T-20 Cricket Analytics

Under the guidance of Don Dalton

Python Software Development [3006-1]

By:
Aakarsh Sagar
Ajay Rawtani

Contents

Introduction

Cricket is one of the most popular sports in the world, played in over 100 countries. One of the most successful and watched cricket leagues globally is the Indian Premier League (IPL), which has been running since 2008. The IPL has broken several records and is now valued at over $6 billion. The league attracts millions of viewers from all over the world and has a significant fan base in India. In this project proposal, we will explore the importance of cricket analytics and how it can be used to build a tournament-winning team in the IPL.

Objectives

The objective of this project is to use cricket analytics to build a tournament-winning team with the right mix of batters, bowlers, and wicketkeepers. Cricket analytics plays an essential role in the modern game, helping teams make data-driven decisions to improve performance and win matches. We will use statistical analysis and data visualizations to analyze player performance and select the best team composition for the Denver Demolishers Cricket Club.

Approach

I. Web Scraping

The T-20 Cricket format, a relatively modern development, gained prominence with the inception of the Indian Premier League (IPL) in 2008. To obtain comprehensive and reliable information about players and their corresponding statistics, the official IPL website serves as the most authoritative source. In order to collect and organize the necessary data for subsequent processing, including data cleaning, analysis, and visualization, Microsoft Power Automate, a software automation tool, was effectively employed to scrape the website and save the extracted information into Microsoft Excel. This systematic approach ensures the availability of accurate and structured data for further scholarly exploration and investigation.

II. Data Cleaning

Upon the completion of the web-scraping process, the acquired dataset was division into two distinct files, namely Batting.xlsx and Bowling.xlsx. These files were subjected to a thorough cleaning procedure, aimed at eliminating any instances of null or dummy values. Additionally, appropriate adjustments were made to rectify the data types of specific variables, ensuring accuracy and consistency throughout the dataset. To facilitate the subsequent analysis, novel variables or columns were derived and manipulated, strategically designed to provide indispensable insights into the data. Notably, the cleaning, manipulation, and wrangling operations were exclusively executed utilizing the Pandas library in Python.

This methodical approach ensures that the data is of high quality, free from any anomalies or irregularities that could potentially compromise the reliability of subsequent analyses. By employing these robust procedures, we establish a solid foundation upon which reliable and meaningful conclusions can be derived from the data.

III. Data Analytics & Visualization

The data cleaning process, having been completed successfully, facilitated the subsequent tasks of data visualization and analysis. For this purpose, the matplotlib library, a widely used tool in Cricket Analytics, was employed to generate visual representations of fundamental metrics. The following are examples of the plots created:

- Batsman plots:
  - Top 10 players with the highest career strike rates.
  - Top 10 players with the highest career batting averages.
  - Top 10 players with the most sixes scored in their careers.
  - Top 10 players with the highest career run totals.

- Wicketkeeper plots:
  - Top 10 wicketkeepers with the most catches.
  - Top 10 wicketkeepers with the most stumpings.
  - Top 10 wicketkeepers with the highest number of dismissals.
  - Top 10 wicketkeepers with the highest average dismissals per match in their careers.

- Bowler plots:
  - Top 10 bowlers with the lowest economy rates.
  - Top 10 bowlers with the lowest strike rates.
  - Top 10 bowlers with the most wickets taken.
  - Top 10 bowlers with the highest average wickets per match in their IPL careers.

While the aforementioned visualizations were effectively created using the Matplotlib library in Python, it was determined that additional insights were necessary to comprehensively evaluate player performance and identify suitable candidates for team selection. To address this need, Tableau, a specialized data visualization tool, was utilized. The implementation of Tableau enabled more detailed analytics and facilitated the identification of key players based on thorough examination and exploration of the data.

IV. Squad Selection

After conducting the analysis, the ultimate objective is to assemble a championship-winning squad by carefully selecting players with a balanced combination of batsmen, bowlers, wicketkeepers, and all-rounders. A team consists of eleven players who actively participate in the game, with additional substitutes available. However, team selection is constrained by a predetermined budget set by the governing body, such as the Board of Cricket Control (BCC), which assigns a specific allocation to each team annually and

stipulates that a team can only have a maximum of 4 International players in the playing eleven. Therefore, our focus will be on choosing the best eleven players while adhering to the stipulated budget. The selection criteria for each specialized role are as follows:

A. Batsman:

Batsmen play a pivotal role in IPL T20 cricket, serving as key contributors to their team's success through run-scoring, partnership building, and providing impetus to achieve competitive totals. To select batsmen for the team, the following criteria are considered:

- A minimum of 50% of the runs scored should be from boundaries (4s or 6s).
- At least 45% of the runs scored should be from non-boundary shots.
- An average of approximately 0.868 sixes per match, as per the analysis.
- A strike rate exceeding 120 and an average of over 20 runs per match.

Name: David Warner				**Batsman**		**Australia**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
174	6256	36.01	139.75	61.45	53.87	1.26	756K

Name: Rajat Patidar				**Batsman**		**India**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
12	404	33.67	144.29	60.89	56.43	1.75	24K

Name: Yashasvi Jaiswal				**Batsman**		**India**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
36	1122	31.17	149.2	74.15	38.56	1.33	484K

Name: Rituraj Gaikwad				**Batsman**		**India**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
48	1632	34	134.21	59.31	54.61	1.33	726K

Name: Suryakumar Yadav				**Batsman**		**India**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
135	3123	23.163	142.99	63.78	51.79	0.8	968K

Name: Aaron Finch				**Batsman**		**Australia**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
92	2091	22.73	128.2	63.32	47.03	.84	182K

Name: David Miller				**Batsman**		**South Africa**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
116	2697	23.25	138.81	55.47	61.81	1.09	360K

Name: Abdul Samad				**Batsman**		**India**	
Matches	Runs	Avg	SR	%RB*	%RNB*	6s/Mat	$ Value
32	391	12.22	137.68	59.85	55.28	.78	480K

B. