A picture containing icon

Description automatically generated­



ANALYSTS

Data Visualization and Exploratory Data Analysis of Indian Premier League

(2008 - 2020)

Anand Prakashbhai Shroff

Ravi Ravi

Santhosh Pachaiyappan

Simranjeet Kaur

Vignesh Dharuman

**TABLE OF CONTENTS**

1. [INTRODUCTION 3](#_TOC_250014)
   1. [Background 3](#_TOC_250013)
   2. [Problem Statement. 3](#_TOC_250012)
   3. [Scope 3](#_TOC_250011)
   4. [Document Overview 4](#_TOC_250010)
2. [PRELIMINARY REQUIREMENTS 4](#_TOC_250009)
3. [TECHNICAL APPROACH 4](#_TOC_250008)
   1. [Analysis 4](#_TOC_250007)
   2. [Requirements Development 5](#_TOC_250006)
   3. [Model Development 5](#_TOC_250005)
   4. [Testing and Evaluation 5](#_TOC_250004)
   5. [Delivery 5](#_TOC_250003)
4. [DESCRIPTIVE STATISTICS 4](#_TOC_250008)
   1. [Approach and Table descriptions 4](#_TOC_250007)
5. [EXPLORATORY DATA ANALYSIS 4](#_TOC_250008)
   1. Match [Analysis 4](#_TOC_250007)
   2. [Players Analysis 5](#_TOC_250006)
6. [EXPECTED RESULTS](#_TOC_250002) 6

7 REFERENCES 7

**1. INTRODUCTION**

**1.1 Background**

After India’s first world cup win in 2007 after 24 years, it was major tournament win for INDIA after a very long time. BCCI (Board of Cricket Control) took this opportunity as a golden chance to start a business of franchise based fast format cricket league which will be known as IPL (Indian Premier League) and the idea was similar to premier league of England and NBA of United states. The tournament has been a successful business venture and excellent instrument for cricket lovers from across the globe and have market value of US $6.3bil dollars in last 13 years which is an explosive growth.

The current dataset is a on which team FIVE ANALYST will be working is a clustered data which is separated in 8 excel data files and data consisting of Ball-by-ball performance players, team performances, Win: Lose rate of home vs away matches and player details.

There are final goals we want to achieve with the help of this project are as follows:

Develop game pre- and post-game analysis report for commentators, team stations, sports analysts, and cricket experts, which will help teams to plan their line-up before game and get better knowledge of terrain they are playing in.

Sportsmen and sportswomen have always been one of the prime choices for marketing any products or taking place of brand ambassadors. With the help this project’s predictive analysis companies will get better idea about selecting best players for their product campaigns.

There are many mobile apps such as Dream 11, MPL, Paytm first games and WINzo where players around the world win big cash prizes for their small investments if their players they select perform well in match. Cricket enthusiasts can find this visualization useful as a direction to put their investments on.

**1.2 Problem Statement**

It is crucial for every franchisee to analyse and evaluate past performances of their team so as to focus and improve on their weak and strong aspects. Our team will be performing a descriptive data analysis of IPL matches played between 2008-2020 in order to gain meaningful insights and recognize patterns and relationships between various attributes such as how many times team winning the toss won the match as well, how batting first or bowling first affects chances of winning, number of matches won at home ground and away, dependency on top order and middle order to perform well in batting, etc.

Scope

This project has a vast amount of data to work with and develop some helpful insights based on different aspects, our project team will be working on the task with data which is vital to predict certain outcomes and will eventually be able to be visualized

**Within Scope:**

Constructing a model based on the players and the average runs they scored to predict the outcome of their performance in the upcoming game.

Comparing the home wins and away wins of multiple teams to visualize which team is stronger on the home ground and which one is stronger on the away ground, and which are all keeping up in both, these visualizations will help figure out the strong winning ratios of teams based upon the ground they play at.

We can only take into consideration of those data which can be visualized and provide wide insight into the task at hand.

**Outside of Scope:**

Data which cannot be visualized with the task for example non striker, match id, innings and so on.

Data which are not in accordance with conditions laid above.

**1.3 Document Overview**

The dataset we are using for project is IPL data from year 2008-2020. The main requirements and results of project are sub-divided into various categories which are explained in this proposal under name of Preliminary requirements, technical approach, expected results and management approach. Each section explains how our team is going to use step by step approach to complete this project by describing problem statement and achieve the scope of this project.

