IPL DATA ANALYSIS

Submitted by

Paramveer Singh Sisodiya

CERTIFICATE

This is to certify that PARAMVEER SINGH SISODIYA bearing Registration no. 11910772 has completed INT217 project titled, **"IPL DATA ANALYSIS"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

DECLARATION

I, Paramveer Singh Sisodiya student of CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Name of the student

PARAMVEER SINGH SISODIYA

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher Ms. Sandeep Kaur, who gave me opportunity to do this amazing project of “IPL DATA ANALYSIS” which also helped me in doing a lot of research and I came to know about lots of new things. I am thankful to them. Secondly, I would also like to thanks my parents and friends who helped me a lot in finalizing this project within the limited time frame. it helped me increase my knowledge and skills

TABLE OF CONTENT

INTRODUCTION

Data Analysis is a process of inspecting, cleaning, modifying, transforming and modeling the data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under **a range** of names, while **being employed in numerous** business, science, and **scientific discipline** domains.

INDIAN PREMIER LEAGUE (IPL) Indian professional T20 league which is founded in 2008 . its based-on Round-Robin group and knockout formatting, has teams in major Indian cities.

The analytics team of INDIAN PREMIER LEAGE (IPL) association and BCCI association anywhere within the world would like to check us through data analysis of every and each match resulting in a well-organized and fruitful information. My analysis contains data on host teams, all stadiums, important of toss, most supported team and teams which are most successful.

The Indian Premier League (IPL), could be a Professional 20-20 cricket league in India contested during April and will of each year by teams representing Indian cities and a few states. The league was founded by the Board of control of cricket India (BCCI) in 2008, and is considered the brain child of Lalit Modi, the founder and former commissioner of the league. IPL has an exclusive window in ICC Future Tours Programs.

The IPL is that the most-attended cricket league within the world and in 2014 ranked sixth by average attendance among all sports leagues. In 2010, the IPL became the primary sporting event within the world to be broadcast continue to exist YouTube. The brand value of IPL in 2017 was US$6.3 billion, in line with Duff and help. per BCCI, the 2015 IPL season contributed ₹11.5 billion (US$182 million) to the GDP of the Indian economy.

Prize Money:

The 2019 season of the IPL offered a complete prize of ₹50 crore (equivalent to ₹530 million or US $7.0 million in 2020), with the winning team netting ₹20 crore (equivalent to ₹21 million or US$2.8 million in 2020). the primary and second runners up received ₹12.5 crore (US$1.7 million) and ₹8.75 crore (US$1.2 million), respectively, with the fourth placed team also winning ₹8.75 crore (US$1.2 million). the opposite teams don't seem to be awarded any prize. The IPL rules mandate that 1/2 the award must be distributed among the players

Match Rules:

PL games utilize television timeouts and hence there's no cut-off date within which teams must complete their innings. However, a penalty is also imposed if the umpires find teams misusing this privilege. Each team is given a two-and-a-half-minute "strategic timeout" during each innings; one must be taken by the bowling team between the ends of the 6th and 9th overs, and one by the batting team between the ends of the 13th to 16th overs. If the match between team become tie, then 1 over match played by each team (super Over) .Since the 2018 season, the Umpire Decision Review System (DRS) is being employed altogether IPL matches, allowing each team one chance to review an on-field umpire's decision per innings.

Insights of data:

IPL DATA ANALYSIS contains the following data fields: -

- Season – The year on which the following match is held.
- Date – Contains date on which the match is held.
- Stadium – Stadium in which match is held.
- City – City in which match is held.
- Home Team Name – Host team name.
- Toss-Winner –Who is the toss winner.
- Toss Decision –What team decided after winning the toss.
- Winner – Who won the match.
- Win by run – This show by how many runs does the team win.
- Win by Wicket – This show by how many wickets does the team win.
- Player of the match – Who won the player of the match.
- Umpire -1 – Name of Umpire 1.
- Umpire -2 – Name of Umpire 2.
- Umpire -3 – Name of Umpire 3.
- Batsman – Batsman who is on strike.
- Non-Striker – Non-striker Batsman.
- ID– This column shows the match number.
- Total Runs – Total runs scored per ball.

SCOPE OF ANALYSIS

IPL DATA ANALYSIS project on IPL Statics of India provides the overall Statistics details of the matches of IPL and teams progress in various aspects from the year 2008 to 2019.

Objectives: -

- To better Understanding of Excel.
- To get to know about different features of excel and friendly with excel.
- To learn the concept of ETL process in EXCEL as well as in Tableau Prep.
- To Know about Pivot Table and pivot chart.
- Learn to make dashboard in Excel**.**
- To get to know different types of graphs in excel.
- To learn how to fetch data from other source to excel in different forms.
- To use of Hyperlinks ion excels.

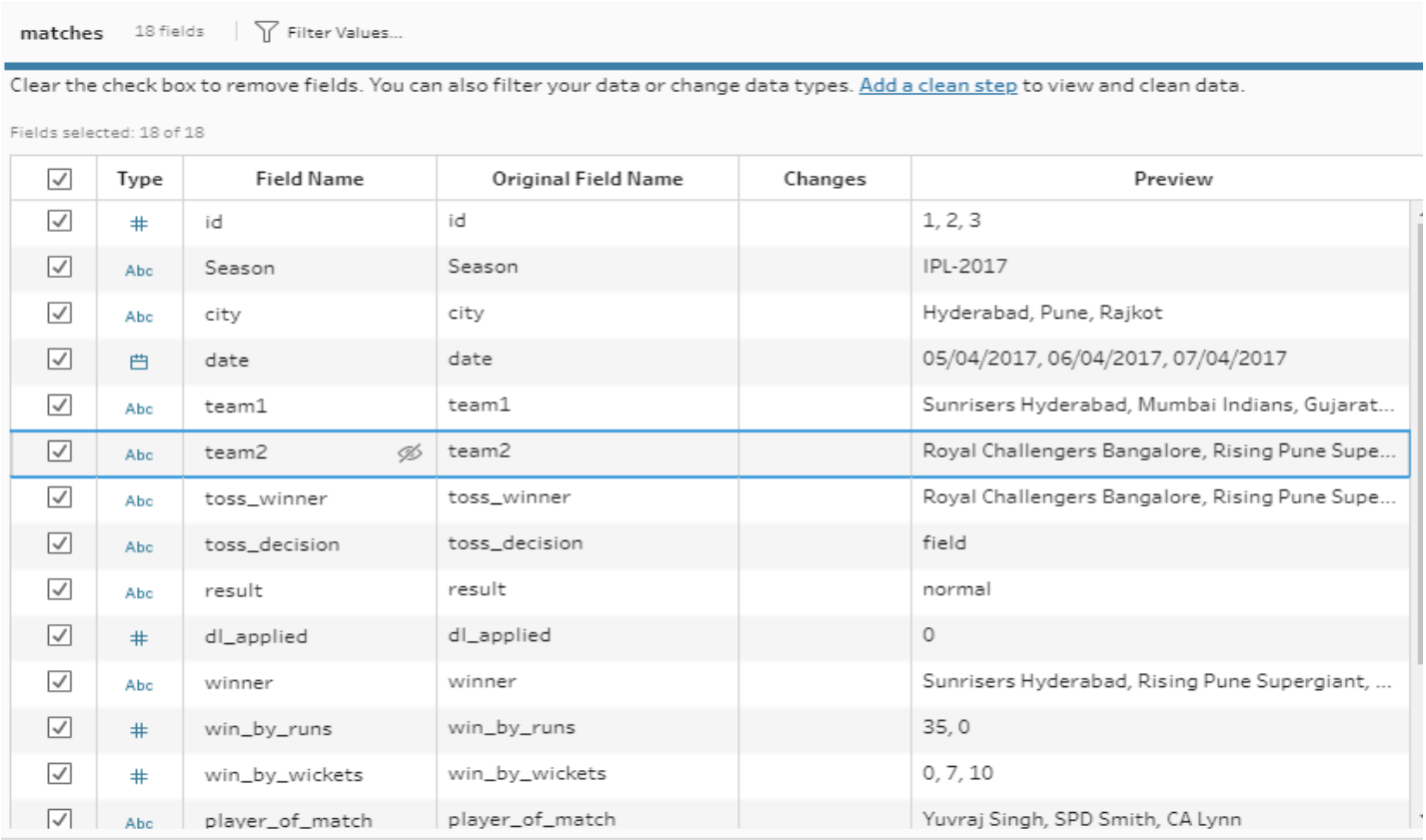
SOURCE OF DATASET

The data is being taken from the Kaggle. Kaggle is an Airbnb for Data Scientists – this is where they spend their nights and weekends. It's a crowd-sourced platform to attract, nurture, train and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems. It has over 536,000 active members from 194 countries and it receives close to 150,000 submissions per month. Started from Melbourne, Australia Kaggle moved to Silicon Valley in 2011, raised some 11 million dollars from the likes of Hal Varian (Chief Economist at Google), Max Lev chin (PayPal), Index and Khosla Ventures and then ultimately been acquired by the Google in March of 2017. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost their Kaggle rankings. There are only 94 Kaggle Grandmasters in the world to this date.

