

2022 | By: Shubham Kumar



Exploratory Data Analysis of IPL Matches

[Using SQL]

EDA of IPL Matches [using SQL]

Introduction

The [dataset](#) consist of data about IPL matches played from the year 2008 to 2019. IPL is a professional Twenty20 cricket league founded by the Board of Control for Cricket in India (BCCI) in 2008. The league has 8 teams representing 8 different Indian cities or states. It enjoys tremendous popularity and the brand value of the IPL in 2019 was estimated to be ₹475 billion (US\$6.7 billion). So, let's analyze IPL through stats.

Objective

- To find the team that won the most number of matches in a season.
- To find the player with the most player of the match awards.
- To find the city that hosted the maximum number of IPL match
- To find the biggest victories in IPL while defending a total and while chasing a total.
- To find who is the most successful bowler
- To find who is the most successful batsman
- To find who is conceded highest runs from boundaries.

We have dataset of all IPL matches from 2008 to 2019 and MYSQL. To begin with analysis firstly Create Database Schema for importing csv file and creating table.

```
create database ipl;
```

Now, create table to store data,

```
(create table matches (id integer,  
    season integer,  
    city varchar(100),  
    d_ate date,  
    team1 varchar(100),  
    team2 varchar(100),  
    toss_winner varchar(100),  
    toss_decision varchar(100),  
    result varchar(100),  
    dl_applied integer,  
    winner varchar(100),  
    win_by_runs integer,  
    win_by_wickets integer,  
    player_of_the_match varchar(100),  
    venue varchar(100),  
    umpire1 varchar(100),  
    umpire2 varchar(100),  
    umpire3 integer
```



```
CREATE TABLE deliveries (  
    matchid integer,  
    inning integer,  
    batting_team varchar(100),  
    bowling_team varchar(100),  
    o_ver integer,  
    ball integer,  
    batsman varchar(100),  
    non_striker varchar(100),  
    bowler varchar(100),  
    is_super_over integer,  
    wide_runs integer,  
    bye_runs integer,  
    legbye_runs integer,  
    noball_runs integer,  
    penalty_runs integer,  
    batsman_runs integer,  
    extra_runs integer,  
    total_runs integer,  
    player_dismissed varchar(100),  
    dismissal_kind varchar(100),  
    fielder varchar(100)
```

Now, Begin with the queries:

Shape of Data

Shapes are combination of rows and columns. To know about table structure (I mean rows and columns)

use below query.

```
select count(*) as No_of_row from matches
```

	No_of_row
▶	756

```
select count(*) as No_of_columns from information_schema.columns  
where TABLE_NAME='matches';
```

	No_of_columns
▶	18

```
select count(*) as No_of_row from deliveries;
```

	No_of_row
▶	179078

```
select count(*) as No_of_columns from information_schema.columns  
where TABLE_NAME='deliveries';
```

	No_of_columns
▶	21

Viewing data

```
Select * from deliveries
limit 5;
```

	matchid	inning	batting_team	bowling_team	o_ver	ball	batsman	non_striker	bowler	is_s
▶	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0
	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0
	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0
	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0
	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0

```
select * from matches
limit 5;
```

	id	season	city	d_ate	team1	team2	toss_winner
▶	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalo
	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant
	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders
	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab
	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalo

Viewing selected columns

```
select m.season,m.city,m.d_ate,m.team1,m.team2,m.winner,m.win_by_runs from
matches m
where m.season='2017'
limit 10;
```

	season	city	d_ate	team1	team2	winner	w
▶	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	35
	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	0
	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	0
	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	0
	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	15
	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	0
	2017	Mumbai	2017-04-09	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	0
	2017	Indore	2017-04-10	Royal Challengers Bangalore	Kings XI Punjab	Kings XI Punjab	0
	2017	Pune	2017-04-11	Delhi Daredevils	Rising Pune Supergiant	Delhi Daredevils	97
	2017	Mumbai	2017-04-12	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	0

Distinct values

```
select distinct season as year from matches m  
order by 1  
limit 15  
;
```

	year
▶	2008
	2009
	2010
	2011
	2012
	2013
	2014
	2015
	2016
	2017
	2018
	2019

```
select count(distinct player_of_match) as 'No of Matches' from matches m  
order by 1;
```

