INT-217

INTRODUCTION TO DATA MANAGEMENT

Project Report

IPL DATA SET (2008-2019)

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

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

Submitted to:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

LOVELY PROFESSIONAL UNIVERSITY

PHAGWARA, PUNJAB

Dated: - December 10, 2021

Submitted by:

Name of the student: Pratik Suhag

Registration Number: 11902729

Signature of the student: 

STUDENT DECLARATION

To whom so ever it may concern

I, **Pratik Suhag, 11902729,** hereby declare that the work done by me on **"Project on IPL DATA SET (2008-2019)"**, 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): Pratik Suhag (11902729)

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

Sr. No.	Topic		Page No.
1	INTRODUCTION OF THE TOPIC		5
	1.1	MICROSOFT EXCEL	5
	1.2	CELL, WORKSHEET AND WORKBOOK	6
	1.3	FEATURES OF MS EXCEL	7
2	TYPES OF CHARTS IN MS EXCEL		9
3	FORMULAS USED		15
3	OBJECTIVES		16
5	SOURCE OF DATASET		17
6	ETL PROCESS		18
7	ANALYSIS ON DATASET		22
8	OBJECTIVE DETAILS		26
9	DASHBOARD		32
10	CONCLUSION		34
11	REFERENCES		35

INTRODUCTION OF THE PROJECT

Microsoft Excel

Microsoft Excel is a spreadsheet software by Microsoft. It comes with Office Suite with several other Microsoft applications, such as Word, PowerPoint, Access, Outlook, and OneNote, etc.

It is an electronic spreadsheet with numerous rows and columns, used for organizing data, graphically represent data(s), and performing different calculations.

It is the industry leading spreadsheet program, used by millions of people all over the world. Excel and other spreadsheet tools are great for data manipulation, analysis, and visualization – you can sort, filter, format, and chart your data all within one program.

Since Excel stores data in tables, it has several rows and columns. By using this way, the data is well organized. Thus, we can find the desired information easily with or without Excel tools.

The first version was released in 1985 and has gone through several changes over the years. However, the main functionality mostly remains the same.

It consists of 1048576 rows and 16383 columns; a row and column together make a cell. Each cell has an address defined by column name and row number example A1, D2, etc. this is also known as a cell reference.

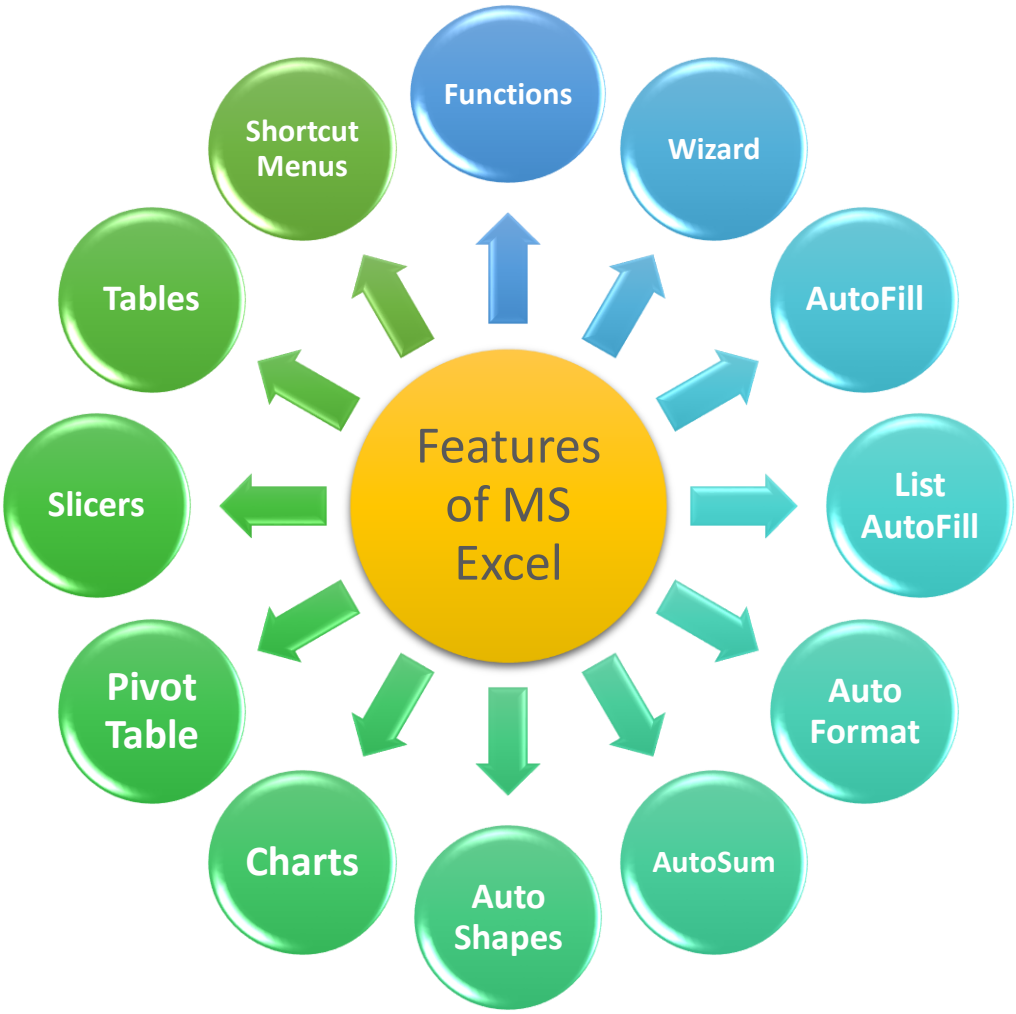
Excel provides us the worksheet to create a new document in it. You can save the Excel file with .xlsx or .xls extension.

Cell, Worksheet and Workbook

Three most important components of Excel are:

- **Cell:** A cell is a smallest but most powerful part of a spreadsheet. You can enter your data into a cell either by typing or by copy-paste. Data can be a text, a number, or a date. You can also customize it by changing its size, font color, background color, borders, etc. Every cell is identified by its cell address, cell address contains its column number and row number (If a cell is on 11th row and on column AB, then its address will be AB11).
- **Worksheet:** A worksheet is made up of individual cells which can contain a value, a formula, or text. It also has an invisible draw layer, which holds charts, images, and diagrams. Each worksheet in a workbook is accessible by clicking the tab at the bottom of the workbook window. In addition, a workbook can store chart sheets; a chart sheet displays a single chart and is accessible by clicking a tab.
- **Workbook:** A workbook is a separate file just like every other application has. Each workbook contains one or more worksheets. You can also say that a workbook is a collection of multiple worksheets or can be a single worksheet. You can add or delete worksheets, hide them within the workbook without deleting them, and change the order of your worksheets within the workbook.

Features of MS Excel

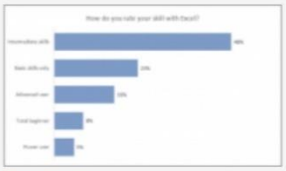


- **Functions:** There are more than 300 built-in formulas to use in the Excel Cells. Verity of the formulas (Text, Date, String, Maths, etc) will help to perform verity of calculations.
- **Wizard:** It guides us to work effectively while we work by displaying several helpful tips and techniques based 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.
- **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.

- **List AutoFill:** It automatically develops cell formatting when a new component is added to the end of a list.
- **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.
- **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.
- **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.
- **Slicers:** Slicers are introduced in Excel 2010; this will help us to connect the multiple pivot tables and filter the data with buttons.
- **Tables:** We can create the tables in the data in records and fields format. This will be helpful to quickly perform further analysis.
- **Shortcut Menus:** The shortcut menu helps users to make the work done through shortcut commands that need a lengthy process.

