INT217

Introduction to Data Management

Project Report on IPL Data Set

Submitted in partial fulfillment of the requirements for the award of degree of

"Integrated B.Tech. – MBA in Computer Science and Engineering"

School of Computer Science and Engineering

Submitted to

LOVELY PROFESSIONAL UNIVERSITY
PHAGWARA, PUNJAB

Dated: - December 12, 2021

Submitted by:

Name of the student: Shruti Tandon

Registration Number: 11902188

Signature of the student:

STUDENT DECLARATION

To whom so ever it may concern

I, **Shruti Tandon, 11902188,** hereby declare that the work done by me on **"Project on IPL Data Set"**, is a record of original work for the partial fulfillment of the requirements for the award of the degree, **Integrated B.Tech.(CSE)-MBA**.

Name of the Student (Registration Number): Shruti Tandon (11902188)

Signature of the student:

Dated: December 12, 2021

ACKNOWLEDGEMENT

Apart from the efforts of ours, the success of any project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project. I would also like to thank Tanima Thakur ma'am for sharing her knowledge and guiding me throughout this project. Without her encouragement and guidance this report would not have been materialized. I take here a great opportunity to express my sincere and deep sense of gratitude to **Lovely Professional University** for giving me an opportunity to work on this project. The guidance and support received from all the members who contributed like my parents and the knowledge I had gained from the Internet and the books was vital for the success of the project. We are grateful for their constant support and help.

TABLE OF CONTENTS

INTRODUCTION OF THE PROJECT

Microsoft Excel is a computer application program written by Microsoft. It mainly comprises tabs, groups of commands, and worksheets. It stores the data in tabular form and allows the users to perform manipulation operations on them. Microsoft Excel is an office use application designed by Microsoft. It comes with Office Suite with several other Microsoft applications, such as Word, PowerPoint, Access, Outlook, and OneNote, etc. It is supported in Windows as well as Mac operating system too.

It is one of the most suitable spreadsheet programs that help us to store and represent the data in tabular form, manage and manipulate data, create optically logical charts, and more. Excel provides us the worksheet to create a new document in it. You can save the Excel file with .xls extension.

What is a Worksheet and Workbook?

A worksheet is made of rows and columns that intersect each other to form cells where data is entered. It can perform multiple tasks like calculations, data analysis, and integrating data. In Excel worksheet, rows are represented by numbers and columns by alphabets. A single Excel workbook can consist of several sheets, named Sheet1, Sheet2, Sheet3… SheetN. You can add one or more sheets to your Excel document.

Features of MS Excel
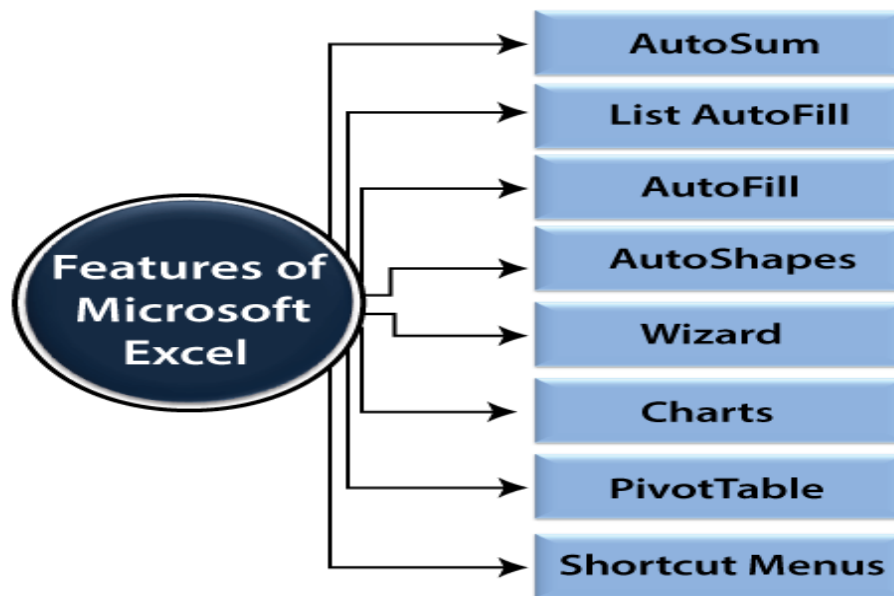
- AutoFormat: It allows the Excel users to use predefined table formatting options.

- AutoSum: AutoSum feature helps us to calculate the sum of a row or column automatically by inserting an addition formula for a range of cells.

- List AutoFill: It automatically develops cell formatting when a new component is added to the end of a list.

- AutoFill: This feature allows us to quickly fill cells with a repetitive or sequential record such as chronological dates or numbers and repeated documents. AutoFill can also be used to copy functions. We can also alter text and numbers with this feature.

- AutoShapes: AutoShapes toolbar will allow us to draw some geometrical shapes, arrows, flowchart items, stars, and more. With these shapes, we can draw our graphs.

- Wizard: It guides us to work effectively while we work by displaying several helpful tips and techniques based on what we are doing. Drag and

Drop feature will help us to reposition the record and text by simply dragging the data with the help of the mouse.

- Charts: This feature will help you to present the data in graphical form by using Pie, Bar, Line charts, and more.

- PivotTable: It flips and sums data in seconds and allows us to execute data analysis and generating documents like periodic financial statements, statistical documents, etc. We can also analyze complex data relationships graphically.

- Shortcut Menus: The shortcut menu helps users to make the work done through shortcut commands that need a lengthy process.

Excel offers the following major chart types:

Line chart	Bar chart	Stacked bar chart	Bullet bar chart
Column chart	Stacked column chart	Pie chart	Pie chart with highlight
Scatterplot chart	Bubble chart	Stacked column volume chc	Stacked column volume wit

Column Chart

A Column Chart typically displays the categories along the horizontal (category) axis and values along the vertical (value) axis. To create a column chart, arrange the data in columns or rows on the worksheet.

A column chart has the following sub-types −

- Clustered Column.

- Stacked Column.

- 100% Stacked Column.

- 3-D Clustered Column.

- 3-D Stacked Column.

- 3-D 100% Stacked Column.

- 3-D Column.

Line Chart

Line charts can show continuous data over time on an evenly scaled Axis. Therefore, they are ideal for showing trends in data at equal intervals, such as months, quarters, or years. In a Line chart Category data is distributed evenly along the horizontal axis and value data is distributed evenly along the vertical axis. To create a Line chart, arrange the data in columns or rows on the worksheet.

A Line chart has the following sub-types −

- Line
- Stacked Line
- 100% Stacked Line
- Line with Markers
- Stacked Line with Markers
- 100% Stacked Line with Markers
- 3-D Line

Pie Chart

Pie charts show the size of items in one data series, proportional to the sum of the items. The data points in a pie chart are shown as a percentage of the whole pie. To create a Pie Chart, arrange the data in one column or row on the worksheet.

A Pie Chart has the following sub-types −

- Pie
- 3-D Pie
- Pie of Pie
- Bar of Pie

Doughnut Chart

A Doughnut chart shows the relationship of parts to a whole. It is like a Pie Chart with the only difference that a Doughnut Chart can contain more than one data series, whereas a Pie Chart can contain only one data series.

A Doughnut Chart contains rings and each ring representing one data series. To create a Doughnut Chart, arrange the data in columns or rows on a worksheet.

Bar Chart

Bar Charts illustrate comparisons among individual items. In a Bar Chart, the categories are organized along the vertical axis and the values are organized along the horizontal axis. To create a Bar Chart, arrange the data in columns or rows on the Worksheet.

A Bar Chart has the following sub-types −

- Clustered Bar

- Stacked Bar

- 100% Stacked Bar

- 3-D Clustered Bar

- 3-D Stacked Bar

- 3-D 100% Stacked Bar

Area Chart

Area Charts can be used to plot the change over time and draw attention to the total value across a trend. By showing the sum of the plotted values, an area chart also shows the relationship of parts to a whole. To create an Area Chart, arrange the data in columns or rows on the worksheet.