	No of Matches
▶	227

Find season winner for each season

```
select winner ,season ,d_ate from matches
where d_ate in (select max(d_ate) from matches
group by season
order by max(d_ate) desc)
order by d_ate
limit 15;
```

	winner	season	d_ate
	Deccan Chargers	2009	2009-05-24
	Chennai Super Kings	2010	2010-04-25
	Chennai Super Kings	2011	2011-05-28
	Kolkata Knight Riders	2012	2012-05-27
	Mumbai Indians	2013	2013-05-26
	Kolkata Knight Riders	2014	2014-06-01
	Mumbai Indians	2015	2015-05-24
	Sunrisers Hyderabad	2016	2016-05-29
	Mumbai Indians	2017	2017-05-21
	Chennai Super Kings	2018	2018-05-27
	Mumbai Indians	2019	2019-05-12

Find venue of 10 most recently played matches


```

select venue,max(d_ate) as date from matches
group by venue
order by date desc
limit 10;

```

	venue	date
►	Rajiv Gandhi Intl. Cricket Stadium	2019-05-12
	ACA-VDCA Stadium	2019-05-10
	M. A. Chidambaram Stadium	2019-05-07
	Wankhede Stadium	2019-05-05
	IS Bindra Stadium	2019-05-05
	M. Chinnaswamy Stadium	2019-05-04
	Feroz Shah Kotla Ground	2019-05-04
	Eden Gardens	2019-04-28
	Sawai Mansingh Stadium	2019-04-27
	Maharashtra Cricket Association Stadium	2018-05-20

Case

```

select batsman,bowler,batsman_runs,ball,
case
  when batsman_runs=0 then "duck"
  when batsman_runs=6 then "six"
  when batsman_runs=4 then "four"
  when batsman_runs=1 then "single"
  else "runs in words"
end batsman_runs
from deliveries;

```

	batsman	bowler	batsman_runs	ball	batsman_runs
	DA Warner	TS Mills	0	2	duck
	DA Warner	TS Mills	4	3	four
	DA Warner	TS Mills	0	4	duck
	DA Warner	TS Mills	0	5	duck
	S Dhawan	TS Mills	0	6	duck
	S Dhawan	TS Mills	0	7	duck
	S Dhawan	A Choudhary	1	1	single
	DA Warner	A Choudhary	4	2	four
	DA Warner	A Choudhary	0	3	duck
	DA Warner	A Choudhary	6	4	six
	DA Warner	A Choudhary	0	5	duck
	MC Henria...	A Choudhary	0	6	duck

Data Aggregation

```
select winner,max(win_by_runs) from matches
group by winner
order by 2 ;
```

winner	max(win_by_runs)
Rising Pune Supergiants	34
Pune Warriors	38
Delhi Capitals	39
Rising Pune Supergiant	61
Deccan Chargers	82
Delhi Daredevils	97
Chennai Super Kings	97
Rajasthan Royals	105
Kings XI Punjab	111
Sunrisers Hyderabad	118
Kolkata Knight Riders	140
Royal Challengers Ban...	144
Mumbai Indians	146

```
select winner,win_by_wickets,season,max(win_by_runs) from matches
where winner='Mumbai Indians'
group by season
order by 3 desc;
```

	winner	win_by_wickets	season	max(win_by_runs)
►	Mumbai Indians	0	2019	46
	Mumbai Indians	0	2018	102
	Mumbai Indians	4	2017	146
	Mumbai Indians	6	2016	80
	Mumbai Indians	0	2015	41
	Mumbai Indians	5	2014	25
	Mumbai Indians	0	2013	65
	Mumbai Indians	8	2012	27
	Mumbai Indians	8	2011	37
	Mumbai Indians	0	2010	98
	Mumbai Indians	0	2009	92
	Mumbai Indians	7	2008	29

How many extra runs have been conceded in ipl?

```
select distinct bowler,sum(extra_runs) from deliveries
group by bowler
having sum(extra_runs)>0
order by 2 desc;
```

	bowler	sum(extra_runs)
►	SL Malinga	293
	P Kumar	236
	UT Yadav	219
	DJ Bravo	201
	B Kumar	197
	I Sharma	194
	RP Singh	181
	SR Watson	171
	DW Steyn	171
	Harbhajan Singh	170
	Z Khan	169
	R Ashwin	167
	JA Morkel	153
	DS Kulkarni	145

