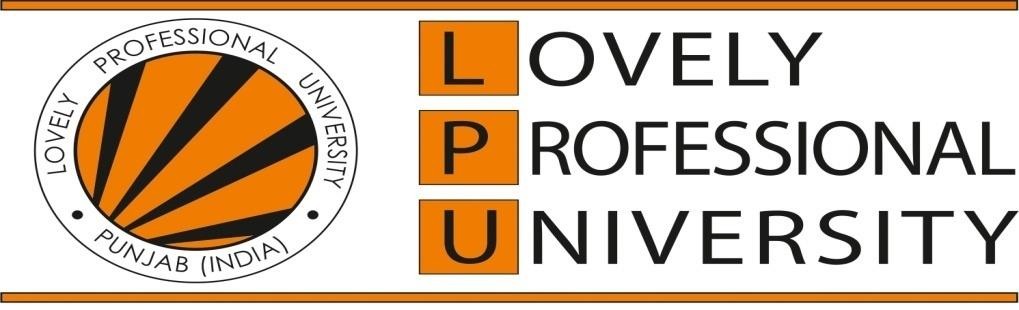
DATA MANAGEMENT PROJECT REPORT

(Project Semester: August-December 2021)



**IPL DATA ANALYSIS**

Submitted By

ROHAN PANDEY

11902091

Btech (CSE), Section-KM011

Course Code: INT217

Under the guidance of

Ms Madhuri -23525

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that **Rohan Pandey** bearing Registration no. **11902091** has completed **INT217** project titled, **“IPL DATA ANALYSIS”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor-**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

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

Text, letter

Description automatically generated

Date: 10.12.2021

Registration No. 11902091 Name: Rohan Pandey

**ACKNOWLEDGEMENT**

I would like to express my special thanks of gratitude to my teacher Ms Madhuri who gave me the golden opportunity to do this wonderful project of “IPL DATA ANALYSIS” which also helped me in doing a lot of research and I came to know about so many new things. I avail this opportunity to extend my heart full thanks and deep respect to faculty member for their able guidance during this project.

I would like to communicate a deep sense of gratitude to all these people without whom my project would not have been such a great learning experience.

**TABLE OF CONTENT**

1. Introduction
2. Objectives/Scope of the Analysis
3. Source of dataset
4. ETL process
5. Analysis on dataset (for each analysis)
6. Introduction
7. General Description
8. Specific Requirements, functions and formulas
9. Analysis results
10. Visualization
11. List of Analysis with results
12. References

**INTRODUCTION**

IPL is a professional Twenty20 cricket league founded by the Board of Control for Cricket in India (BCCI) in 2008. The league has 8 teams representing 8 different Indian cities or states. It enjoys tremendous popularity and the brand value of the IPL in 2019 was estimated to be ₹475 billion (US$6.7 billion). IPL has an exclusive window in ICC Future Tours Programs.

There have been fourteen seasons of the IPL tournament. The current IPL title holders are the CSK, who won the 2021 season. The most successful franchises in the tournament are the CSK and MI with 3 tournament wins each. We will analyse the data of seasons through 2008 to 2020.

Dataset

INDIAN PREMIER LEAGE ANALYSIS contains the following data fields: -

• Date – Contains date on which the match is held.

• Stadium – Stadium in which match is held.

• City – City in which match is held.

• Home Team Name – Host team name.

• Toss-Winner –Who is the toss winner.

• Toss Decision –What team decided after winning the toss.

• Winner – Who won the match.

• Result – This show by runs/wicket team won.

• Win margin– This show by how many runs/wickets does the team win.

• Player of the match – Who won the player of the match.

• ID– This column shows the match number.

**SCOPE OF ANALYISIS**

This project on Indian Premiere League Statistics of India provides the overall Statistics details of the matches of IPL and teams progress in various aspects from the year 2008 to 2020.

**Objectives:**

* To find the team that won the most number of matches in a season.
* Does winning toss increases the chances of victory.
* To find the player with the most player of the match awards.
* To find the venue that hosted the maximum number of IPL matches.
* To find the trend of toss decisions taken by the teams

Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively and efficiently.

At the end of this project we

• get good hands on excel.

• learn to use different feature and get friendly with the excel.

• how to use pivot table and pivot chart.

• Learn to make dashboard in excel.

• To make different type of graphs in excel.

• To learn how to fetch data from other source to excel in different formats.

**Source of Data**

* The dataset is taken from the Kaggle with the name “IPL complete dataset”

<https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020>

* Author- Prateek Bhardwaj

**ETL Process**

In computing, extract, transform, load (ETL) is a process in database usage to prepare data for analysis, especially in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions. ETL stands for Extract, Transform and Load.