An Area Chart has the following sub-types −

- Area

- Stacked Area

- 100% Stacked Area

- 3-D Area

- 3-D Stacked Area

- 3-D 100% Stacked Area

XY (Scatter) Chart

XY (Scatter) charts are typically used for showing and comparing numeric values, like scientific, statistical, and engineering data.

A Scatter chart has two Value Axes − Horizontal (x) Value Axis and Vertical (y) Value Axis. It combines x and y values into single data points and displays them in irregular intervals, or clusters. To create a Scatter chart, arrange the data in columns and rows on the worksheet.

Place the x values in one row or column, and then enter the corresponding y values in the adjacent rows or columns.

A Scatter chart has the following sub-types −

- Scatter

- Scatter with Smooth Lines and Markers

- Scatter with Smooth Lines

- Scatter with Straight Lines and Markers

- Scatter with Straight Lines

Bubble Chart

A Bubble chart is like a Scatter chart with an additional third column to specify the size of the bubbles it shows to represent the data points in the data series.

A Bubble chart has the following sub-types −

- Bubble
- Bubble with 3-D effect

Radar Chart

Radar charts compare the aggregate values of several data series. To create a Radar chart, arrange the data in columns or rows on the worksheet.

A Radar chart has the following sub-types −

- Radar
- Radar with Markers
- Filled Radar

Combo Chart

Combo charts combine two or more chart types to make the data easy to understand, especially when the data is widely varied. It is shown with a secondary axis and is even easier to read. To create a Combo chart, arrange the data in columns and rows on the worksheet.

A Combo chart has the following sub-types −

- Clustered Column – Line
- Clustered Column – Line on Secondary Axis
- Stacked Area – Clustered Column
- Custom Combination

OBJECTIVES

- Analyse the ratio of home win and away win.

- Top 10 "Powe Hitter" batsman of the IPL with their strike rate and average.

- Analyse the dependency of Toss win on Match win.

- Respected winning team by making toss decision as bat or field.

- Total no. of wickets in each over in IPL.

- Total matches won and played by each team.

- Total no. of players from India.

SOURCE OF DATASET

- **Kaggle:** https://www.kaggle.com/ramjidoolla/ipl-data-set

- **COUNTIF**

  It is one of the statistical functions, to count the number of cells that meet a criterion; for example, to count the number of times a particular city appears in a customer list.

  Syntax: —
  =COUNTIF (range, criteria)

  Usage Example: —
  =COUNTIF('Matches   Analyzed'!$K$2:'Matches   Analyzed'!$K$757, teamwise_home_and_away!A2)

  It was used to calculate Total matches won by a team.

ETL PROCESS

- The mechanism of extracting information from source systems and bringing it into the data warehouse is commonly called ETL, which stands for Extraction, Transformation and Loading.

- The ETL process requires active inputs from various stakeholders, including developers, analysts, testers, top executives and is technically challenging.

- Three steps make up the ETL process and enable data to be integrated from source to destination. These are data extraction, data transformation, and data loading.

ETL Process

Source Systems → Extract → Transform → Load → Destination
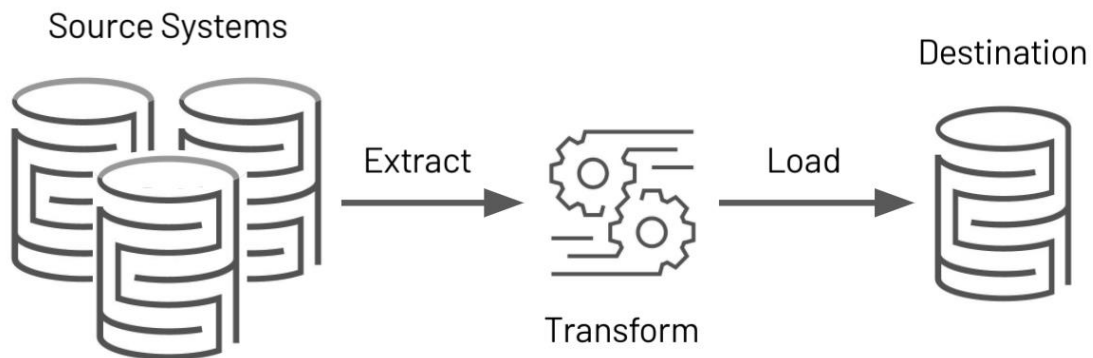
- **Extraction**

    In the first step, extracted data sets come from a source into a staging area. Structured and unstructured data is imported and consolidated into a single repository. Raw data can be extracted from a wide range of sources.

Here input of dataset i.e., Extraction is taken from Kaggle dataset called IPL_data_set which include raw data of players, teams, matches, deliveries, most_runs_average_Strikerate, teamwise_home_and_away from 2008 to 2019.

- **Transformation**

The data cleaning and organization stage is the transformation stage. All that data from multiple source systems will be normalized and converted to a single system format — improving data quality and compliance.

ETL yields transformed data through these methods:

- Cleansing — inconsistencies and missing values in the data are resolved.

- Standardization — formatting rule are applied to the data set.

- Deduplication — redundant data is excluded or discarded.

- Verification — unusable data is removed, and anomalies are flagged.

- Sorting — data is organized according to type.

- Other tasks — any additional/optional rules can be applied to improve data quality.

Transformation is generally considered to be the most important part of the ETL process. Data transformation improves data integrity and helps ensure that data arrives at its new destination fully compatible and ready to use.

S2 | ✓ fx =G2=K2

	A	B	C	D	E	F	G	H	I	K	S	T
1	id	Season	city	date	team1	team2	toss_winner	toss_decision	result	winner	toss win = match win	
2	1	IPL-2017	Hyderabad	4/5/2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	Sunrisers Hyderabad	FALSE	
3	2	IPL-2017	Pune	4/6/2017	Mumbai Indians	Rising Pune Supergiants	Rising Pune Supergiants	field	normal	Rising Pune Supergiants	TRUE	
4	3	IPL-2017	Rajkot	4/7/2017	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	Kolkata Knight Riders	TRUE	
5	4	IPL-2017	Indore	4/8/2017	Rising Pune Supergiants	Kings XI Punjab	Kings XI Punjab	field	normal	Kings XI Punjab	TRUE	
6	5	IPL-2017	Bangalore	4/8/2017	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	Royal Challengers Bangalore	TRUE	
7	6	IPL-2017	Hyderabad	4/9/2017	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	Sunrisers Hyderabad	TRUE	
8	7	IPL-2017	Mumbai	4/9/2017	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	Mumbai Indians	TRUE	
9	8	IPL-2017	Indore	4/10/2017	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	normal	Kings XI Punjab	FALSE	
10	9	IPL-2017	Pune	4/11/2017	Delhi Daredevils	Rising Pune Supergiants	Rising Pune Supergiants	field	normal	Delhi Daredevils	FALSE	
11	10	IPL-2017	Hyderabad	4/12/2017	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	Mumbai Indians	TRUE	
12	11	IPL-2017	Kolkata	4/13/2017	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	Kolkata Knight Riders	TRUE	
13	12	IPL-2017	Bangalore	4/14/2017	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	normal	Mumbai Indians	TRUE	
14	13	IPL-2017	Rajkot	4/14/2017	Rising Pune Supergiants	Gujarat Lions	Gujarat Lions	field	normal	Gujarat Lions	TRUE	
15	14	IPL-2017	Kolkata	4/15/2017	Kolkata Knight Riders	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	Kolkata Knight Riders	FALSE	
16	15	IPL-2017	Delhi	4/15/2017	Delhi Daredevils	Kings XI Punjab	Delhi Daredevils	bat	normal	Delhi Daredevils	TRUE	
17	16	IPL-2017	Mumbai	4/16/2017	Gujarat Lions	Mumbai Indians	Mumbai Indians	field	normal	Mumbai Indians	TRUE	
18	17	IPL-2017	Bangalore	4/16/2017	Rising Pune Supergiants	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	Rising Pune Supergiants	FALSE	
19	18	IPL-2017	Delhi	4/17/2017	Delhi Daredevils	Kolkata Knight Riders	Delhi Daredevils	bat	normal	Kolkata Knight Riders	FALSE	
20	19	IPL-2017	Hyderabad	4/17/2017	Sunrisers Hyderabad	Kings XI Punjab	Kings XI Punjab	field	normal	Sunrisers Hyderabad	FALSE	
21	20	IPL-2017	Rajkot	4/18/2017	Royal Challengers Bangalore	Gujarat Lions	Gujarat Lions	field	normal	Royal Challengers Bangalore	FALSE	
22	21	IPL-2017	Hyderabad	4/19/2017	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	Sunrisers Hyderabad	TRUE	
23	22	IPL-2017	Indore	4/20/2017	Kings XI Punjab	Mumbai Indians	Mumbai Indians	field	normal	Mumbai Indians	TRUE	
24	23	IPL-2017	Kolkata	4/21/2017	Kolkata Knight Riders	Gujarat Lions	Gujarat Lions	field	normal	Gujarat Lions	TRUE	
25	24	IPL-2017	Mumbai	4/22/2017	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	field	normal	Mumbai Indians	FALSE	

