



Daffodil
International
University

PROJECT REPORT

COURSE CODE: CSE 317

COURSE NAME: Software Project V

Submitted To

Fariha Jahan

Lecturer, Department of Computer Science & Engineering

Daffodil International University

Submitted By

TEAM QUANTUM_X

Md. Foyzal Ahmed (191-15-12486)

Rubaiya Tasnim Sholi (191-15-12361)

S.M Khasrul Alam Shakil (191-15-12180)

Rasel Ahmed (191-15-12460)



Indian Premier League

Season: 2008-2019

Abstract:

This report is based on the dataset of **Indian Premier League season 2008-2019**. The goal was to visualize the Indian Premier League datasets using the programming language of **Python** and some of its library (Numpy, Pandas, Matplotlib) and this project was successfully completed by us. Visualization results are shown using some charts (Bar chart, Horizontal bar chart, Pie chart, Line plot, Scatter plot and Histogram). Of all the teams, Mumbai Indians have the most wins in season 2008-2019.

Introduction:

The project was to visualize data from datasets. Data visualization can be said as 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.

The dataset used in this project was taken from [kaggle.com](https://www.kaggle.com). The dataset contains 4 CSV files (matches, players, teams, most_runs_avg_strikerate). Since the data was in raw format some pre-processing were made while working with the dataset. Python's Pandas Library has been used for data pre-processing. Since our goal was to visualize the datasets using some graphs and charts using python which we have successfully done without any hassle. Finally, the visualization results are shown through the chart. The whole project has been kept at GitHub.

Methodology:

The name of the dataset was Indian Premier League Season 2008-2019. It was quite a small dataset consists of 4 CSV files (matches, players, teams, most_runs_avg_strikerate). In each CSV there are different number of rows and columns connected to each other. Among the datasets **matches** having 745 rows and 17 columns (id, season, city, date, team1, team2, toss_winner, toss_decision, result, dl_applied, winner, win_by_runs, win_by_wickets, player_of_the_match, venue, umpire1, umpire2).

Teams contain the details of total number of teams consist in 2 columns (Team_Id, Teams) and 14 rows. On the other hand, **Players** having the information about all the players who have participated in IPL 2008-2019 season. It contains 471 rows and 5 columns (Player_name, DOB, Batting_Hand, Bowling_Skill, Country). Lastly, **most_runs_avg_strikerate** is having 516 rows and 6 columns (batsman, total_runs, num_of_outs, average, strikerate).

The combination of these 4 CSV files explain the whole information about the Indian Premier League in the season of 2008-2019 such as winning rate of each individual teams, information about Indian and Foreign players.

For data visualization Python has been used along with its different library. Python is an interpreted, high-level and general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python provides an excellent way to visualize any data set using some of its data visualization library for example: pandas, numpy, matplotlib and so on.

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. Pandas provide extremely streamlined forms of data representation. This helps to analyze and understand data better. Simpler data representation facilitates better results for data science projects. It can efficiently handle large amount of data.

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib can be used in Python scripts, the Python and IPython shell, web application servers, and various graphical user interface toolkits.

Results & Discussion:

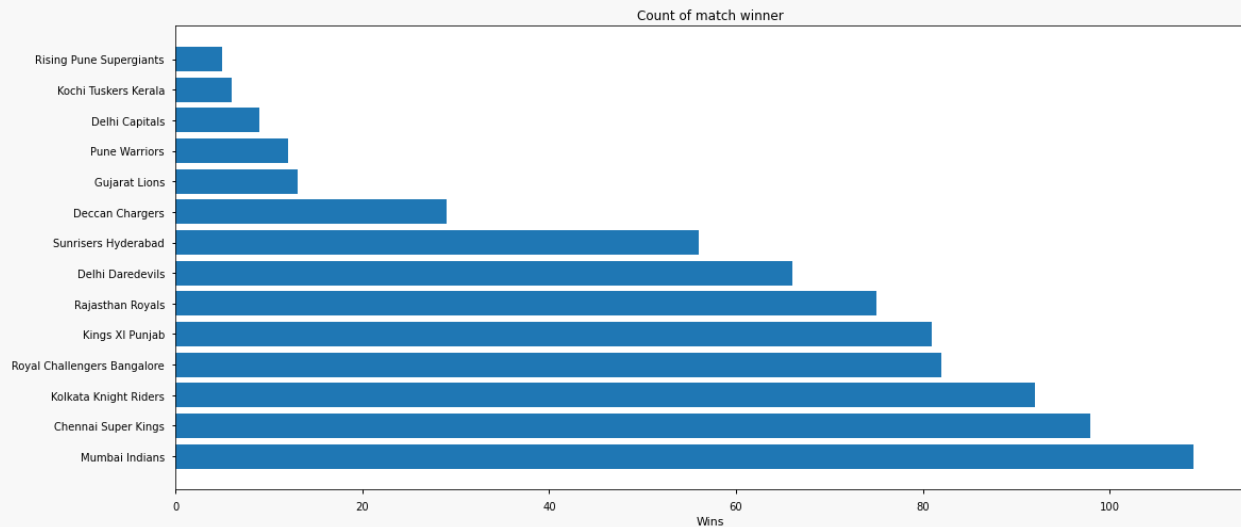


Figure 1: Horizontal Bar chart on count of winner by Teams

Above Bar chart shows that among all the teams, Mumbai Indians achieved highest number of wins in the season of 2008-2019 and total the number of wins they have achieved was 109.

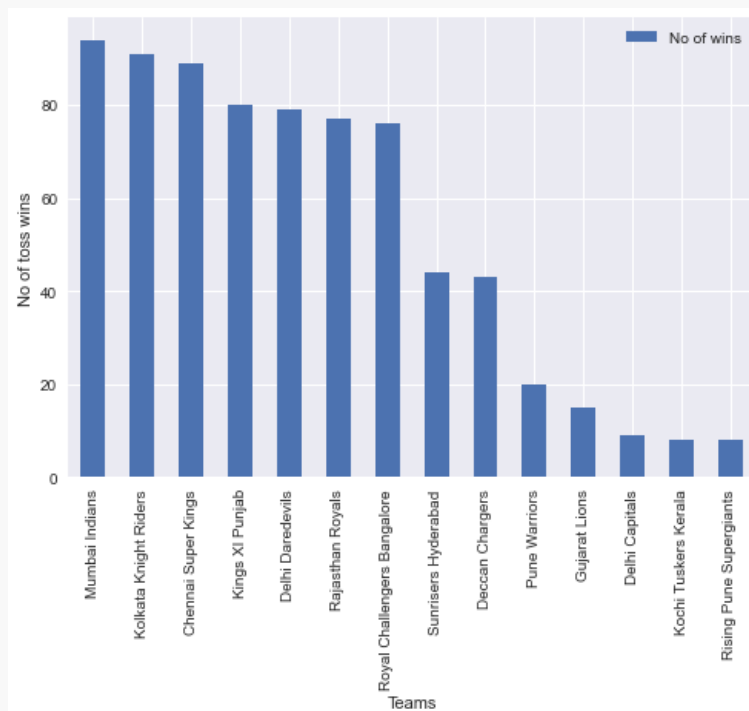


Figure 2: Bar chart on number of toss wins by each team

Above bar chart shows that the number of toss wins by all the teams. The visualization shows that Mumbai Indians have won the toss the most number of times.

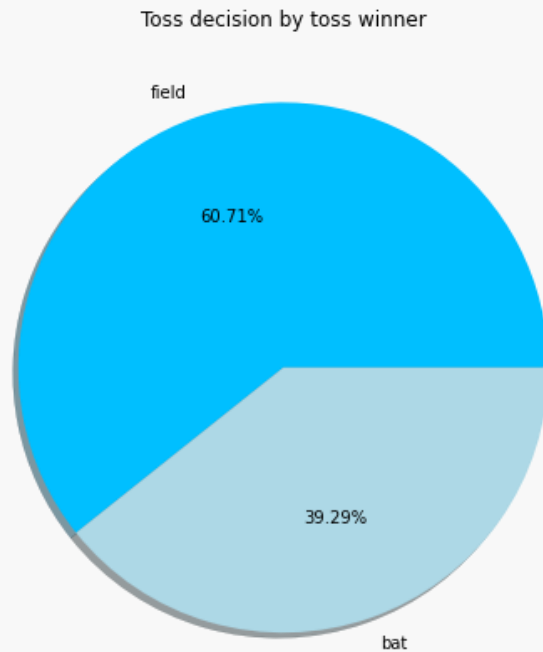


Figure 4: Pie chart on number of toss decisions by each teams

The above pie chart shows that most of the time the toss winner team decided fielding over batting.