**2. PRELIMINARY REQUIREMENTS**

**Data Requirement:**

The complete data of India premier league from 2008 to 2020, it should have Bowling stats, batting stats, Season details, ball by ball match information, venue details etc.

The dataset is collected from below links

[**https://www.kaggle.com/ramjidoolla/ipl-data-set**](https://www.kaggle.com/ramjidoolla/ipl-data-set)

[**https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020**](https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020)

[**https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all\_season\_batting\_card.csv**](https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all_season_batting_card.csv)

[**https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all\_season\_bowling\_card.csv**](https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all_season_bowling_card.csv)

**Business Requirement**

|  |  |  |
| --- | --- | --- |
| **S.no** | **Functional requirement** | **Comments** |
| 1 | Visualize the data in graphs of each season winners, toss winners and total number of wins |  |
| 2 | Geographical visualization of venues, player progress, team progress and history of venues |  |
| 3 | Plot graphs, bar chart, bubble chart, pie chart of all the data provided |  |
| 4 | Pictorial representation of player stats from top to least |  |
| 5 | Represent dashboard with tabs such as Team, venue, Player, and seasons |  |
| 6 | Home dashboard with all key details and links to details information that navigates to more stats |  |
| 7 | Complete dashboard in tableau which should look like a website |  |

**3.TECHNICAL APPROACH**

Diagram

Description automatically generated

Figure 1.Technical approach used in this project

**3.1 Analysis**

In order to carry out this analysis successfully team has to understand the following points:

Learn to perform the ETL (Extract Transform Load) basics before starting to analyze data, how to implement the variables in the visualization.

Learn about the relation between all csv sheets and how to join key variables across the dataset, which variable to choose for dataset displaying and which to ignore.

Team also has to learn software like tableau and excel analysis and how to use them to draw visualizations from the clustered dataset.

**3.2 Requirements Development**

Knowledge of Microsoft Office tools which will be used for understanding the data and its attributes thoroughly. It will help in scoping and identifying potential relationships to be explored within the dataset.

Understanding and working of MS Excel and Python for data cleaning, preparation, and visualization.

Working with Tableau, where the dashboard will be developed based on the project

**3.3 Model Development**

The first thing to consider in the development of the model is to take the inputs which are needed, which can be achieved by analysing the complete data set and by taking all the aspects into consideration. The player average and the winning ratios of the teams will be visualized by patterns of graphs.

Further visualizations will be done by sorting out the data in Excel by combing through each and every aspect of individual rows and columns and then with the use of python we will be manipulating the data to figure out more useful insights as it gets to the point where the visualization can be done, we will be using tableau to do the final visualization.

**3.4 Testing and Evaluation**

An evaluation of whether the data model visualize the correct information and in right manner is dashboard. Testing and evaluation of product is categorized into multiple process such as Cosmetics bugs which includes spelling, grammatical and presentation. And data testing, plotting and graphs also tested and evaluated. Input data and output result also evaluated in each step.

Testing and evaluation can be split into multiple phases as follows

* **Unit Testing:**

In this phase, data is tested, and input and output of each plotting is evaluated. Testing methods used are Component testing, UI testing and Sanity testing

* **Integration Testing:**

In this phase, Plots and graphs are integrated one by one and evaluated. Testing methods used are integration testing, Regression testing, Smoke testing and Sanity testing

* **System Testing:**

In this phase, Plots and graphs are integrated into one dashboard in local server and evaluated. Testing methods used are integration testing, Regression testing, Smoke testing and Sanity testing, compatibility test, load test, Performance test.

* **User acceptance testing:**

In this phase, the pilot product is prepared and is hosted online and evaluated as user. Testing methods used are Regression testing, Sanity testing, load test, Performance test and connectivity testing and UAT testing.

**3.5 Delivery**

Final delivery will be done as a tableau dashboard which will be the in the form of dashboard charts which will clearly depict the conclusions drawn by the dataset. The results will have information about relation of Ground conditions with player performances, win ratio, player stats, player growth and will also work on match win prediction, and player performance stats in relation the terrain (ground) or certain bowler types.

