INT217 PROJECT REPORT

(Project Semester August-December 2021)

Analysing IPL dataset (2008-2020)

Submitted by

Shaik ShoyabAktar

Registration No **11906144**

Programme and Section **Bachelor of Technology**

Course Code **INT217**

Under the Guidance of

Sandeep Kaur: 23614

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

ACKNOWLEDGEMENT

Primarily I'd thank God for being able to complete my project with success. Then I'd like to thank my mentor **Ms. Sandeep Kaur**, whose valuable guidance has been the ones that helped me patch this project and make it full proof success in contribution towards the completion of this project.

Finally, I would like to thank Lovely Professional University, and my parent's inspiration, who gave me this golden opportunity to learn many new things, to learn another aspect of life.

CERTIFICATE

This is to certify that Shaik Shoyab Aktar bearing Registration no 11906144 has completed INT217 project titled, **"Analyzing IPL dataset (2008-2020)"** 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.

Signature

Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering.
Lovely Professional University
Phagwara, Punjab.

Date: 10/12/2021

DECLARATION

I, **Shaik Shoyab Aktar** student at **Lovely Professional University** 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.

Date: 10/12/2021

Registration No. **11906144**

Name of the student: **Shaik Shoyab Aktar**

TABLE OF CONENTS

Introduction

Data management is an administrative process that includes acquiring, validating, storing protecting, and processing required data to ensure the accessibility, reliability, and timeliness of the data for its users

Organizations and enterprises are making use of Big Data more than ever before to inform business decisions and gain deep insights into customer behavior, trends, and opportunities for creating extraordinary customer experiences. To make sense of the vast quantities of data that enterprises are gathering analyzing and storing today, companies turn to data management solutions and platforms. Data management solutions make processing validation, and other essential functions simpler and less time intensive.

Fluency in Microsoft Excel is one of the most valuable soft skill in any professional's life. Excel's broad applicability and user-friendly interface make it so it can be utilized by the masses for a vast number of tasks, from simple data entry to complex data analysis and querying. No matter where you fall or plan to fall on Excel's utilization spectrum, it is of cardinal importance that you have proper guidance in order to witness how you can use

Excel to leverage your professional toolkit. The hundreds of formulas that excel offers range from basic organization tools like the COUNTIF, SUMIF, Index, Match functions to formulas designed to derive key metrics in statistical analysis. For more complicated formulas, there are courses that will provide users a step-by-step walkthrough troubleshooting and customization techniques.

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps. data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

There are many options for Excel focused on formulas which help you to automate your spreadsheets. Charts, graphs, and models can all be used to add a visual component to your data With Excel Charts, users will be able to find the best ways to display their data that could drive more effective decision making. Not only can you optimize the way which your data is presented through courses in Excel Charts, but you can also learn customization techniques so you can coordinate your Excel figures with associated documents in Microsoft Office.

In this project I chose IPL as my project theme. Because IPL was started in 2008 and if we analyze the data we can get the idea about teams, players, winners, places(stadiums) etc. I made a dashboard using pivot charts and pivot tables.

An Excel dashboard is one-pager that helps managers and business leaders tracking key or metrics and take a decision based on it. It contains charts/tables/views that are backed by data. A dashboard is often called a report, however, not all reports are dashboards.

When making a dashboard using dataset on excel objectives are taking main place. When we are handling big data set, we need to make proper objectives to visualize the data. In this project I used

5 objectives to visualize data on the dashboard.

Objectives

1.Matches won by team 1) By Batting first 2) By Bowling first

2.More times player of the match

3.(A)Winning Toss and Winning match (B)Winning Toss and Losing Match (C)Losing both Toss and Match (D)Losing Toss and Winning Match

4. Number of matches won by a team

5.Matches won by Team by Huge Margin (number of runs Scored).

In this dataset I used pivot tables and pivot charts to visualize the data on my dataset. In dashboard year wise, team wise, city wise data are visualized.

Source of Dataset

A data set is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files. In this project I used a dataset which is including the details of IPL (2008-2020). And I found out this data set from Kaggle.com.

Kaggle allows users to find and publish data sets, explore, and build models in a web-based data science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Inside Kaggle you'll find all the code & data you need to do your data science work. Use over 19.000 public datasets and 200,000 public notebooks to conquer any analysis in no time. Apparently, they have a community of around 3 million people on the platform. In Kaggle.com there are lot of data sets available. Among them I chose the dataset IPL (2008-2020).

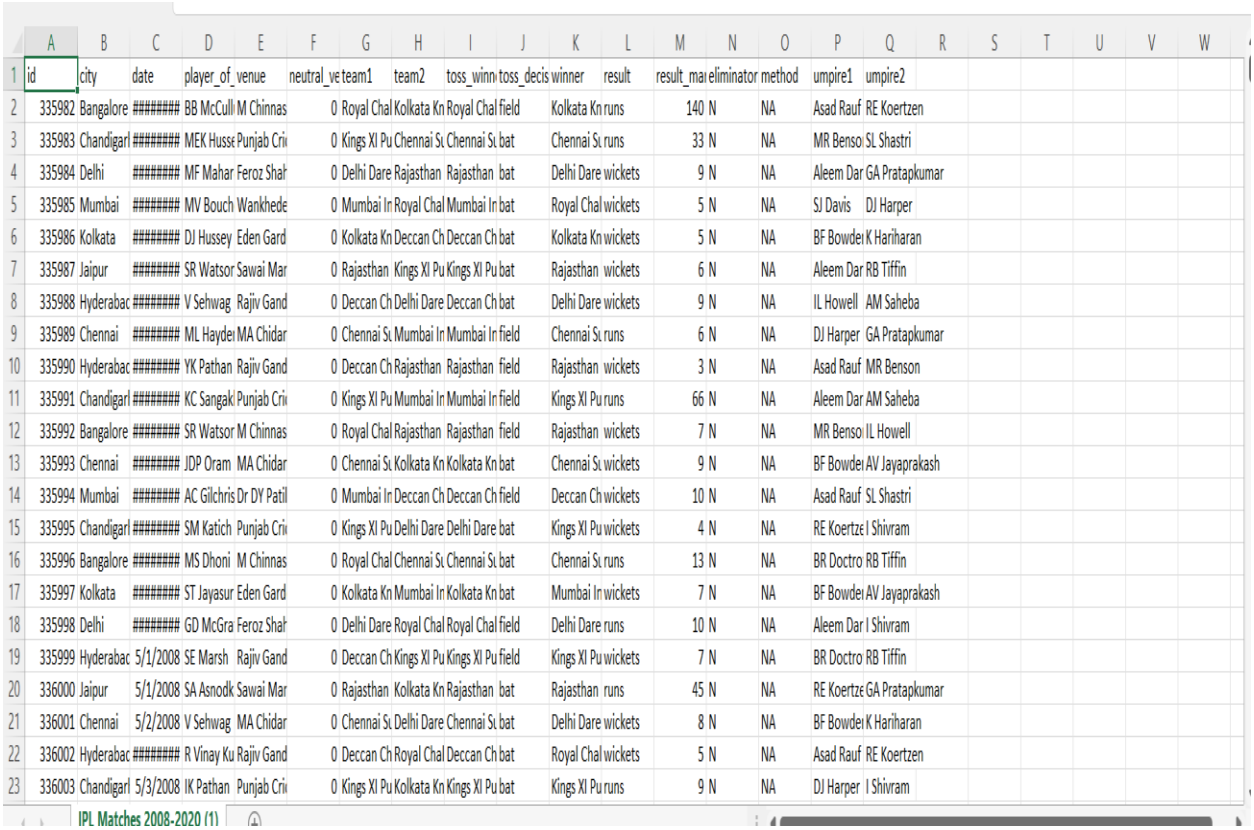
ETL PROCESS

ETL is a process that extracts the data from different source systems, then transforms the data like applying calculations, concatenations and finally loads the data into the Data Warehouse system. Full form of ETL is Extract, Transform and Load.

It's tempting to think a creating a Data warehouse is simply extracting data from multiple sources and loading into database of a Data warehouse. This is far from the truth and requires a complex ETL process. The ETL process requires active inputs from various stakeholders including developers, analysts, testers, top executives and is technically challenging.

In order to maintain its value as a tool for decision-makers, Data warehouse system needs to change with business changes. ETL is a recurring activity of a Data warehouse system and needs to be agile, automated, and well documented.