In this IPL dataset data is cleaned manually as most of it was already cleaned for example some entries missing in the country field of players worksheet were manually filled, formatting was standardized throughout the worksheets, process of deduplication was carried out as there were some cases of duplicate values in teams, data was organized and required field were calculated and added, like, toss win = match win, total matches played and won fields were calculated.

H2 | ✓ fx =COUNTIF('Matches Analyzed'!$E$2:'Matches Analyzed'!$E$757,teamwise_home_and_away!A2)

	A	B	C	D	E	F	G	H	I	J	K
1	team	home_wins	away_wins	home_matches	away_matches	home_win_percentage	away_win_percentage	Team 1 "position"	Team 2 "position"	Total Matches Played	Total Matches Won
2	Rising Pune Supergiants	5	5	8	8	62.5	62.5	15	15	30	15
3	Mumbai Indians	58	51	101	86	57.42574257	59.30232558	101	86	187	109
4	Chennai Super Kings	51	49	89	75	57.30337079	65.33333333	89	75	164	100
5	Delhi Capitals	3	7	6	10	50	70	6	10	16	10
6	Sunrisers Hyderabad	30	28	63	45	47.61904762	62.22222222	63	45	108	58
7	Rajasthan Royals	29	46	67	80	43.28358209	57.5	67	80	147	75
8	Deccan Chargers	18	11	43	32	41.86046512	34.375	43	32	75	29
9	Kings XI Punjab	38	44	91	85	41.75824176	51.76470588	91	85	176	82
10	Royal Challengers Bangalore	35	49	85	95	41.17647059	51.57894737	85	95	180	84
11	Kolkata Knight Riders	34	58	83	95	40.96385542	61.05263158	83	95	178	92
12	Delhi Daredevils	25	42	72	89	34.72222222	47.19101124	72	89	161	67
13	Pune Warriors	6	6	20	26	30	23.07692308	20	26	46	12
14	Kochi Tuskers Kerala	2	4	7	7	28.57142857	57.14285714	7	7	14	6
15	Gujarat Lions	1	12	14	16	7.142857143	75	14	16	30	13
16											
17											
18											
19											
20											

Y2 | ✓ fx =COUNTIF($K$2:$K$757,U2)

	S	T	U	V	W	X	Y	Z
1	toss win = match win		Teams	Team 1 "Position"	Team 2 "Position"	Total Matches Played	Total Matches Won	
2	FALSE		Mumbai Indians	101	86	187	109	
3	TRUE		Kings XI Punjab	91	85	176	82	
4	TRUE		Chennai Super Kings	89	75	164	100	
5	TRUE		Royal Challengers Bangalore	85	95	180	84	
6	TRUE		Kolkata Knight Riders	83	95	178	92	
7	TRUE		Delhi Daredevils	72	89	161	67	
8	TRUE		Rajasthan Royals	67	80	147	75	
9	FALSE		Sunrisers Hyderabad	63	45	108	58	
10	FALSE		Deccan Chargers	43	32	75	29	
11	TRUE		Pune Warriors	20	26	46	12	
12	TRUE		Rising Pune Supergiants	15	15	30	15	
13	TRUE		Gujarat Lions	14	16	30	13	
14	TRUE		Kochi Tuskers Kerala	7	7	14	6	
15	FALSE		Delhi Capitals	6	10	16	10	
16	TRUE		Grand Total	756	756	1512	752	
17	TRUE							
18	FALSE							
19	FALSE							
20	FALSE							

- **Loading**

The final step in the ETL process is to load the newly transformed data into a new destination. Data that has been extracted to a staging area and transformed is loaded into your data warehouse.

Depending upon your business needs, data can be loaded in batches or all at once. Data can be loaded all at once (full load) or at scheduled intervals (incremental load). The exact nature of the loading will depend upon the data source, ETL tools, and various other factors.

Here data was not to be loaded in any data warehouse, it was simply transformed with all the required modifications and saved in work area/staging area in excel.

Introduction about the Dataset

My dataset is on IPL statistics from 2008-2019. The Indian Premier League (IPL) is a professional Twenty20 cricket league, contested by eight teams based out of eight Indian cities. The league was founded by the Board of Control for Cricket in India (BCCI) in 2007. It is usually held between March and May of every year and has an exclusive window in the ICC Future Tours Programme. The IPL is the most-attended cricket league in the world and in 2014 was ranked sixth by average attendance among all sports leagues.

The dataset consists of these 6 worksheets:

- Players Worksheet: It consists of the name of the players, their date of birth, country to which they belong, their batting hand and bowling skills.

	Player_Name	DOB	Batting_Hand	Bowling_Skill	Country
1	**Player_Name**	**DOB**	**Batting_Hand**	**Bowling_Skill**	**Country**
2	A Ashish Reddy	24-Feb-91	Right_Hand	Right-arm medium	India
3	A Chandila	05-Dec-83	Right_Hand	Right-arm offbreak	India
4	A Chopra	19-Sep-77	Right_Hand	Right-arm offbreak	India
5	A Choudhary		Right_Hand	Left-arm fast-medium	India
6	A Dananjaya		Right_Hand	Right-arm offbreak	Sri Lanka
7	A Flintoff	06-Dec-77	Right_Hand	Right-arm fast-medium	England
8	A Hales		Right_Hand	Right-arm medium	England
9	A Joseph		Right_Hand	Right-arm fast	West Indies
10	A Kumble	17-Oct-70	Right_Hand	Legbreak googly	India
11	A Mishra	24-Nov-82	Right_Hand	Legbreak	India
12	A Mithun	25-Oct-89	Right_Hand	Right-arm medium	India
13	A Mukund	06-Jan-90	Right_Hand	Legbreak googly	India
14	A Nehra	29-Apr-79	Right_Hand	Left-arm medium-fast	India
15	A Nel	15-Jul-77	Right_Hand	Right-arm fast-medium	South Africa
16	A Roy		Left_Hand	Slow left-arm orthodox	India
17	A Singh	21-Jun-81	Right_Hand	Right-arm medium-fast	India
18	A Symonds	09-Jun-75	Right_Hand	Right-arm offbreak	Australia
19	A Turner		Right_Hand	Right-arm offbreak	Australia
20	A Uniyal	21-Nov-81	Right_Hand	Left-arm medium	India
21	A Zampa	31-Mar-92	Right_Hand	Legbreak googly	Australia
22	AA Bilakhia	31-May-86	Right_Hand	Right-arm medium	India
23	AA Chavan	28-Oct-85	Right_Hand	Slow left-arm orthodox	India
24	AA Jhunjhunwala	01-Dec-82	Right_Hand	Right-arm offbreak	India
25	AA Kazi	29-Sep-89	Right_Hand	Slow left-arm orthodox	India
26	AA Noffke	30-Apr-77	Right_Hand	Right-arm fast-medium	Australia
27	AB Agarkar	04-Dec-77	Right_Hand	Right-arm fast	India
28	AB Barath	14-Apr-90	Left_Hand	Right-arm offbreak	India
29	AB de Villiers	17-Feb-84	Right_Hand	Right-arm medium	South Africa
30	AB Dinda	25-Mar-84	Right_Hand	Right-arm medium-fast	India
31	AB McDonald	05-Jun-81	Right_Hand	Right-arm fast-medium	Australia
32	Abdur Razzak	15-Jun-82	Left_Hand	Slow left-arm orthodox	Bangladesh
33	AC Blizzard	27-Jun-84	Left_Hand	Left-arm medium	Australia