In June 2017, Kaggle announced that it passed 1 million registered users, or Kaggle's, and as of 2021 has over 8 million registered users. The community spans 194 countries. it's a various community, starting from those just starting intent on many of the world's best-known researchers. Kaggle competitions regularly attract over m teams and individuals. Kaggle's community has thousands of public datasets and code snippets (called "kernels" on Kaggle). Many of those researchers publish papers in peer-reviewed journals supported their performance in Kaggle competitions.

ETL PROCESS

n computing, extract, transform, load (ETL) could be a process in database usage to organize data for analysis, especially in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a correct storage format/structure for the needs of querying and analysis; finally, data loading describes the insertion of knowledge into the ultimate target database like an operational data store, an information mart, or a knowledge warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so separate sources is used together, and at last delivers data in a very presentation-ready format so application developers can build applications and end users can make decisions. Precisely, ETL is defined as a process that extracts the information from different RDBMS source systems, then transforms the information (like applying calculations, concatenations, etc.) and at last loads the info into the info Warehouse system. ETL stands for Extract, Transform and cargo. Before ETL, the dataset seemed like this. This data is taken from Kaggle.



matches 18 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. Add a clean step to view and clean data.

Fields selected: 18 of 18

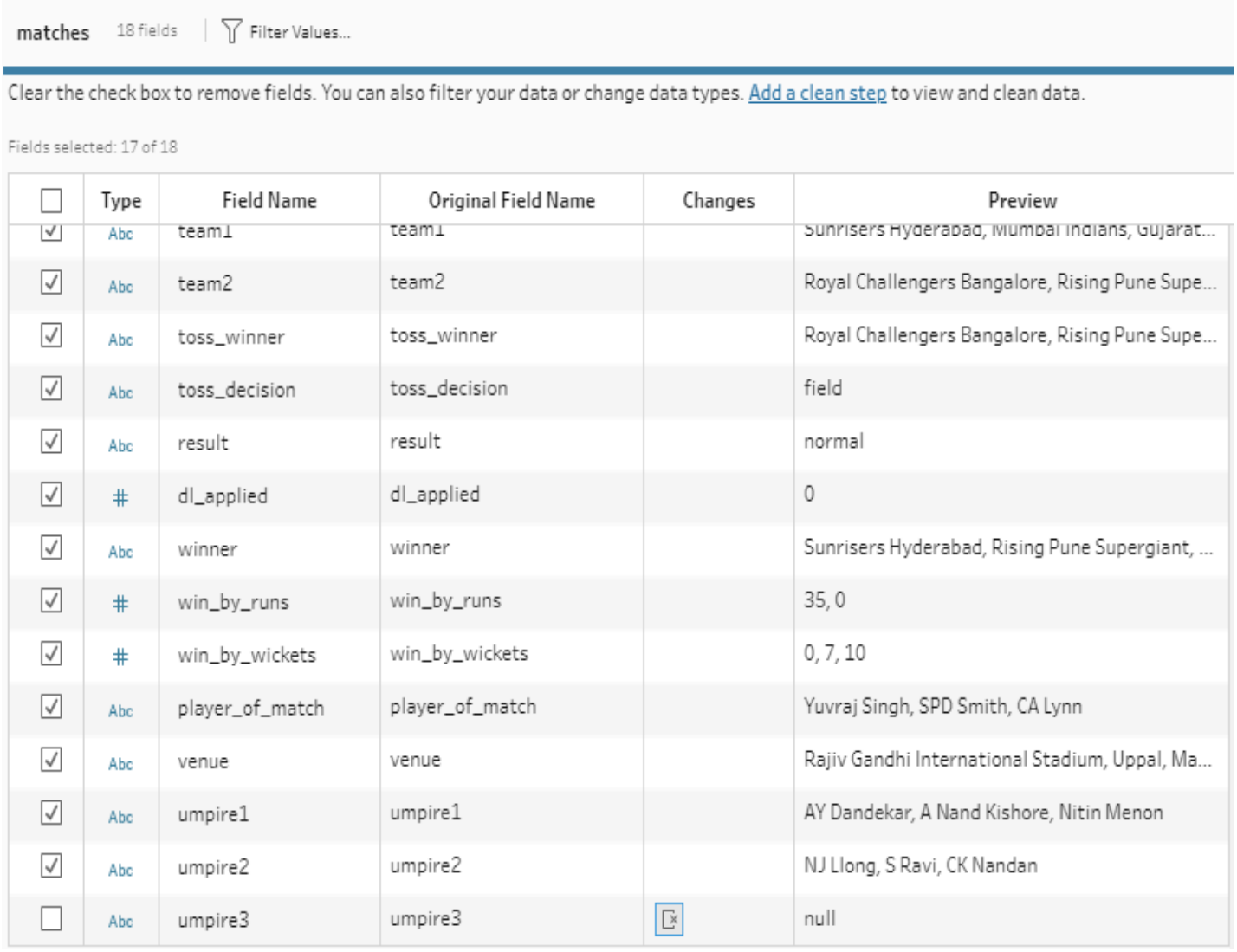
Type	Field Name	Original Field Name	Changes	Preview
#	id	id		1, 2, 3
Abc	Season	Season		IPL-2017
Abc	city	city		Hyderabad, Pune, Rajkot
date	date	date		05/04/2017, 06/04/2017, 07/04/2017
Abc	team1	team1		Sunrisers Hyderabad, Mumbai Indians, Gujarat...
Abc	team2	team2		Royal Challengers Bangalore, Rising Pune Supe...
Abc	toss_winner	toss_winner		Royal Challengers Bangalore, Rising Pune Supe...
Abc	toss_decision	toss_decision		field
Abc	result	result		normal
#	dl_applied	dl_applied		0
Abc	winner	winner		Sunrisers Hyderabad, Rising Pune Supergiant, ...
#	win_by_runs	win_by_runs		35, 0
#	win_by_wickets	win_by_wickets		0, 7, 10
Abc	player_of_match	player_of_match		Yuvraj Singh, SPD Smith, CA Lynn

Through the process of ETL, we are going to clean the dataset and bring all the entries to their proper format which help to easy to read and use.

For following ETL process we are using tableau prep.

Step 1: Removing the columns which are not properly defined or not crucial to our analysis.

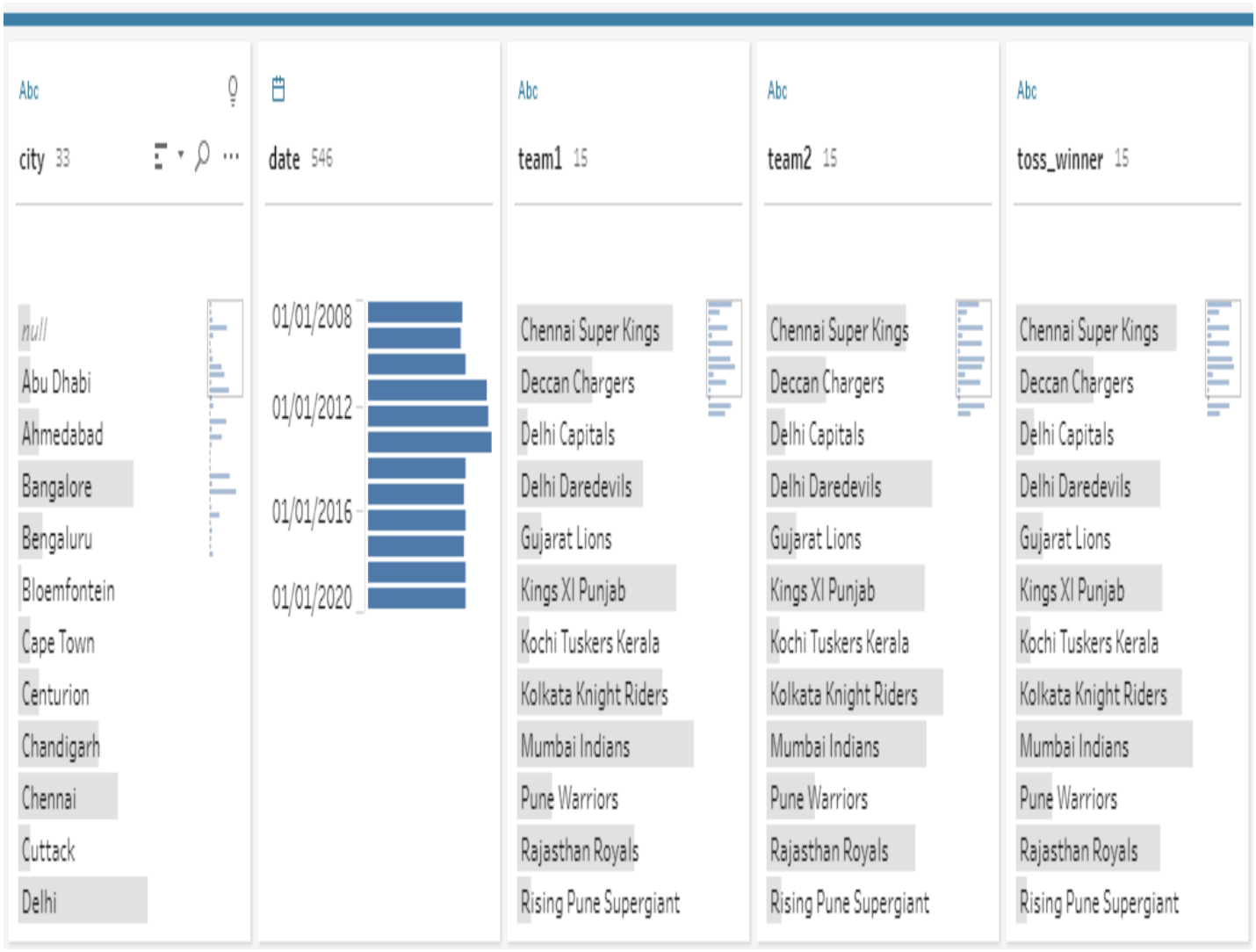
For this we will columns which are redundant like the column with just the index numbers. For this we will select that particular column and then go to delete option in the home tag and then select Delete Columns from the drop-down menu. In this data set I am deleting umpire 3 column.



Type	Field Name	Original Field Name	Changes	Preview
Abc	team1	team1		Sunrisers Hyderabad, Mumbai Indians, Gujarat...
Abc	team2	team2		Royal Challengers Bangalore, Rising Pune Supe...
Abc	toss_winner	toss_winner		Royal Challengers Bangalore, Rising Pune Supe...
Abc	toss_decision	toss_decision		field
Abc	result	result		normal
#	dl_applied	dl_applied		0
Abc	winner	winner		Sunrisers Hyderabad, Rising Pune Supergiant, ...
#	win_by_runs	win_by_runs		35, 0
#	win_by_wickets	win_by_wickets		0, 7, 10
Abc	player_of_match	player_of_match		Yuvraj Singh, SPD Smith, CA Lynn
Abc	venue	venue		Rajiv Gandhi International Stadium, Uppal, Ma...
Abc	umpire1	umpire1		AY Dandekar, A Nand Kishore, Nitin Menon
Abc	umpire2	umpire2		NJ Llong, S Ravi, CK Nandan
Abc	umpire3	umpire3		null

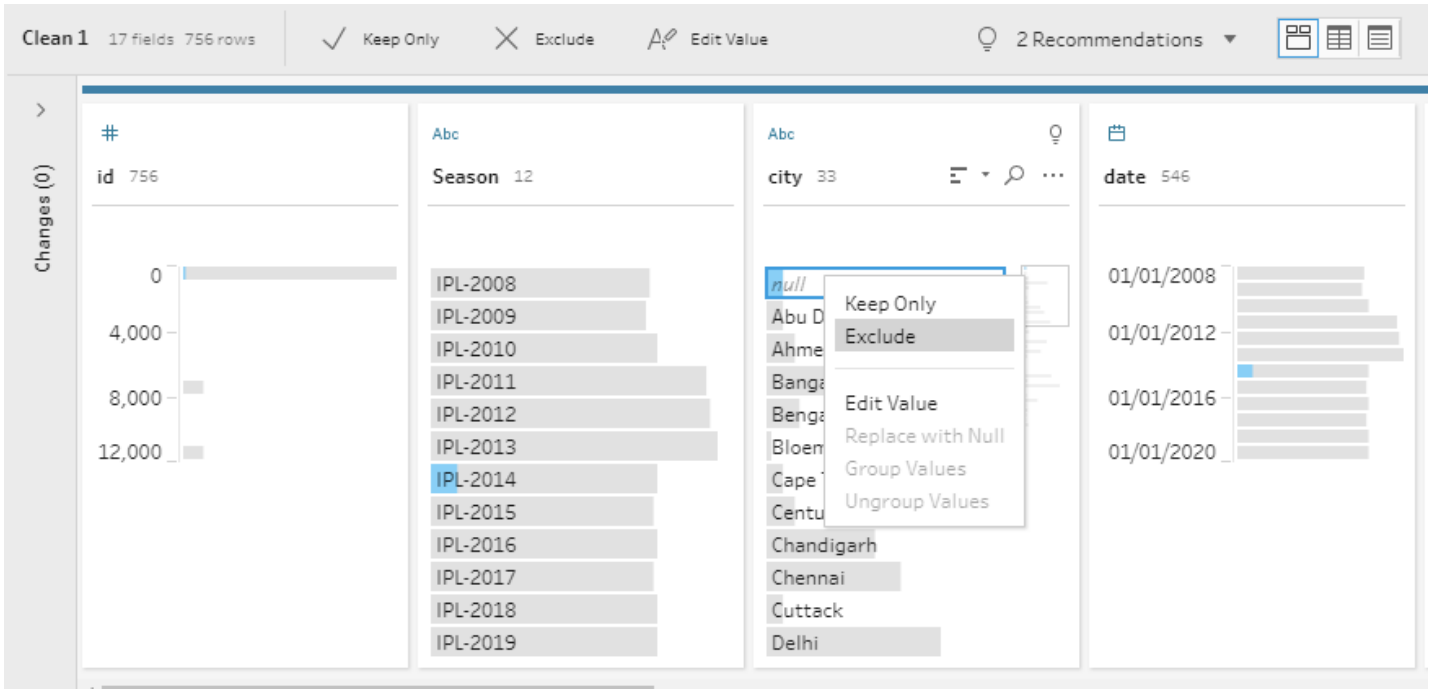
STEP 2: Giving proper name of column:

The dataset does not have proper columns so our next step would be to giver proper column names to the columns wherever required. Which help us to understand about field data.



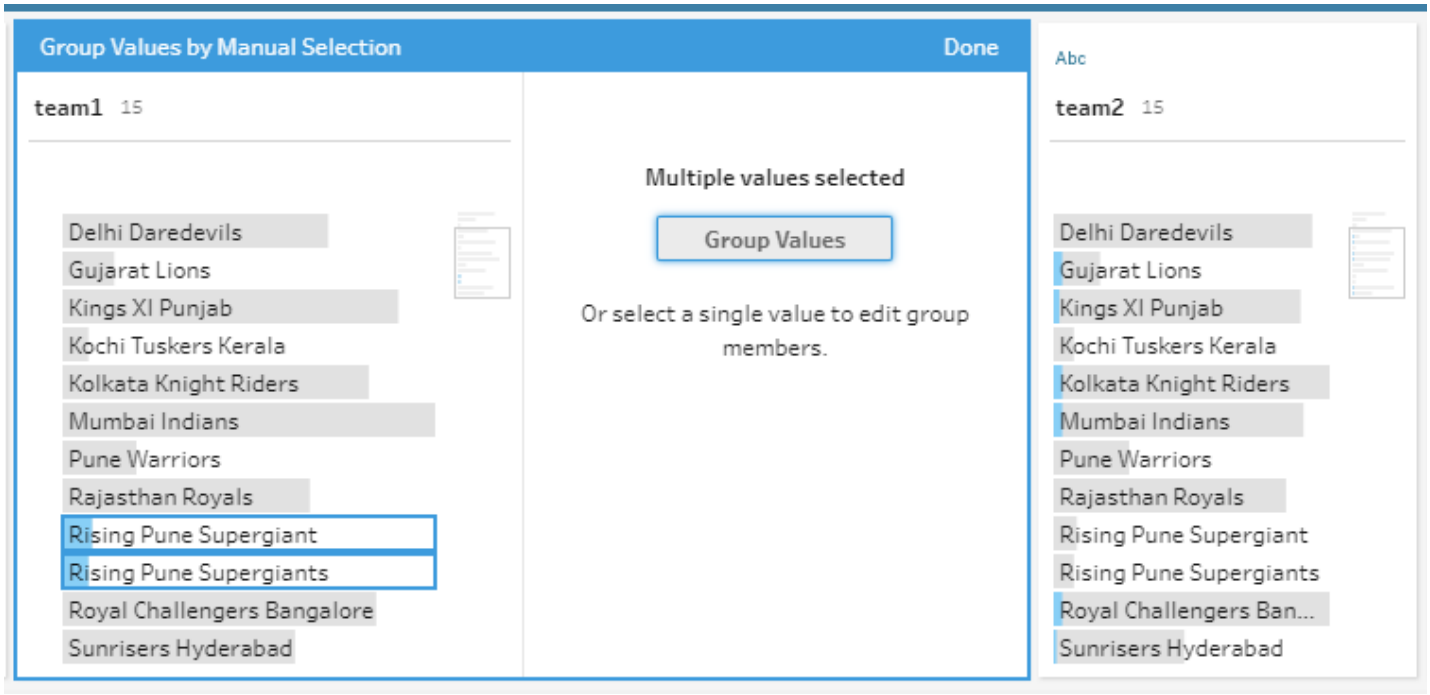
Step 3: Excluding Null Values :