**4. DESCRIPTIVE STATISTICS**

**4.1 Approach and Data descriptions**

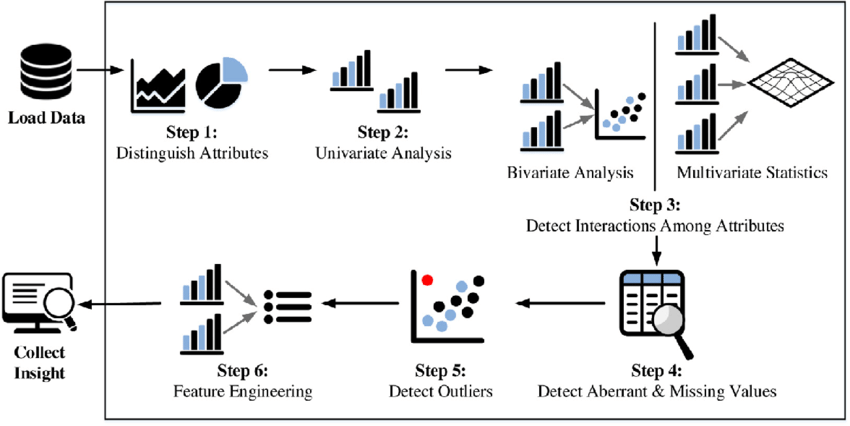


Figure 2. Steps followed in Descriptive statistics

The above steps are followed in identifying the dataset details which are categorized in match stats and player stats

Diagram

Description automatically generated

Figure 3. Variables split up

Table 1. IPL Match (2008-2020) Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column name | Data Type | Count | Frequency (Unique) | Null values | NA values | Variable type |
| id | Int64.Type | 816 | 816 | 0 | 0 | Continuous |
| city | type text | 816 | 33 | 0 | 13 | Categorical |
| date | type date | 816 | 596 | 0 | 0 | Continuous |
| player\_of\_match | type text | 816 | 234 | 0 | 4 | Categorical |
| venue | type text | 816 | 36 | 0 | 0 | Categorical |
| neutral\_venue | Int64.Type | 816 | 2 | 0 | 0 | Categorical |
| team1 | type text | 816 | 15 | 0 | 0 | Categorical |
| team2 | type text | 816 | 15 | 0 | 0 | Categorical |
| toss\_winner | type text | 816 | 15 | 0 | 0 | Categorical |
| toss\_decision | type text | 816 | 2 | 0 | 0 | Categorical |
| winner | type text | 816 | 16 | 0 | 4 | Categorical |
| result | type text | 816 | 4 | 0 | 4 | Categorical |
| result\_margin | type text | 816 | 92 | 0 | 17 | Categorical |
| eliminator | type text | 816 | 3 | 0 | 4 | Categorical |
| method | type text | 816 | 2 | 0 | 797 | Categorical |
| umpire1 | type text | 816 | 48 | 0 | 0 | Categorical |
| umpire2 | type text | 816 | 47 | 0 | 0 | Categorical |

Table 2. Teamwise\_home\_and\_away.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column name | Data Type | Count | Frequency (Unique) | Null values | NA values | Variable type |
| team | type text | 14 | 14 | 0 | 0 | Categorical |
| home\_wins | Int64.Type | 14 | 14 | 0 | 0 | Continuous |
| away\_wins | Int64.Type | 14 | 13 | 0 | 0 | Continuous |
| home\_matches | Int64.Type | 14 | 14 | 0 | 0 | Continuous |
| away\_matches | Int64.Type | 14 | 13 | 0 | 0 | Continuous |
| home\_win\_percentage | type number | 14 | 14 | 0 | 0 | Continuous |
| away\_win\_percentage | type number | 14 | 14 | 0 | 0 | Continuous |

Table 3. Descriptive statistics of teamwise\_home\_and\_away

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Stats | *Home wins* | *Away wins* | *Home matches* | *Away matches* | *Home win percentage* | *Away win percentage* |
| Mean | 23.92857143 | 29.42857143 | 53.5 | 53.5 | 41.73766314 | 55.57428267 |
| Standard Error | 4.993913092 | 5.563533431 | 9.573745004 | 9.590374066 | 3.767120029 | 3.653375494 |
| Median | 27 | 35 | 65 | 60 | 41.80935344 | 58.40116279 |
| Mode | #N/A | 49 | #N/A | 95 | #N/A | #N/A |
| Standard Deviation | 18.68551181 | 20.81683596 | 35.82167371 | 35.88389397 | 14.09527248 | 13.6696794 |
| Sample Variance | 349.1483516 | 433.3406593 | 1283.192308 | 1287.653846 | 198.6767064 | 186.860135 |
| Kurtosis | -0.946104413 | -1.97836982 | -1.736496419 | -1.948633501 | 1.723415964 | 1.496333299 |
| Skewness | 0.280532405 | -0.086871449 | -0.255905168 | -0.154059133 | -0.906750499 | -1.139749427 |
| Range | 57 | 54 | 95 | 88 | 55.35714286 | 51.92307692 |
| Minimum | 1 | 4 | 6 | 7 | 7.142857143 | 23.07692308 |
| Maximum | 58 | 58 | 101 | 95 | 62.5 | 75 |
| Sum | 335 | 412 | 749 | 749 | 584.3272839 | 778.0399574 |
| Count | 14 | 14 | 14 | 14 | 14 | 14 |

