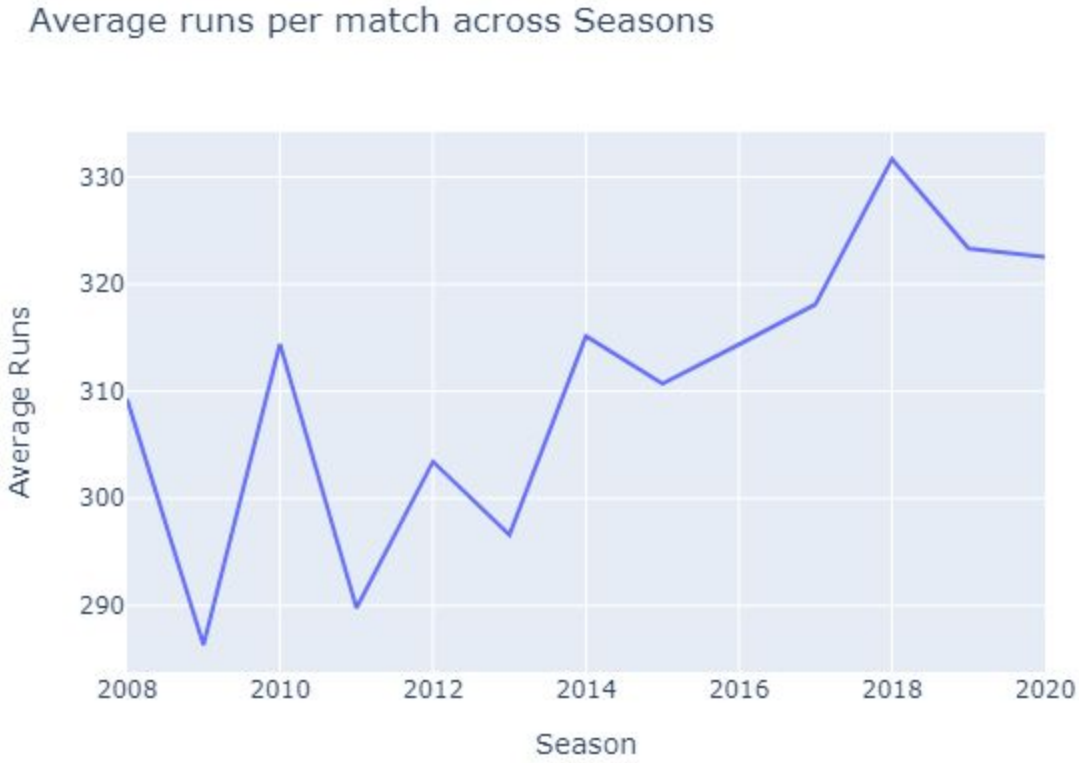

Toss Win By Teams
No. of Toss Wins
0
20
40
60
80
100
MI
KKR
CSK
RR
RCB
KXIP
DD
SRH
DC
PW
Delhi Capitals
GL
RPS
KTK
Teams
Teams
MI
KKR
CSK
RR
RCB
KXIP
DD
SRH
DC
PW