· Initially, the raw dataset was arranged as shown in given picture:



id	city	date	player_of_	venue	neutral_ve	team1	team2	toss_winn	toss_decis	winner	result	result_mai	eliminator	method	umpire1	umpire2
335982	Bangalore	########	BB McCull	M Chinnas	0	Royal Chal	Kolkata Kn	Royal Chal	field	Kolkata Kn	runs	140	N	NA	Asad Rauf	RE Koertzen
335983	Chandigarl	########	MEK Husse	Punjab Cri	0	Kings XI Pu	Chennai Su	Chennai Su	bat	Chennai Su	runs	33	N	NA	MR Benso	SL Shastri
335984	Delhi	########	MF Mahar	Feroz Shat	0	Delhi Dare	Rajasthan	Rajasthan	bat	Delhi Dare	wickets	9	N	NA	Aleem Dar	GA Pratapkumar
335985	Mumbai	########	MV Bouch	Wankhede	0	Mumbai In	Royal Chal	Mumbai In	bat	Royal Chal	wickets	5	N	NA	SJ Davis	DJ Harper
335986	Kolkata	########	DJ Hussey	Eden Gard	0	Kolkata Kn	Deccan Ch	Deccan Ch	bat	Kolkata Kn	wickets	5	N	NA	BF Bowde	K Hariharan
335987	Jaipur	########	SR Watsor	Sawai Mar	0	Rajasthan	Kings XI Pu	Kings XI Pu	bat	Rajasthan	wickets	6	N	NA	Aleem Dar	RB Tiffin
335988	Hyderabac	########	V Sehwag	Rajiv Gand	0	Deccan Ch	Delhi Dare	Deccan Ch	bat	Delhi Dare	wickets	9	N	NA	IL Howell	AM Saheba
335989	Chennai	########	ML Hayde	MA Chidar	0	Chennai Su	Mumbai In	Mumbai In	field	Chennai Su	runs	6	N	NA	DJ Harper	GA Pratapkumar
335990	Hyderabac	########	YK Pathan	Rajiv Gand	0	Deccan Ch	Rajasthan	Rajasthan	field	Rajasthan	wickets	3	N	NA	Asad Rauf	MR Benson
335991	Chandigarl	########	KC Sangak	Punjab Cri	0	Kings XI Pu	Mumbai In	Mumbai In	field	Kings XI Pu	runs	66	N	NA	Aleem Dar	AM Saheba
335992	Bangalore	########	SR Watsor	M Chinnas	0	Royal Chal	Rajasthan	Rajasthan	field	Rajasthan	wickets	7	N	NA	MR Benso	IL Howell
335993	Chennai	########	JDP Oram	MA Chidar	0	Chennai Su	Kolkata Kn	Kolkata Kn	bat	Chennai Su	wickets	9	N	NA	BF Bowde	AV Jayaprakash
335994	Mumbai	########	AC Gilchris	Dr DY Patil	0	Mumbai In	Deccan Ch	Deccan Ch	field	Deccan Ch	wickets	10	N	NA	Asad Rauf	SL Shastri
335995	Chandigarl	########	SM Katich	Punjab Cri	0	Kings XI Pu	Delhi Dare	Delhi Dare	bat	Kings XI Pu	wickets	4	N	NA	RE Koertze	I Shivram
335996	Bangalore	########	MS Dhoni	M Chinnas	0	Royal Chal	Chennai Su	Chennai Su	bat	Chennai Su	runs	13	N	NA	BR Doctro	RB Tiffin
335997	Kolkata	########	ST Jayasur	Eden Gard	0	Kolkata Kn	Mumbai In	Kolkata Kn	bat	Mumbai In	wickets	7	N	NA	BF Bowde	AV Jayaprakash
335998	Delhi	########	GD McGra	Feroz Shat	0	Delhi Dare	Royal Chal	Royal Chal	field	Delhi Dare	runs	10	N	NA	Aleem Dar	I Shivram
335999	Hyderabac	5/1/2008	SE Marsh	Rajiv Gand	0	Deccan Ch	Kings XI Pu	Kings XI Pu	field	Kings XI Pu	wickets	7	N	NA	BR Doctro	RB Tiffin
336000	Jaipur	5/1/2008	SA Asnodk	Sawai Mar	0	Rajasthan	Kolkata Kn	Rajasthan	bat	Rajasthan	runs	45	N	NA	RE Koertze	GA Pratapkumar
336001	Chennai	5/2/2008	V Sehwag	MA Chidar	0	Chennai Su	Delhi Dare	Chennai Su	bat	Delhi Dare	wickets	8	N	NA	BF Bowde	K Hariharan
336002	Hyderabac	########	R Vinay Ku	Rajiv Gand	0	Deccan Ch	Royal Chal	Deccan Ch	bat	Royal Chal	wickets	5	N	NA	Asad Rauf	RE Koertzen
336003	Chandigarl	5/3/2008	IK Pathan	Punjab Cri	0	Kings XI Pu	Kolkata Kn	Kings XI Pu	bat	Kings XI Pu	runs	9	N	NA	DJ Harper	I Shivram

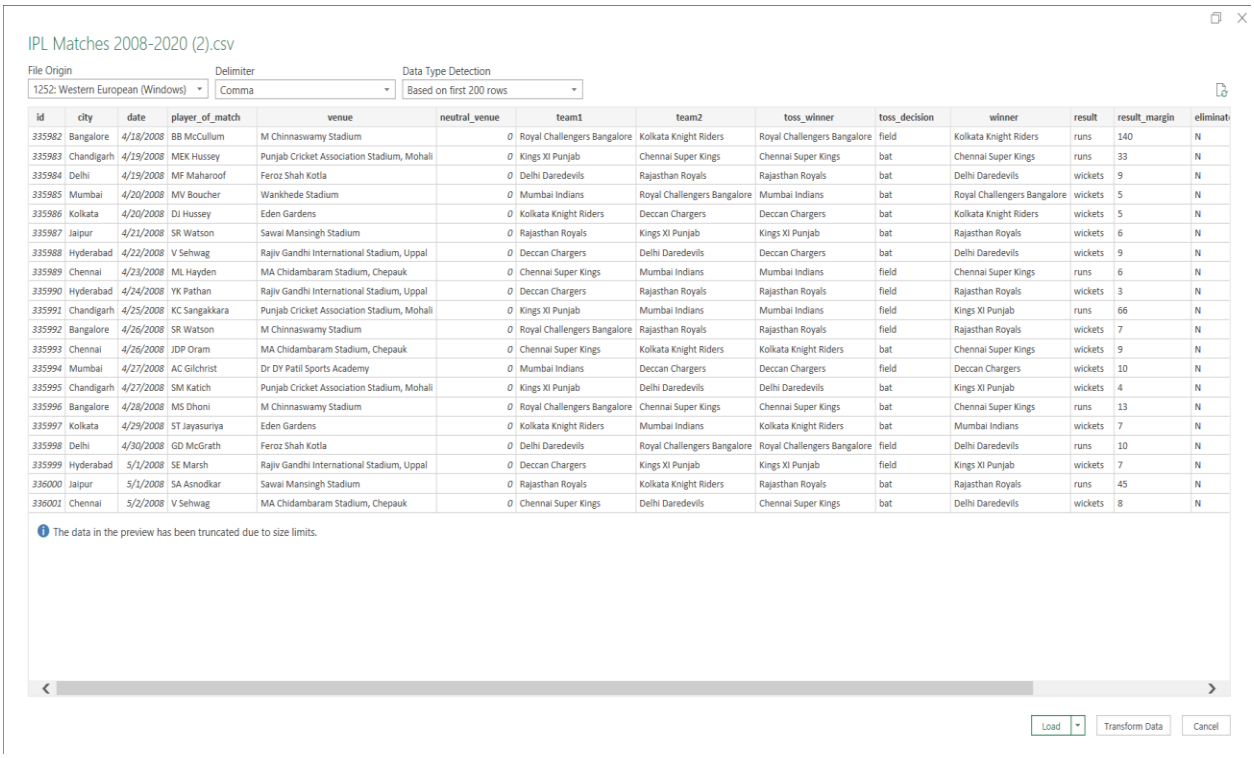
IPL Matches 2008-2020 (1)

- Extraction

In this step, data is extracted from the source system into the staging area. Transformations if any are done in staging area so that performance of source system in not degraded. Also, if corrupted data is copied directly from the source into Data warehouse database, rollback

will be a challenge. Staging area gives an opportunity to validate extracted data before warehouse. moves into the Data.

Step1: To extract a csv file go to data tab and select from CSV/text in Get&Transform data Which takes us to the Query editor and then to clean the data select transform.