TYPES OF CHARTS IN MS EXCEL


Bar chart


Column chart


Stacked Column Chart

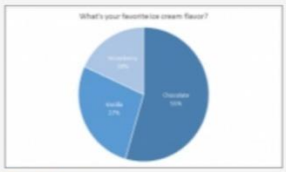

Clustered Column Chart

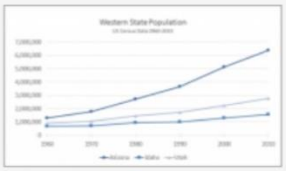

100% Stacked Column Chart

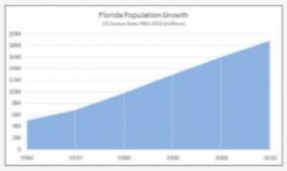

100% Stacked Bar Chart

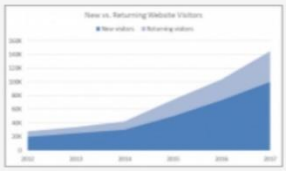

Stacked Bar Chart


Clustered Bar Chart

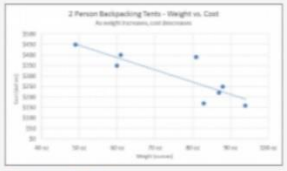

Pie Chart

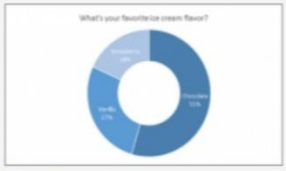

Line Chart

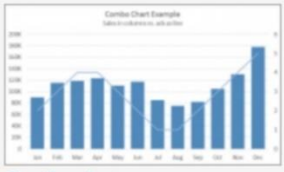

Area Chart

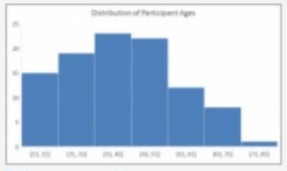

Stacked Area Chart


100% Stacked Area Chart


Scatter Plot


Doughnut Chart


Combo chart


Histogram Chart


Radar chart

Excel offers the following major chart types:

Column Chart

A column chart is a primary Excel chart type, with data series plotted using vertical columns. Column charts are a good way to show change over time because it's easy to compare column lengths. Like bar charts, column charts can be used to plot both nominal data and ordinal data, and they can be used instead of a pie chart to plot data with a part-to-whole relationship. Column charts work best where data points are limited (like, 12 months, etc.). With more data points, you can switch to a line graph. 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

A line chart is a built-in Excel chart type, with each data series plotted as a separate line. Line charts are a good way to show change or trends over time. In contrast to column or bar charts, line charts can handle more categories and more data points without becoming too cluttered. Line charts can be customized to show or hide data markers of various shapes and sizes.

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

Bar Chart

A bar chart is one of Excel's primary chart types and a good choice for categorical data. Bar charts plot data using horizontal bars, so they are very easy to read because the human eye can easily compare bars. Also, because of the horizontal layout, bar charts have room to accommodate longer category names.

Bar charts are also versatile. They can be used to plot both nominal data and ordinal data, and they can be used instead of a pie chart to plot data with a part-to-whole relationship. If you're trying to decide on a chart type, a bar chart is a good first choice.

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

Pie Chart

The Pie Chart is a primary chart type in Excel. Pie charts are meant to express a "part to whole" relationship, where all pieces together represent 100%. Pie charts work best to display data with a small number of categories (2-5). For example, survey questions in yes/no format, data split by gender (male/female), new and returning visitors to a website, etc.

Pie charts should be avoided when there are many categories, or when categories do not total 100%. The human eye has trouble comparing the relative size of slices in a pie chart, so pie charts should also be avoided when slices are similar, unless similarity is the point.

A Pie Chart has the following sub-types −

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

Doughnut Chart

The Doughnut Chart is a built-in chart type in Excel. Doughnut charts are meant to express a "part-to-whole" relationship, where all pieces together represent 100%. Doughnut charts work best to display data with a small number of categories. For example, you could use a doughnut chart to plot survey questions with a small number of answers, data split by gender, Windows vs. Mac users, or other data where categories are limited. Doughnut charts should be avoided when there are many categories, or when categories do not sum to 100%.

Radar Chart

The Radar Chart is a built-in chart type in Excel. Radar charts, sometimes called spider charts, have one axis per category which all use the same scale. The axes of a radar chart radiate out from the centre of the chart and data points are plotted on each axis using a common scale. The result is a geometric shape that shows "at-a-glance" performance across all categories.

Radar charts can be used to plot the performance of employees, athletes, products, and companies in various categories. They can be used for performance evaluations and satisfaction surveys.

A Radar chart has the following sub-types −

- Radar
- Radar with Markers
- Filled Radar

Area Chart

An area chart is a primary Excel chart type, with data series plotted using lines with a filled area below. Area charts are a good way to show change over time with one data series. They offer a simple presentation that is easy to interpret briefly.

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

A scatter plot is a built-in chart type in Excel meant to show the relationship between two variables. A scatter plot works by placing one variable on the vertical axis and a different variable on the horizontal axis. Each piece of data is then plotted as a discrete point on the chart. In a scatter plot, both the X and Y axis display values – an XY chart has no category axis.

By convention, the X axis represents arbitrary values that do not depend on another variable, referred to as the independent variable. Y values are placed on the vertical axis and represent the dependent variable.

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

The Bubble Chart is a built-in chart type in Excel. Bubble charts are a special kind of XY chart that can display another data series which is used to scale the bubble (marker) plotted at X and Y values. You can think of a bubble chart as "X versus Y, scaled by Z". Like a regular XY scatter chart, both axes are used to plot values – there is no category axis.

A Bubble chart has the following sub-types −

- Bubble
- Bubble with 3-D effect

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

FORMULAS USED

- **IF**

  The IF function runs a logical test and returns one value for a TRUE result, and another for a FALSE result.

  Syntax: —

  =IF (logical_test, [value_if_true], [value_if_false])

- **SORT**

  The Excel SORT function sorts the contents of a range or array in ascending or descending order. Values can be sorted by one or more columns. SORT returns a dynamic array of results.

  Syntax: —

  =SORT (array, [sort_index], [sort_order], [by_col])

- **COUNTIF**

  The COUNTIF function is one of the statistical functions, to count the number of cells from a range that meet a criterion.

  Syntax: —

  =COUNTIF (range, criteria)

- **UNIQUE**

  The Excel UNIQUE function returns a list of unique values in a list or range. Values can be text, numbers, dates, times, etc.

  Syntax: —

  =UNIQUE (array, [by_col], [exactly_once])

OBJECTIVES

- Top 15 batsman of the IPL with their strike rate, average and number of times dismissed.
- Analysis of Match win dependency on Toss win.
- Analyse Total Runs scored in each over in IPL.
- Analyse the Home and Away - Matches played and Matches won.
- Total Matches Played and Won by each team.
- Analyse Players on basis of their Country and Batting Type

SOURCE OF DATASET

- **Kaggle (IPL Dataset 2008-2019):**

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

ETL PROCESS

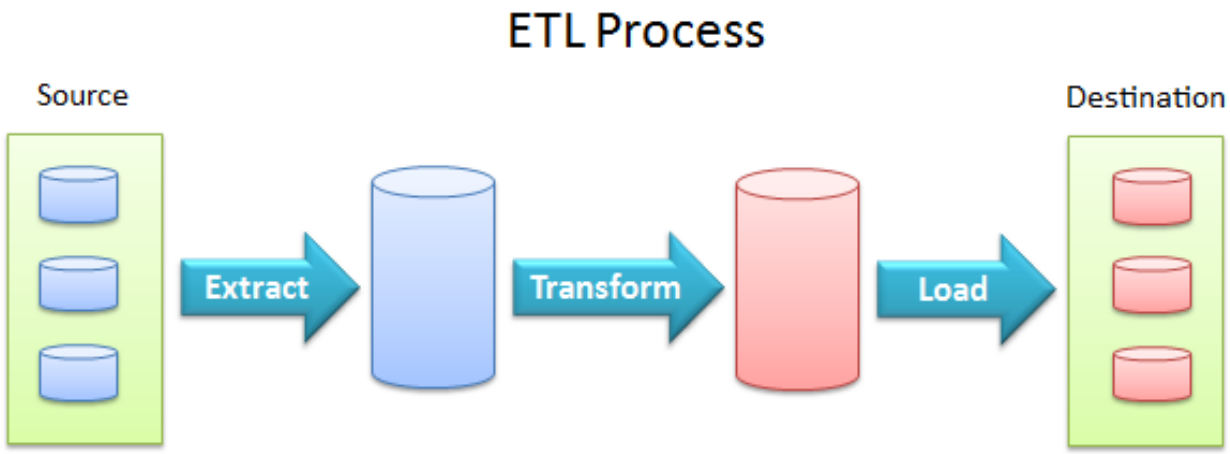
ETL stands for Extract Transform and Load. Just like the name applies ETL tool Extracts data from the source. Transforms the data while in transit and then it loads the data in to Specified database.

The mechanism of extracting information from source systems and bringing it into the data warehouse is commonly called ETL.

The ETL process requires inputs from various sources, which can be anything like, developers, analysts, testers, top executives and is technically challenging.

Data from sources can homogenous or heterogenous, and from a single source or a combination of multiple sources.

There are three steps in ETL process which enable data to be integrated from source to destination. These are data extraction, data transformation, and data loading.



- **Extraction**

  Extraction is the first step of ETL. The data sets are extracted from a source into a staging area. Raw data can be extracted from a wide range of sources.

  Structured and unstructured data is imported and consolidated into a single repository.

  In the project the input of dataset i.e., Extraction is taken from Kaggle dataset named IPL_data_set which include raw data in from of worksheets and .csv files. This raw data included players, teams, matches, deliveries, most_runs_average_Strikerate, teamwise_home_and_away from 2008 to 2019.