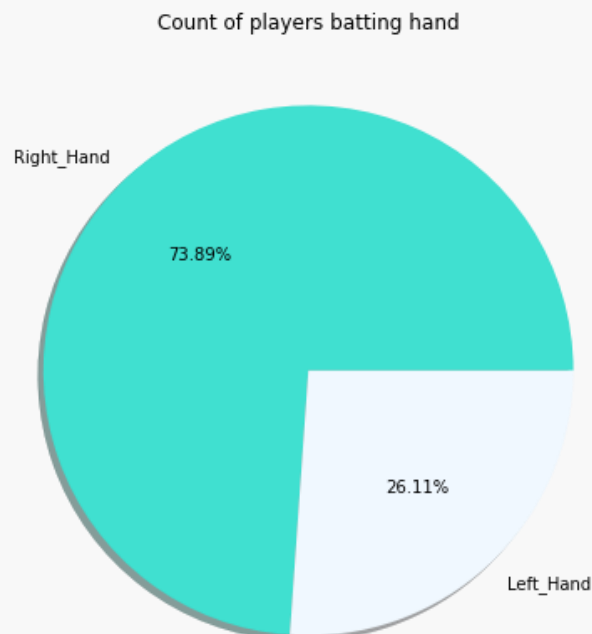


Figure 5: Pie chart on batting hand of batsman

The above pie chart shows that most of the players who were precipitated in IPL were right-hand batsman.

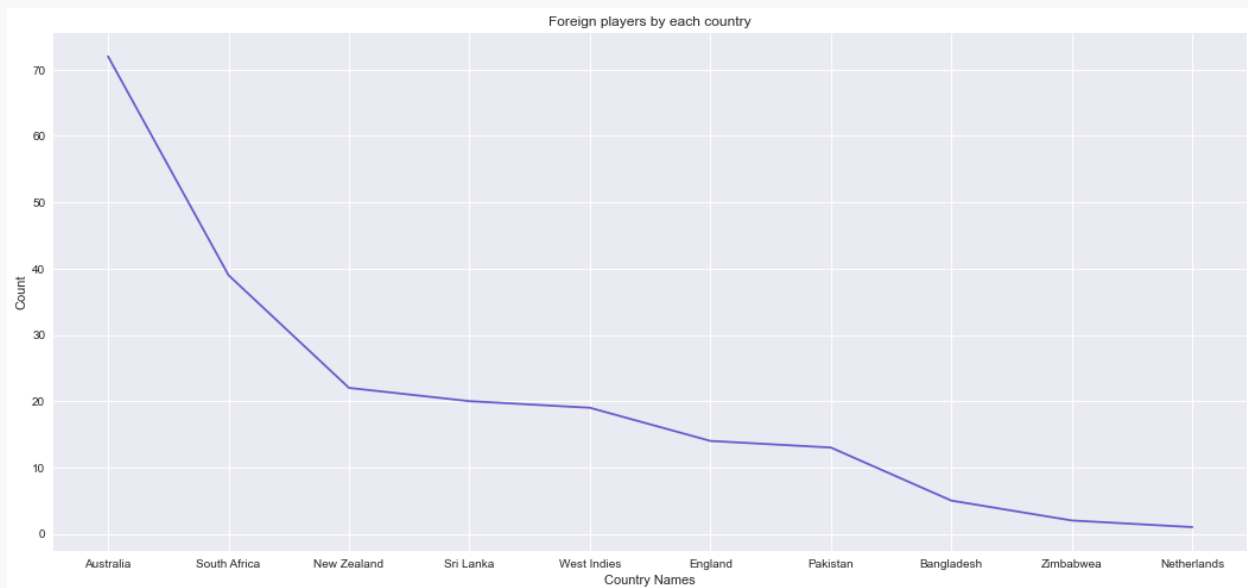


Figure 6: Line plot on Foreign players by each country

Above Tree map shows that number of participating players from each country outside India. Among the countries most of the participating players are from Australia.



Figure 7: Scatter plot on Total Foreign & Indian Players contribution by runs

Above scatter plot shows the total contribution of runs by Foreign and Indian players where all the foreign players scored a total of 90k while the Indian players scored 124K. Due to the large number of Indian players, their total scores are usually higher. However, the total score of foreign players is also noticeable.