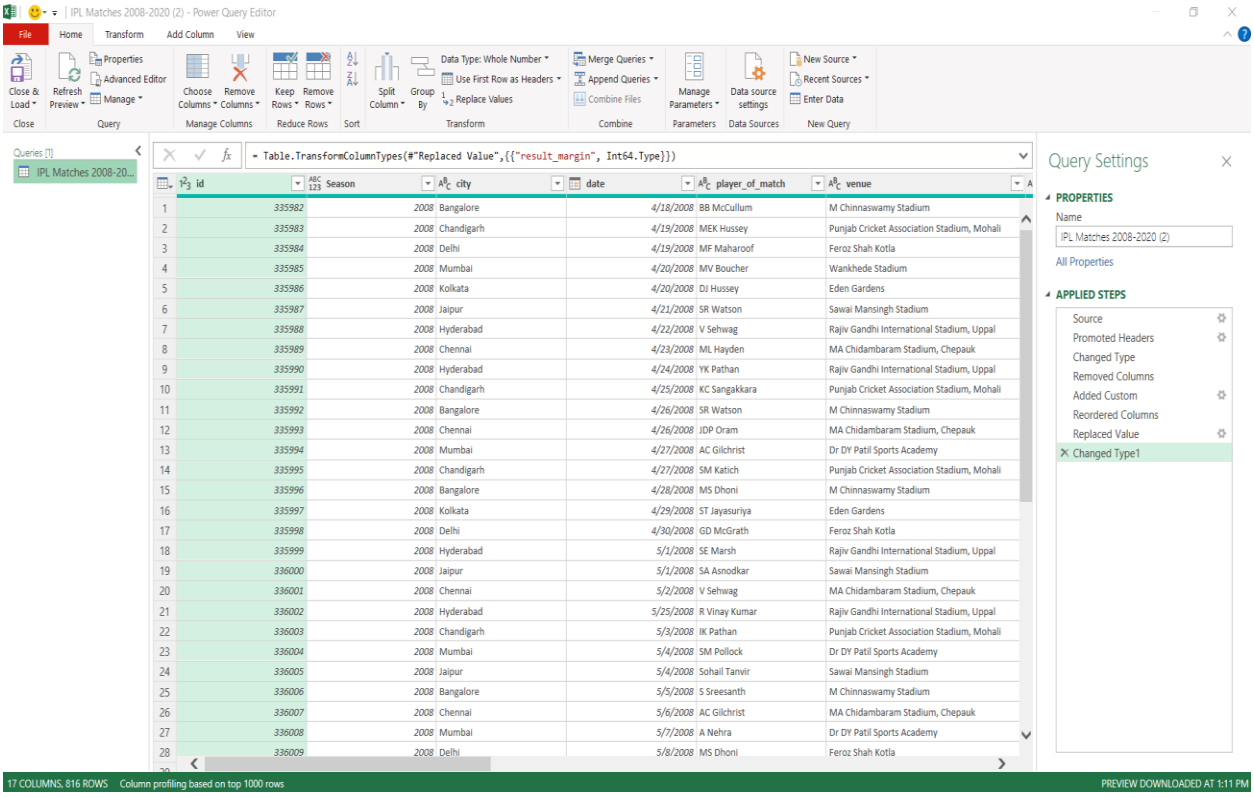
id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_margin	eliminat
335982	Bangalore	4/18/2008	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Riders	runs	140	N
335983	Chandigarh	4/19/2008	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kings	runs	33	N
335984	Delhi	4/19/2008	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils	wickets	9	N
335985	Mumbai	4/20/2008	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers Bangalore	wickets	5	N
335986	Kolkata	4/20/2008	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Riders	wickets	5	N
335987	Jaipur	4/21/2008	SR Watson	Sawai Mansingh Stadium	0	Rajasthan Royals	Kings XI Punjab	Kings XI Punjab	bat	Rajasthan Royals	wickets	6	N
335988	Hyderabad	4/22/2008	V Sehwag	Rajiv Gandhi International Stadium, Uppal	0	Deccan Chargers	Delhi Daredevils	Deccan Chargers	bat	Delhi Daredevils	wickets	9	N
335989	Chennai	4/23/2008	ML Hayden	MA Chidambaram Stadium, Chepauk	0	Chennai Super Kings	Mumbai Indians	Mumbai Indians	field	Chennai Super Kings	runs	6	N
335990	Hyderabad	4/24/2008	YK Pathan	Rajiv Gandhi International Stadium, Uppal	0	Deccan Chargers	Rajasthan Royals	Rajasthan Royals	field	Rajasthan Royals	wickets	3	N
335991	Chandigarh	4/25/2008	KC Sangakkara	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Mumbai Indians	Mumbai Indians	field	Kings XI Punjab	runs	66	N
335992	Bangalore	4/26/2008	SR Watson	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	Rajasthan Royals	wickets	7	N
335993	Chennai	4/26/2008	JDP Oram	MA Chidambaram Stadium, Chepauk	0	Chennai Super Kings	Kolkata Knight Riders	Kolkata Knight Riders	bat	Chennai Super Kings	wickets	9	N
335994	Mumbai	4/27/2008	AC Gilchrist	Dr DY Patil Sports Academy	0	Mumbai Indians	Deccan Chargers	Deccan Chargers	field	Deccan Chargers	wickets	10	N
335995	Chandigarh	4/27/2008	SM Katich	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Delhi Daredevils	Delhi Daredevils	bat	Kings XI Punjab	wickets	4	N
335996	Bangalore	4/28/2008	MS Dhoni	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kings	runs	13	N
335997	Kolkata	4/29/2008	ST Jayasuriya	Eden Gardens	0	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	bat	Mumbai Indians	wickets	7	N
335998	Delhi	4/30/2008	GD McGrath	Feroz Shah Kotla	0	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Delhi Daredevils	runs	10	N
335999	Hyderabad	5/1/2008	SE Marsh	Rajiv Gandhi International Stadium, Uppal	0	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab	wickets	7	N
336000	Jaipur	5/1/2008	SA Asnodkar	Sawai Mansingh Stadium	0	Rajasthan Royals	Kolkata Knight Riders	Rajasthan Royals	bat	Rajasthan Royals	runs	45	N
336001	Chennai	5/2/2008	V Sehwag	MA Chidambaram Stadium, Chepauk	0	Chennai Super Kings	Delhi Daredevils	Chennai Super Kings	bat	Delhi Daredevils	wickets	8	N

- Transformation

Data extracted from source server is raw and not usable in its original form. Therefore it needs to be cleansed, mapped and transformed. In fact, this is the key step where ETL process adds value and changes data. In this step. you apply a set of functions on extracted

data. Data that does not require any transformation is called as direct move or pass through data.

Step 2: Transform the data by removing the unneeded columns, changing datatypes ,removing null values and adding a new column known as season as shown in the figure.