Players | Teams | Matches | Deliveries | Most_runs_average_Strikerate | teamwise_home_and_away | M

- Teams worksheet: It consists of the names of all the IPL teams.
- Matches worksheet: This worksheet consists of the match id, IPL season, city in which the match was played, date of the match, teams between which the match was played, which team won the toss and what was the toss decision, winner of the match, player of the match, venue where the match was played and many more fields.

id	Season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_mat
1	IPL-2017	Hyderabad	05-04-2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh
2	IPL-2017	Pune	06-04-2017	Mumbai Indians	Rising Pune Supergiants	Rising Pune Supergiants	field	normal	0	Rising Pune Supergiants	0	7	SPD Smith
3	IPL-2017	Rajkot	07-04-2017	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CA Lynn
4	IPL-2017	Indore	08-04-2017	Rising Pune Supergiants	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	GJ Maxwell
5	IPL-2017	Bangalore	08-04-2017	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav
6	IPL-2017	Hyderabad	09-04-2017	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	0	9	Rashid Khan
7	IPL-2017	Mumbai	09-04-2017	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	N Rana
8	IPL-2017	Indore	10-04-2017	Royal Challengers Bangalore	Kings XI Punjab	Royal Challengers Bangalore	bat	normal	0	Kings XI Punjab	0	8	AR Patel
9	IPL-2017	Pune	11-04-2017	Delhi Daredevils	Rising Pune Supergiants	Rising Pune Supergiants	field	normal	0	Delhi Daredevils	97	0	SV Samson
10	IPL-2017	Mumbai	12-04-2017	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	JJ Bumrah
11	IPL-2017	Kolkata	13-04-2017	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	8	SP Narine
12	IPL-2017	Bangalore	14-04-2017	Royal Challengers Bangalore	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	KA Pollard
13	IPL-2017	Rajkot	14-04-2017	Rising Pune Supergiants	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	7	AJ Tye
14	IPL-2017	Kolkata	15-04-2017	Kolkata Knight Riders	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Kolkata Knight Riders	17	0	RV Uthappa
15	IPL-2017	Delhi	15-04-2017	Delhi Daredevils	Kings XI Punjab	Delhi Daredevils	bat	normal	0	Delhi Daredevils	51	0	CJ Anderson
16	IPL-2017	Mumbai	16-04-2017	Gujarat Lions	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	6	N Rana
17	IPL-2017	Bangalore	16-04-2017	Rising Pune Supergiants	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Rising Pune Supergiants	27	0	BA Stokes
18	IPL-2017	Delhi	17-04-2017	Delhi Daredevils	Kolkata Knight Riders	Delhi Daredevils	bat	normal	0	Kolkata Knight Riders	0	4	NM Coulter-Ni
19	IPL-2017	Hyderabad	17-04-2017	Sunrisers Hyderabad	Kings XI Punjab	Kings XI Punjab	field	normal	0	Sunrisers Hyderabad	5	0	B Kumar
20	IPL-2017	Rajkot	18-04-2017	Royal Challengers Bangalore	Gujarat Lions	Gujarat Lions	field	normal	0	Royal Challengers Bangalore	21	0	CH Gayle
21	IPL-2017	Hyderabad	19-04-2017	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	15	0	KS Williamson
22	IPL-2017	Indore	20-04-2017	Kings XI Punjab	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	8	JC Buttler
23	IPL-2017	Kolkata	21-04-2017	Kolkata Knight Riders	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	4	SK Raina
24	IPL-2017	Mumbai	22-04-2017	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	field	normal	0	Mumbai Indians	14	0	MJ McClenagh
25	IPL-2017	Pune	22-04-2017	Sunrisers Hyderabad	Rising Pune Supergiants	Rising Pune Supergiants	field	normal	0	Rising Pune Supergiants	0	6	MS Dhoni
26	IPL-2017	Rajkot	23-04-2017	Kings XI Punjab	Gujarat Lions	Gujarat Lions	field	normal	0	Kings XI Punjab	26	0	HM Amla
27	IPL-2017	Kolkata	23-04-2017	Kolkata Knight Riders	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Kolkata Knight Riders	82	0	NM Coulter-Ni
28	IPL-2017	Mumbai	24-04-2017	Rising Pune Supergiants	Mumbai Indians	Mumbai Indians	field	normal	0	Rising Pune Supergiants	3	0	BA Stokes
29	IPL-2017	Pune	26-04-2017	Rising Pune Supergiants	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	7	RV Uthappa
30	IPL-2017	Bangalore	27-04-2017	Royal Challengers Bangalore	Gujarat Lions	Gujarat Lions	field	normal	0	Gujarat Lions	0	7	AJ Tye
31	IPL-2017	Kolkata	28-04-2017	Delhi Daredevils	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	7	G Gambhir
32	IPL-2017	Chandigarh	28-04-2017	Sunrisers Hyderabad	Kings XI Punjab	Kings XI Punjab	field	normal	0	Sunrisers Hyderabad	26	0	Rashid Khan
33	IPL-2017	Pune	29-04-2017	Rising Pune Supergiants	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Rising Pune Supergiants	61	0	LH Ferguson
34	IPL-2017	Rajkot	29-04-2017	Gujarat Lions	Mumbai Indians	Gujarat Lions	bat	tie	0	Mumbai Indians	0	0	KH Pandya
35	IPL-2017	Chandigarh	30-04-2017	Delhi Daredevils	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	10	Sandeep Sharn

Players Teams **Matches** Deliveries Most_runs_average_Strikerate teamwise_home_and_away Matches Analy... ...

- Deliveries Worksheet: This worksheet consists of the details of all the deliveries made, the match id, inning, batting team, bowling team, over number, batsman and non-striker name, the name of the bowler who bowled, player dismissed details and many more fields.

match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	wide_runs	bye_runs	legbye_runs	noball_runs	penalty_runs
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	2	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	6	S Dhawan	DA Warner	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	7	S Dhawan	DA Warner	TS Mills	0	0	0	1	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	1	S Dhawan	DA Warner	A Choudhary	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	2	DA Warner	S Dhawan	A Choudhary	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	3	DA Warner	S Dhawan	A Choudhary	0	0	0	0	1	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	4	DA Warner	S Dhawan	A Choudhary	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	5	DA Warner	S Dhawan	A Choudhary	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	6	MC Henriques	S Dhawan	A Choudhary	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	7	MC Henriques	S Dhawan	A Choudhary	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	1	S Dhawan	MC Henriques	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	2	MC Henriques	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	3	MC Henriques	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	4	MC Henriques	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	5	S Dhawan	MC Henriques	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	6	MC Henriques	S Dhawan	TS Mills	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	1	MC Henriques	S Dhawan	YS Chahal	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	2	MC Henriques	S Dhawan	YS Chahal	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	3	S Dhawan	MC Henriques	YS Chahal	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	4	S Dhawan	MC Henriques	YS Chahal	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	5	MC Henriques	S Dhawan	YS Chahal	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	6	S Dhawan	MC Henriques	YS Chahal	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	1	S Dhawan	MC Henriques	S Aravind	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	2	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	3	S Dhawan	MC Henriques	S Aravind	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	4	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	5	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	6	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0

Players Teams Matches **Deliveries** Most_runs_average_Strikerate teamwise_home_and_away Matches Analy... ...

