Capstone Project
Background and Objectives

SPORTS ANALYTICS - IPL

Background

1. Background

Sports analytics involves using data and statistical analysis to gain insights into various aspects of sports performance, strategy, and player behavior, ultimately aiding decision-making for teams and athletes.

This Capstone project is all about using sports analytics to understand and improve performance in the Indian Premier League (IPL) cricket matches. By digging into the numbers, we want to figure out how teams and players have been doing over the years.

This project brings together sports and data to uncover interesting insights in cricket

Background

2. Primary Objectives

1. To analyze team performance over the years (Team Performance
    Analytics)

2. To analyze players performance over the years(Player
    Performance Analytics)

3.   To predict "score at the end of first innings (after 20 overs)"
based on first 6 overs situation

Background

3. Data

The following datasets are available:

1. IPL_Ball_by_Ball_2008_2022

2. IPL_Matches_2008_2022

3. each_ball_records_2023

4. each_match_records_2023

Data : IPL_Ball_by_Ball_2008_2022

Content

IPL_Ball_by_Ball_2008_2022 data contains data for each ball for the entire tournament from 2008 to 2022.

ID	innings	overs	ballnumber	batter	bowler	non-striker	extra_type	batsman_run	extras_run
1312200	1	0	1	YBK Jaiswal	Mohammed Shami	JC Buttler	NA	0	0
1312200	1	0	2	YBK Jaiswal	Mohammed Shami	JC Buttler	legbyes	0	1
1312200	1	0	3	JC Buttler	Mohammed Shami	YBK Jaiswal	NA	1	0
1312200	1	0	4	YBK Jaiswal	Mohammed Shami	JC Buttler	NA	0	0
1312200	1	0	5	YBK Jaiswal	Mohammed Shami	JC Buttler	NA	0	0
1312200	1	0	6	YBK Jaiswal	Mohammed Shami	JC Buttler	NA	0	0
1312200	1	1	1	JC Buttler	Yash Dayal	YBK Jaiswal	NA	0	0
1312200	1	1	2	JC Buttler	Yash Dayal	YBK Jaiswal	NA	0	0

Columns	Description	Type	Possible values
ID	Match ID	Integer	
innings	Inning Number	Integer	1 to 6(3 to 6 if super over)
overs	Over Number	Integer	0 to 19
ballnumber	Ball Number	Integer	
batter	Batsman on Strike	Character	
bowler	Bowler name	Character	
non-striker	Non-Striker Batsman	Character	
extra_type	If the ball is an extra, then type of extra	Factor	byes/legbyes/wides/noballs/penalty
batsman_run	Run Scored by the batsman for the ball	Integer	
	Runs given due to extras for the		

Data : IPL_Ball_by_Ball_2008_2022

Content

IPL_Ball_by_Ball_2008_2022 data contains data for each ball for the entire tournament from 2008 to 2022.

total_run	non_boundary	isWicketDelivery	player_out	kind	fielders_involved	BattingTeam
0	0	0	NA	NA	NA	Rajasthan Royals
1	0	0	NA	NA	NA	Rajasthan Royals
1	0	0	NA	NA	NA	Rajasthan Royals
0	0	0	NA	NA	NA	Rajasthan Royals
0	0	0	NA	NA	NA	Rajasthan Royals
0	0	0	NA	NA	NA	Rajasthan Royals

Columns	Description	Type	Possible values
total_run	Total Runs for the ball	Integer	
non_boundary	The 4 or 6 scored in batsman_run column did not result from a boundary, such as all runs or overthrows; 1 if due to an overthrow	Factor	0/1
isWicketDelivery	If a wicket fell on the delivery	Factor	0/1
player_out	Batsman who got out	Character	
kind	Type of wicket taking ball	Factor	bowled/caught/caught and bowled/hit wicket/lbw/obstructing the field/retired hurt/retired out/run out/stumped

DATA SCIENCE INSTITUTE

Data : IPL_Matches_2008_2022

Content

IPL_Matches_2008_2022 data contains summary of each match for IPL 2008-2022. The datasets capture