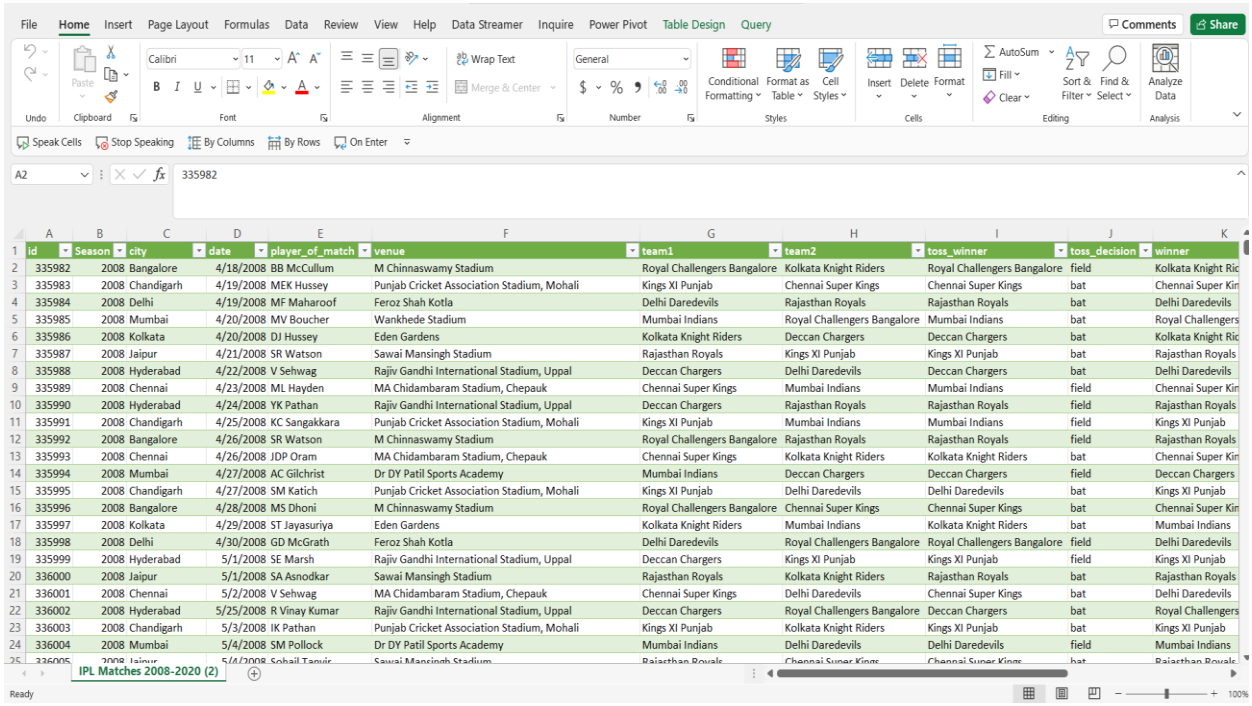


- Loading

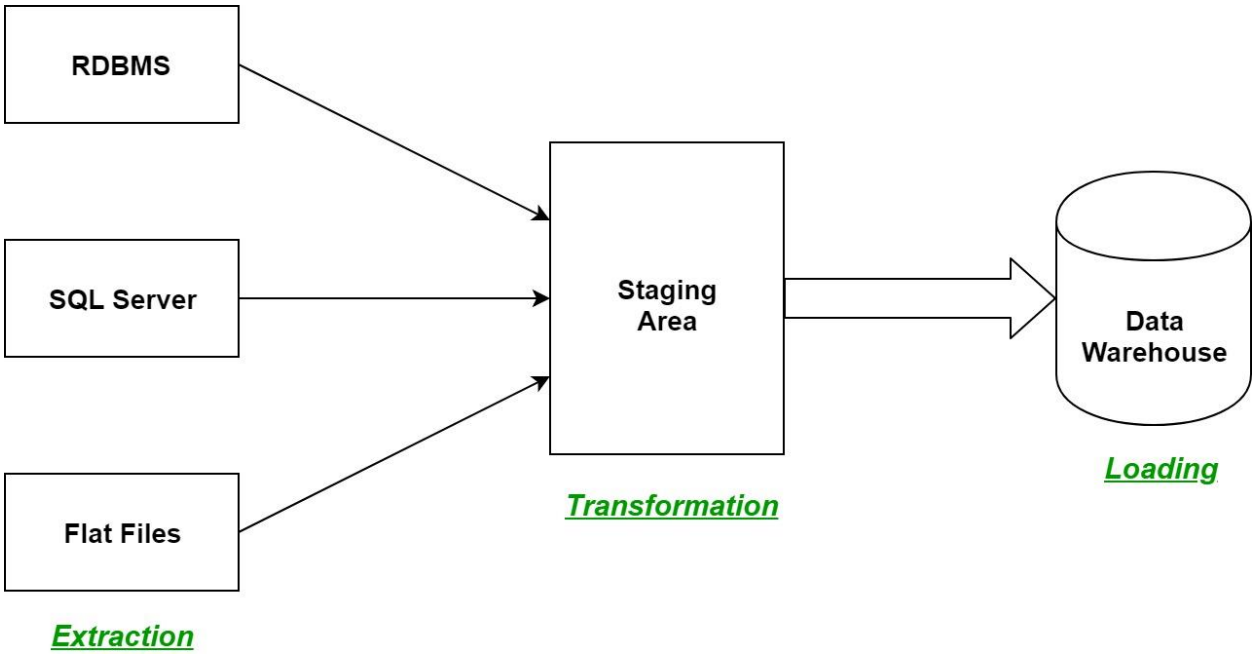
Loading data into the target data warehouse database is the last step of the ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in

a relatively short period (nights). Hence, load process should be optimized for performance.

Step 3: After pressing load in the power query editor the output will be as follows



id	Season	city	date	player_of_match	venue	team1	team2	toss_winner	toss_decision	winner
335982	2008	Bangalore	4/18/2008	BB McCullum	M Chinnaswamy Stadium	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Ric
335983	2008	Chandigarh	4/19/2008	MEK Hussey	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kin
335984	2008	Delhi	4/19/2008	MF Maharoof	Feroz Shah Kotla	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils
335985	2008	Mumbai	4/20/2008	MV Boucher	Wankhede Stadium	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers
335986	2008	Kolkata	4/20/2008	DJ Hussey	Eden Gardens	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Ric
335987	2008	Jaipur	4/21/2008	SR Watson	Sawai Mansingh Stadium	Rajasthan Royals	Kings XI Punjab	Kings XI Punjab	bat	Rajasthan Royals
335988	2008	Hyderabad	4/22/2008	V Sehwag	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Delhi Daredevils	Deccan Chargers	bat	Delhi Daredevils
335989	2008	Chennai	4/23/2008	ML Hayden	MA Chidambaram Stadium, Chepauk	Chennai Super Kings	Mumbai Indians	Mumbai Indians	field	Chennai Super Kin
335990	2008	Hyderabad	4/24/2008	YK Pathan	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Rajasthan Royals	Rajasthan Royals	field	Rajasthan Royals
335991	2008	Chandigarh	4/25/2008	KC Sangakkara	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Mumbai Indians	Mumbai Indians	field	Kings XI Punjab
335992	2008	Bangalore	4/26/2008	SR Watson	M Chinnaswamy Stadium	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	Rajasthan Royals
335993	2008	Chennai	4/26/2008	JDP Oram	MA Chidambaram Stadium, Chepauk	Chennai Super Kings	Kolkata Knight Riders	Kolkata Knight Riders	bat	Chennai Super Kin
335994	2008	Mumbai	4/27/2008	AC Gilchrist	Dr DY Patil Sports Academy	Mumbai Indians	Deccan Chargers	Deccan Chargers	field	Deccan Chargers
335995	2008	Chandigarh	4/27/2008	SM Katich	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Delhi Daredevils	Delhi Daredevils	bat	Kings XI Punjab
335996	2008	Bangalore	4/28/2008	MS Dhoni	M Chinnaswamy Stadium	Royal Challengers Bangalore	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kin
335997	2008	Kolkata	4/29/2008	ST Jayasuriya	Eden Gardens	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	bat	Mumbai Indians
335998	2008	Delhi	4/30/2008	GD McGrath	Feroz Shah Kotla	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Delhi Daredevils
335999	2008	Hyderabad	5/1/2008	SE Marsh	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab
336000	2008	Jaipur	5/1/2008	SA Asnodkar	Sawai Mansingh Stadium	Rajasthan Royals	Kolkata Knight Riders	Rajasthan Royals	bat	Rajasthan Royals
336001	2008	Chennai	5/2/2008	V Sehwag	MA Chidambaram Stadium, Chepauk	Chennai Super Kings	Delhi Daredevils	Chennai Super Kings	bat	Delhi Daredevils
336002	2008	Hyderabad	5/25/2008	R Vinay Kumar	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Royal Challengers Bangalore	Deccan Chargers	bat	Royal Challengers
336003	2008	Chandigarh	5/3/2008	IK Pathan	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Kolkata Knight Riders	Kings XI Punjab	bat	Kings XI Punjab
336004	2008	Mumbai	5/4/2008	SM Pollock	Dr DY Patil Sports Academy	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	field	Mumbai Indians



Analysis of Data Set

i. Introduction

A dashboard is a visual interface that provides at-a-glance views into key measures relevant to a particular objective or business process. Excel dashboards make it easy to perform quick overviews of data reports rather than going through large volumes of data. Overviews help in making quick and urgent decisions since one can skim through a lot of information at once and within a short time.

Excel dashboards include various elements such as charts, tables, figures, and gauges that help in presenting the data. They can handle any type of data from different market and purposes, and the information can be used for marketing, financial, or other business projects. The dashboard is most applicable to large volumes of data since it would otherwise be hectic to go through such large volumes of data, especially with limited time.