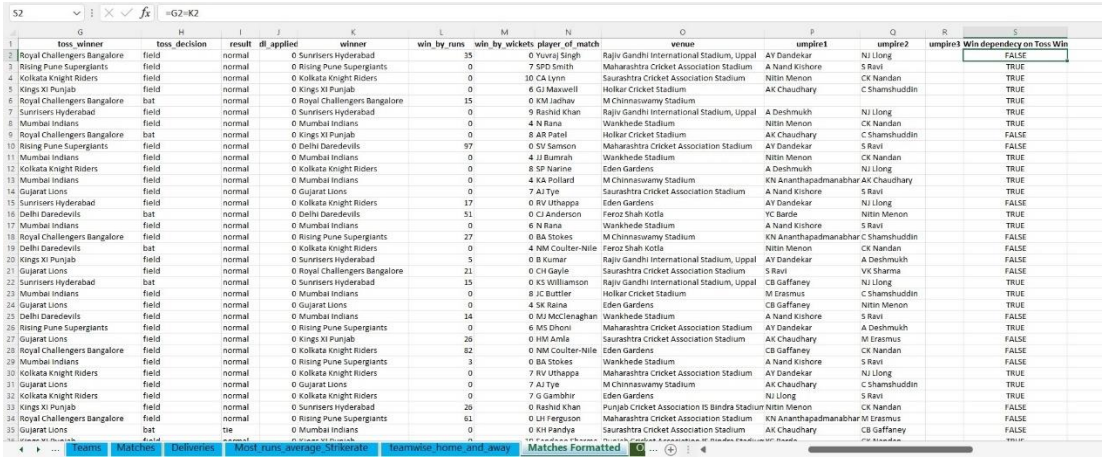
- **Transformation**

  The Transformation of raw data into final data is done by cleaning and organizing. All that data from multiple source systems will be normalized and converted to a single system format. This helps in improving data quality and compliance.

  ETL yields transformed data through these methods:

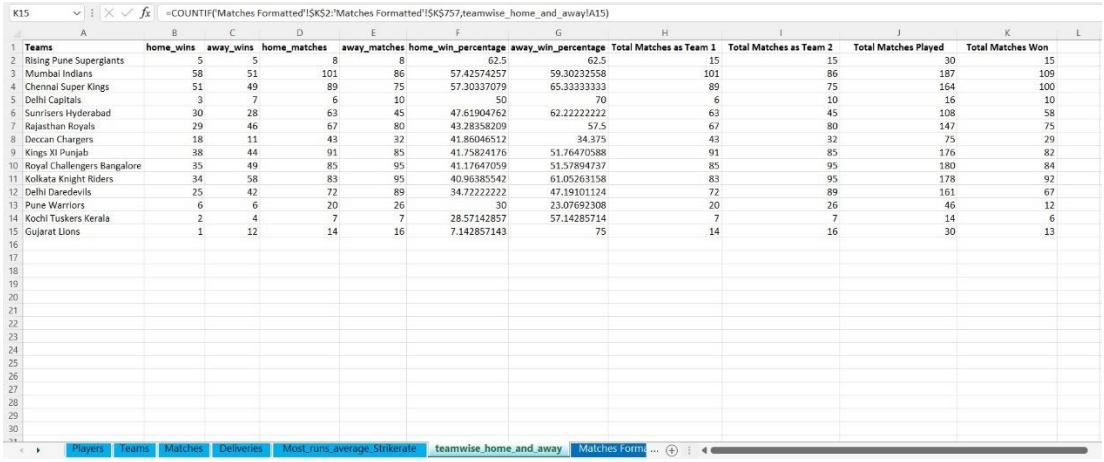
  - Standardization: - formatting rule is applied to the data set.
  - Verification: - unusable data is removed, and anomalies are flagged.
  - Cleansing: - inconsistencies and missing values in the data are resolved.
  - Deduplication: - redundant data is excluded or discarded.
  - Sorting: - data is organized according to type.
  - Other tasks: - any additional/optional rules can be applied to improve data quality.

Transformation is main and the most important 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.



	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire2	umpire3	Win dependecy on Toss Win
2	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	NJ Llong		FALSE
3	Rising Pune Supergiants	field	normal	0	Rising Pune Supergiants	0	7	SPD Smith	Maharashtra Cricket Association Stadium	A Nand Kishore	S Ravi		TRUE
4	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon	CK Nandan		TRUE
5	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	GJ Maxwell	Holkar Cricket Stadium	AK Chaudhary	C Shamshuddin		TRUE
6	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium				TRUE
7	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	0	9	Rashid Khan	Rajiv Gandhi International Stadium, Uppal	A Deshmukh	NJ Llong		TRUE
8	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	N Rana	Wankhede Stadium	Nitin Menon	CK Nandan		TRUE
9	Royal Challengers Bangalore	bat	normal	0	Kings XI Punjab	0	8	AR Patel	Holkar Cricket Stadium	AK Chaudhary	C Shamshuddin		FALSE
10	Rising Pune Supergiants	field	normal	0	Delhi Daredevils	97	0	SV Samson	Maharashtra Cricket Association Stadium	AY Dandekar	S Ravi		FALSE
11	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	JJ Bumrah	Wankhede Stadium	Nitin Menon	CK Nandan		TRUE
12	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	8	SP Narine	Eden Gardens	A Deshmukh	NJ Llong		TRUE
13	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	KA Pollard	M Chinnaswamy Stadium	KN Ananthapadmanabhar	AK Chaudhary		TRUE
14	Gujarat Lions	field	normal	0	Gujarat Lions	0	7	AJ Tye	Saurashtra Cricket Association Stadium	A Nand Kishore	S Ravi		TRUE
15	Sunrisers Hyderabad	field	normal	0	Kolkata Knight Riders	17	0	RV Uthappa	Eden Gardens	AY Dandekar	NJ Llong		FALSE
16	Delhi Daredevils	bat	normal	0	Delhi Daredevils	51	0	CJ Anderson	Feroz Shah Kotla	YC Barde	Nitin Menon		TRUE
17	Mumbai Indians	field	normal	0	Mumbai Indians	0	6	N Rana	Wankhede Stadium	A Nand Kishore	S Ravi		TRUE
18	Royal Challengers Bangalore	field	normal	0	Rising Pune Supergiants	27	0	BA Stokes	M Chinnaswamy Stadium	KN Ananthapadmanabhar	C Shamshuddin		FALSE
19	Delhi Daredevils	bat	normal	0	Kolkata Knight Riders	0	4	NM Coulter-Nile	Feroz Shah Kotla	Nitin Menon	CK Nandan		FALSE
20	Kings XI Punjab	field	normal	0	Sunrisers Hyderabad	5	0	B Kumar	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	A Deshmukh		FALSE
21	Gujarat Lions	field	normal	0	Royal Challengers Bangalore	21	0	CH Gayle	Saurashtra Cricket Association Stadium	S Ravi	VK Sharma		FALSE
22	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	15	0	KS Williamson	Rajiv Gandhi International Stadium, Uppal	CB Gaffaney	NJ Llong		TRUE
23	Mumbai Indians	field	normal	0	Mumbai Indians	0	8	JC Buttler	Holkar Cricket Stadium	M Erasmus	C Shamshuddin		TRUE
24	Gujarat Lions	field	normal	0	Gujarat Lions	0	4	SK Raina	Eden Gardens	CB Gaffaney	Nitin Menon		TRUE
25	Delhi Daredevils	field	normal	0	Mumbai Indians	14	0	MJ McClenaghan	Wankhede Stadium	A Nand Kishore	S Ravi		FALSE
26	Rising Pune Supergiants	field	normal	0	Rising Pune Supergiants	0	6	MS Dhoni	Maharashtra Cricket Association Stadium	AY Dandekar	A Deshmukh		TRUE
27	Gujarat Lions	field	normal	0	Kings XI Punjab	26	0	HM Amla	Saurashtra Cricket Association Stadium	AK Chaudhary	M Erasmus		FALSE
28	Royal Challengers Bangalore	field	normal	0	Kolkata Knight Riders	82	0	NM Coulter-Nile	Eden Gardens	CB Gaffaney	CK Nandan		FALSE
29	Mumbai Indians	field	normal	0	Rising Pune Supergiants	3	0	BA Stokes	Wankhede Stadium	A Nand Kishore	S Ravi		FALSE
30	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	7	RV Uthappa	Maharashtra Cricket Association Stadium	AY Dandekar	NJ Llong		TRUE
31	Gujarat Lions	field	normal	0	Gujarat Lions	0	7	AJ Tye	M Chinnaswamy Stadium	AK Chaudhary	C Shamshuddin		TRUE
32	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	7	G Gambhir	Eden Gardens	NJ Llong	S Ravi		TRUE
33	Kings XI Punjab	field	normal	0	Sunrisers Hyderabad	28	0	Rashid Khan	Punjab Cricket Association IS Bindra Stadium	Nitin Menon	CK Nandan		FALSE
34	Royal Challengers Bangalore	field	normal	0	Rising Pune Supergiants	61	0	LH Ferguson	Maharashtra Cricket Association Stadium	KN Ananthapadmanabhar	M Erasmus		FALSE
35	Gujarat Lions	bat	tie	0	Mumbai Indians	0	0	KH Pandya	Saurashtra Cricket Association Stadium	AK Chaudhary	CB Gaffaney		FALSE

In the project, the 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.