- Most_runs_average_Strikerate Worksheet: It consists of the batsman's name, and the total number of runs he has made since 2008 till 2019, Number of times he got dismissed, how many number of balls he has played till 2019, and what is his average and strike rate.

	batsman	total_runs	out	numberofballs	average	strikerate
2	V Kohli	5426	152	4111	35.69736842	131.987351
3	SK Raina	5386	160	3916	33.6625	137.5383044
4	RG Sharma	4902	161	3742	30.44720497	130.9994655
5	DA Warner	4717	114	3292	41.37719298	143.2867558
6	S Dhawan	4601	137	3665	33.58394161	125.5388813
7	CH Gayle	4525	110	2972	41.13636364	152.2543742
8	MS Dhoni	4450	118	3206	37.71186441	138.8022458
9	RV Uthappa	4420	156	3381	28.33333333	130.7305531
10	AB de Villiers	4414	104	2902	42.44230769	152.1019986
11	G Gambhir	4219	134	3400	31.48507463	124.0882353
12	AM Rahane	3834	117	3133	32.76923077	122.3747207
13	KD Karthik	3669	138	2813	26.58695652	130.4301458
14	SR Watson	3590	115	2566	31.2173913	139.9064692
15	AT Rayudu	3313	114	2616	29.06140351	126.6437309
16	YK Pathan	3222	110	2240	29.29090909	143.8392857
17	BB McCullum	2886	106	2185	27.22641509	132.0823799
18	PA Patel	2864	127	2352	22.5511811	121.7687075
19	MK Pandey	2855	95	2352	30.05263158	121.3860544
20	KA Pollard	2772	102	1879	27.17647059	147.5252794
21	Yuvraj Singh	2755	109	2121	25.27522936	129.8915606
22	V Sehwag	2728	98	1746	27.83673469	156.2428408
23	M Vijay	2589	93	2104	27.83870968	123.0513308
24	SE Marsh	2489	65	1866	38.29230769	133.3869239
25	JH Kallis	2427	89	2219	27.26966292	109.3735917
26	DR Smith	2385	81	1758	29.44444444	135.665529
27	SR Tendulkar	2334	71	1943	32.87323944	120.1235203
28	SV Samson	2215	79	1693	28.03797468	130.8328411
29	R Dravid	2174	77	1877	28.23376623	115.823122
30	AC Gilchrist	2069	76	1492	27.22368421	138.6729223
31	SPD Smith	2038	59	1570	34.54237288	129.8089172
32	JP Duminy	2031	49	1633	41.44897959	124.3723209
33	KL Rahul	1988	49	1428	40.57142857	139.2156863

Sheet tabs: Players | Teams | Matches | Deliveries | **Most_runs_average_Strikerate** | teamwise_home_and_away | M...

- Teamwise_home_and_away Worksheet: It consists of the details of how many times a team played in its hometown and how many times it played away. It also consists of the total number of home and away win and the home and away win percentage.
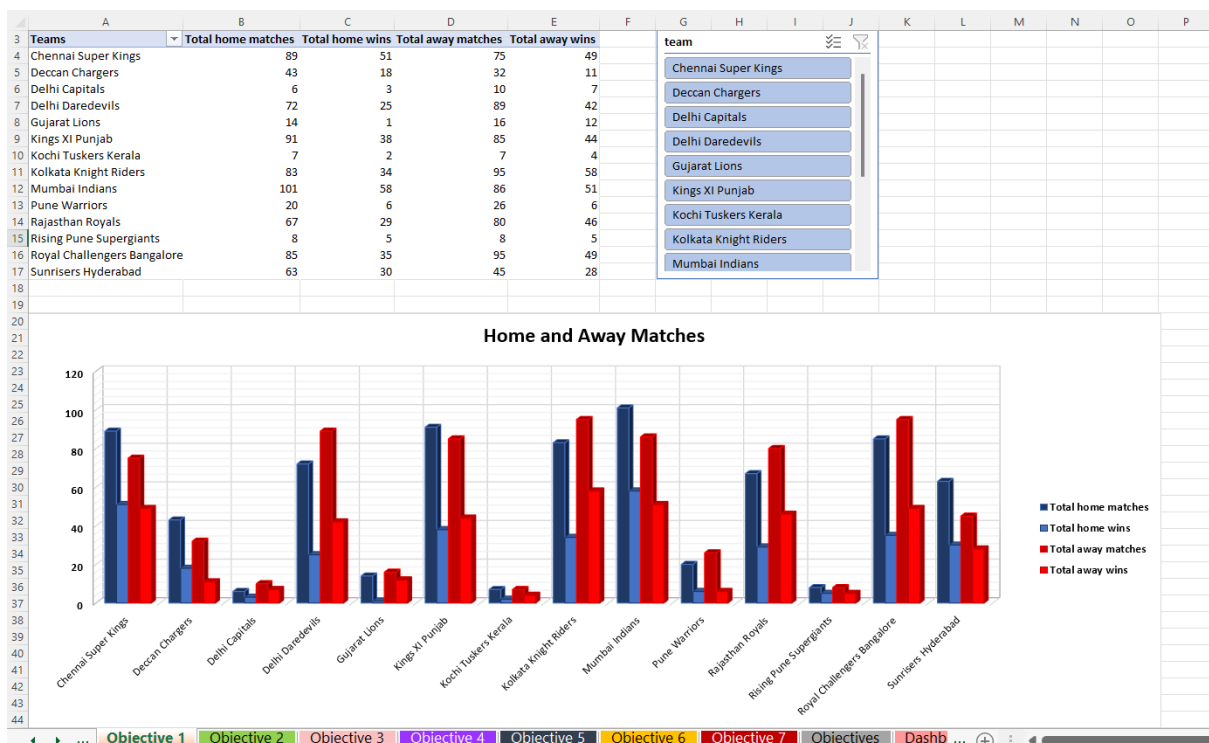
|	team	home_wins	away_wins	home_matches	away_matches	home_win_percentage	away_win_percentage	Team 1 "position"	Team 2 "position"	Total Matches Played	Total Matches Won
2	Rising Pune Supergiants	5	5	8	8	62.5	62.5	15	15	30	15
3	Mumbai Indians	58	51	101	86	57.42574257	59.30232558	101	86	187	109
4	Chennai Super Kings	51	49	89	75	57.30337079	65.33333333	89	75	164	100
5	Delhi Capitals	3	7	6	10	50	70	6	10	16	10
6	Sunrisers Hyderabad	30	28	63	45	47.61904762	62.22222222	63	45	108	58
7	Rajasthan Royals	29	46	67	80	43.28358209	57.5	67	80	147	75
8	Deccan Chargers	18	11	43	32	41.86046512	34.375	43	32	75	29
9	Kings XI Punjab	38	44	91	85	41.75824176	51.76470588	91	85	176	82
10	Royal Challengers Bangalore	35	49	85	95	41.17647059	51.57894737	85	95	180	84
11	Kolkata Knight Riders	34	58	83	95	40.96385542	61.05263158	83	95	178	92
12	Delhi Daredevils	25	42	72	89	34.72222222	47.19101124	72	89	161	67
13	Pune Warriors	6	6	20	26	30	23.07692308	20	26	46	12
14	Kochi Tuskers Kerala	2	4	7	7	28.57142857	57.14285714	7	7	14	6
15	Gujarat Lions	1	12	14	16	7.142857143	75	14	16	30	13

Objective 1: Analyse the ratio of home win and away win.