On an average, teams won by how many runs in ipl

```
select winner,avg(win_by_runs) from matches
group by winner
having avg(win_by_runs)>0
order by 2 desc;
```

	winner	avg(win_by_runs)
▶	Chennai Super Kings	17.7800
	Mumbai Indians	17.1193
	Royal Challengers Bangalore	14.9048
	Deccan Chargers	14.5172
	Rising Pune Supergiant	12.3000
	Sunrisers Hyderabad	12.2931
	Rajasthan Royals	11.9333
	Kolkata Knight Riders	11.8043
	Pune Warriors	11.5833
	Kings XI Punjab	11.2805
	Rising Pune Supergiants	10.6000
	Delhi Daredevils	10.0746
	Delhi Capitals	9.2000
	Kochi Tuskers Kerala	3.8333

How many extra runs were conceded in ipl by SK Warne?

```
select bowler,sum(extra_runs) from deliveries
where bowler='SK Warne'
group by bowler;
```

	bowler	sum(extra_runs)
▶	SK Warne	57

How many boundaries (4s or 6s) have been hit in IPL by Mumbai Indians team

```
select m.winner, d.total_runs, count(d.total_runs) from deliveries d
inner join matches m on m.id=d.matchid
where d.total_runs in (4,6)
and m.winner='Mumbai Indians'
group by m.winner, d.total_runs;
```

	winner	total_runs	count(d.total_runs)
▶	Mumbai Indians	4	2955
	Mumbai Indians	6	1232

How many balls did SK Warne bowl to batsman SR Tendulkar?

```
select batsman, bowler, count(ball) from deliveries
where bowler='SK Warne' and batsman='SR Tendulkar'
group by batsman, bowler;
```

	batsman	bowler	count(ball)
▶	SR Tendulkar	SK Warne	39

How many matches were played in the month of April?

```
select count(*) from matches
where month(d_ate)='4';
-- or
select count(*) from matches
where extract(month from d_ate)=4;
```

	count(*)
▶	364

```
select count(*) from matches
where month(d_ate) in ('3','6');
```

	count(*)
▶	43

Total number of wickets taken in ipl (count not null values)

```
select count(player_dismissed) as 'Wicket' from deliveries
where player_dismissed <>"";
```

	Wicket
▶	8834

No of dismissal by a player of each kind

```
select dismissal_kind,player_dismissed, count(*) from deliveries
where player_dismissed <>""
group by dismissal_kind,player_dismissed
order by 3 desc;
```

	dismissal_kind	player_dismissed	count(*)
▶	caught	SK Raina	112
	caught	RV Uthappa	108
	caught	RG Sharma	103
	caught	V Kohli	95
	caught	KD Karthik	85
	caught	Yuvraj Singh	84
	caught	G Gambhir	80
	caught	S Dhawan	79
	caught	YK Pathan	77
	caught	CH Gayle	75

Pattern Match (Like operators % _)

select distinct) from ipl.matches where player of match like 'K_ P%';

```
select distinct player_of_match from matches where player_of_match
like 'K_ P%';
```

	player_of_match
▶	KA Pollard
	KH Pandya
	KP Pietersen

Maximum runs by which any team won a match per season

```
select season,max(win_by_runs) from matches
group by season
order by 1;
```

	season	max(win_by_runs)
▶	2008	140
	2009	92
	2010	98
	2011	111
	2012	86
	2013	130
	2014	93
	2015	138
	2016	144
	2017	146
	2018	102
	2019	118

Create score card for each match Id

```
select batting_team,batsman,sum(batsman_runs) from deliveries
group by batting_team,batsman;
```

	batting_team	batsman	sum(batsman_runs)
▶	Sunrisers Hyderabad	DA Warner	3306
	Sunrisers Hyderabad	S Dhawan	2550
	Sunrisers Hyderabad	MC Henriques	755
	Sunrisers Hyderabad	Yuvraj Singh	488
	Sunrisers Hyderabad	DJ Hooda	384
	Sunrisers Hyderabad	BCJ Cutting	116
	Royal Challengers Bangalore	CH Gayle	3175

Top 10 players with max boundaries (4 or 6)

```
select DISTINCT batsman,count(total_runs) from deliveries
where total_runs in (4,6)
group by batsman
order by 2 desc
limit 10;
```

	batsman	count(total_runs)
▶	CH Gayle	705
	SK Raina	686
	V Kohli	676
	DA Warner	641
	S Dhawan	624
	RG Sharma	624
	RV Uthappa	599
	AB de Villiers	572
	G Gambhir	557
	SR Watson	523