Dashboards have three main attributes:

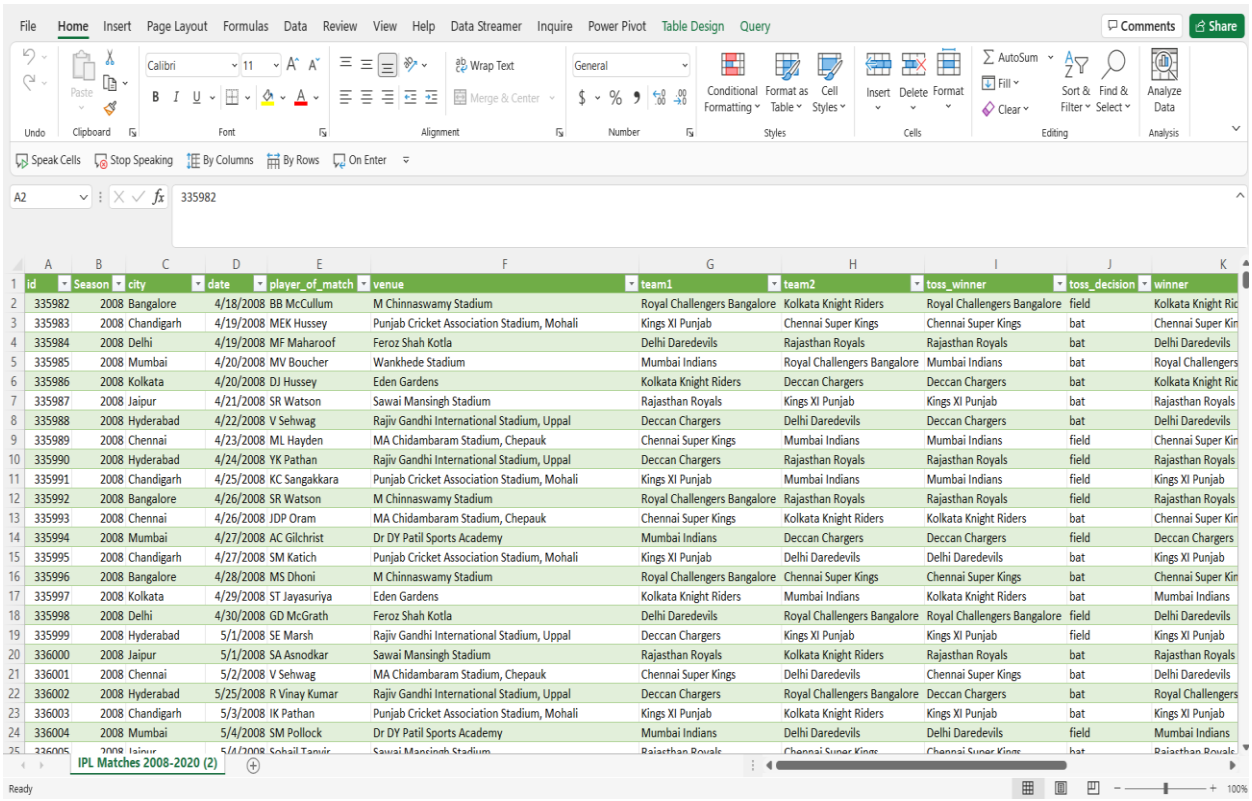
- Dashboards are typically graphical in nature, providing visualizations that help focus attention on key trends, comparisons, and exceptions

- Dashboards often display only data that are relevant to the goal of the dashboard

  Because dashboards are designed with a specific purpose or goal, they inherently contain predefined conclusions that relieve the end user from performing his own analysis

ii. General Description

In this project I used a data set IPL(2008-2020) from the kaggle.com. In the data set there are 17 columns & 817 rows Data set is containing all data for 2008-2020.



id	Season	city	date	player_of_match	venue	team1	team2	toss_winner	toss_decision	winner
335982	2008	Bangalore	4/18/2008	BB McCullum	M Chinnaswamy Stadium	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Ric
335983	2008	Chandigarh	4/19/2008	MEK Hussey	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kin
335984	2008	Delhi	4/19/2008	MF Maharoof	Feroz Shah Kotla	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils
335985	2008	Mumbai	4/20/2008	MV Boucher	Wankhede Stadium	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers
335986	2008	Kolkata	4/20/2008	DJ Hussey	Eden Gardens	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Ric
335987	2008	Jaipur	4/21/2008	SR Watson	Sawai Mansingh Stadium	Rajasthan Royals	Kings XI Punjab	Kings XI Punjab	bat	Rajasthan Royals
335988	2008	Hyderabad	4/22/2008	V Sehwag	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Delhi Daredevils	Deccan Chargers	bat	Delhi Daredevils
335989	2008	Chennai	4/23/2008	ML Hayden	MA Chidambaram Stadium, Chepauk	Chennai Super Kings	Mumbai Indians	Mumbai Indians	field	Chennai Super Kin
335990	2008	Hyderabad	4/24/2008	YK Pathan	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Rajasthan Royals	Rajasthan Royals	field	Rajasthan Royals
335991	2008	Chandigarh	4/25/2008	KC Sangakkara	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Mumbai Indians	Mumbai Indians	field	Kings XI Punjab
335992	2008	Bangalore	4/26/2008	SR Watson	M Chinnaswamy Stadium	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	Rajasthan Royals
335993	2008	Chennai	4/26/2008	JDP Oram	MA Chidambaram Stadium, Chepauk	Chennai Super Kings	Kolkata Knight Riders	Kolkata Knight Riders	bat	Chennai Super Kin
335994	2008	Mumbai	4/27/2008	AC Gilchrist	Dr DY Patil Sports Academy	Mumbai Indians	Deccan Chargers	Deccan Chargers	field	Deccan Chargers
335995	2008	Chandigarh	4/27/2008	SM Katich	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Delhi Daredevils	Delhi Daredevils	bat	Kings XI Punjab
335996	2008	Bangalore	4/28/2008	MS Dhoni	M Chinnaswamy Stadium	Royal Challengers Bangalore	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kin
335997	2008	Kolkata	4/29/2008	ST Jayasuriya	Eden Gardens	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	bat	Mumbai Indians
335998	2008	Delhi	4/30/2008	GD McGrath	Feroz Shah Kotla	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	Delhi Daredevils
335999	2008	Hyderabad	5/1/2008	SE Marsh	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab
336000	2008	Jaipur	5/1/2008	SA Asnodkar	Sawai Mansingh Stadium	Rajasthan Royals	Kolkata Knight Riders	Rajasthan Royals	bat	Rajasthan Royals
336001	2008	Chennai	5/2/2008	V Sehwag	MA Chidambaram Stadium, Chepauk	Chennai Super Kings	Delhi Daredevils	Chennai Super Kings	bat	Delhi Daredevils
336002	2008	Hyderabad	5/25/2008	R Vinay Kumar	Rajiv Gandhi International Stadium, Uppal	Deccan Chargers	Royal Challengers Bangalore	Deccan Chargers	bat	Royal Challengers
336003	2008	Chandigarh	5/3/2008	IK Pathan	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab	Kolkata Knight Riders	Kings XI Punjab	bat	Kings XI Punjab
336004	2008	Mumbai	5/4/2008	SM Pollock	Dr DY Patil Sports Academy	Mumbai Indians	Delhi Daredevils	Delhi Daredevils	field	Mumbai Indians

In this data set there are 17 columns named as id, city, date, team 1, team 2, toss winner, toss decision, result, al applied winner, win by runs, win by wickets, player of the match,venue, umpire 1. umpire 2. umpire

In dashboard mainly I used 5 charts to Visualize the data and they change according to year,city.

(iii)Specific Requirements-Functions and Formulas

Making dashboard I used pivot tables and pivot charts to analyze the data. With the help of them we can analyse the data and visualize the data very well. There are lot of advantages of using pivot tables and pivot charts. This is effective way of representing data in a pictorial manner, makes the process of data visualization of data easier. And I used 1 slicer to visualize data more attractive. And for the Toss

vs Match objective I used the Index, Match , count, count if functions and a Dropbox.

✓Pivot Tables

Pivot tables are one of Excel's most powerful features. A pivot table allows you to extract the significance from a large, detailed data set. A pivot table is a statistics tool that summarizes and reorganizes selected columns and rows of data in a spreadsheet or database table to obtain a desired report. The tool does not actually change the spreadsheet or database itself, it simply "pivots" or tums the data to view it from different perspectives.

Pivot tables are especially useful with large amounts of data that would be time-consuming to calculate by hand. A few data processing functions a pivot table can perform include identifying sums, averages, ranges, and others. The table then arranges this information in a simple, meaningful layout that draws attention to key values. Pivot table is a generic term. This refers to a tool specific to Excel for creating pivot tables.

✓Pivot Charts

Pivot Charts complement PivotTables by adding visualizations to the summary data in a Pivot Table, and allow you to easily see comparisons, patterns, and trends. A pivot chart is especially useful when dealing with tremendous amounts of data.

For example, a society having many employees is maintaining the working hours of each pupil through Excel, such that at the end of each month, the employee with the highest number of working hours, would be provided a bonus due to the sincerity and devotion to the society.