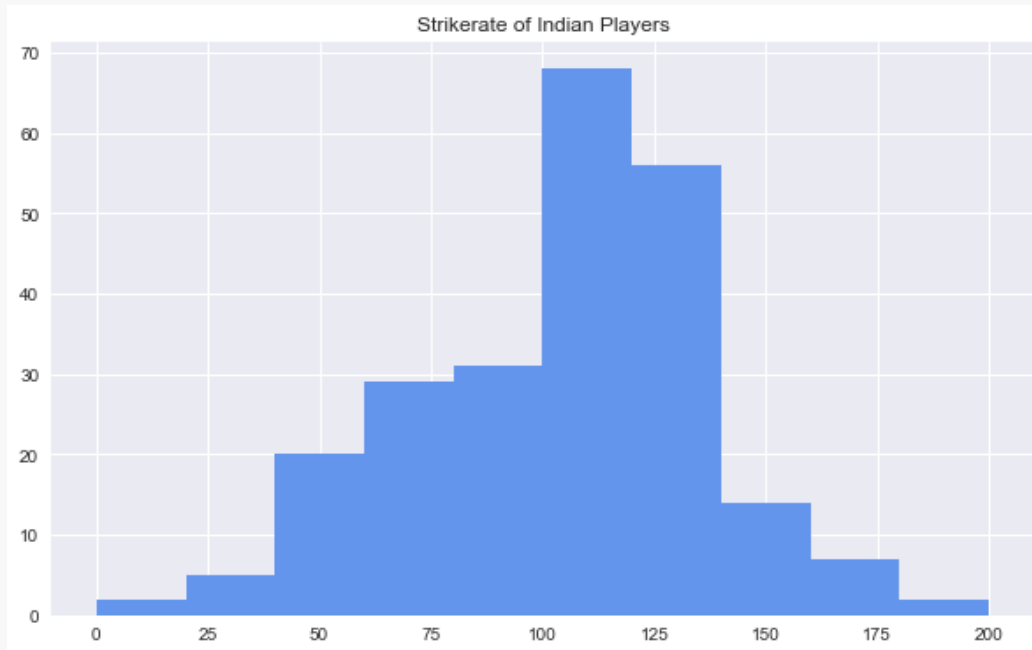


Figure 8: Histogram on Strike rate of Indian players

Above histogram shows the average strike rate of Indian players in the season of IPL 2008-2019. From the visualization it is observed that most of the players maintained strike rate between 125.

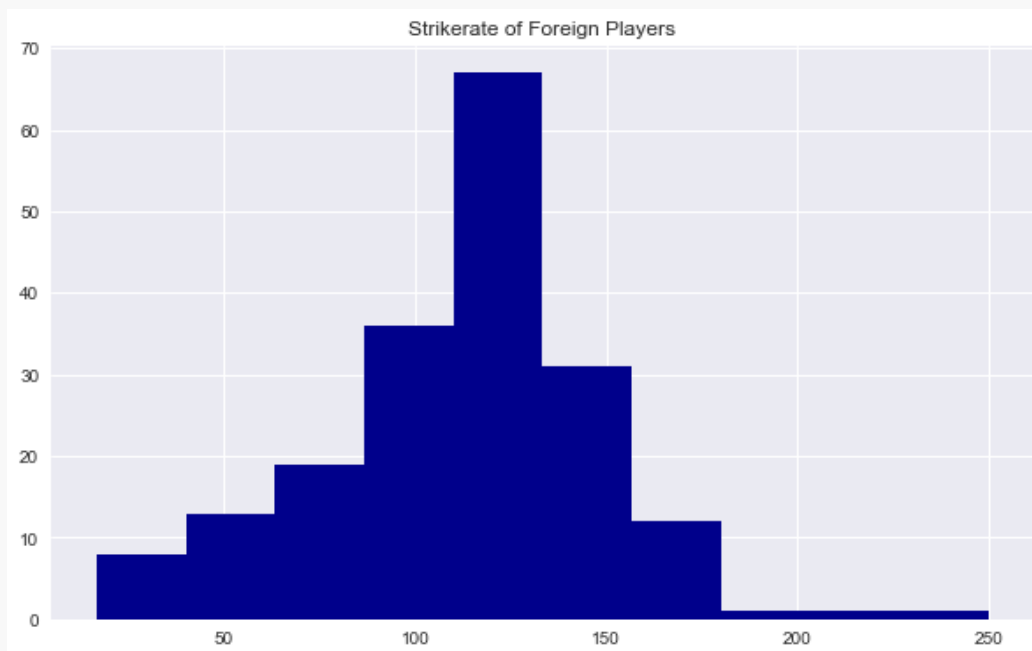


Figure 8: Histogram on Strike rate of Foreign players

Above histogram shows the average strike rate of Foreign players in the season of IPL 2008-2019. From the visualization it is observed that most of the players maintained strike rate between 125 to 150.

Conclusion:

The aim of this project was to analyze the dataset of Indian Premier League season 2008-2019 using python programming language. 8 visuals or charts were created by analyzing the entire dataset. Looking at the visuals, it is understood that Mumbai Indians is the leading team in the IPL season of 2008-2019 as they have won most of the matches. Besides Indian players are dominating the IPL which is a good sign for upcoming Tournaments. However, it can be said that other franchises need to pay more attention to their game.

References:

- [1] “Indian Premier League Season 2008-2019” dataset collected from [kaggle.com](https://www.kaggle.com).
- [2] “Python” – information collected from [Wikipedia](https://en.wikipedia.org).
- [3] “Pandas” – information collected from [Pandas.pydata](https://pandas.pydata.org).
- [4] “Matplotlib” – documentation collected from matplotlib.org.
- [5] “Jupyter notebook” – Using [anaconda](https://www.anaconda.com) platform.