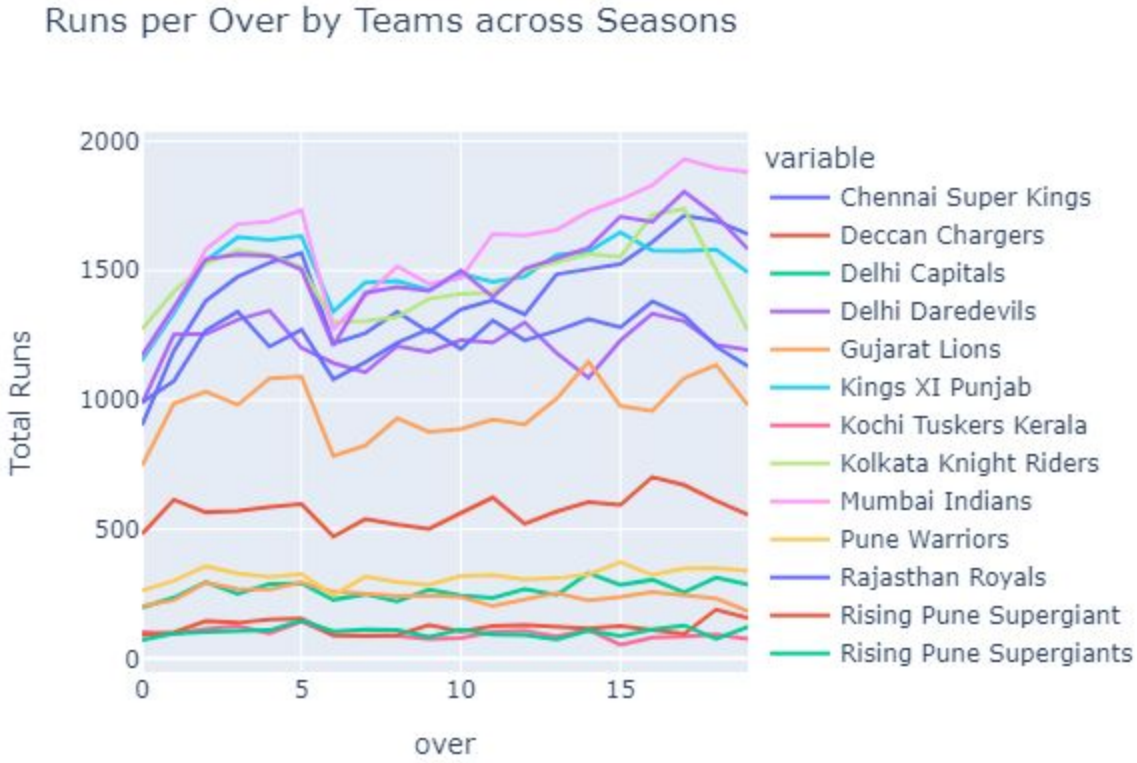

Total Matches vs Wins
No. of Matches
200
150
100
50
0
MI
RCB
KKR
KXIP
CSK
DD
RR
SRH
DC
PW
Delhi Capitals
GL
RPS
KTK
Team
Match Result
wins
lost