Table 4. IPL Ball-by-Ball 2008-2020

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column name | Datatype | Count | Frequency (Unique) | Null Values | NA Values | Variable type |
| id | Int64.Type | 193468 | 816 | 0 | 0 | continuous |
| inning | Int64.Type | 193468 | 2 | 0 | 0 | Numerical |
| over | Int64.Type | 193468 | 20 | 0 | 0 | Numerical |
| ball | Int64.Type | 193468 | 9 | 0 | 0 | continuous |
| batsman | Type.Text | 193468 | 537 | 0 | 0 | categorical |
| non\_striker | Type.Text | 193468 | 530 | 0 | 0 | categorical |
| bowler | Type.Text | 193468 | 420 | 0 | 0 | categorical |
| batsmans\_runs | Int64.Type | 193468 | 7 | 0 | 0 | continuous |
| extra\_runs | Int64.Type | 193468 | 7 | 0 | 0 | continuous |
| total\_runs | Int64.Type | 193468 | 8 | 0 | 0 | continuous |
| non\_boundary | Int64.Type | 193468 | 2 | 0 | 0 | continuous |
| is\_wicket | Int64.Type | 193468 | 2 | 0 | 0 | categorical |
| dismissal\_kind | Type.Text | 193468 | 10 | 0 | 183973 | categorical |
| player\_dismissed | Type.Text | 193468 | 507 | 0 | 183973 | categorical |
| fielder | Type.Text | 193468 | 880 | 0 | 186684 | categorical |
| extras\_type | Type.Text | 193468 | 6 | 0 | 183235 | categorical |
| batting\_team | Type.Text | 193468 | 15 | 0 | 0 | categorical |
| bowling\_team | Type.Text | 193468 | 16 | 0 | 191 | categorical |

Table 5. All Season Bowling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column Name | Data Type | Count | Frequency | Null Values | NA values | Variable type |
| Season | Int64.Type | 9967 | 13 | 0 | 0 | Numeric(Int) |
| Match\_id | Int64.Type | 9967 | 816 | 0 | 0 | Numeric(Int) |
| Match\_name | type text | 9967 | 111 | 0 | 0 | Categorical |
| Home\_Team | type date | 9967 | 12 | 0 | 0 | Categorical |
| Away\_Team | type text | 9967 | 12 | 0 | 0 | Categorical |
| Bowling\_team | type text | 9967 | 12 | 0 | 0 | Categorical |
| Venue | type text | 9967 | 12 | 0 | 0 | Categorical |
| city | type text | 9967 | 12 | 0 | 0 | Categorical |
| country | type text | 9967 | 33 | 0 | 0 | Categorical |
| innings\_id | Int64.Type | 9967 | 12 | 0 | 0 | Categorical |
| name | type text | 9967 | 33 | 0 | 0 | Categorical |
| bowling\_style | type text | 9967 | 31 | 0 | 0 | Categorical |
| bowler\_country | type text | 9427 | 33 | 0 | 0 | Categorical |
| full\_name | type text | 9967 | 31 | 0 | 0 | Categorical |
| overs | Int64.Type | 9967 | 3 | 0 | 0 | Numeric(Float) |
| maidens | Int64.Type | 9967 | 31 | 0 | 0 | Numeric(Int) |
| conceded | Int64.Type | 9967 | 3 | 0 | 0 | Numeric(Int) |
| wickets | Int64.Type | 9967 | 2 | 0 | 0 | Numeric(Int) |
| economyRate | Int64.Type | 9966 | 3 | 1 | 0 | Numeric(Float) |
| dots | Int64.Type | 9967 | 2 | 0 | 0 | Numeric(Int) |
| fours conceded | Int64.Type | 9967 | 419 | 0 | 0 | Numeric(Int) |
| sixes conceded | Int64.Type | 9967 | 2 | 0 | 0 | Numeric(Int) |
| wides | Int64.Type | 9967 | 419 | 0 | 0 | Numeric(Int) |
| no balls | Int64.Type | 9967 | 13 | 0 | 0 | Numeric(Int) |
| captian | Type Boolean | 9967 | 419 | 0 | 0 | Boolean |
| herf | type text | 9967 | 13 | 0 | 0 | String |