Before ETL, the dataset looked like this. This data is taken from [Kaggle](https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020).

Graphical user interface

Description automatically generated

Through the process of ETL, we are going to clean the dataset and bring all the entities to their proper data format.

1. **Getting the data from the downloaded csv file in excel.**

From the data tab in excel under get and transform data->chosen get data->from csv file.

A screenshot of a computer

Description automatically generated

1. **Selecting the transform option to removing columns which are not properly defined or not crucial to our analysis.**

Graphical user interface, application, table, Excel

Description automatically generated

1. **Loading the transformed data to the excel sheet**

A screenshot of a computer

Description automatically generated

**ANALYSIS OF DATA**

* **Season winners**

**Description:**

No. of matches won by different teams in each season. Here we compare wins and thus year wise seasons corresponding to the no. of matches the teams won.

**Specific function and requirements:**

1)Pivot table of Team winner vs year

2)Stacked column chart

**Analysis Result**-

**A picture containing chart

Description automatically generated**We can see the no. of matches won by each team in different seasons with help of slicer.

**Visualization-** Created a stacked column bar graph

**A picture containing chart

Description automatically generated**

* **Venue**

**Description:**

No. of matches hosted by different stadiums. Therefore, graph between stadium and total no. of matches played in each stadium.

**Specific function and requirements:**

1)Pivot table of venue

2)Clustered bar

**Analysis result:**

Here we see the most no. of matches hosted by each stadium. Eden gardens stadium hosted most number of matches throughout all seasons.

**Visualization:**

**Chart

Description automatically generated**

* **Win Probability if won toss**

**Description:**

Checking the probability of a team winning if the team has won the toss.

**Specific function and requirements:**

1)Table with columns team, total no of wins, total no. of toss wins of the corresponding team

2)(for wins) =COUNTIF function

3)(for toss wins) =COUNTIFS function

4)division formula for probability

5)pivot chart

6)Stacked area chart

Table

Description automatically generated**Analysis Result:**

Observing the chart and calculating the percentage **we find that chance of winning is 51.09% if the team has won the toss**.

**Visualization:**

**A picture containing chart

Description automatically generated**

* **Toss Decisions:**

**Description:**

Understand the trend in the decisions of team captains to either chose to bat or to field

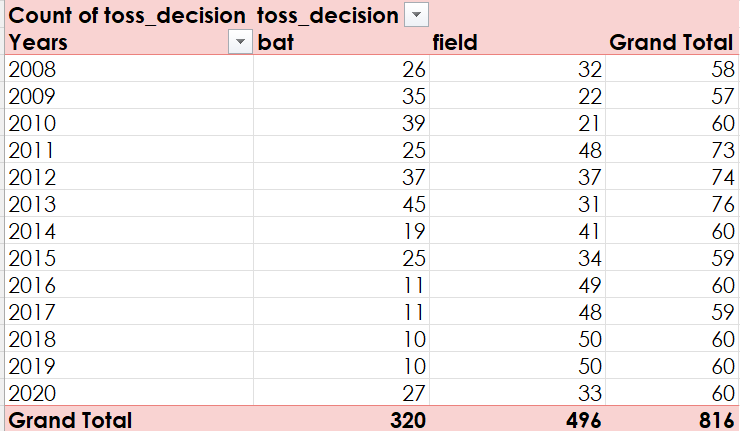
**Specific function and requirements:**

1)Pivot table

2)Line with markers chart

**Analysis Result:**

We see before 2014 teams tended to choose batting but recently the teams more often choose to field first.

****

**Visualization:**

Chart, line chart

Description automatically generated

* **Player of the match:**

**Description-** See the players winning most no of titles of player of the match.

**Specific function and requirements:**

1)Pivot table

2)Clustered bar chart

**Analysis Result and Vizualization:**

We observe that AB de Villers has won this title most number of times followed by Chris Gayle.

**Chart

Description automatically generated with medium confidence** A picture containing text

Description automatically generated

**Final Dashboard**

**Graphical user interface

Description automatically generated**

**List of Analysis with Results**

* Probability to win a match if the team has won the toss is 51.23%
* Eden gardens stadium has hosted most number of ipl matches i.e. 77 till 2020
* Ab devillers has won the title of player of the match the greatest number of times followed by Chris Gayle.
* We see before 2014 teams tended to choose batting but recently the teams more often choose to field first.

**References**

* Kaggle(<https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020>)
* Stock image for background
* Youtube