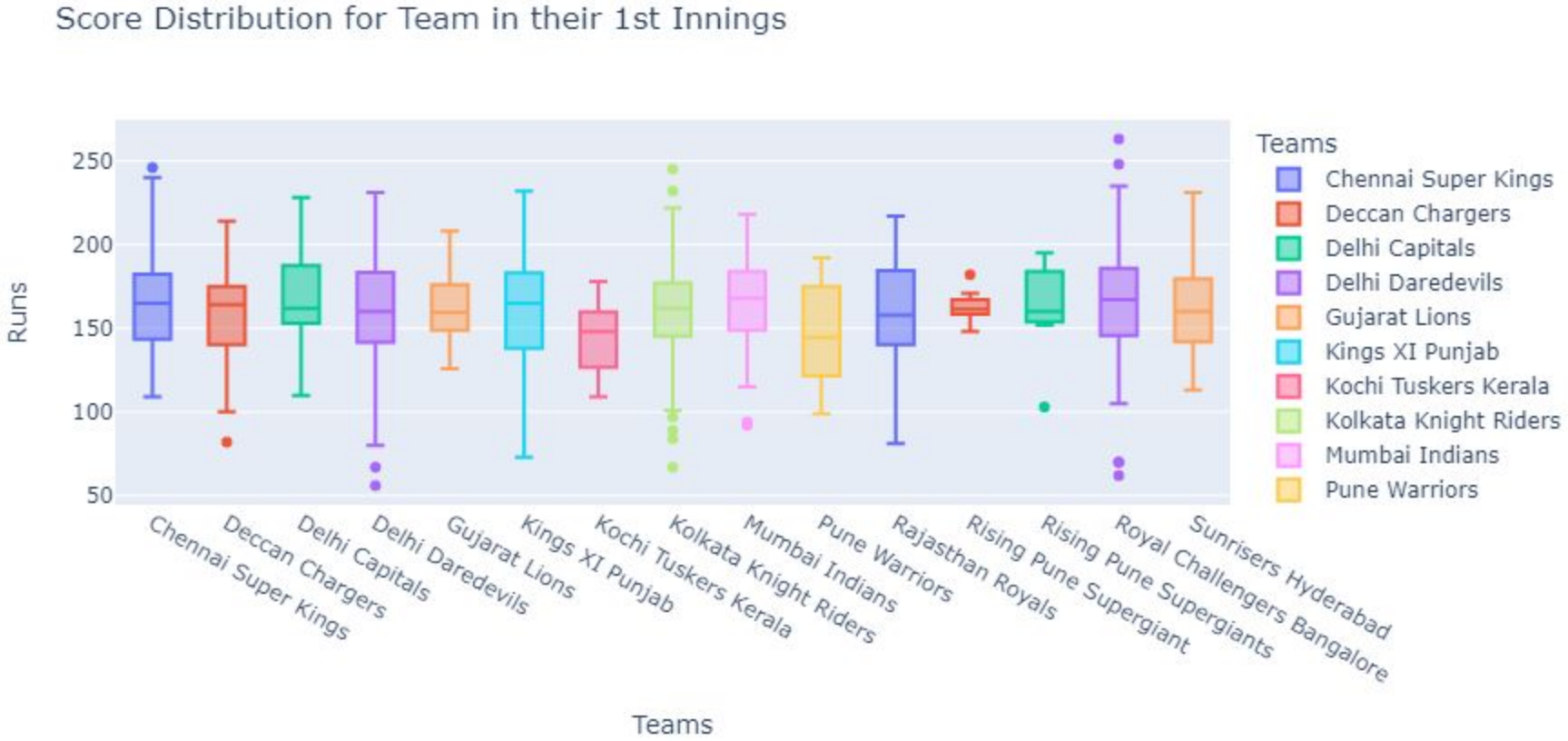
Is Toss Winner also the Match winner?