Table 6.Ball by Ball Descriptive statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Runs* | *Fours* | *Sixes* | *Balls Faced* |
| Mean | 19.31458 | 1.762779 | 0.71593 | 15.10183 |
| Standard Error | 0.18879 | 0.020524 | 0.011699 | 0.122125 |
| Median | 12 | 1 | 0 | 11 |
| Mode | 0 | 0 | 0 | 1 |
| Standard Deviation | 21.04223 | 2.287568 | 1.303937 | 13.61184 |
| Sample Variance | 442.7756 | 5.232966 | 1.700251 | 185.2821 |
| Kurtosis | 2.787055 | 4.27037 | 12.85916 | 0.914698 |
| Skewness | 1.599561 | 1.857772 | 2.936472 | 1.217949 |
| Range | 175 | 19 | 17 | 73 |
| Minimum | 0 | 0 | 0 | 0 |
| Maximum | 175 | 19 | 17 | 73 |
| Sum | 239945 | 21899 | 8894 | 187610 |
| Count | 12423 | 12423 | 12423 | 12423 |
| Confidence Level (95.0%) | 0.370057 | 0.04023 | 0.022932 | 0.239383 |

Table 7. Player details

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column name | Data Type | Count | Frequency (Unique) | Null values | NA values | Variable type |
| Player\_Name | Text | 566 | 566 | 0 | 0 | Categorical |
| DOB | Date | 471 | 457 | 95 | 0 | Continuous |
| Batting\_Hand | Text | 563 | 2 | 3 | 0 | Categorical |
| Bowling\_Skill | Text | 502 | 14 | 64 | 0 | Categorical |
| Country | Text | 471 | 11 | 95 | 0 | Categorical |

Table 8.Bowling descriptive statistics

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *overs* | *maidens* | *conceded* | *wickets* | *economyRate* | *dots* | *foursConceded* | *sixesConceded* | *wides* | *noballs* |
| Mean | 3.210096204 | 0.028447295 | 25.63711596 | 0.888486604 | 8.266773226 | 7.391538223 | 2.265025344 | 0.92024413 | 0.605979104 | 0.079238647 |
| Standard Error | 0.010467467 | 0.001697265 | 0.108242136 | 0.00997303 | 0.031593638 | 0.039467659 | 0.01696108 | 0.010862507 | 0.009021044 | 0.003180653 |
| Median | 4 | 0 | 25 | 1 | 8 | 7 | 2 | 1 | 0 | 0 |
| Mode | 4 | 0 | 24 | 0 | 8 | 8 | 1 | 0 | 0 | 0 |
| Standard Deviation | 1.029170778 | 0.16687665 | 10.6424646 | 0.980557348 | 3.106154308 | 3.88049588 | 1.667628682 | 1.068011461 | 0.88695715 | 0.312724688 |
| Skewness | -1.022719151 | 5.74048163 | 0.263723613 | 1.056122497 | 1.085529882 | 0.138820402 | 0.725186694 | 1.360744295 | 1.709129068 | 4.657206243 |
| Range | 4 | 2 | 70 | 6 | 36 | 20 | 11 | 7 | 6 | 4 |
| Minimum | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Maximum | 4 | 2 | 70 | 6 | 36 | 20 | 11 | 7 | 6 | 4 |
| Sum | 31032 | 275 | 247834 | 8589 | 79906.63 | 71454 | 21896 | 8896 | 5858 | 766 |
| Count | 9667 | 9667 | 9667 | 9667 | 9666 | 9667 | 9667 | 9667 | 9667 | 9667 |