	Teams	home_wins	away_wins	home_matches	away_matches	home_win_percentage	away_win_percentage	Total Matches as Team 1	Total Matches as Team 2	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

- **Loading**

  Loading final transformed data into a new destination stage is the last step in the ETL.

  Depending upon the requirements, data can be loaded. Data can be loaded

  - all at once (full load) or
  - scheduled intervals (incremental load).

  The exact nature of the loading will depend upon the data source, ETL tools, and various other factors.

  In the project, data was to be analysed on the PC using Excel, so it was not required to load the data in any data warehouse.
  It was simply transformed with all the required modifications and saved in work area/staging area in excel.

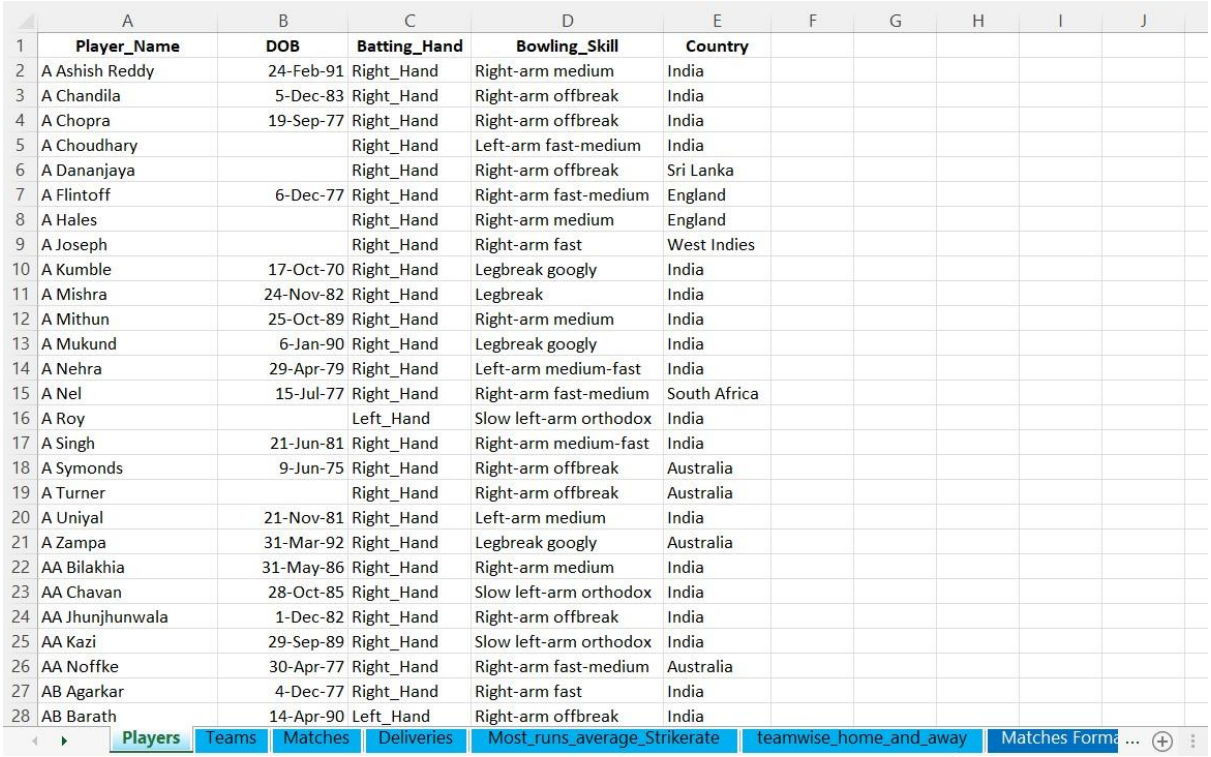
ANALYSIS ON DATASET

Introduction about the Dataset

The selected dataset from Kaggle 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:

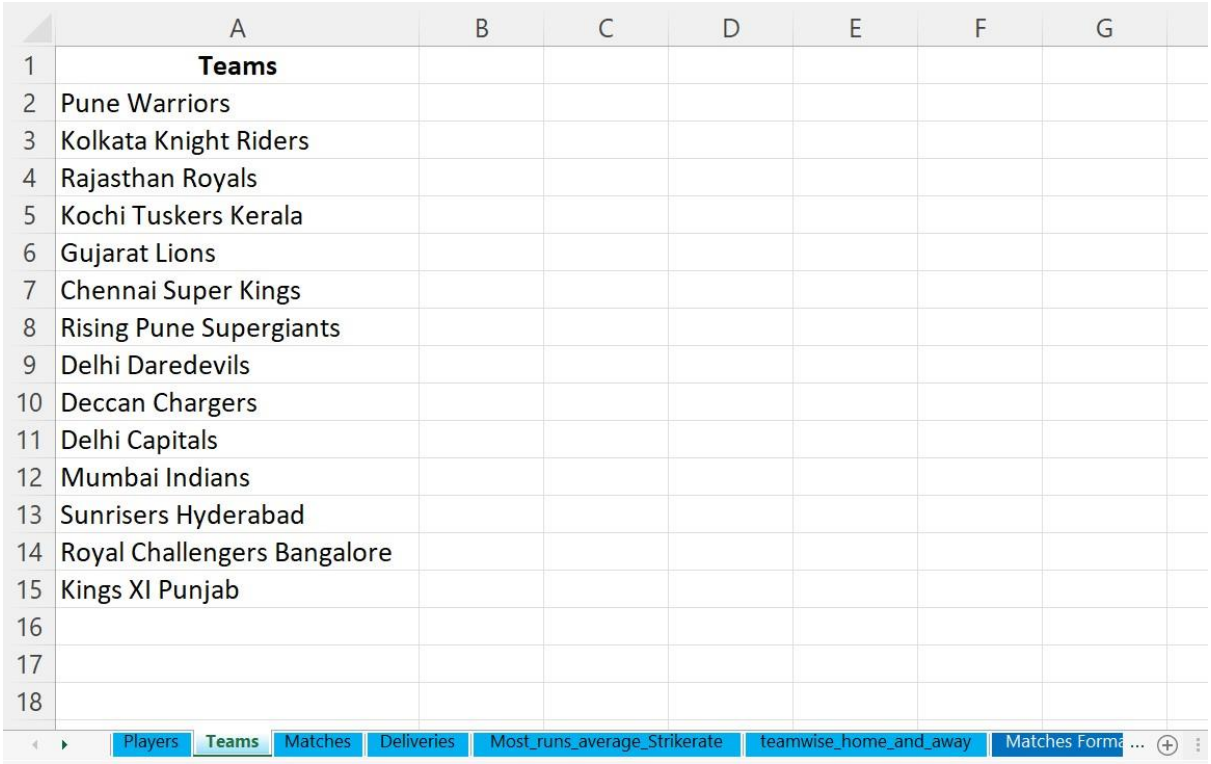
- **Players:** It includes the columns with name of the players, date of birth, batting hand, bowling skills and country.