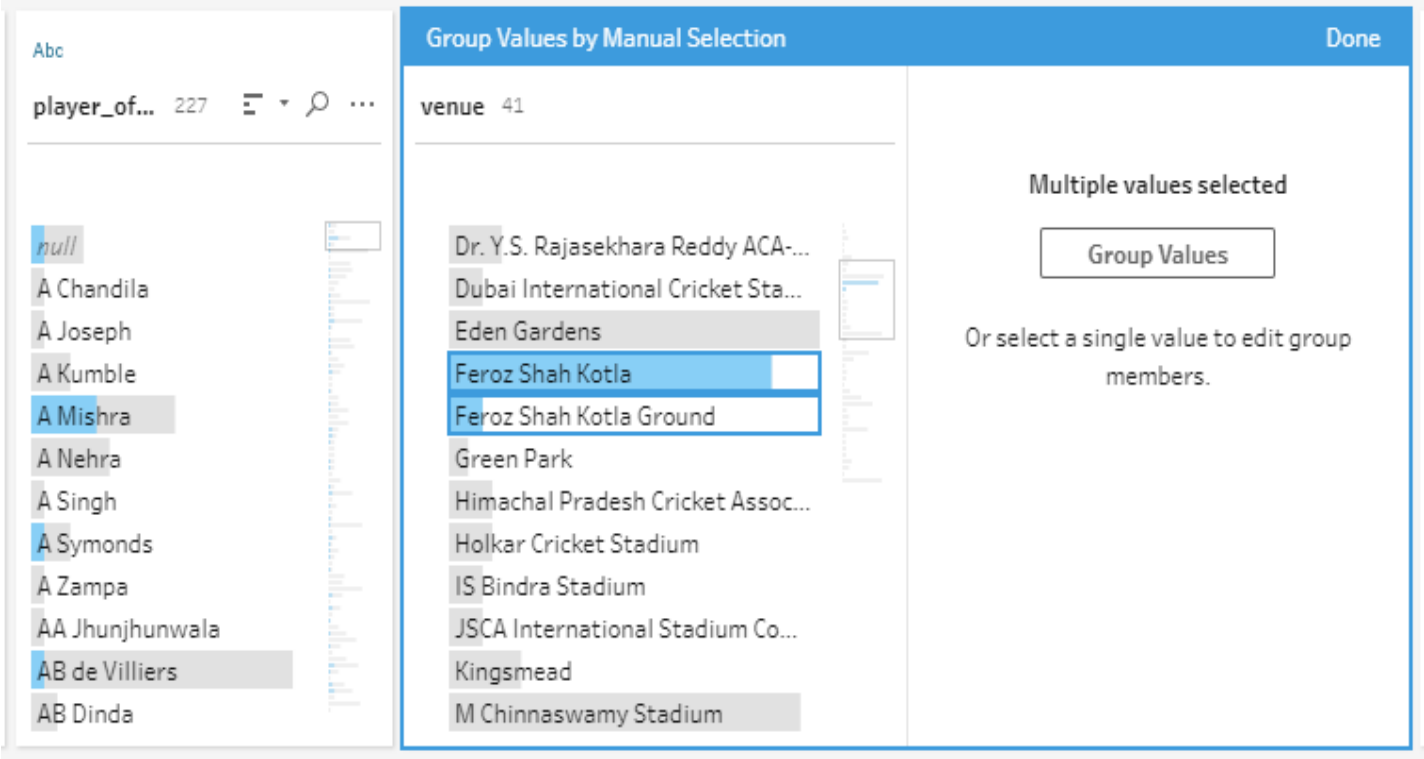
We'll be using Tableau prep for this work as it'll make the work simple and faster because we might not know how many null values could be there in this huge data set. Tableau helps us doing one step cleaning with ease. Its help to make table or chart easier to understand



Step 4: Improving proper data Formatting.

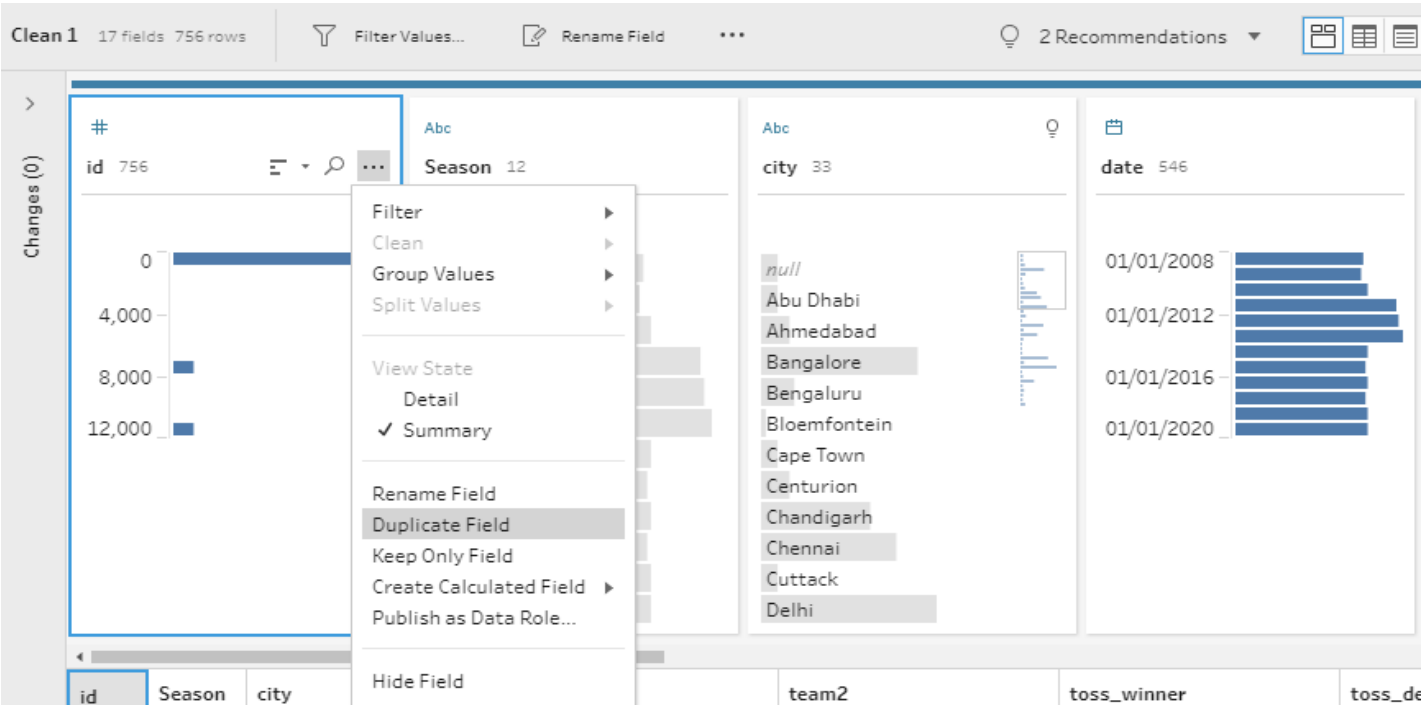
Without proper Data Formatting, proper analysis will not take place. So, we will bring down certain columns to their proper format. For example, the dates should be in the date format and price and sales should be in currency format for better results.





Step 5: Removal of duplicate Value:

it might be possible that our data may be containing duplicate values which may hinder in precise analysis. So, our last task in ETL will be removing duplicate values and making our data perfect for analysis. Now we can go to next step.



ANALYSIS OF DATASET

1. *Venue of match per Season:*

Description:

In this objective we will finding the number of matches held in different cricket stadium for this we will be using total result in each venue and in each season.

Specific function and requirement:

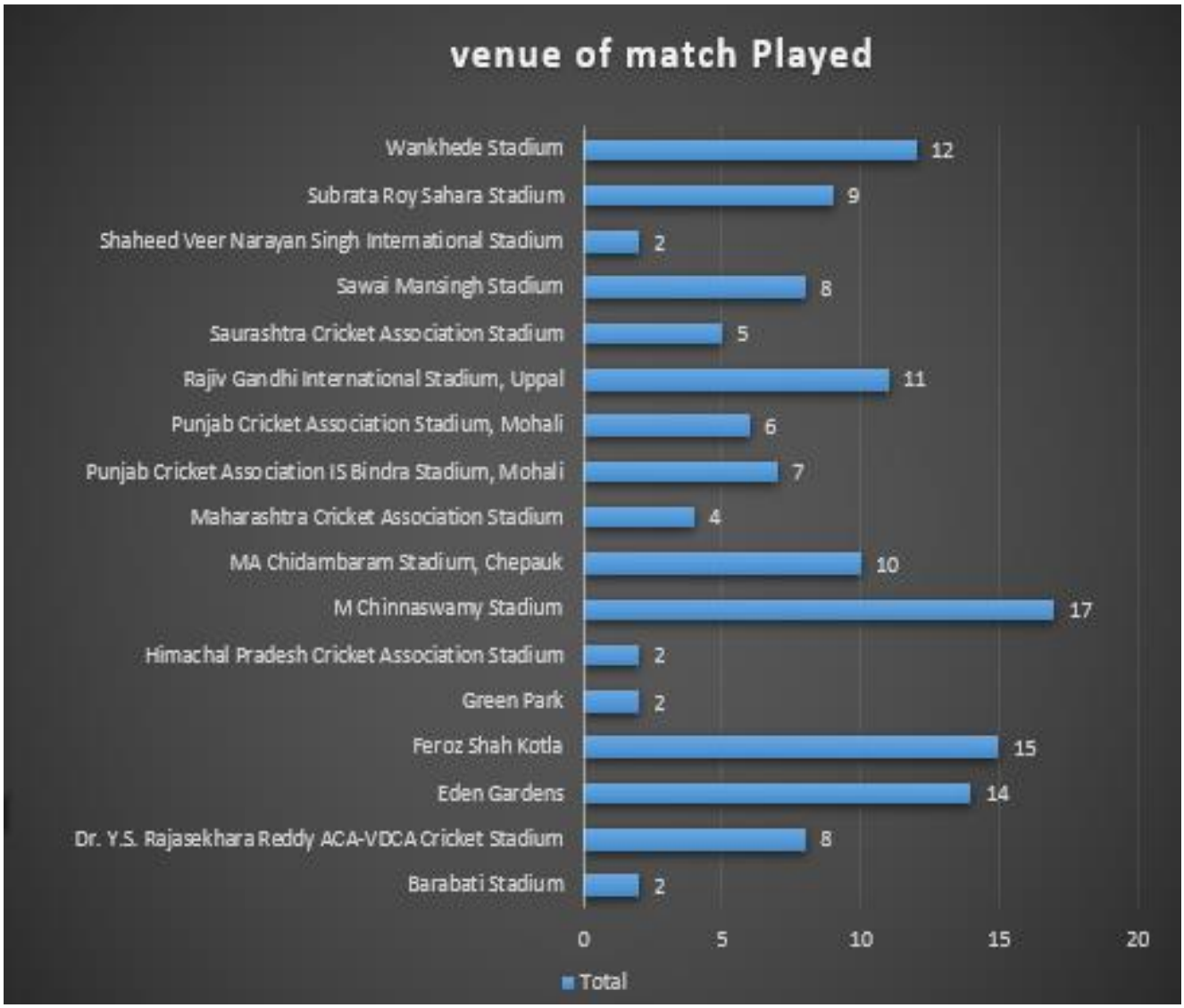
We have to create pivot table to determine the different the number of venue and visualize it on graph.