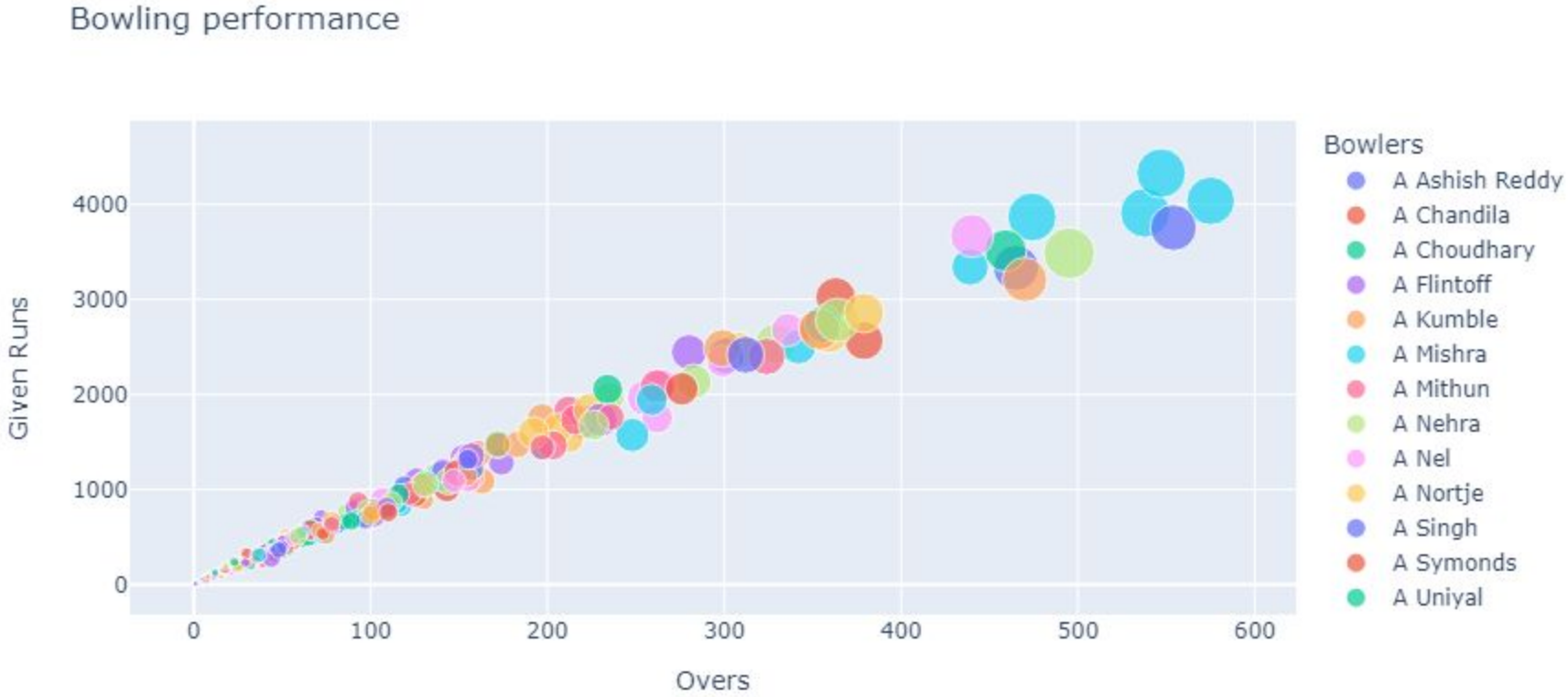

Total Runs Scored in each Season
Total Runs scored
22k
21k
20k
19k
18k
17k
16k
2008
2010
2012
2014
2016
2018
2020
Season