	A	B	C	D	E
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	5-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	6-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	6-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	9-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	1-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	4-Dec-77	Right_Hand	Right-arm fast	India
28	AB Barath	14-Apr-90	Left_Hand	Right-arm offbreak	India

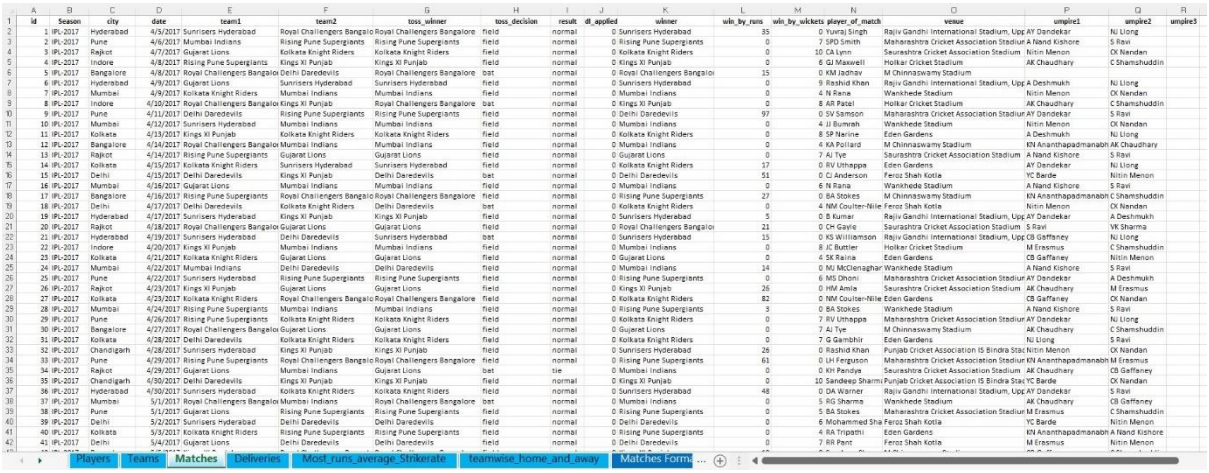
Players | Teams | Matches | Deliveries | Most_runs_average_Strikerate | teamwise_home_and_away | Matches Forma ...

- **Teams:** It includes the names of all the IPL teams.

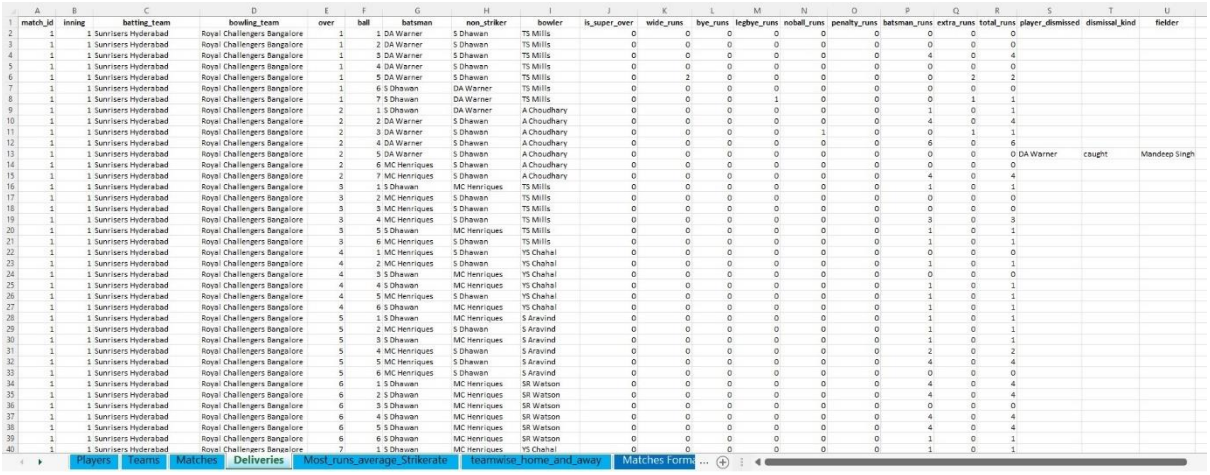


	A
1	**Teams**
2	Pune Warriors
3	Kolkata Knight Riders
4	Rajasthan Royals
5	Kochi Tuskers Kerala
6	Gujarat Lions
7	Chennai Super Kings
8	Rising Pune Supergiants
9	Delhi Daredevils
10	Deccan Chargers
11	Delhi Capitals
12	Mumbai Indians
13	Sunrisers Hyderabad
14	Royal Challengers Bangalore
15	Kings XI Punjab
16	
17	
18	

- **Matches:** This includes columns with the match id, season, city in which the match was played, date of the match played, teams between which the match was played, team winning toss, toss decision, result of the match, Duckworth Lewis applied or not, win by runs or wickets, player of the match, venue of the match and umpires.



- **Deliveries:** This includes of the details of all the deliveries during all seasons from 2008-2019. The match id, inning, batting team, bowling team, over and ball, batsman, non-striker, bowler, whether super over or not, runs and extras, player dismissed and their details.



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	batsman_runs	extra_runs	total_runs	player_dismissed	dismissal_kind	fielder
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	0	0	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	0	0	0			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	0	0	0	0	0	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	0	0	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	0	2	2			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	6	S Dhawan	DA Warner	TS Mills	0	0	0	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	0	1	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	1	S Dhawan	DA Warner	A Choudhary	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	2	DA Warner	S Dhawan	A Choudhary	0	0	0	0	0	0	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	3	DA Warner	S Dhawan	A Choudhary	0	0	0	0	1	0	0	1	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	4	DA Warner	S Dhawan	A Choudhary	0	0	0	0	0	0	6	0	6			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	5	DA Warner	S Dhawan	A Choudhary	0	0	0	0	0	0	0	0	0	DA Warner	caught	Mandeep Singh
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	2	6	MC Henriques	S Dhawan	A Choudhary	0	0	0	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	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	1	S Dhawan	MC Henriques	TS Mills	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	2	MC Henriques	S Dhawan	TS Mills	0	0	0	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	0	0	0			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	4	MC Henriques	S Dhawan	TS Mills	0	0	0	0	0	0	3	0	3			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	5	S Dhawan	MC Henriques	TS Mills	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	3	6	MC Henriques	S Dhawan	TS Mills	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	1	MC Henriques	S Dhawan	YS Chahal	0	0	0	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	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	3	S Dhawan	MC Henriques	YS Chahal	0	0	0	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	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	5	MC Henriques	S Dhawan	YS Chahal	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	4	6	S Dhawan	MC Henriques	YS Chahal	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	1	S Dhawan	MC Henriques	S Aravind	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	2	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	3	S Dhawan	MC Henriques	S Aravind	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	4	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0	2	0	2			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	5	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	5	6	MC Henriques	S Dhawan	S Aravind	0	0	0	0	0	0	0	0	0			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	6	1	S Dhawan	MC Henriques	SR Watson	0	0	0	0	0	0	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	6	2	S Dhawan	MC Henriques	SR Watson	0	0	0	0	0	0	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	6	3	S Dhawan	MC Henriques	SR Watson	0	0	0	0	0	0	0	0	0			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	6	4	S Dhawan	MC Henriques	SR Watson	0	0	0	0	0	0	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	6	5	S Dhawan	MC Henriques	SR Watson	0	0	0	0	0	0	4	0	4			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	6	6	S Dhawan	MC Henriques	SR Watson	0	0	0	0	0	0	1	0	1			
1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	7	1	S Dhawan	MC Henriques	YS Chahal	0	0	0	0	0	0	1	0	1			

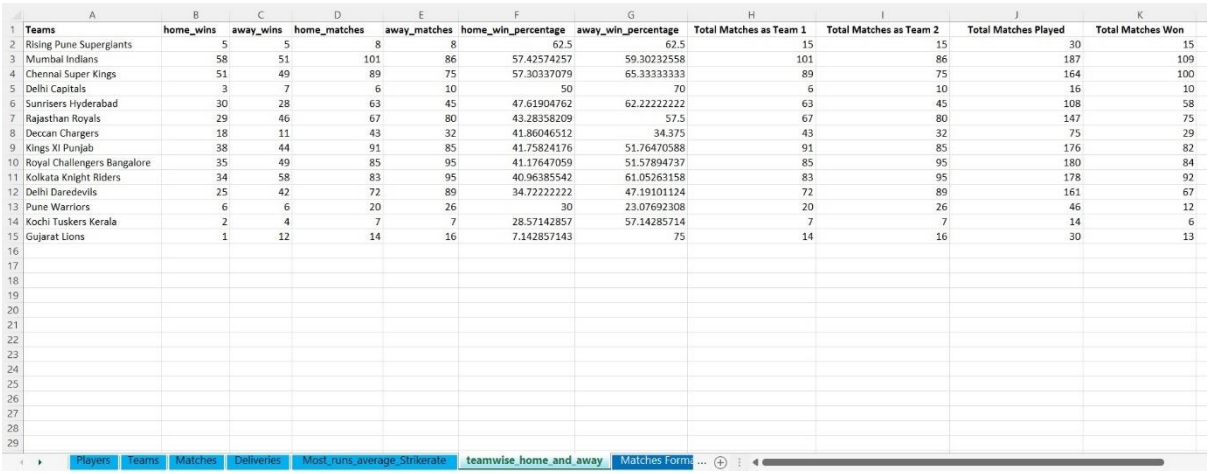
Players | Teams | Matches | Deliveries | Most_runs_average_Strikerate | teamwise_home_and_away | Matches Form

- **Most_runs_average_Strikerate:** It includes the batsman's name, total runs, number of times dismissed, number of balls he has played, average and strike rate of the batsman.