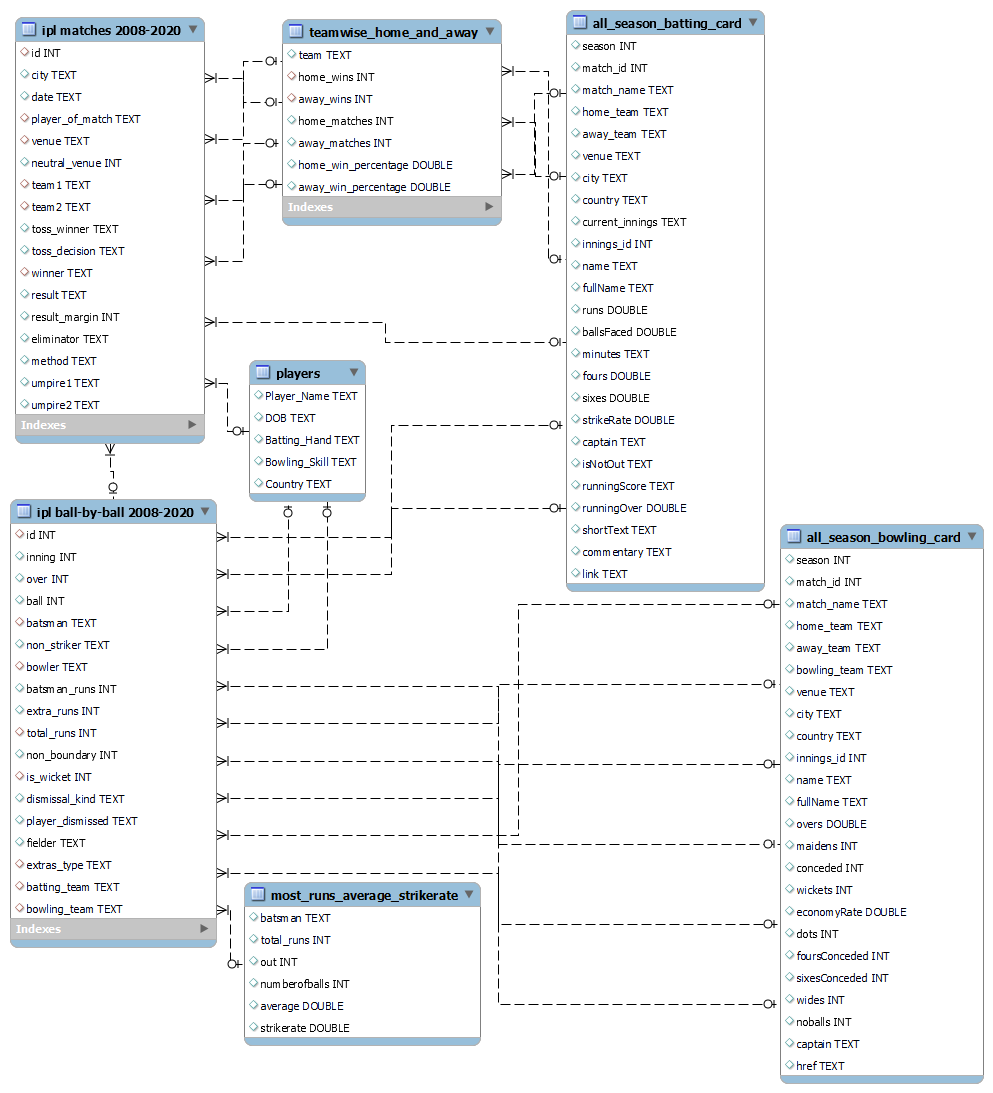


Figure 4. SCHEMA of IPL Dataset

**Outliers**

The IPL is mostly focused on Indian player, so the participants for Indian players will be more compared to foreign players. so, outliers related to players is not considered in this dataset. And the Venues are more concentrated in India though some matches are conducted out of India therefore outliers related to venue and place are still valid.

*Note: NA values in IPL Ball-by-Ball 2008-2020 Table doesn’t means it is not available, it actually linked with this column “dismissal\_kind,* *player\_dismissed, fielder, extras\_type”. For example, if one column has value other 3 columns will be NA.*

**5. EXPLORATORY DATA ANALYSIS**

* 1. **Match Analysis**

**Match and Toss Winners**

IPL is conducted every year where top 8 teams are selected for each season and each team will have minimum 14 matches in a season. Below chart show the number of matches won by each team in all seasons. In this it clearly shows that Mumbai Indian won highest number of match where the Kochi Tuskers Kerala. It also states that top 5 teams are participating in all seasons.