Top 10 bowlers who conceded highest extra runs

```

select bowler,sum(extra_runs) as 'highest extra runs' from deliveries
group by bowler
order by 2 desc
limit 10;

```

	bowler	highest extra runs
►	SL Malinga	293
	P Kumar	236
	UT Yadav	219
	DJ Bravo	201
	B Kumar	197
	I Sharma	194
	RP Singh	181
	SR Watson	171
	DW Steyn	171
	Harbhajan Singh	170

Top 10 wicket takers

```

select bowler,count(player_dismissed) as NoWicket_Taken,dismissal_kind from
deliveries
where dismissal_kind <>""
group by bowler
order by NoWicket_Taken desc
limit 10;

```

bowler	NoWicket_Taken	dismissal_kind
SL Malinga	188	caught
DJ Bravo	168	caught
A Mishra	165	caught
Harbhajan Singh	161	caught
PP Chawla	156	caught
B Kumar	141	caught
R Ashwin	138	caught and bowled
SP Narine	137	lbw
UT Yadav	136	caught
R Vinay Kumar	127	caught

Top 2 player of match for each season

```
select season, player_of_match, CNT from
(
  select row_number() over (partition by season) as
  rn,season,player_of_match,Cnt
  from (
    select season,player_of_match, count(player_of_match)
    as Cnt
    from matches
    group by season,player_of_match
    order by 1 asc,3 desc
  ) rw
) Temp
where Temp.rn<3;
```

	season	player_of_match	CNT
▶	2008	SE Marsh	5
	2008	SR Watson	4
	2009	YK Pathan	3
	2009	SR Tendulkar	2
	2010	SR Tendulkar	4
	2010	JH Kallis	3
	2011	CH Gayle	6
	2011	MEK Hussey	3
	2012	CH Gayle	5
	2012	AB de Villiers	3
	2013	MEK Hussey	5
	2013	A Mishra	4
	2014	GJ Maxwell	4
	2014	RA Jadeja	3
	2015	DA Warner	4
	2015	A Nehra	3

Window Functions- (CTE)

Combine column date from matches with table deliveries to get data by year

```
with
t1 as (select id,season,d_ate,city,team1,team2,winner from matches),
t2 as (select matchid,batting_team,bowling_team from deliveries)
select distinct
t1.season,t1.d_ate,t1.city,t1.team1,t1.team2,t2.batting_team,t2.bowling_team,
t1.winner
from t1 inner join t2 on t1.id=t2.matchid;
/*
```

	season	d_ate	city	team1	team2	batting_team	
▶	2017	2017-04-05	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	^
	2017	2017-04-05	Hyderabad	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	
	2017	2017-04-06	Pune	Mumbai Indians	Rising Pune Supergiant	Mumbai Indians	
	2017	2017-04-06	Pune	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	
	2017	2017-04-07	Rajkot	Gujarat Lions	Kolkata Knight Riders	Gujarat Lions	
	2017	2017-04-07	Rajkot	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	
	2017	2017-04-08	Indore	Rising Pune Supergiant	Kings XI Punjab	Rising Pune Supergiant	
	2017	2017-04-08	Indore	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	
	2017	2017-04-08	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	
	2017	2017-04-08	Bangalore	Royal Challengers Bangalore	Delhi Daredevils	Delhi Daredevils	
	2017	2017-04-09	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Gujarat Lions	
	2017	2017-04-09	Hyderabad	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	
	2017	2017-04-09	Mumbai	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	▼

Inferences and Conclusion

Let's summarize the important observations we made during Exploratory Data Analysis:

- Mumbai Indians is the most successful team in IPL.
- When defending a total, the biggest victory was by 146 runs (Mumbai Indians defeated Delhi Daredevils by 146 runs on 06 May 2017 at Feroz Shah Kotla stadium, Delhi).
- When chasing a target, the biggest victory was by 10 wickets (without losing any wickets) and there were 11 such instances.
- The Mumbai city has hosted the most number of IPL matches.
- Chris Gayle has won the maximum number of player of the match title.
- Eden Gardens has hosted the maximum number of IPL matches.
- Till 2019, 40 venues have hosted 756 IPL matches
- SL Malinga has been conceded highest 293 extra runs in ipl
- SL Malinga is the most successful bowler in ipl
- Chris Gayle secured highest runs from boundaries