My first objective is about the ratio of home win and away win for each team in IPL. In this I have calculated the total number of home matches played by a respective team and the number of times that team won in its home ground. I have also calculated the total number of away matches played by a team and the number of wins in away matches. I have displayed the data using 3-D clustered column chart type and have also used a slicer to display the results of a selected team.
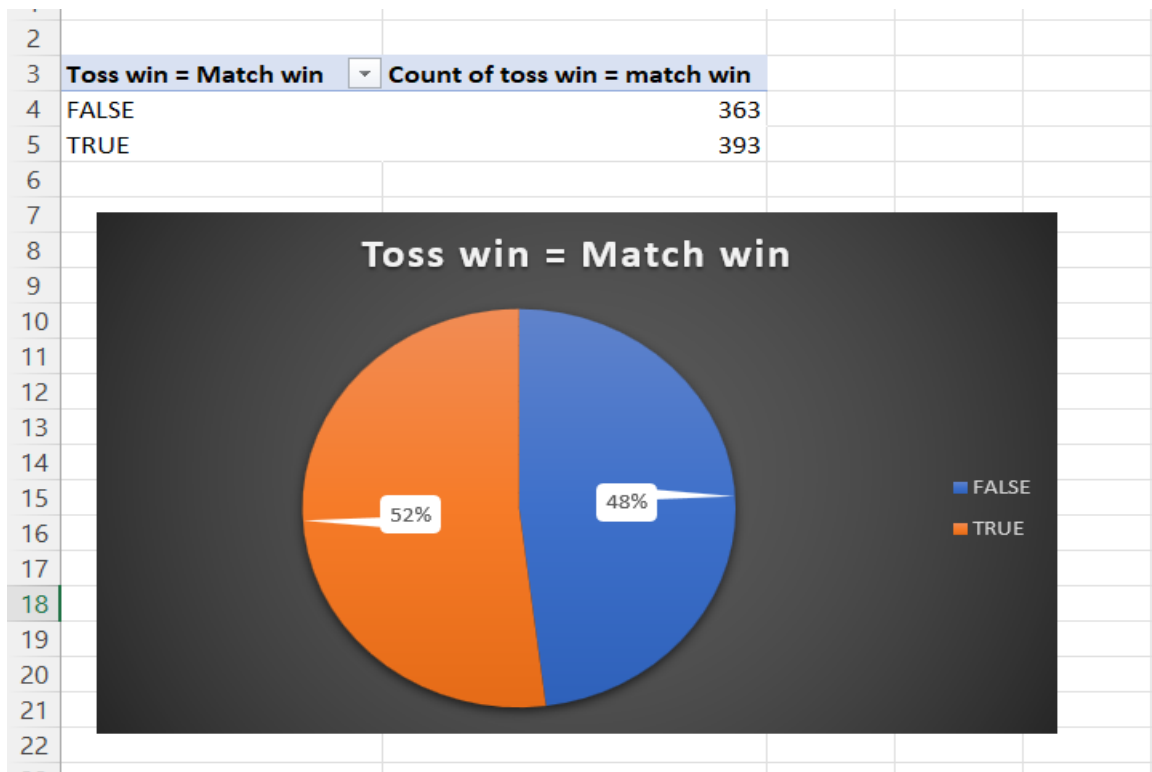
Teams	Total home matches	Total home wins	Total away matches	Total away wins
Chennai Super Kings	89	51	75	49
Deccan Chargers	43	18	32	11
Delhi Capitals	6	3	10	7
Delhi Daredevils	72	25	89	42
Gujarat Lions	14	1	16	12
Kings XI Punjab	91	38	85	44
Kochi Tuskers Kerala	7	2	7	4
Kolkata Knight Riders	83	34	95	58
Mumbai Indians	101	58	86	51
Pune Warriors	20	6	26	6
Rajasthan Royals	67	29	80	46
Rising Pune Supergiants	8	5	8	5
Royal Challengers Bangalore	85	35	95	49
Sunrisers Hyderabad	63	30	45	28

team

- Chennai Super Kings
- Deccan Chargers
- Delhi Capitals
- Delhi Daredevils
- Gujarat Lions
- Kings XI Punjab
- Kochi Tuskers Kerala
- Kolkata Knight Riders
- Mumbai Indians

Home and Away Matches

Objective 2: Top 10 "Powe Hitter" batsman of the IPL with their strike rate and average.

This objective filter outs the top 10 "Power Hitter" batsman of the IPL by analysing the total number of runs made throughout, the number of balls played, number of times a player got dismissed and the average and strike rate of a player. For displaying the results, I have used Combo chart having Clustered column and marked line graph.
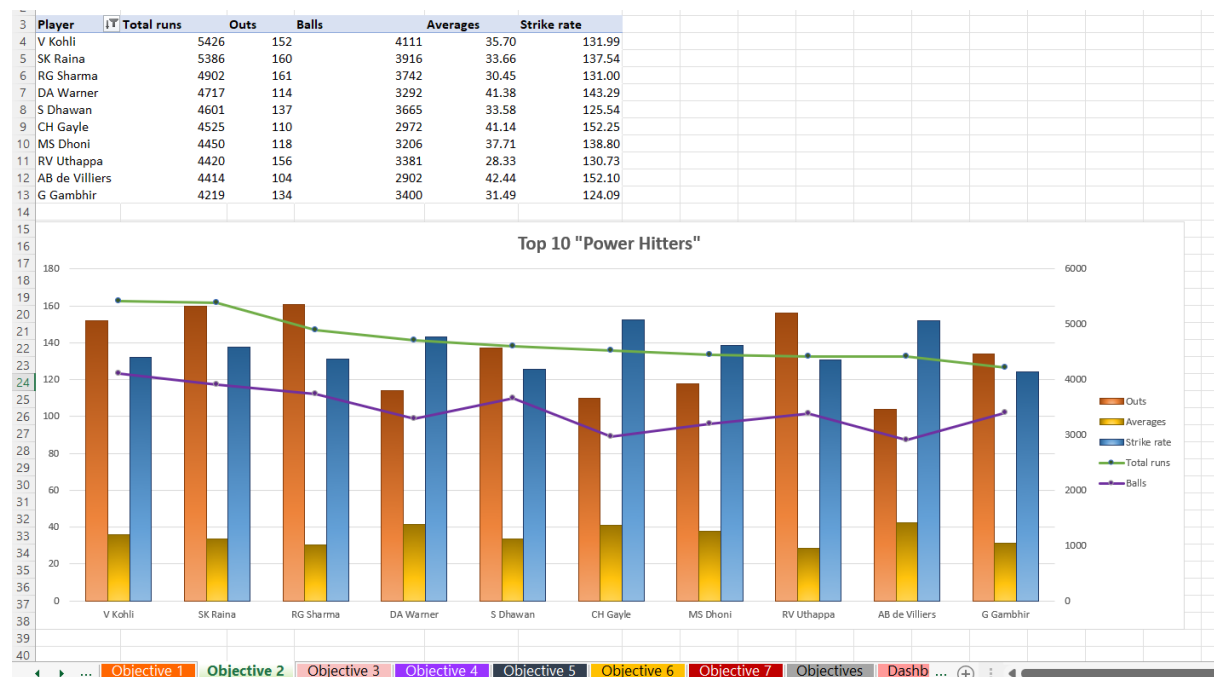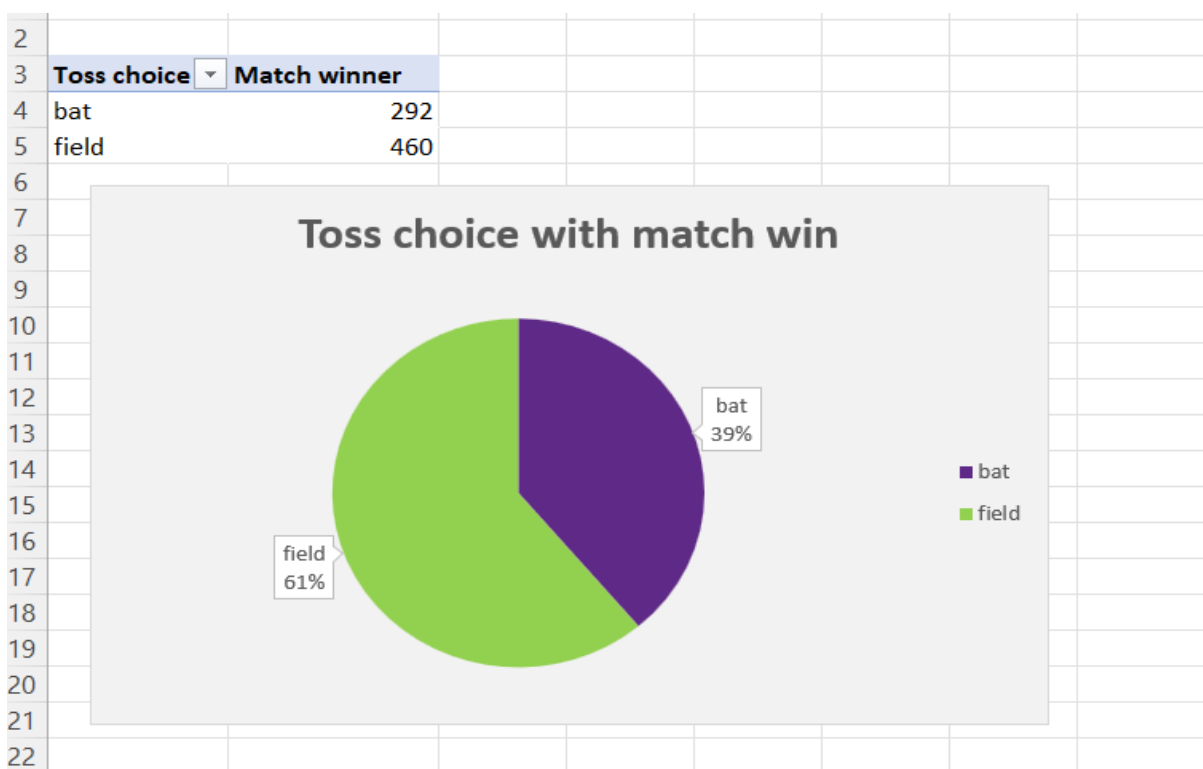