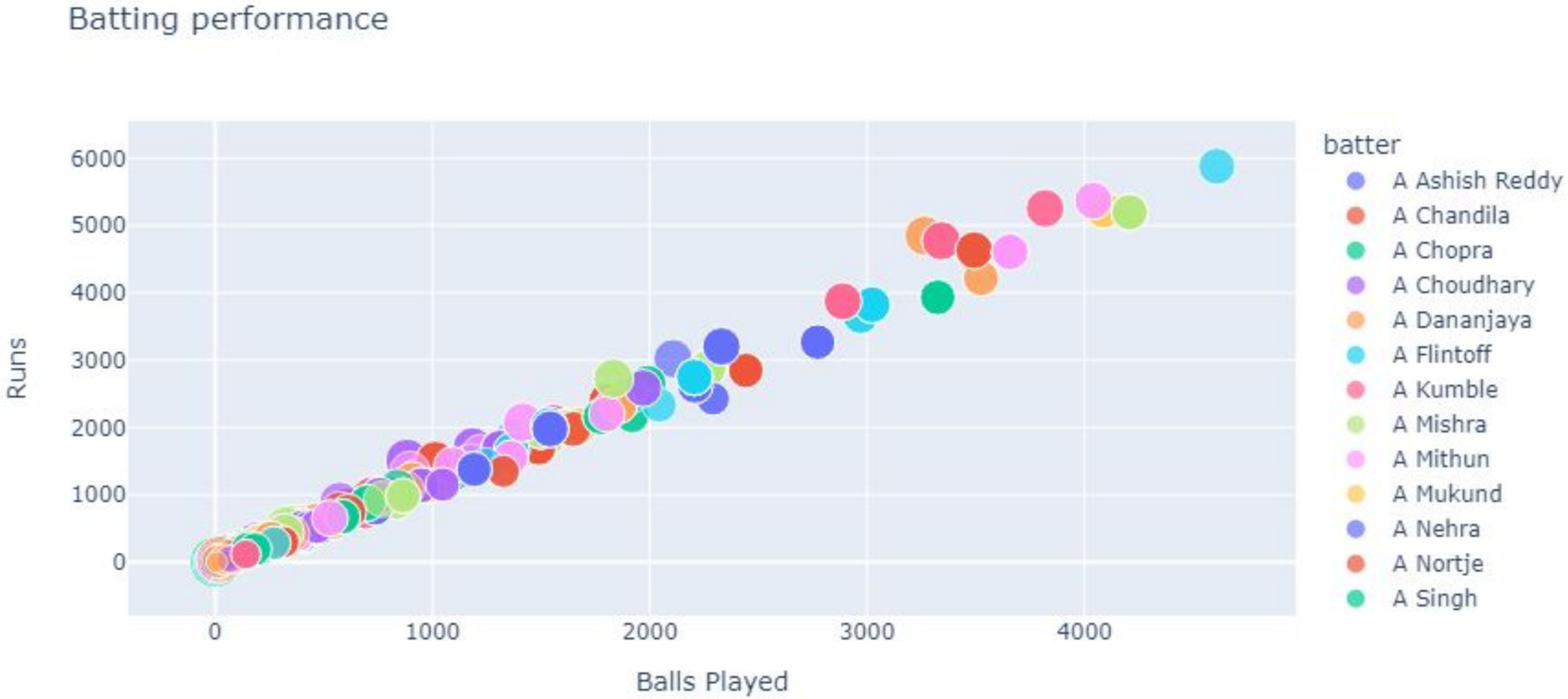

Average runs per match across Seasons
Average Runs
330
320
310
300
290
2008
2010
2012
2014
2016
2018
2020
Season


Runs per Over by Teams across Seasons
Total Runs
2000
1500
1000
500
0
0
5
10
15
over
variable
Chennai Super Kings
Deccan Chargers
Delhi Capitals
Delhi Daredevils
Gujarat Lions
Kings XI Punjab
Kochi Tuskers Kerala
Kolkata Knight Riders
Mumbai Indians
Pune Warriors
Rajasthan Royals
Rising Pune Supergiant
Rising Pune Supergiants


Score Distribution for Team in their 1st Innings
Runs
250
200
150
100
50
Chennai Super Kings
Deccan Chargers
Delhi Capitals
Delhi Daredevils
Gujarat Lions
Kings XI Punjab
Kochi Tuskers Kerala
Kolkata Knight Riders
Mumbai Indians
Pune Warriors
Rajasthan Royals
Rising Pune Supergiant
Rising Pune Supergiants
Royal Challengers Bangalore
Sunrisers Hyderabad
Teams
Teams
Chennai Super Kings
Deccan Chargers
Delhi Capitals
Delhi Daredevils
Gujarat Lions
Kings XI Punjab
Kochi Tuskers Kerala
Kolkata Knight Riders
Mumbai Indians
Pune Warriors


Score Distribution for Team in their 2nd Innings
Runs
200
150
100
50
0
Chennai Super Kings
Deccan Chargers
Delhi Capitals
Delhi Daredevils
Gujarat Lions
Kings XI Punjab
Kochi Tuskers Kerala
Kolkata Knight Riders
Mumbai Indians
Pune Warriors
Rajasthan Royals
Rising Pune Supergiant
Rising Pune Supergiants
Royal Challengers Bangalore
Sunrisers Hyderabad
Teams
Teams
Chennai Super Kings
Deccan Chargers
Delhi Capitals
Delhi Daredevils
Gujarat Lions
Kings XI Punjab
Kochi Tuskers Kerala
Kolkata Knight Riders
Mumbai Indians
Pune Warriors


Bowling performance
Given Runs
4000
3000
2000
1000
0
0
100
200
300
400
500
600
Overs
Bowlers
A Ashish Reddy
A Chandila
A Choudhary
A Flintoff
A Kumble
A Mishra
A Mithun
A Nehra
A Nel
A Nortje
A Singh
A Symonds
A Uniyal


Batting performance
Runs
6000
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
Balls Played
batter
A Ashish Reddy
A Chandila
A Chopra
A Choudhary
A Dananjaya
A Flintoff
A Kumble
A Mishra
A Mithun
A Mukund
A Nehra
A Nortje
A Singh