Result:

Season	(Multiple Items)
Row Labels	**Count of result**
Barabati Stadium	2
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium	8
Eden Gardens	14
Feroz Shah Kotla	15
Green Park	2
Himachal Pradesh Cricket Association Stadium	2
M Chinnaswamy Stadium	17
MA Chidambaram Stadium, Chepauk	10
Maharashtra Cricket Association Stadium	4
Punjab Cricket Association IS Bindra Stadium, Mohali	7
Punjab Cricket Association Stadium, Mohali	6
Rajiv Gandhi International Stadium, Uppal	11
Saurashtra Cricket Association Stadium	5
Sawai Mansingh Stadium	8
Shaheed Veer Narayan Singh International Stadium	2
Subrata Roy Sahara Stadium	9
Wankhede Stadium	12
Grand Total	**134**

Visualization:

The result are the visualization in the form of bar graph for each venue and total count of result.



2. *Player of match award top 10 players*

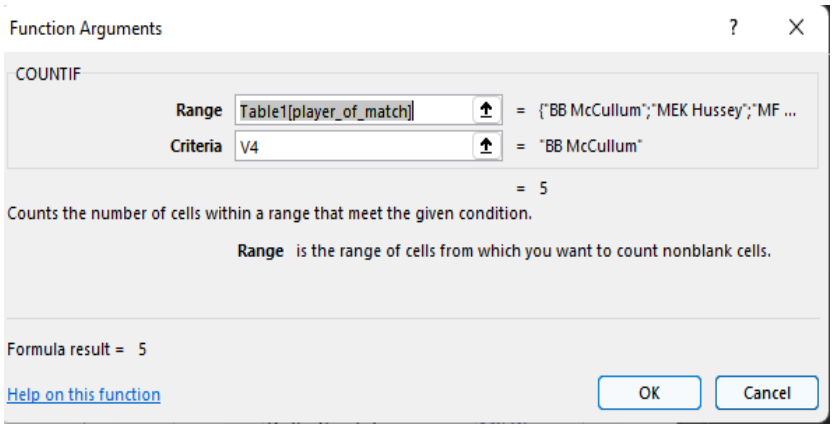
Description:

Top 10 player who got player of match award up to year 2019. So we need Player_of_match Column and we are sort by our use.

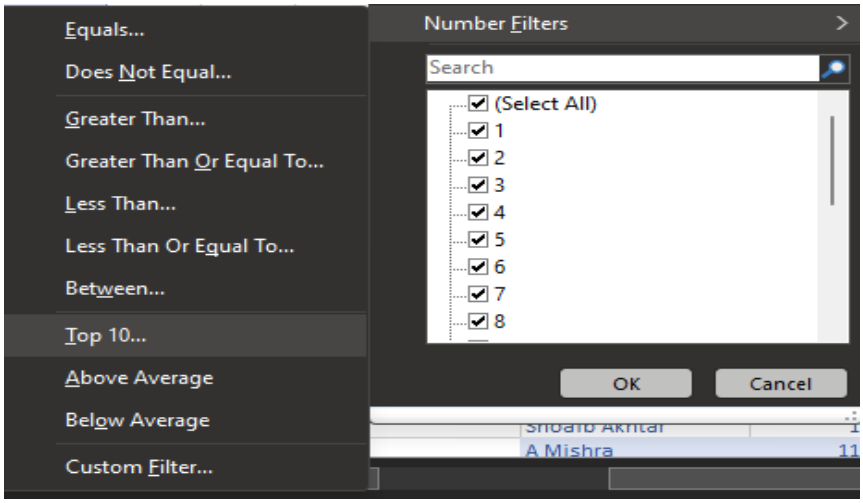
Specific Function and requirement:

First we need to remove duplicate for player of match column with the help of advanced filter then we have to add new column and use count function to count how many time the player won MOM award

=COUNTIF(Table1[player_of_match],V4)



Then after we are using format as table in new created columns with the help of format as table we can easily sort data for our priority we are only using top 10 player who got MOM award then sort data.

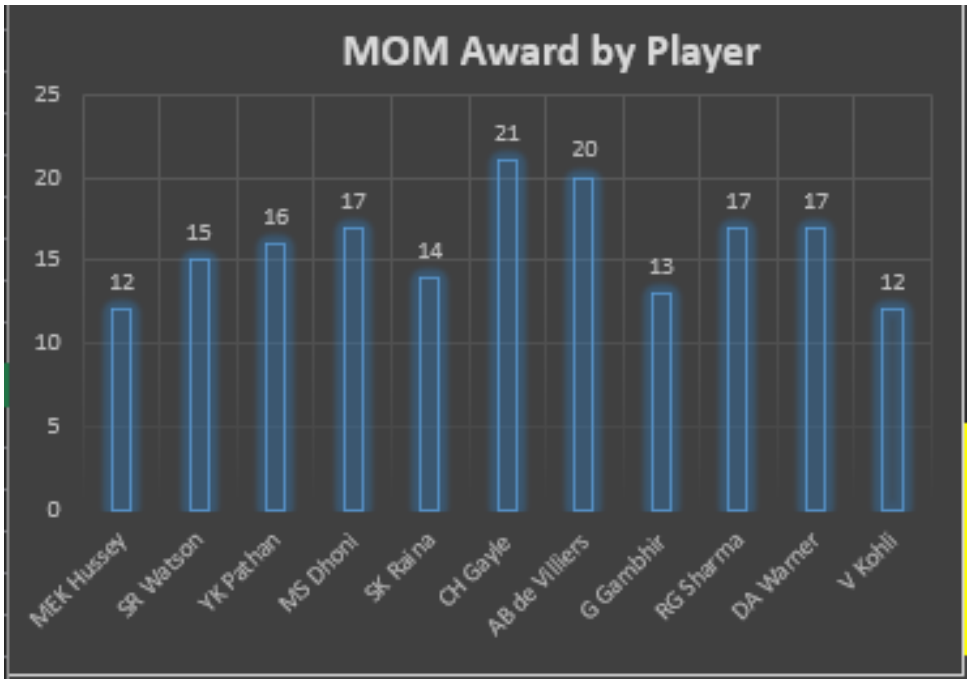


Result:

Top 10 Player of Match Award Winner		
Name of player		**Mom**
MEK Hussey		12
SR Watson		15
YK Pathan		16
MS Dhoni		17
SK Raina		14
CH Gayle		21
AB de Villiers		20
G Gambhir		13
RG Sharma		17
DA Warner		17
V Kohli		12

Visualizations:

We will use a clustered column to visualization the distribution.



3. *Player of match Award by each team.*

Description:

we can find total MOM award won by each team

Specific Function and Analysis:

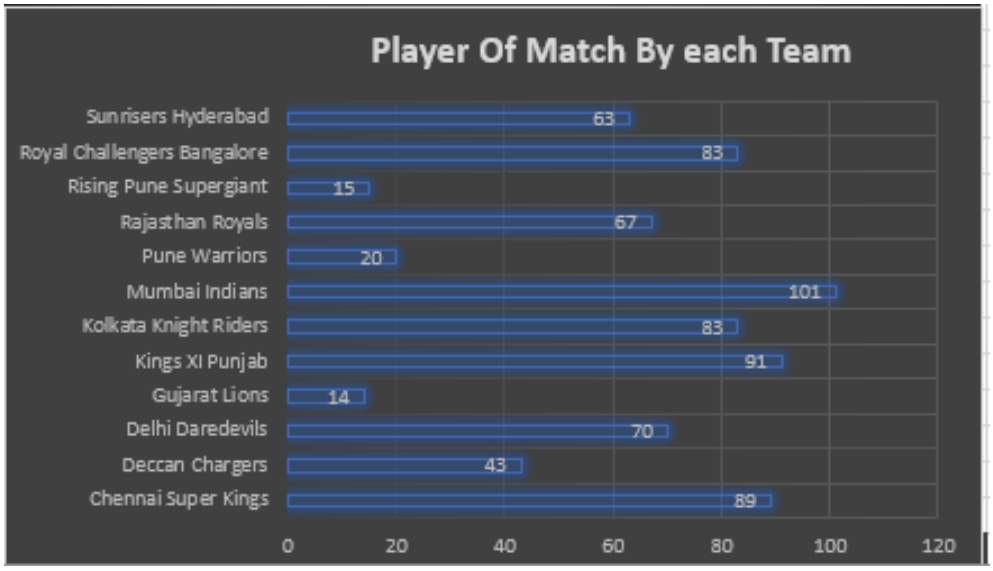
To calculate MOM Award got by each team we are using pivot table to count the no of player of match.