batsman	total_runs	out	numberofballs	average	strikerate
V Kohli	5426	152	4111	35.69736842	131.987351
SK Raina	5386	160	3916	33.6625	137.5383044
RG Sharma	4902	161	3742	30.44720497	130.9994655
DA Warner	4717	114	3292	41.37719298	143.2867558
S Dhawan	4601	137	3665	33.58394161	125.5388813
CH Gayle	4525	110	2972	41.13636364	152.2543742
MS Dhoni	4450	118	3206	37.71186441	138.8022458
RV Uthappa	4420	156	3381	28.33333333	130.7305531
AB de Villiers	4414	104	2902	42.44230769	152.1019986
G Gambhir	4219	134	3400	31.48507463	124.0882353
AM Rahane	3834	117	3133	32.76923077	122.3747207
KD Karthik	3669	138	2813	26.58695652	130.4301458
SR Watson	3590	115	2566	31.2173913	139.9064692
AT Rayudu	3313	114	2616	29.06140351	126.6437309
YK Pathan	3222	110	2240	29.29090909	143.8392857
BB McCullum	2886	106	2185	27.22641509	132.0823799
PA Patel	2864	127	2352	22.5511811	121.7687075
MK Pandey	2855	95	2352	30.05263158	121.3860544
KA Pollard	2772	102	1879	27.17647059	147.5252794
Yuvraj Singh	2755	109	2121	25.27522936	129.8915606

Players | Teams | Matches | Deliveries | Most_runs_average_Strikerate | teamwise_home_and_away | Matches Forma

- **Teamwise_home_and_away:** It includes columns with the name of the team, matches played in its home and away, the number of home and away win and the home and away win percentage.



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

Players | Teams | Matches | Deliveries | Most_runs_average_Strikerate | teamwise_home_and_away | Matches Forma ...

OBJECTIVE DETAILS

<u>Objective 1</u>: - Top 15 batsman of the IPL with their strike rate, average and number of times dismissed.

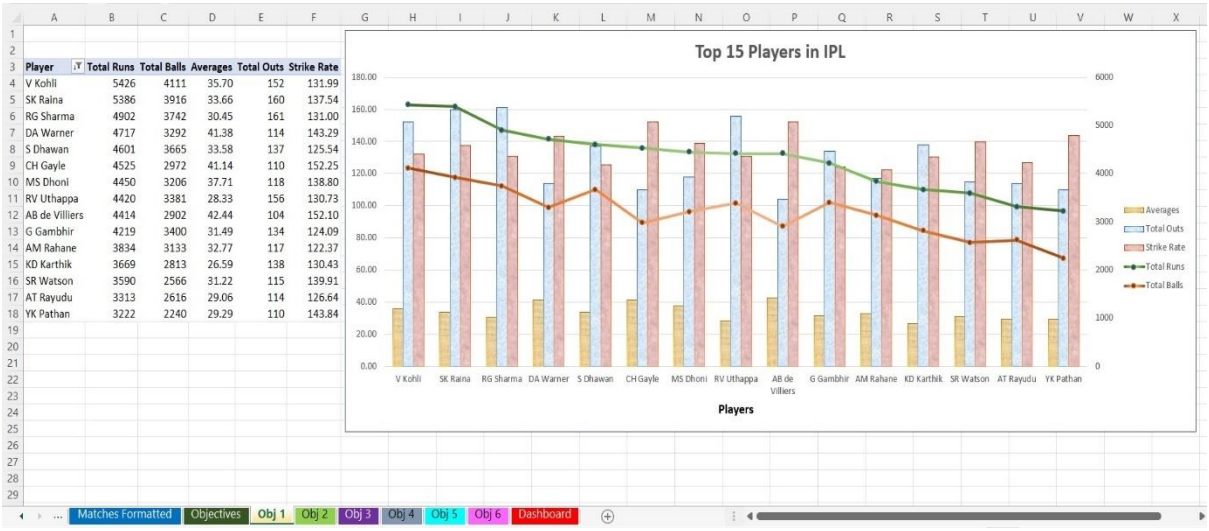
The objective is to filter outs the top 15 batsman of the IPL based on their total number of runs from 2008-2019, total balls played, number of dismissals, average and strike rate of the player.

The representation of the data filtered is done through Combo chart. It includes:

- Clustered column :

  For total dismissals, average and strike rate.
- Marked line graph :

  For total runs scored and total balls played.



Player	Total Runs	Total Balls	Averages	Total Outs	Strike Rate
V Kohli	5426	4111	35.70	152	131.99
SK Raina	5386	3916	33.66	160	137.54
RG Sharma	4902	3742	30.45	161	131.00
DA Warner	4717	3292	41.38	114	143.29
S Dhawan	4601	3665	33.58	137	125.54
CH Gayle	4525	2972	41.14	110	152.25
MS Dhoni	4450	3206	37.71	118	138.80
RV Uthappa	4420	3381	28.33	156	130.73
AB de Villiers	4414	2902	42.44	104	152.10
G Gambhir	4219	3400	31.49	134	124.09
AM Rahane	3834	3133	32.77	117	122.37
KD Karthik	3669	2813	26.59	138	130.43
SR Watson	3590	2566	31.22	115	139.91
AT Rayudu	3313	2616	29.06	114	126.64
YK Pathan	3222	2240	29.29	110	143.84

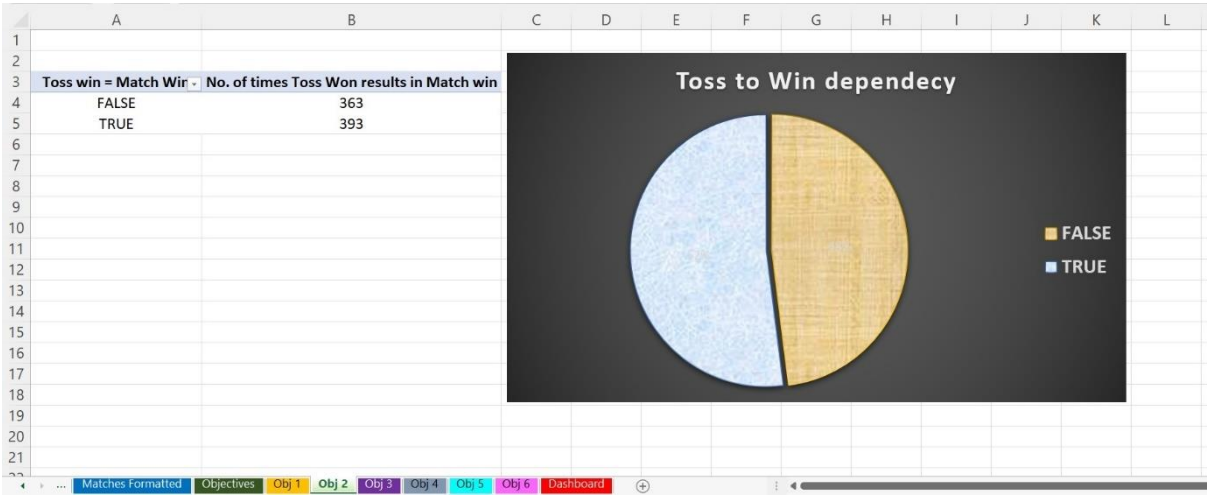
<u>Objective 2</u>: - Analysis of Match win dependency on Toss win.

In this objective I analysed the dependency of Match win on Toss win. It determines the number of times a team won the toss and then their decision leads to winning of the team.

The analysis of data shows that:

- 393(i.e., 52%) matches were won by the team which won the toss.
- 363(i.e., 48%) matches were won by the team which lost the toss.

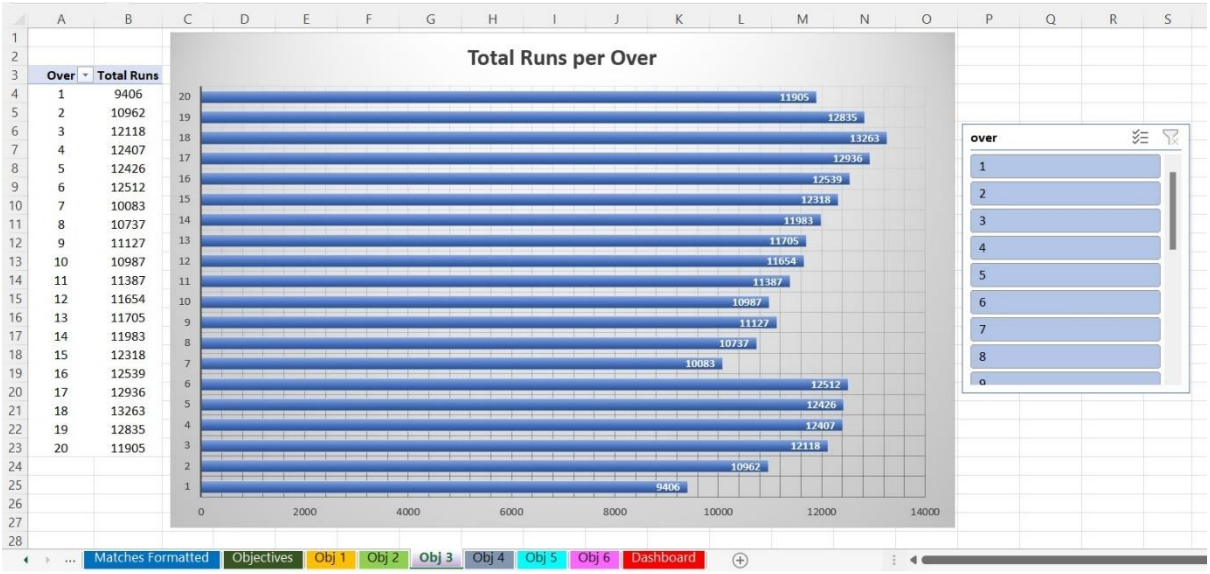
The results were represented by a Pie chart.