Objective 3: Analyse the dependency of Toss win on Match win.

This objective analyses the dependency of Toss win on Match win. How much the winning of toss and then deciding whether to bowl or bat first influences the chances of match winning for a team is depicted using a pie chart. It shows that 52% matches were won by the team who won the toss. I have displayed the results using a Pie chart.

	Toss win = Match win	▼	Count of toss win = match win
2			
3	**Toss win = Match win**		**Count of toss win = match win**
4	FALSE		363
5	TRUE		393

Toss win = Match win

52%   48%

■ FALSE
■ TRUE

Objective 4: Analyse the respected winning team by making toss decision as bat or field. These objective analyses that after winning a toss how many times a team won the match by making the decision of fielding first or to bat first. 61% times a team won by deciding to field or bowl first.

	Toss choice	▼	Match winner
2			
3	**Toss choice**		**Match winner**
4	bat		292
5	field		460

Toss choice with match win

bat
39%

field
61%

■ bat
■ field

Objective 5: Total number of wickets in each over in IPL.

This objective analyses the number of wickets in each over in IPL. It counts the number of players dismissed per over. It tells us that in the first over the number of players that get dismissed is least as compared to the maximum number of wickets in the last i.e., 20th over. I have used Clustered column chart type to display the result from first to last over.
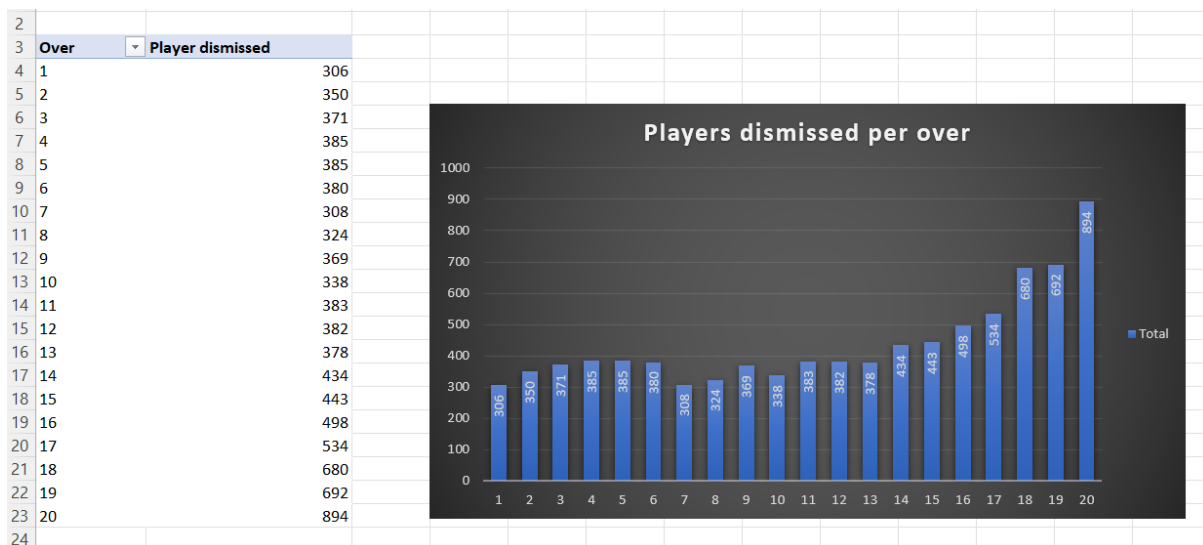
Objective 6: Total matches won and played by each team.