Result:

Row Labels	Count of player_of_match
Chennai Super Kings	89
Deccan Chargers	43
Delhi Capitals	6
Delhi Daredevils	70
Gujarat Lions	14
Kings XI Punjab	91
Kochi Tuskers Kerala	7
Kolkata Knight Riders	83
Mumbai Indians	101
Pune Warriors	20
Rajasthan Royals	67
Rising Pune Supergiant	15
Royal Challengers Bangalore	83
Sunrisers Hyderabad	63
Grand Total	**752**

Visualization:

We will using bar graph for visualizing count of MOM award team wise.



4. *Total runs by player and dismissal for against each team.*

Description:

Runs scored by each player in each season which he played. So the required columns are the name of the player and the all the season and the total runs for which we will be needing the different data of types of runs scored from a player. And for dismissal we need dismissed column to calculate.

Specific function and requirement:

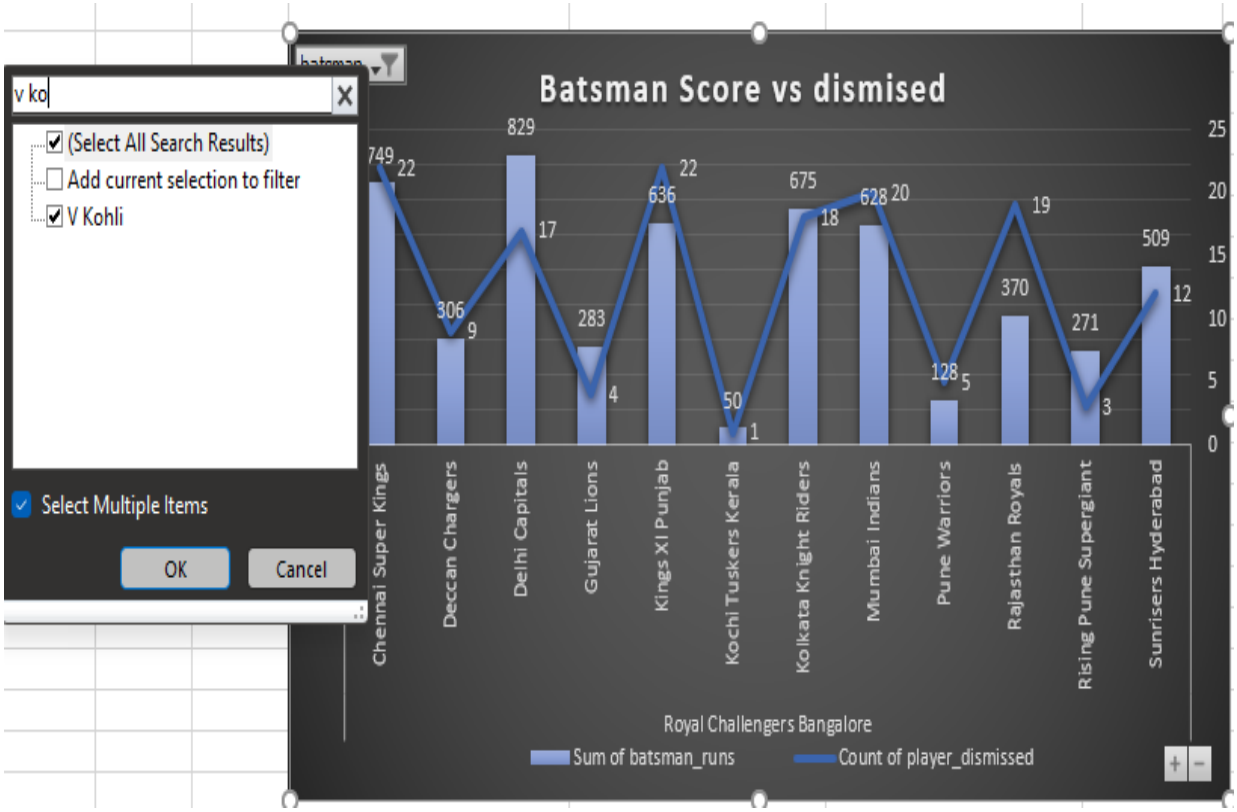
We will be using batsman name as filter and sum of values only for batsman run and dismissed and in row we are adding batting team and bowling team in fileld.

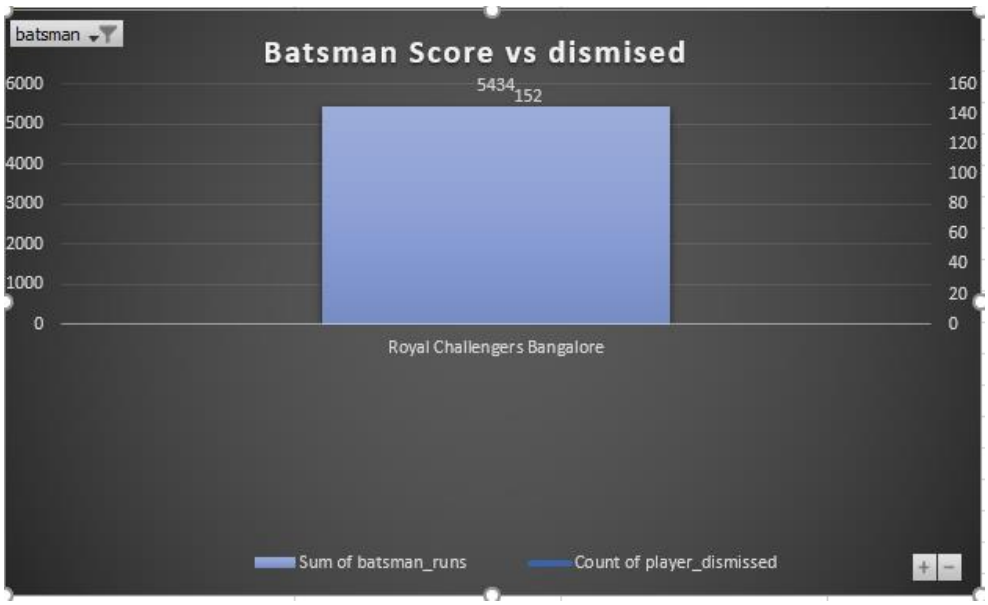
Result:

batsman	V Kohli	
Row Labels	**Sum of batsman_runs**	**Count of player_dismissed**
Royal Challengers Bangalore	**5434**	**152**
Chennai Super Kings	749	22
Deccan Chargers	306	9
Delhi Capitals	829	17
Gujarat Lions	283	4
Kings XI Punjab	636	22
Kochi Tuskers Kerala	50	1
Kolkata Knight Riders	675	18
Mumbai Indians	628	20
Pune Warriors	128	5
Rajasthan Royals	370	19
Rising Pune Supergiant	271	3
Sunrisers Hyderabad	509	12
Grand Total	**5434**	**152**

Visualization:

We will using clustered column for total runs score by particular player and line graph for count of dismissal by player.





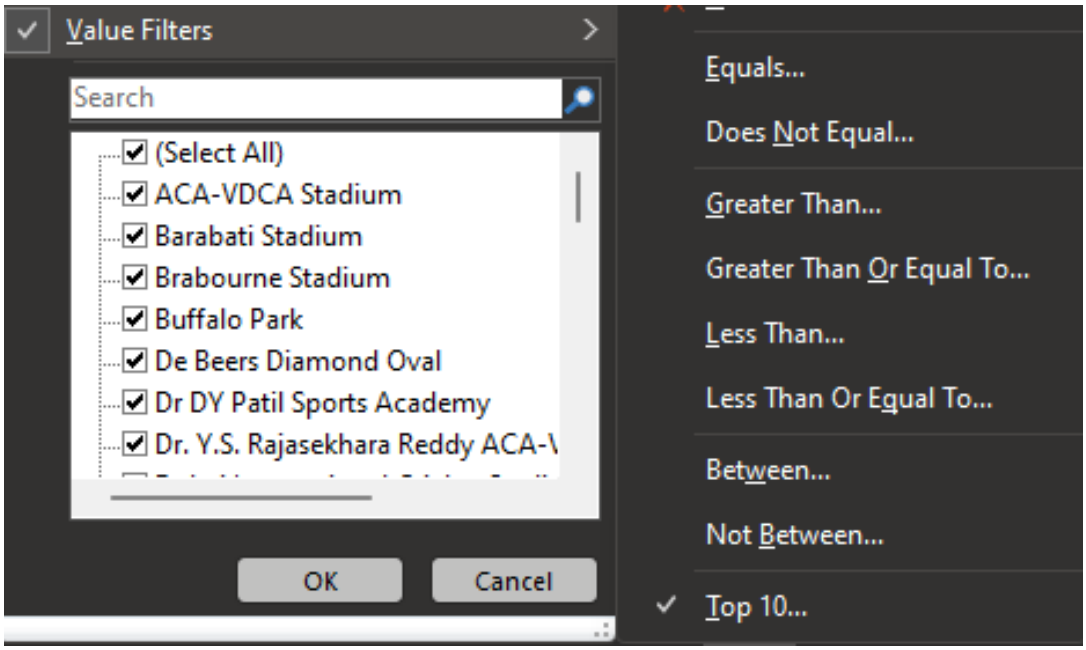
5. *Team won on different stadium by choosing bat or field first.*

Description.