Objective 3: - Analyse Total Runs scored in each over in IPL.

In this objective I analysed the total number of runs scored in each over throughout IPL 2008-2019. The results show that most runs are scored in powerplays and least run scored are in the last over and in the middle overs.

The analysed result is represented by using Bar chart all overs.



<u>Objective 4</u>: - Analyse the Home and Away - Matches played and Matches won.

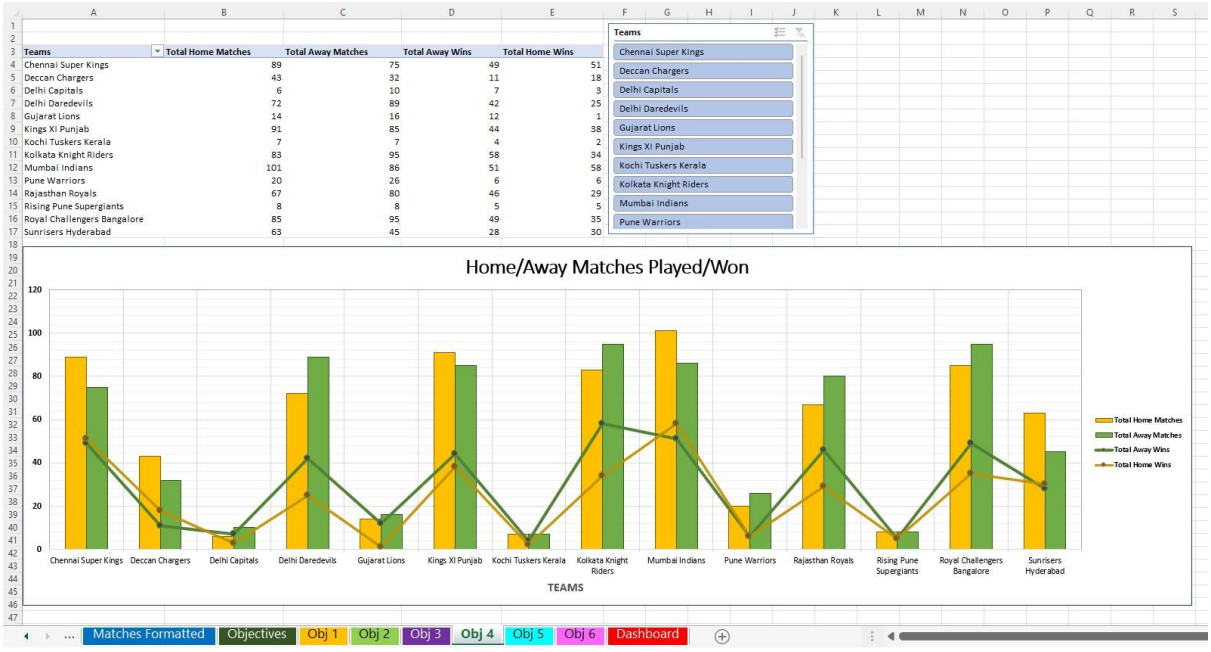
This objective is to analyse the number of home and away matches played and home and away matches won for each team in IPL.

The representation of the analysed data is done using Combo Chart:

- Clustered Column Chart :

  For the total number of home and away matches played.
- Marked Line Chart :

  For the total number of home and away matches won.



Teams	Total Home Matches	Total Away Matches	Total Away Wins	Total Home Wins
Chennai Super Kings	89	75	49	51
Deccan Chargers	43	32	11	18
Delhi Capitals	6	10	7	3
Delhi Daredevils	72	89	42	25
Gujarat Lions	14	16	12	1
Kings XI Punjab	91	85	44	38
Kochi Tuskers Kerala	7	7	4	2
Kolkata Knight Riders	83	95	58	34
Mumbai Indians	101	86	51	58
Pune Warriors	20	26	6	6
Rajasthan Royals	67	80	46	29
Rising Pune Supergiants	8	8	5	5
Royal Challengers Bangalore	85	95	49	35
Sunrisers Hyderabad	63	45	28	30

Objective 5: - Total Matches Played and Won by each team.

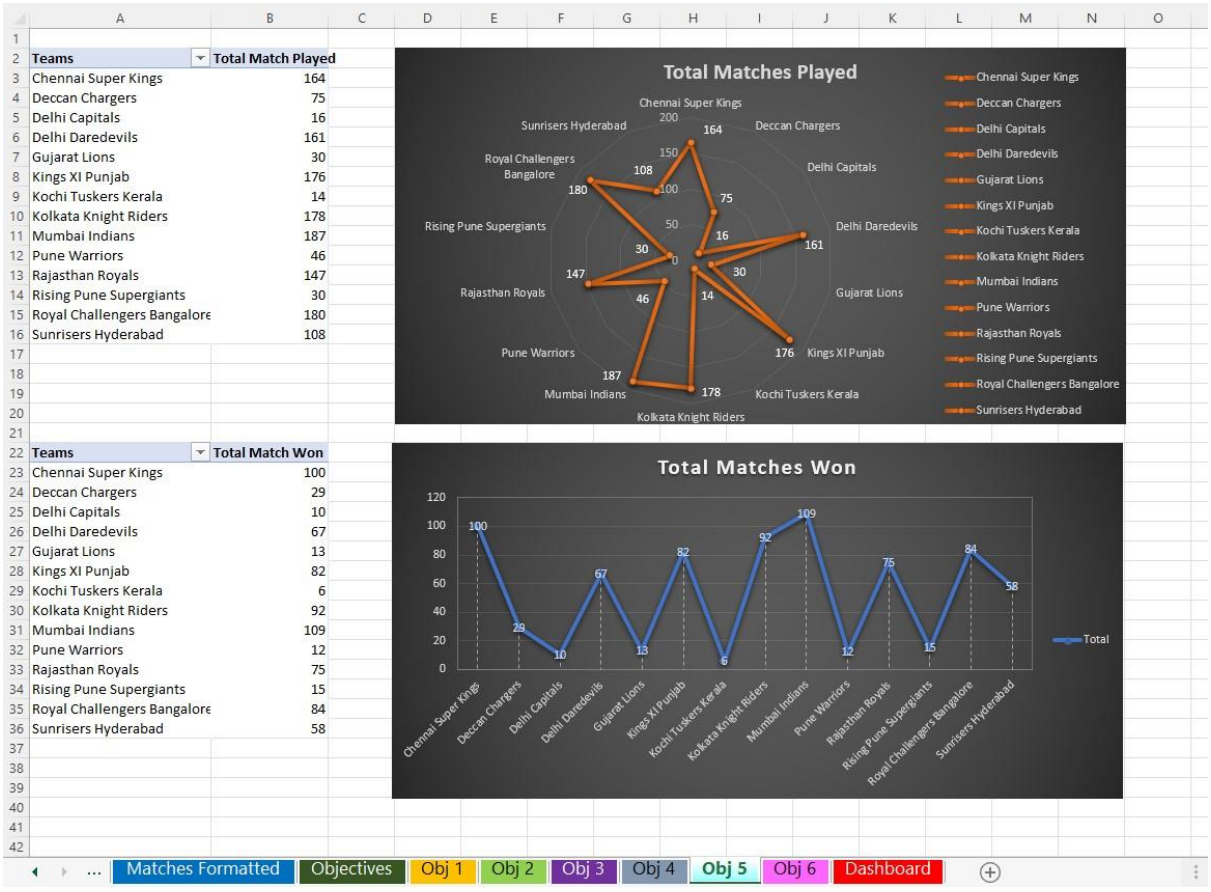
In his objective I have calculated the total number of matches played by each team by counting for each team playing as team 1 and team 2 in the dataset and then adding them. Then I calculated the total number of matches won by each team by counting it from winner team name's column in IPL from 2008-2019. I calculated the data in teamwise_home_and_away worksheet.