Figure 5. Match Winners

Figure 6. Toss Winners

The above chart describes the number of times each team won toss in each season. Mumbai Indians is the highest and followed by Chennai super kings and Kochi tuskers Kerala are the least. When we compare both chart it clearly shows that team who wins the toss has the higher chances of winning the match.

**Comparison of home and away wins**

The below table and chart describe the number of each team won in home and away ground(venue). Mumbai Indians has good records in home ground whereas Kolkata Knight Riders has good record in Away grounds. On comparing both home and away ground, Mumbai Indians has played a greater number of matches in home ground than the other teams.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Teams** | **Sum of home\_wins** | **Sum of home\_matches** |  | **Row Labels** | **Sum of away\_wins** | **Sum of away\_matches** | | Delhi Capitals | 3 | 6 |  | Kochi Tuskers Kerala | 4 | 7 | | Kochi Tuskers Kerala | 2 | 7 |  | Rising Pune Supergiant | 5 | 8 | | Rising Pune Supergiant | 5 | 8 |  | Pune Warriors | 6 | 26 | | Gujarat Lions | 1 | 14 |  | Delhi Capitals | 7 | 10 | | Pune Warriors | 6 | 20 |  | Deccan Chargers | 11 | 32 | | Deccan Chargers | 18 | 43 |  | Gujarat Lions | 12 | 16 | | Sunrisers Hyderabad | 30 | 63 |  | Sunrisers Hyderabad | 28 | 45 | | Rajasthan Royals | 29 | 67 |  | Delhi Daredevils | 42 | 89 | | Delhi Daredevils | 25 | 72 |  | Kings XI Punjab | 44 | 85 | | Kolkata Knight Riders | 34 | 83 |  | Rajasthan Royals | 46 | 80 | | Royal Challengers Bangalore | 35 | 85 |  | Chennai Super Kings | 49 | 75 | | Chennai Super Kings | 51 | 89 |  | Royal Challengers Bangalore | 49 | 95 | | Kings XI Punjab | 38 | 91 |  | Mumbai Indians | 51 | 86 | | Mumbai Indians | 58 | 101 |  | Kolkata Knight Riders | 58 | 95 | |  |

Figure 7. Home Matches vs Home wins

Figure 8. Away Matches vs Away Wins

**Over analysis**

This data contains all the details of the balls bowled in both innings which also includes the bowler, batsmen, non-striker, fielder, wicket, and boundary details. This sheet also includes columns with NA values which are Player dismissed, dismissal kind, fielder and extras type, the reason behind that is all those data are depending on the outcome of the certain other column for example there will not be any player dismissed if there is no wicket in that ball.

Figure 9. Top 10 Bowlers with Most Extra Runs

This above chart depicts the data of the top 10 bowlers who has given most extra runs in the period of 2008 to 2020, S.L Malinga is the number one player who has given most runs which almost contributes to 34% and this chart can give an idea about how to deploy which player on which over which could be crucial at times when there should be no extra runs to control the unnecessary runs.

Figure 10. Top player's Runs based on overs

This above chart contains detailed data of total runs scored per over for both the innings from which we can identify the innings in which most of the runs is scored and also it will give a clear idea of runs scored in the overs to analyse which overs should be maintained to keep the overall scored leveled and this stat clearly shows which player has good stats based on the over. David warner has highest run percentage in first 5 overs and MS Dhoni has great run percentage in last 5 overs.

* 1. **Players Analysis**

**Batting**

**1. Number of players per Country**

IPL (Indian premier league) is a cricket league hosted by India but players from all countries participate in it under different team names. The chart below shows number of players from different countries which participated in IPL from 2008-2019. From chart it is clear that India has the greatest number of players (264) followed by Australia (72) and South Africa (39).

Figure 11. Number Of Players per Country

**2. Count of Players per country distributed according to batting style**

Figure 12. Country wise Batting style of Players

This chart shows the players distribution per country based on their batting. It’s clear from chart that Australia has most right-hand players after India. In charts it is shown that Netherlands and Zimbabwe have most right-hand players but count of players from both countries is 1 and 2 respectively which is less than Australia.

**3. Number of Boundaries per team**

In this we looked at the boundaries of each team which include Fours and sixes of each team.

Figure 13. Number of Boundaries per Team

This chart depicts data from 2018 – 2020 from where we can interpret that DC had most fours which is 658 and KXIP had least fours which is 545. When we look at sixes of league both KKR and MI have same number of sixes which is 359 and RR and SRH have least sixes which is again same 248.