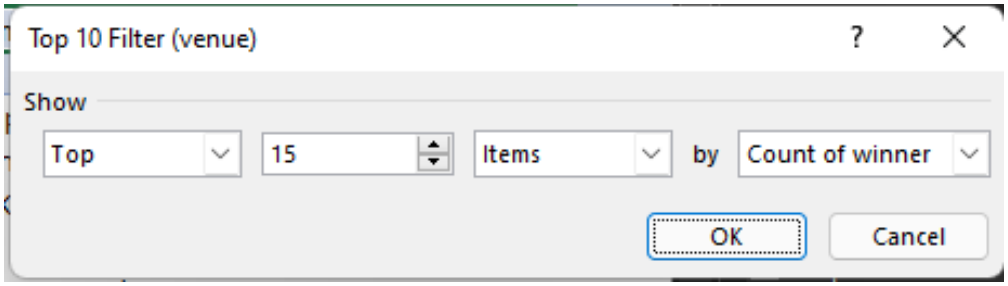
In this objective we will using toss decision in column and venue in rows and count the winner in each stadium. how many times bat choose or how many time filed choosen team won on different venue.

Specific function and requirement:

We will have to create pivot table . and for better visualization we are only calculation the data for top 15 stadium.



In top 10 we are modify top 10 to top 15 venue filed.

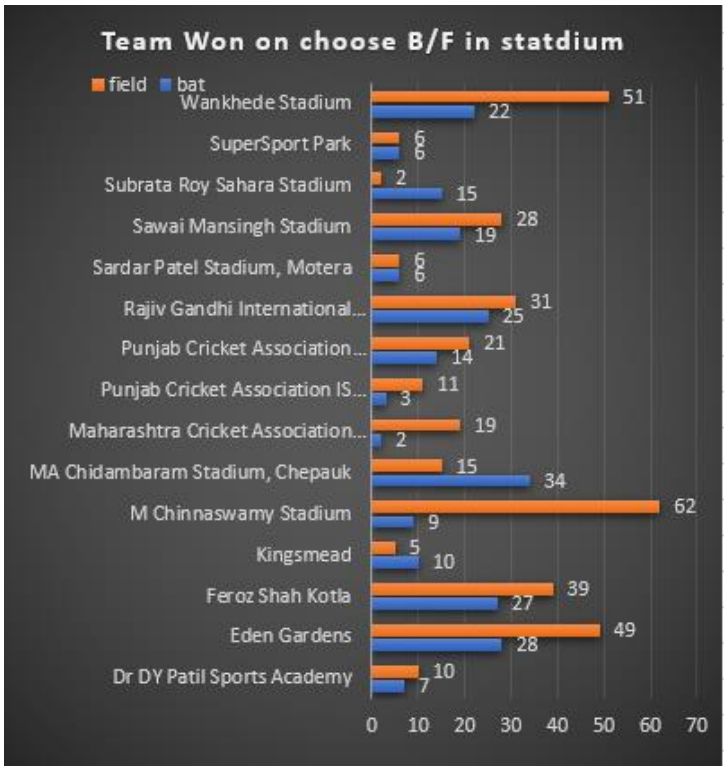


Result:

Count of winner	Column Labels		
Row Labels	bat	field	Grand Total
Dr DY Patil Sports Academy	7	10	17
Eden Gardens	28	49	77
Feroz Shah Kotla	27	39	66
Kingsmead	10	5	15
M Chinnaswamy Stadium	9	62	71
MA Chidambaram Stadium, Chepauk	34	15	49
Maharashtra Cricket Association Stadium	2	19	21
Punjab Cricket Association IS Bindra Stadium, Mohali	3	11	14
Punjab Cricket Association Stadium, Mohali	14	21	35
Rajiv Gandhi International Stadium, Uppal	25	31	56
Sardar Patel Stadium, Motera	6	6	12
Sawai Mansingh Stadium	19	28	47
Subrata Roy Sahara Stadium	15	2	17
SuperSport Park	6	6	12
Wankhede Stadium	22	51	73
Grand Total	**227**	**355**	**582**

Visualization:

We will using two cluster column for each bat and filed column.



Extra objective for detail toss effect on match results

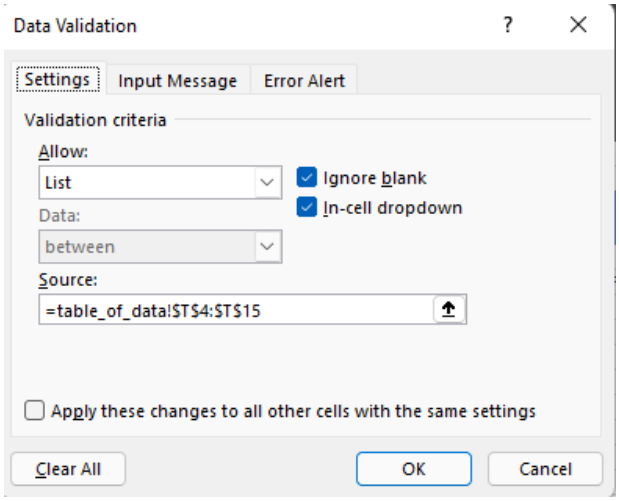
Description:

In this objective we will analysis in detail how the roll of toss effect on our result on winning and losing the match .

Specific requirement Function and formula:

In this object we will using only COUNTIFS() function. And for better visualization we will using dropdown box to change the IPL season.

With the help of advanced filter first we only keep distingue value of each season and also ream name then with the help of data validation we will create dropdown and add distingue season value.



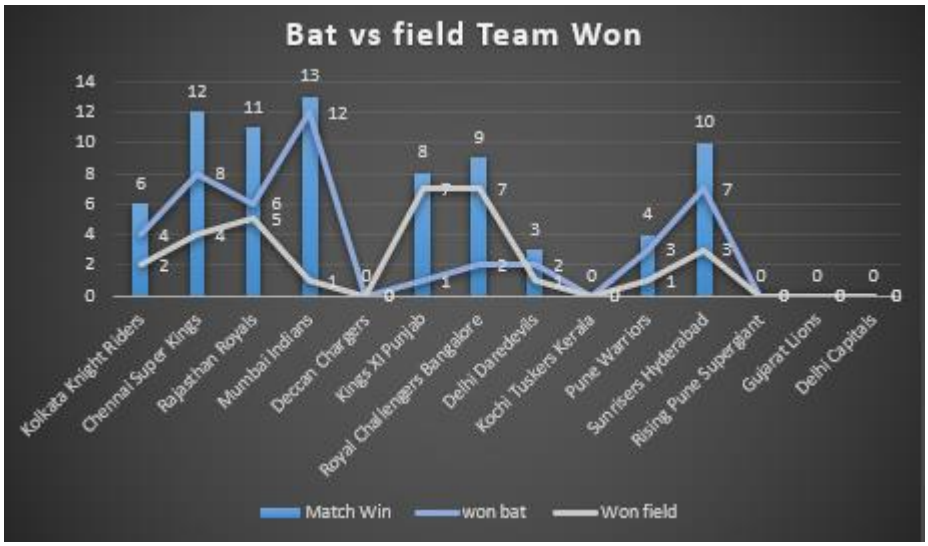
Result:

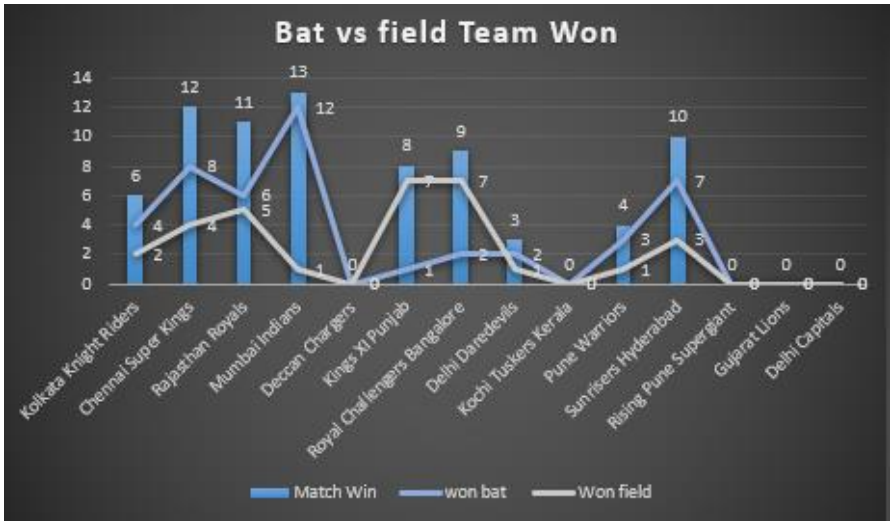
Full toss analysis along with win.