The representation of the analysed data from is done using:

- Radar Chart :

  For total number of matches played by each team.
- Marked Line Chart :

  For total number of matches won by each team.



Objective 6: - Analyse Players on basis of their Country and Batting Type.

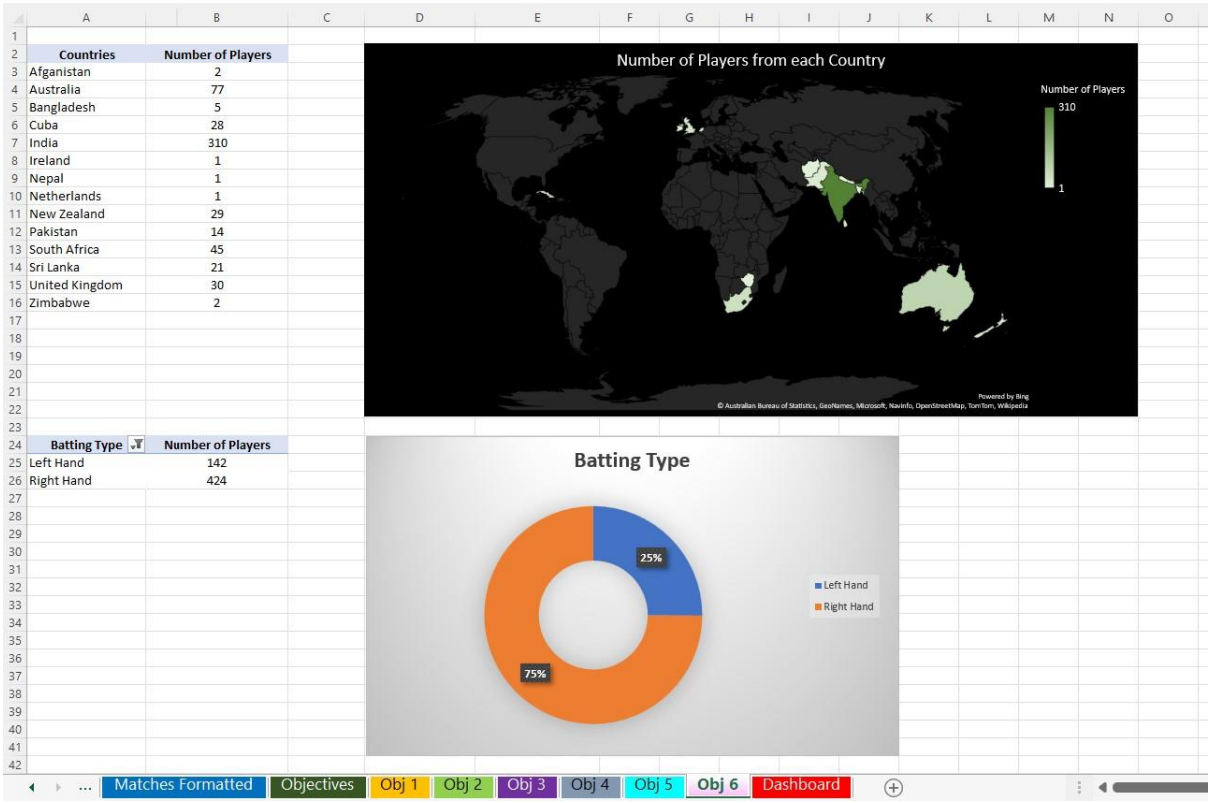
In this objective I have analysed the players from player worksheet by their countries. For that I sorted all the unique values from country column and then counting number of players from each country. Further, I analysed players based on their batting type and checking if a player is left-handed batsman or right-handed.

The charts used for the representation of the analysed data are:

- Filled Maps :

  For showing the number of players from each country on the map.
- Doughnut Chart :

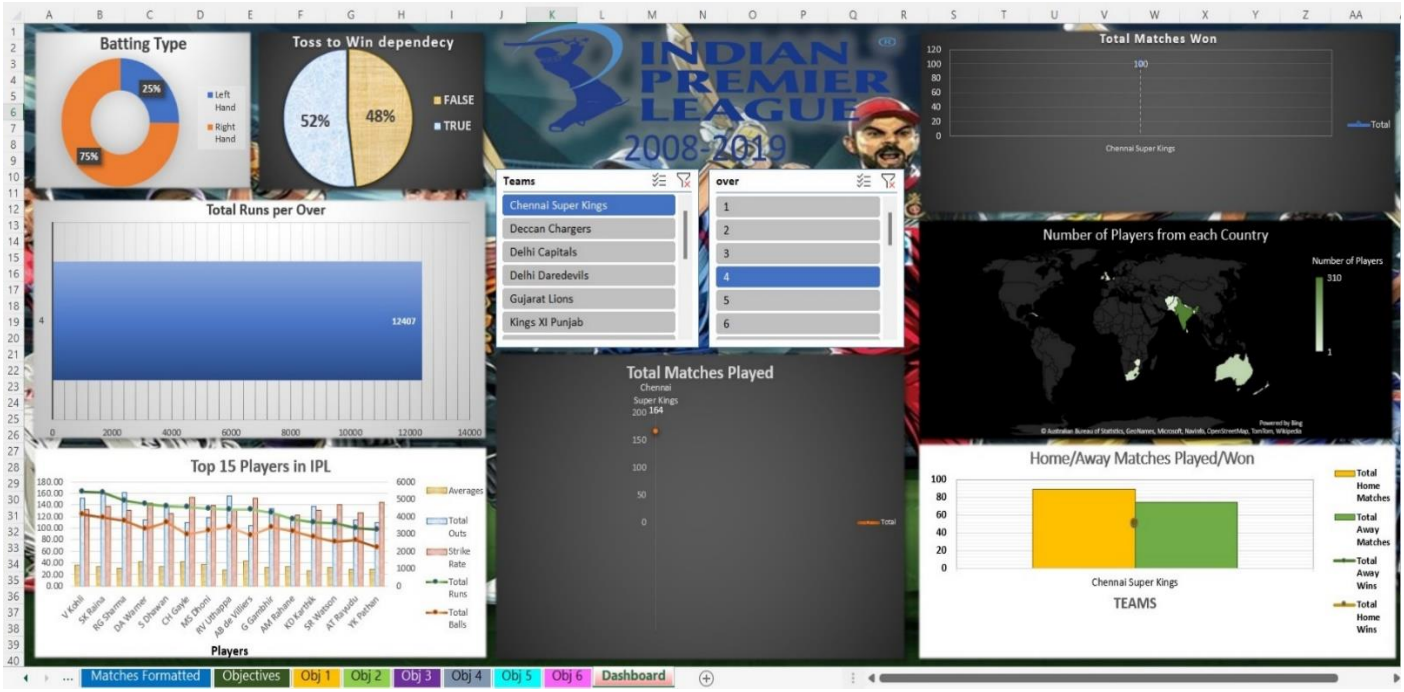
  For players categorised based on their batting type.



DASHBOARD

- Main look of the Dashboard Worksheet (Screenshot).

- **<u>Using Slicers</u>**: - We can get filtered data that we want using slicers on the dashboard. For Example, select a Team “Chennai Super Kings” and select over number “4”. So, data of “Chennai Super Kings” is shown for Home and Away – Matches and won, Matches Played and Matches. And Total runs scored in “4th” over is displayed.



CONCLUSION

This project was valuable and rewarding. I have gained some useful insights about data analysis and representation in MS Excel. I was able to conclude some insights from IPL dataset . I analysed information from the dataset like Top 15 batsman of the IPL with their strike rate, average and number of times dismissed, Match win dependency on Toss win, Total Runs scored in each over in IPL, Home and Away - Matches played and Matches won, Total Matches Played and Won by each team, Players on basis of their Country and Batting Type.

I used ETL process make proper use of the dataset by cleaning and organising it. It helped to develop data management and analytic skills. I also gained knowledge of the scope of using MS Excel its formulas and features and practically applying them to analyse information from a dataset. I worked with different features like, 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 a dashboard for representing all the analysed information and used slicers to filter the data get personalised extracted information from the dataset. The dashboard offered a common, user friendly and interactive stage to present the data.

Finally, this was very informative, and I got hands-on experience on a lot of the data management and analysis techniques and skills on the MS Excel software. I explored workbooks, worksheets, formulas, features like, Pivot tables, charts, etc. This is a data driven world, and with this I am well-verse with the 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 Dataset 2008-2019):** https://www.kaggle.com/ramjidoolla/ipl-data-set
- https://blog.appliedinformaticsinc.com/etl-extract-transform-and-load-process-concept/
- https://exceljet.net/chart-types
- 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