This objective counts the total number of matches played and won by each team in IPL from 2008-2019. I have taken the data from teamwise_home_and_away table. For displaying the result, I have used Clustered column which shows the total number of matches played by a team and the number of matches won by that team.
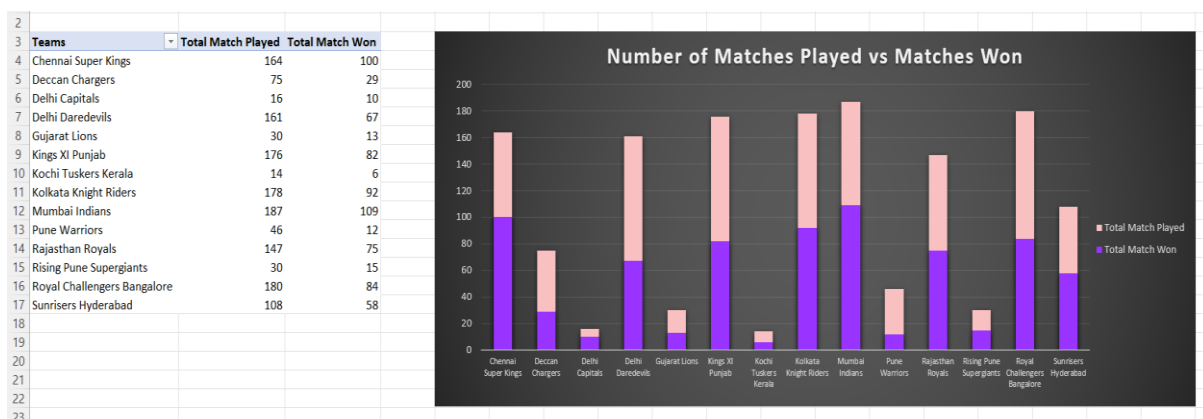
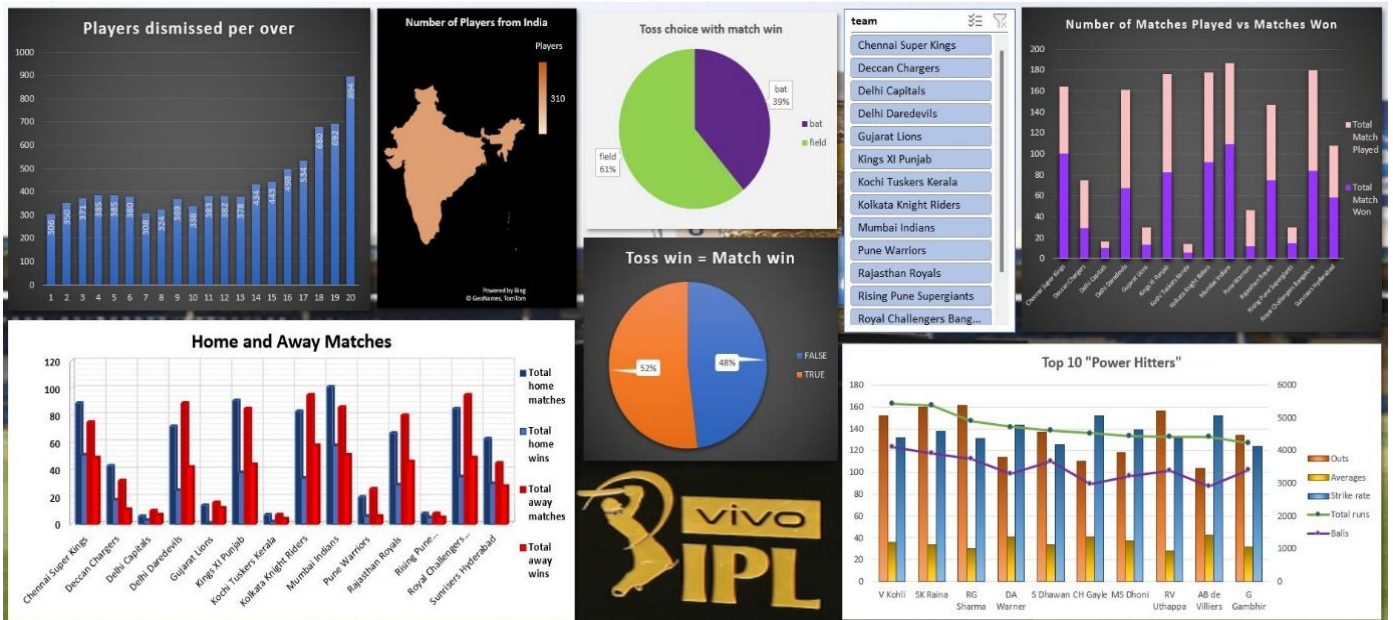**Objective 7:** Total no. of players from India.

In this objective I have analysed the total number of players which are from India. For this I have used the chart type Filled Map. There are a total of 310 Indian players in different IPL teams.
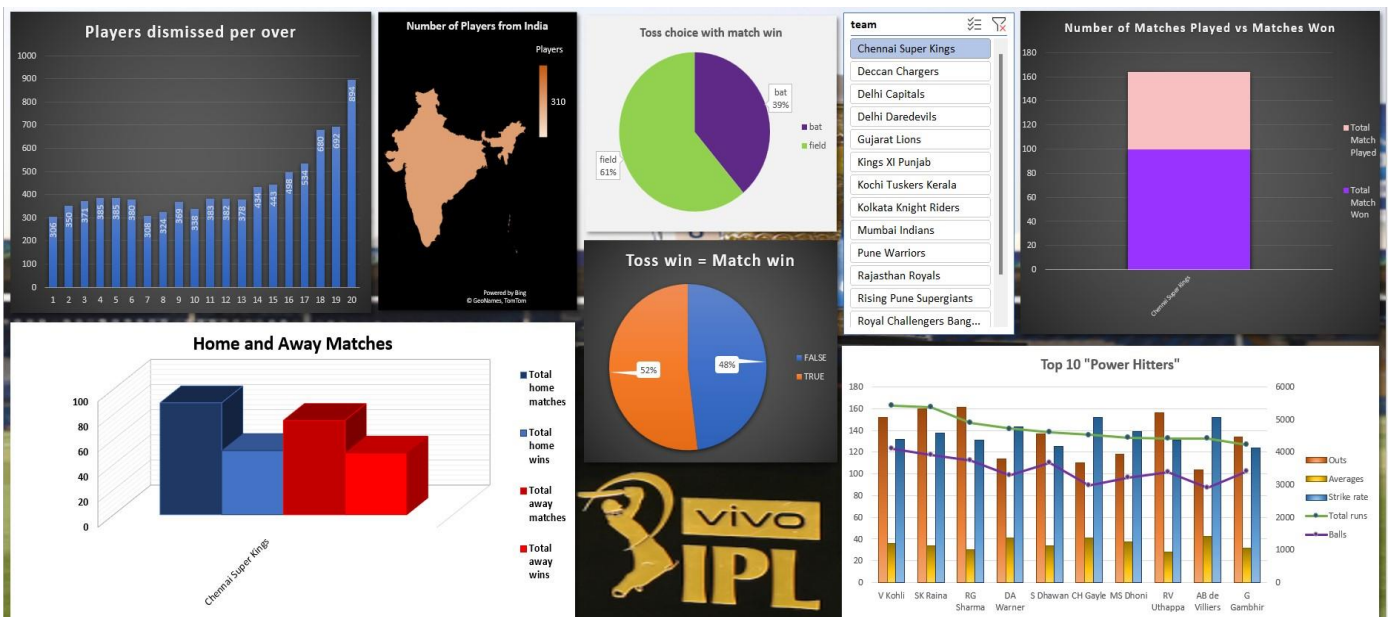
- View of Dashboard Worksheet

- After clicking the Slicer and Selecting a Team say, "Chennai Super Kings". Data of "Chennai Super Kings" is shown for "Home and Away Matches 3-D Column Chart" and "Number of Matches Played vs Matches won Stacked Column Chart".

CONCLUSION

Overall, this project was a useful experience. I have gained some useful knowledge about data analysis and MS Excel. I was able to conclude some insights from IPL dataset . I analysed information from the dataset like the ratio of home win and away win, Top 10 "Powe Hitter" batsman of the IPL with their strike rate and average, dependency of toss win on match win, respected winning team by making toss decision as bat or field, Over-wise wickets, total matches played and won by each team and number of players from India.

I also learned scope of using MS Excel its formulas and features and practically applying them to analyse information from a dataset. I worked with Pivot table to analyse data from worksheets. I explored different types of charts and maps, and design them to give a presentable representation of analysed information. I used ETL process make proper use of the dataset by cleaning and organising it. It helped to develop data management and analysis skills.

I represented all the analysed insights on a dashboard and used slicers to filter the data get personalised extracted information from the dataset. The dashboard offered a common, user friendly and interactive place for a good user interface.

In conclusion this was helpful, and I got to practice a lot of the data management and analysis techniques and skills and got a good hands-on experience on the MS Excel software, explored workbooks, worksheets, formulas, features like, Pivot tables, charts, etc. In this data driven world, I am well-verse basics of data analysis and MS Excel software which will help in the Data Science minor. It strengthened my professional ability to work with datasets.

REFERENCES

- **Kaggle IPL_data_set:**

    https://www.kaggle.com/ramjidoolla/ipl-data-set

- http://radar.oreilly.com/2010/12/quick-starts-for-charts.html

- https://www.tutorialspoint.com/excel_charts/excel_charts_types.htm

- https://www.geeksforgeeks.org/etl-process-in-data-warehouse/

- https://www.javatpoint.com/etl-process-in-data-warehouse

- https://databricks.com/glossary/extract-transform-load

- https://www.talend.com/resources/what-is-etl/

- https://www.javatpoint.com/excel-tutorial