Season	Team	Match Pla	Match Wi	Match Los	Toss Win	Toss Loss	Select Bat	Select Fie	WMWT	WMLT	LMWT	LMLT	won bat	Won field
IPL-2014	Kolkata Knight Rider	16	6	10	12	4	6	6	4	2	8	2	4	2
	Chennai Super Kings	18	12	6	8	10	5	3	7	5	1	5	8	4
	Rajasthan Royals	18	11	7	11	7	6	5	6	5	5	2	6	5
	Mumbai Indians	19	13	6	12	7	10	2	8	5	4	2	12	1
	Deccan Chargers	0	0	0	0	0	0	0	0	0	0	0	0	0
	Kings XI Punjab	16	8	8	7	9	2	5	4	4	3	5	1	7
	Royal Challengers Ba	16	9	7	5	11	2	3	4	5	1	6	2	7
	Delhi Daredevils	16	3	13	5	11	3	2	0	3	5	8	2	1
	Kochi Tuskers Kerala	0	0	0	0	0	0	0	0	0	0	0	0	0
	Pune Warriors	16	4	12	9	7	5	4	2	2	7	5	3	1
	Sunrisers Hyderabad	17	10	7	7	10	6	1	1	9	6	1	7	3
	Rising Pune Supergi	0	0	0	0	0	0	0	0	0	0	0	0	0
	Gujarat Lions	0	0	0	0	0	0	0	0	0	0	0	0	0
	Delhi Capitals	0	0	0	0	0	0	0	0	0	0	0	0	0

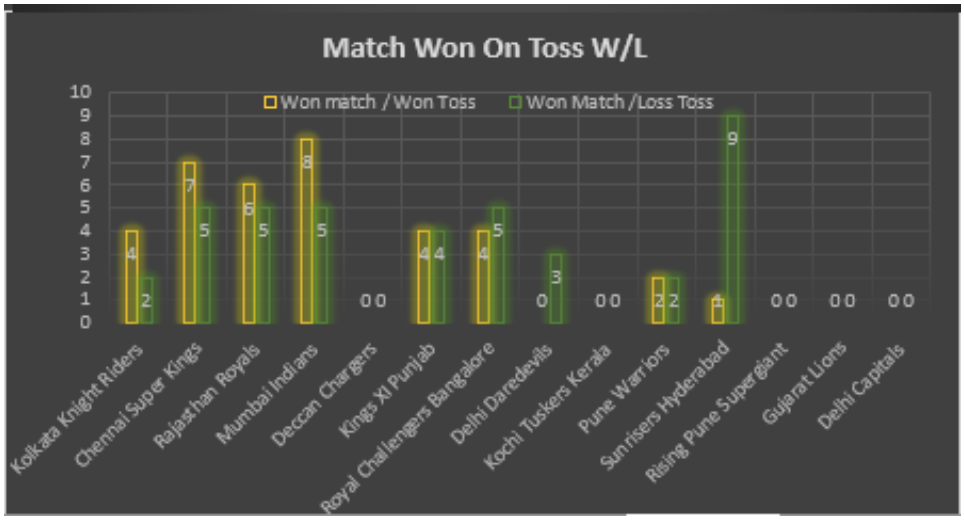
Visualization:

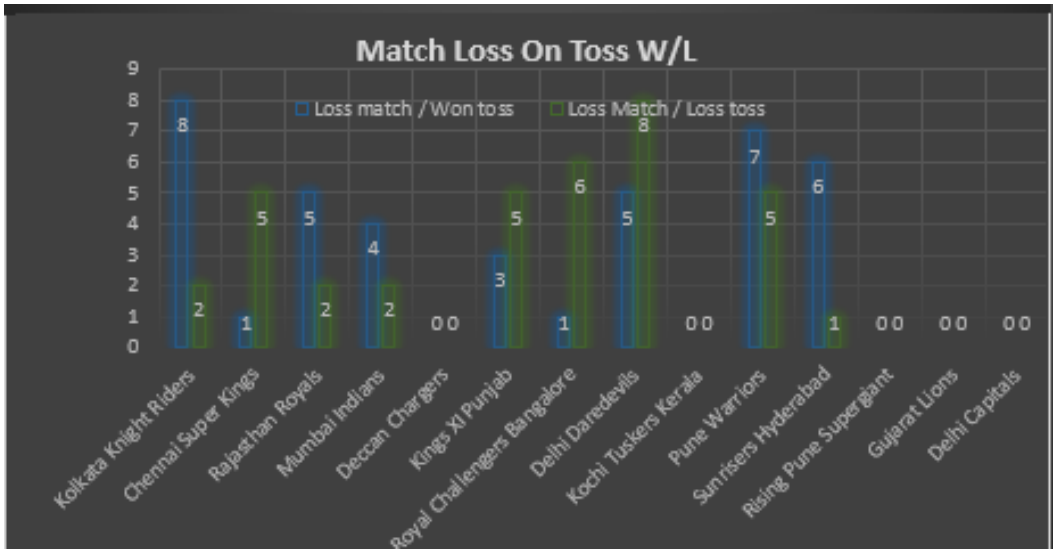
We will using clustered column for total win and line graph for bat vs field.



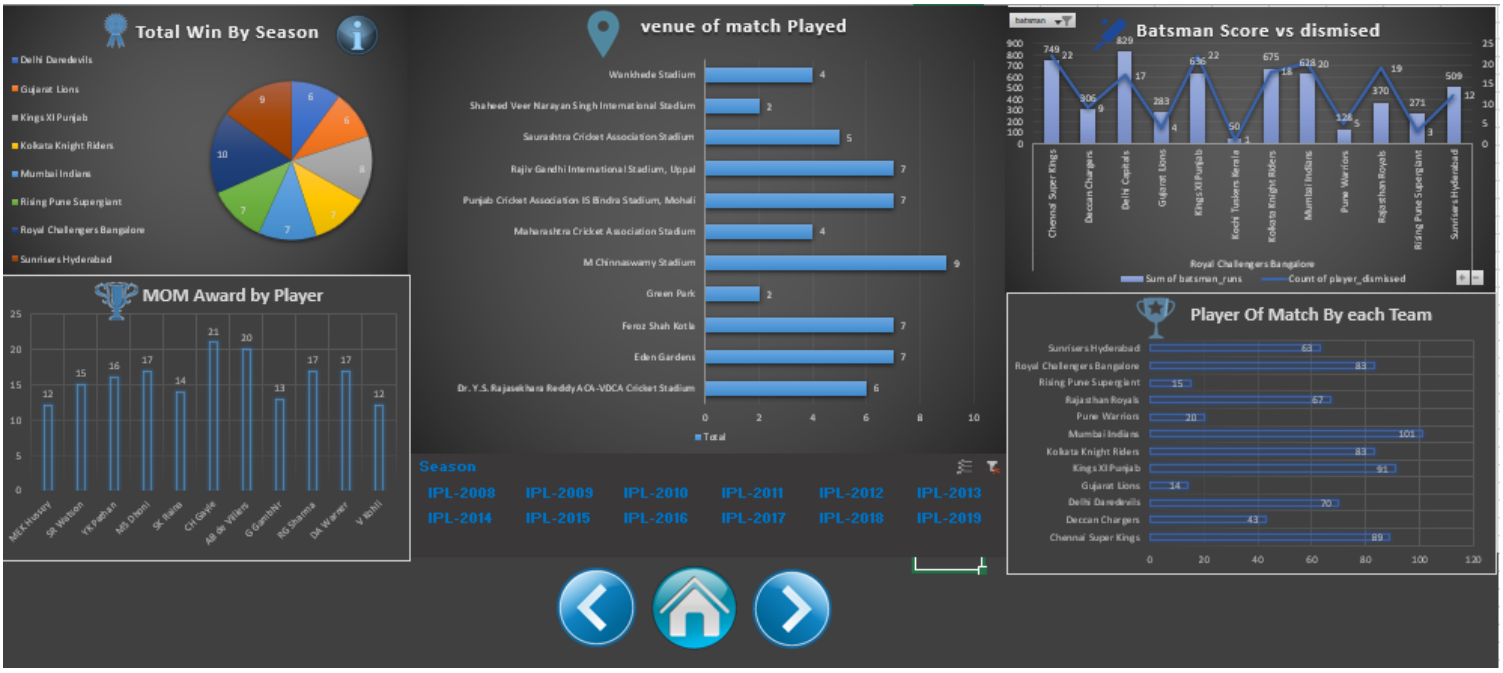


For match won on toss win or loss we will use clustered column





DASHBOARD



Dashboard Will be giving all the brief details of the of work done on the project.

FUTURE SCOPE

The Indian Premiere league is one amongst the simplest entertaining tournament of the planet. per annum of IPL come up with a brand new launching style.

IPL to me looks a awfully good option as far as providing the exposure to the young lot is anxious having played with the biggies of the games likes of Sachin, Kumble, Ponting etc. still hanging around with teams. As far because the skills is anxious, it doesn't dig deep into skill-testing especially for the batsmen, bowlers can still show their intelligence and skill by sticking to basics and containing the batter.

But with IPL the most important issue is that the crowd, since it takes place in major Indian cities which are already hosts of test matches and a few One-dyers it provides the group the choice to decide on between matches. So, this decreases the attendance within the stadium during IPL since the gang wants to determine more or tough competitive cricket instead of friendly ventures.

And coming to the longer term, it's great scope and potential and with BCCI taking care of it so well it doesn't seem the trophy is going to lose any shine in near future.

REFEERENCE

- Kaggle
- Wikipedia
- Microsoft.Excel.com
- IPL.com
- YouTube
- stack overflow

BIBLIOGRAPHY

- Microsoft Excel 2016 Bible