ID	City	Date	Season	MatchNumber	Team1	Team2	Venue
1312200	Ahmedabad	29-05-2022	2022	Final	Rajasthan Royals	Gujarat Titans	Narendra Modi Stadium, Ahmedabad
1312199	Ahmedabad	27-05-2022	2022	Qualifier 2	Royal Challengers Bangalore	Rajasthan Royals	Narendra Modi Stadium, Ahmedabad
1312198	Kolkata	25-05-2022	2022	Eliminator	Royal Challengers Bangalore	Lucknow Super Giants	Eden Gardens, Kolkata
1312197	Kolkata	24-05-2022	2022	Qualifier 1	Rajasthan Royals	Gujarat Titans	Eden Gardens, Kolkata
1304116	Mumbai	22-05-2022	2022	70	Sunrisers Hyderabad	Punjab Kings	Wankhede Stadium, Mumbai
1304115	Mumbai	21-05-2022	2022	69	Delhi Capitals	Mumbai Indians	Wankhede Stadium, Mumbai
1304114	Mumbai	20-05-2022	2022	68	Chennai Super Kings	Rajasthan Royals	Brabourne Stadium, Mumbai

Columns	Description	Type	Possible values
ID	Unique match ID	Integer	
City	Host City	Factor	Ahmedabad, Kolkata, Mumbai etc
Date	Date	Date	
Season	Season year	Integer	2008,2009 to 2022
MatchNumber	Match number	Character	
Team1	Home team(except for seasons 2009,2020-22 and some matches of 2014	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc
Team2	Away team(except for seasons 2009,2020-22 and some matches of 2014	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc

DATA SCIENCE INSTITUTE

Data : IPL_Matches_2008_2022

Content

IPL_Matches_2008_2022 data contains summary of each match for IPL 2008-2022. The datasets capture various aspects of each IPL match, offering a comprehensive overview of the tournament over the specified period.

TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	Team1Players	Team2Players	Umpire1	Umpire2
Rajasthan Royals	bat	N	Gujarat Titans	Wickets		NA	HH Pandya	er', 'R Ashwin', 'R watia', 'Rashid K		CB Gaffaney	Nitin Menon
Rajasthan Royals	field	N	Rajasthan Royals	Wickets	7	NA	JC Buttler	arthik', 'Shahbaz A	'R Parag', 'R A	CB Gaffaney	Nitin Menon
Lucknow Super Giants	field	N	al Challengers Bangal	Runs	14	NA	RM Patidar	arthik', 'Shahbaz Awis', 'KH Pandy		Madanagopal	MA Gough
Gujarat Titans	field	N	Gujarat Titans	Wickets	7	NA	DA Miller	er', 'R Parag', 'R Aswatia', 'Rashid		BNJ Oxenford	VK Sharma
Sunrisers Hyderabad	bat	N	Punjab Kings	Wickets	5	NA	Harpreet Brar	/ashington Sundatone', 'JM Sharn		AK Chaudhary	NA Patwardhan
Mumbai Indians	field	N	Mumbai Indians	Wickets	5	NA	JJ Bumrah	, 'R Powell', 'AR Pa	'andeep Singh'	Nitin Menon	Tapan Sharma

Columns	Description	Type	Possible values
TossWinner	Team that won the toss	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc
TossDecision	Decision taken by the team winning the toss	Factor	bat/field
SuperOver	If the match went into a super over or not	Factor	N/Y
WinningTeam	Team that won the match	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc
WonBy	Way in which the team won the match	Factor	NoResults/Runs/SuperOver/Wickets
Margin	Margin of runs/wickets by which team won	Integer	
method	Was there Duckworth Lewis method in the match	Factor	D/L, NA
Player_of_Match	Player who received player of the match award (Best performing player of that match)	Character	
Team1Players	Players that featured in the starting lineup of team 1	Character	

Data : each_ball_records_2023

Content

each_ball_records_2023 data contains data for each ball for the entire tournament for the year 2023.

match_no	ballnumber	inningno	over	outcome	batter	bowler	comment	score
1	1	1	0.1	0	Devon Conway	Mohammed Shami	mmed Shami to Devon Conway, r	0
1	2	1	0.2	1lb	Devon Conway	Mohammed Shami	med Shami to Devon Conway, 1 l	1
1	3	1	0.3	0	Ruturaj Gaikwad	Mohammed Shami	nmed Shami to Ruturaj Gaikwad,	0
1	4	1	0.4	1	Ruturaj Gaikwad	Mohammed Shami	mmed Shami to Ruturaj Gaikwad,	1
1	5	1	0.5	0	Devon Conway	Mohammed Shami	mmed Shami to Devon Conway, r	0
1	6	1	0.6	0	Devon Conway	Mohammed Shami	mmed Shami to Devon Conway, r	0

Columns	Description	Type	Possible values
match_no	Match number	Integer	
ballnumber	Ball Number	Integer	1 to 6(3 to 6 if super over)
inningno	Inning Number	Integer	0 to 19
over	Over Number	Integer	
outcome	Outcome of the delivery	Character	
batter	Batsman on Strike	Character	
bowler	Bowler name	Character	
comment	Description of delivery(what happended on the delivery)	Factor	byes/legbyes/wides/noballs/penalty
score	Runs given for the ball	Integer	