(iv) Analysis Results with Visualization

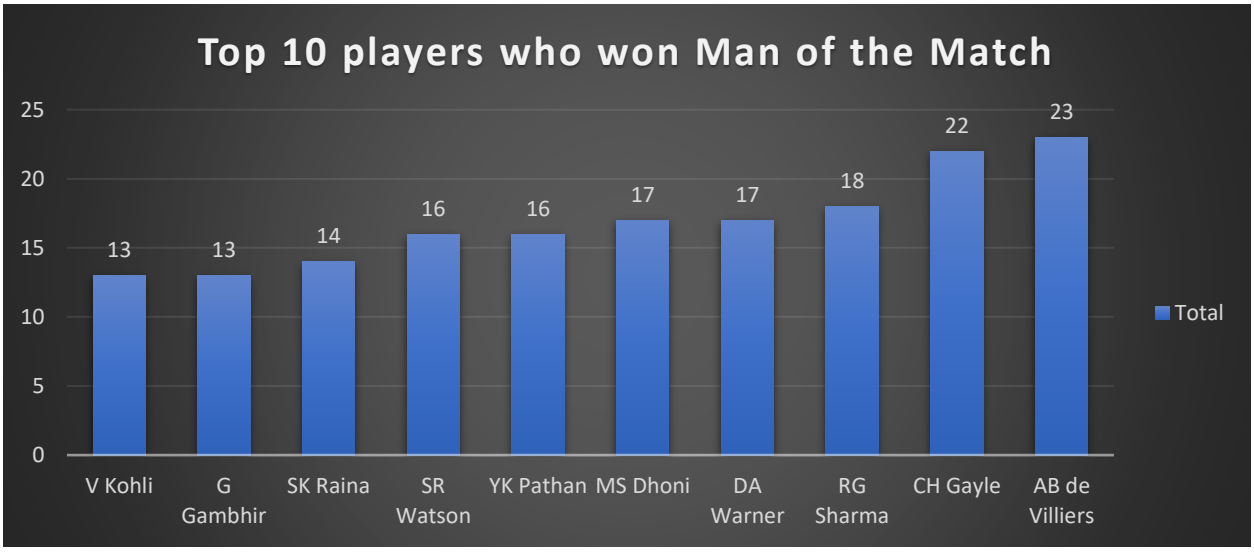
1.Top 10 players who won Man of the Match:

a)Introduction: In this objective we have listed the top 10 players who won more number of times the Man of the Match player .

b)Specific Requirements:

- **i.** Inserted a Pivot table from the data and filtered the Top10 players who won more number of Man of the Match.
- **ii.** With the help of data Column chart is plotted.
- **iii.** With the help of slicer of city and season we can know which player won more man of the matches in a particular city or column.

c) Analysis Results: AB de Villiers is the player who is more number of times Man of the match in all seasons and in all the cities.



2.Total Number of matches won by each team

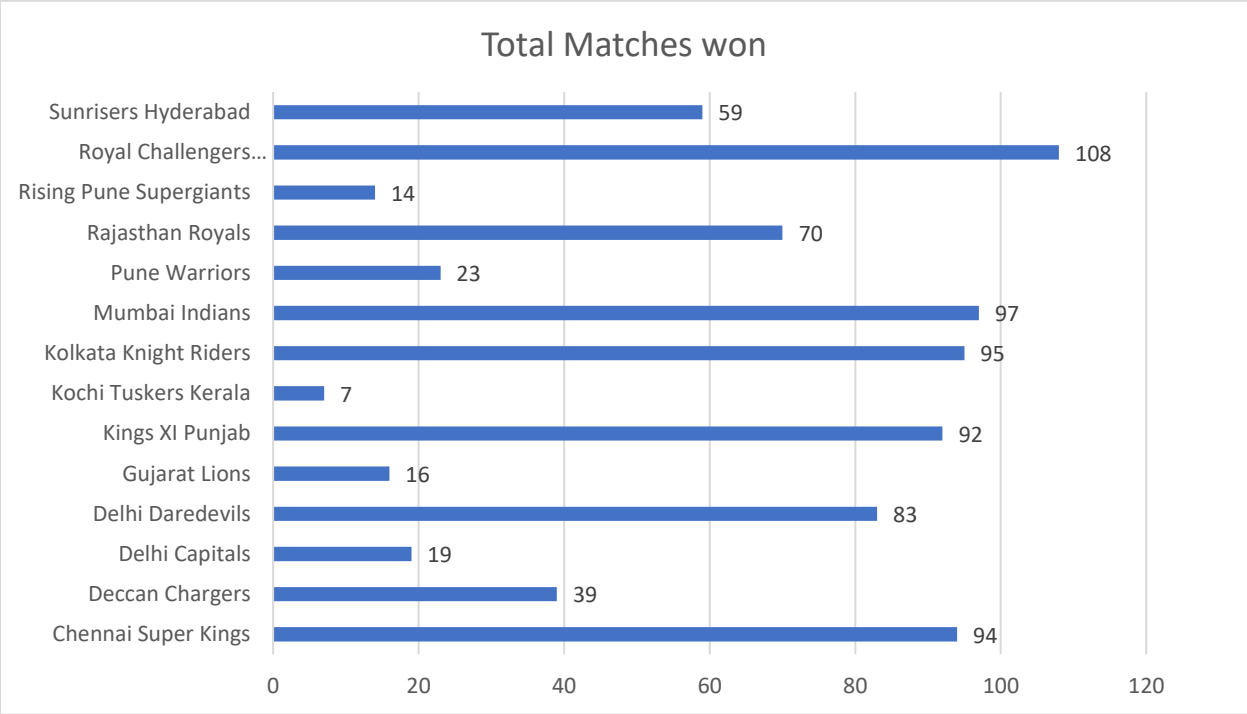
a) Introduction: In this objective I had showed that the count of winnings of each team

b) Specific Requirements:

i. Inserted a Pivot table from the data and filtered the Top10 players who won more number of Man of the Match.
ii. With the help of data Bar chart is plotted.
iii. With the help of slicer of city and season we can know the total winning of each team

c) Analysis Results:

RCB is the team who won most number of matches in all seasons and cities.



3.Total Number of matches won by each team

a) Introduction: In this objective I had showed that the count of winnings of each team by choosing bowling first and by batting first.

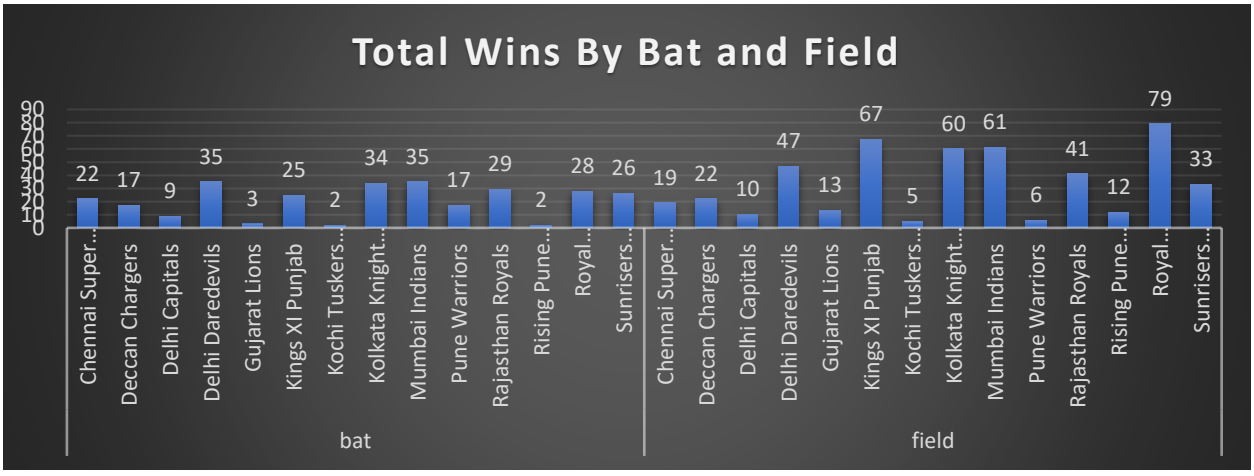
b) Specific Requirements:

- **i.** Inserted a Pivot table from the data and filtered the Top10 players who won a greater number of Man of the Match.
- **ii.** With the help of data Bar chart is plotted.
- **iii.** With the help of slicer of city and season we can know the total winning of each team by bowling first and batting first.

c) Analysis Results:

RCB is the team who won more number of matches by choosing bowling first

CSK is the team who won more number of matches by choosing batting first.