**4. Highest scores by players**

The 10 highest scorer of IPL league are represented in this pivot chart from 2018 - 2020. KL Rahul scored highest runs in IPL (1922). At second place Shikar Dhawan have 1636. When we notice the pattern in chart the difference between first two highest scorer of runs is more than 200 but after that the difference between players scores is less. The least runs are made by David Warner (1240).

Figure 14. Highest Run Scorers

**6. Total boundaries vs total runs**

Figure 15. Total Boundaries VS Total Runs

This chart depicts data from 2018 – 2020 from where it’s clear that MI have both highest runs (7451) and boundaries (1015). Both RCB and RR have same boundaries, but the runs scored by RR are more than RCB. Again, on detail analysis we can note that KKR have more boundaries 1004 than CSK 913 but have less runs than CSK.

**7. Batting Average of Highest Run Scorers**

In this we noticed about the batting average of highest run scorer.

Figure 16. Batting Average of Highest Run Scorers

Virat Kohli have highest run (5426), but his batting average is less than most of players. AB de Villers have highest batting average, but his runs are at second least (4414) for top 10 players.

**Bowling**

**1. Top Bowlers of the season**

Figure 17. TOP BOWLERS OF THE SEASON

The above chart shows the Top 10 bowlers of IPL (Indian Premier League) from 2018-2020. Y-axis of the chart shows name of the bowlers and on X-axis shows the number of times the bowlers have dismissed the batsman during the league. JJ Bumrah and Rashid Khan are leading with 63 and 58 dismissals.

**2. Top Bowlers of the season**

Figure 18. Sucess Rate of Bowling types in IPL

The above chart shows the success rate of any bowler in IPL which is measured by 2 factors Number of times bowlers have taken wickets (dismissed batsman) and economy rate (Number of runs given by bowlers per over, the lesser the value is the better). The chart depicts on the left Y axis shows the number of dismissals, right y axis shows the economy rate and X axis shows the type of bowlers in IPL. According to the chart left\_arm\_fast medium and Leg break googly are the most economic bowler types and right\_arm\_fast\_medium and right\_arm\_medium is highest wicket taking bowler types, but Leg break googly is overall best with 3rd highest wicket taking type with 2nd best economic rate.

**2. Percentage of wickets taken vs Overs**

Figure 19. %OF WICKETS TAKEN VS OVERS

The final chart takes input of the bowler’s name, and it shows the blower’s strength overs and in which overs a particular bowler has dismissed most of the batsman and can also see bowlers’ strength overs. The following chart shows the bowling record of top wicket taker of IPL JJ BUMRAH from 2018-2020 and according to the chart he has taken most of his wickets in 17th and 19th over of the games with 20% of his total wickets followed by 15th over with 9% dismisses which shows he is really good and effective bowler in the final phases of match where batsman look forward to score more runs.

**6. EXPECTED RESULTS**

The project will yield three major deliverables:

**Methods** **developed:**

In this team will provide detailed summary of the methods used, how they are chosen to do a specific task and how they are used to get best result out of the dataset. Team will also provide insight of how methods are used to know about relation between csv sheets, tableau, and visualization with help of data analysis.

**Model to predict:**

Project will deliver the various ways to correlate variables of data and will provide a great overview of data. Our team will also provide different charts in form of visualization to understand analysis easily for future reference.

**Analysis of data**:

In the end we will provide a well-managed dashboard on tableau with reference to other documents and material used to carry out this project. our team will provide a detailed narrative summary of each step of project.

**6. References**

**Datasets links**

[**https://www.kaggle.com/ramjidoolla/ipl-data-set**](https://www.kaggle.com/ramjidoolla/ipl-data-set)

[**https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020**](https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020)

[**https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all\_season\_batting\_card.csv**](https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all_season_batting_card.csv)

[**https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all\_season\_bowling\_card.csv**](https://www.kaggle.com/datasets/rajsengo/indian-premier-league-ipl-all-seasons?select=all_season_bowling_card.csv)

**Document preparation references**

https://www.pngkit.com/view/u2e6y3t4e6u2q8u2\_data-analytics-methodology-diseo-de-una-infografia

https://www.researchgate.net/figure/The-fundamental-steps-of-the-exploratory-data-analysis-process\_fig3\_32993077