DATA SCIENCE INSTITUTE

Data : each_match_records_2023

each_match_records_2023 data contains summary of each match for IPL 2008-2023. The datasets capture various aspects of each IPL match, offering a comprehensive overview of the tournament over the specified period. of the tournament over the specified period.

season	date	match_number	match_type	venue	location	team1	team2	toss_won
2023	31-03-2023	1	Group	rendra Modi Stadiu	Ahmedabad	hennai Super Kin	Gujarat Titans	Gujarat Titans
2023	01-04-2023	2	Group	et Association IS Bi	Chandigarh	Punjab Kings	olkata Knight Rider	Kolkata Knight Riders
						Super Gia	Delhi Capitals	Delhi Capitals
						an Royal	unrisers Hyderabad	Sunrisers Hyderabad
						ai Indiansl	Challengers Bang	yal Challengers Banga
						Super Kin	ucknow Super Gian	Lucknow Super Giant

Columns	Description	Type	Possible values
season	Season year	Integer	2023
date	Date	Date	
match_number	Match number	Integer	
match_type	Type of Match	Factor	Group stage/Eliminator/Qualifier/Final
venue	Stadium/Ground where match is played	Factor	
location	Host City	Factor	Ahmedabad, Kolkata, Mumbai etc
team1	Home team	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc
team2	Away team	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc
toss_won	Team that won the toss	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc

DATA SCIENCE INSTITUTE

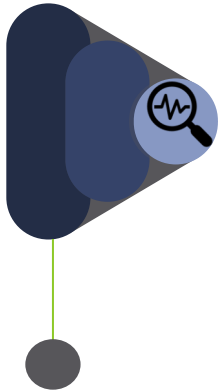# Data : each_match_records_2023

Content

each_match_records_2023 data contains summary of each match for IPL 2008-2023. The datasets capture

toss_decision	umpire1	umpire2	reserve_umpire	match_referee	winner	winner_runs	winner_wickets	man_of_match
field	Nitin Menon	HAS Khalid	A Bengeri	J Srinath	Gujarat Titans		5	Rashid Khan
field	BNJ Oxenford	YC Barde	PM Joshi	M Nayyar	Punjab Kings	7		Arshdeep Singh
field	AK Chaudhary	NA Patwardhan	M Kuppuraj	DS Manohar	cknow Super Gia	50		MA Wood
field	N Ananthapadmanabh	R Pandit	Abhijit Bhattacharya	V Narayan Kutty	Rajasthan Royals	72		JC Buttler
field	Nitin Menon	Tapan Sharma	A Bengeri	J Srinath	Challengers Bangalore		8	F du Plessis
field	A Totre	BNJ Oxenford	PM Joshi	M Nayyar	hennai Super King	12		MM Ali

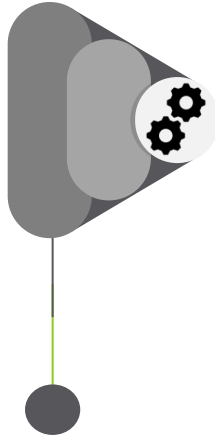| Columns	Description	Type	Possible values
toss_decision	Decision taken by the team winning the toss(bat/field)	Factor	bat/field
umpire1	Umpire 1	Character	
umpire2	Umpire 2	Character	
reserve_umpire	Reserve Umpire	Character	
match_referee	Match Referee	Character	
winner	Team that won the match	Factor	Rajasthan Royals, Royal Challengers Bangalore, Mumbai Indians etc
winner_runs	Margin of runs by which team won	Integer	
winner_wickets	Margin of wickets by which team won	Integer	
man_of_match	Player who received player of the match award (Best performing player of that match)	Character	

DATA SCIENCE INSTITUTE

Next steps

Data management

Compile 2008-2022 datasets using match ID

Data cleaning , Handling missing values and completing Basic Data checks

Check if any variables needed to be feature coded i.e made into groups or want to be left as continuous variables

Descriptive Statistics & Data visualization

Explore data for team performance analytics and player performance analytics.

How can this data be presented better visually ?

Once again post Data visualization check if any variable needs to be feature coded

Predictive modelling

Develop a model to predict end of first inning score.

Using different Predictive model techniques to find Significant variables

Ensure you follow all steps like Train and test data , checking for Multicollinearity

Check if any other ML technique fits better

DATA SCIENCE INSTITUTE

12

* You can go through the Questions and Hint ppt to understand each section better / cross check if you have done all the steps mentioned