4.Toss Vs Match:

a)Introduction: In this Objective, We can see the Number of Matches :

1)Won Toss and Won Match

2)Won Toss and Lost Match

3)Lost Toss and Won Match

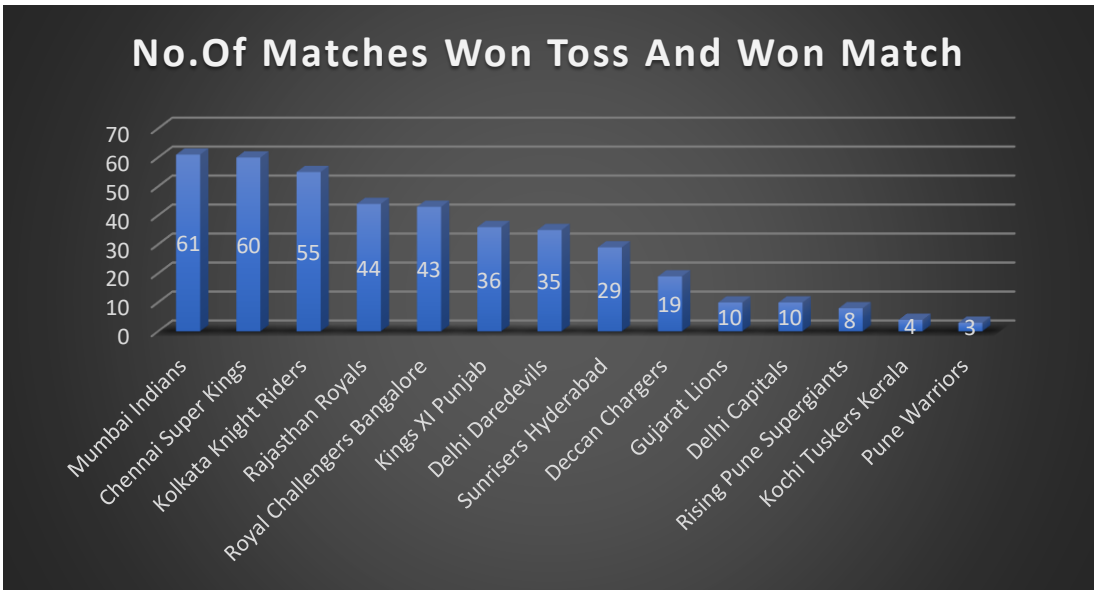
4)Lost Toss and Won Match by respective Teams.

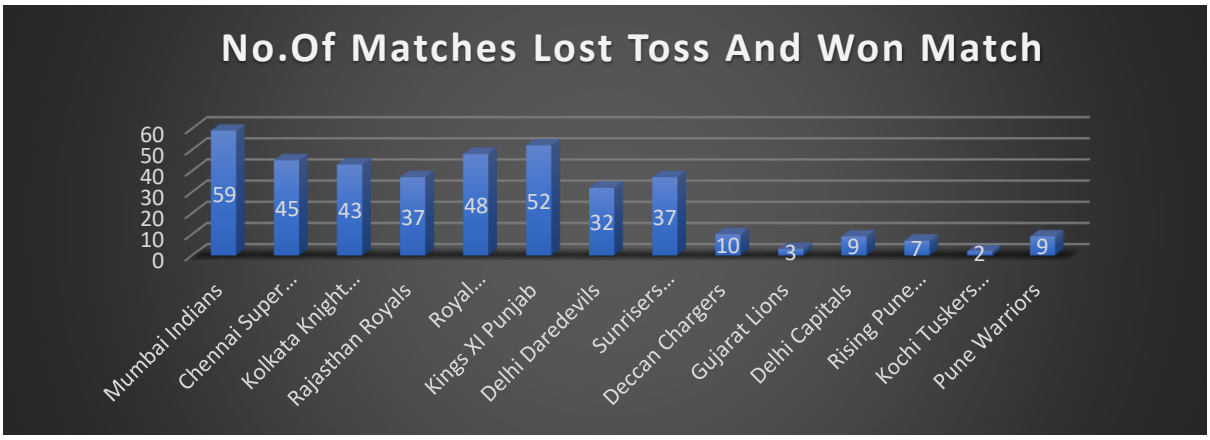
b) Specific Requirements/Functions and Formulas:

- Index,Match,CountIf functions are used.
- I have created a two columns of Toss Looser and Game Looser with the help of If function.
- With the help of **COUNTIFS** function I have got Number of matches 1)Won toss and Won match 2) Lost Toss and Lost Match 3) Won Toss and Lost Match 4) Lost Toss and Won Match by the respective teams.
- I have created a Dropdown List in which we can select the above four listed categories and with the help of **Index** and **Match** functions I created a formula which will display the count of matches played by respective teams who
- 1)won toss and won match
- 2)lost toss and won match
- 3)lost toss and lost match
- 4) won toss and lost match.

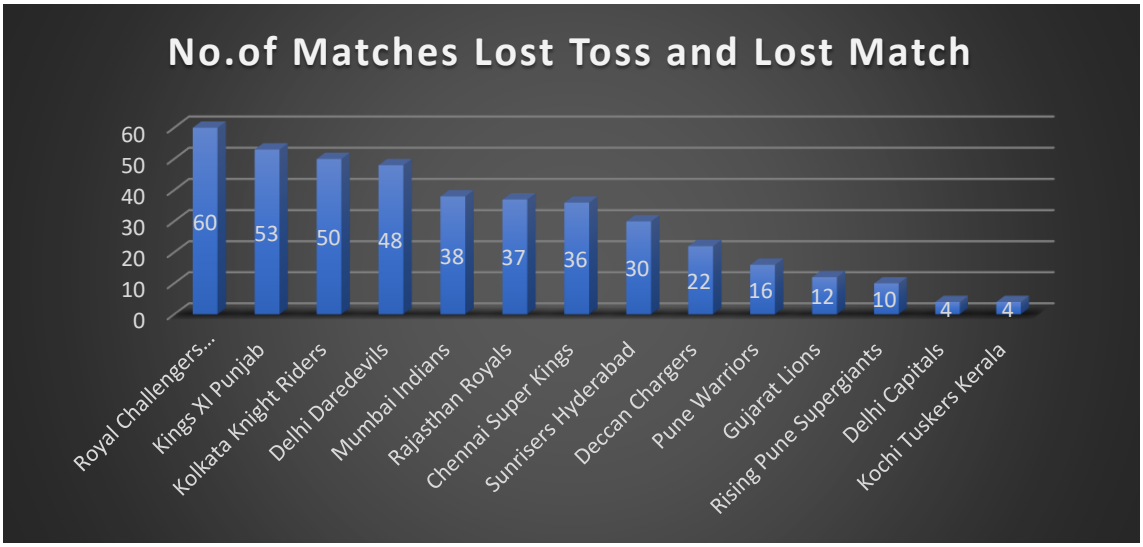
c) Analysis Results:

- Mumbai Indians had won more Number of Matches Winning Toss and also By Loosing Toss.

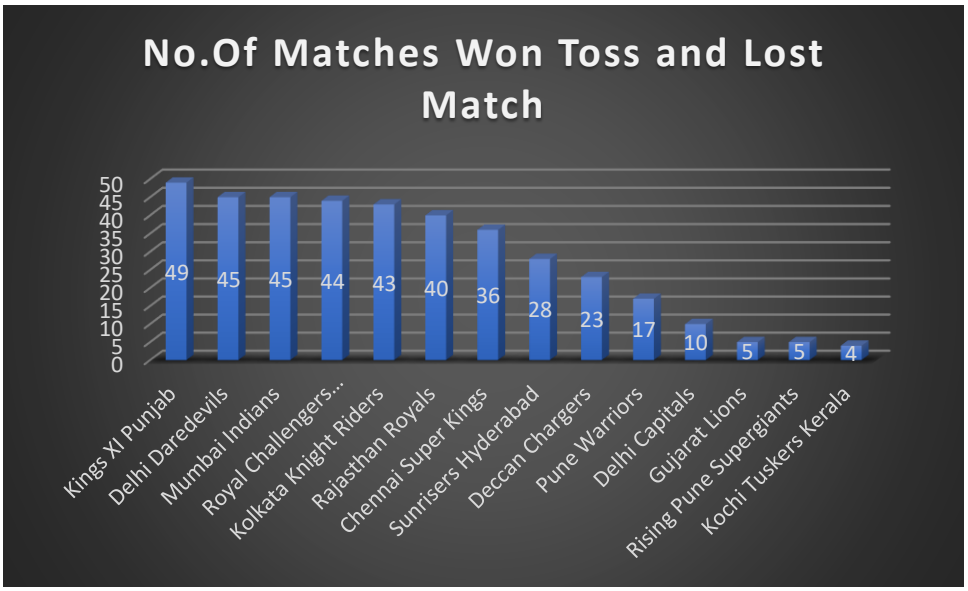




- Royal Challengers Bangalore had Lost more number of matches By Loosing Toss.



- Kings XI Punjab had Lost many Matches although By Winning Toss.



5.Matches won by Team by Huge Margin : (By runs and By wickets**)**

a) Introduction: The analysis shows the Teams who won matches By huge margins (i.e, by more runs and wickets) and the Teams who won matches by Low margin (i.e.., by least runs) while defending. And the Teams who won matches by more wickets and a smaller number of wickets while chasing.

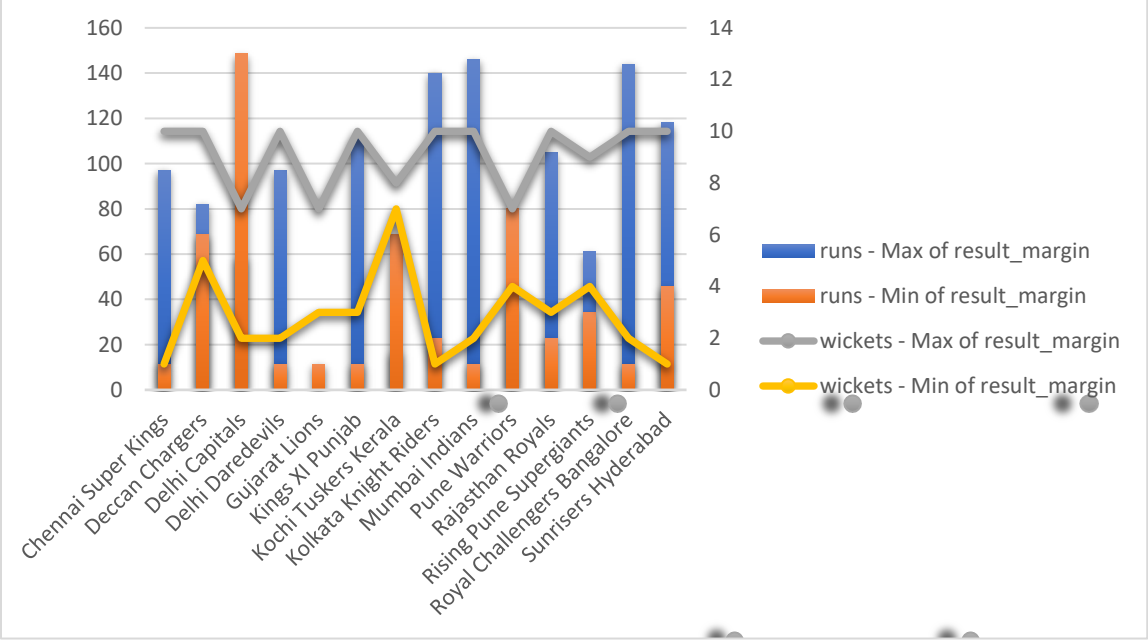
b)Specific Requirements/Functions and Formulas:

i. Pivot table of the data of Import

ii. With the help of data Combined Chart (i.e.., column and line chart) is plotted.

c) Analysis Results:

Chennai Super Kings is the team who won by huge margin of 97 runs while defending, while chasing also it has won the match without loosing a single wicket.



6.Top 10 Umpire Standings:

a) Introduction: The analysis shows the Top10 Umpire standings as Umpire1 and in other Top10 Umpire standings as Umpire 2 in IPL matches 2008-2020.

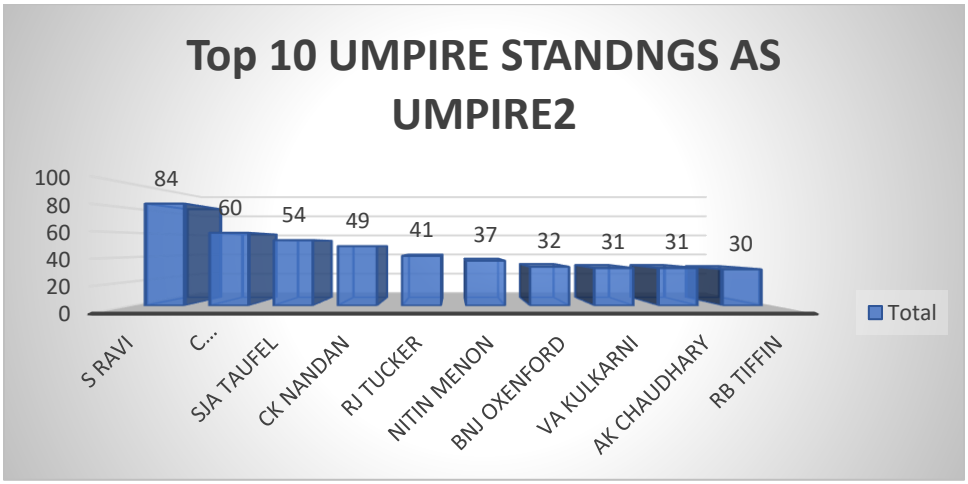
b) Specific Requirements/Functions and Formulas:

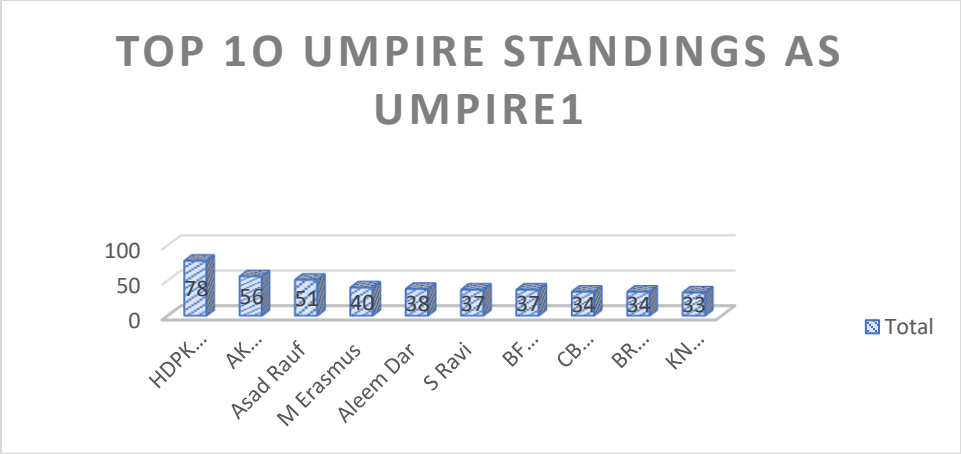
iv. Inserted a Pivot table from the data and filtered the Top10 umpire standings as umpire1 and umpire2.

v. With the help of data Column chart is plotted.

c)Analysis Results with Visualization:

- **HDPK DHARMASENA** is the name of umpire who did the umpiring 78 times as umpire and he is the Topmost umpire in IPL matches 2008-2020
- **S RAVI** is the name of umpire who did the umpiring 84 times as umpire2 and he is the Topmost umpire in IPL matches 2008-2020.





List of Analysis with results

- AB de Villiers is the player who is a greater number of times Man of the match in all seasons and in all the cities.
- RCB is the team who won the greatest number of matches in all seasons and cities.
- Chennai Super Kings is the team who won by huge margin of 97 runs while defending, while chasing also it has won the match without losing a single wicket.
- Mumbai Indians had won a greater Number of Matches Winning Toss and By Loosing Toss.
- Royal Challengers Bangalore had Lost a greater number of matches By Loosing Toss.

- Kings XI Punjab had Lost many Matches although By Winning Toss.

- **HDPK DHARMASENA** is the name of umpire who did the umpiring 78 times as umpire and he is the Topmost umpire in IPL matches 2008-2020
- **S RAVI** is the name of umpire who did the umpiring 84 times as umpire2 and he is the Topmost umpire in IPL matches 2008-

Conclusion

From this subject I learned lot of things such that how to manage data, visualizing data, preparing data etc. Using excel, tableau how we can handle big data in these platforms.

This whole subject is containing lot of things such that ETL process in data warehouse, cleaning data, preparing data, visualizing data, formulas. Using these methods, we can manage; handle big data and visualize the data.

In this project I came to know that how to make attractive dashboard using pivot tables and pivot charts. Using these methods and functions we can do data management easily in near future.

References

1. **www.kaggle.com**
2. **www.youtube.com**
3. **www.github.com**
4. **www.google.com**

Bibliography

- Following book is used to have an idea about Excel Topics:

- Microsoft Excel 2016 Bible: The Comprehensive Tutorial Resource by John Walkenbach, Wiley

- Following Links have been used for completion of this project:

- www.kaggle.com
- www.geeksforgeeks.com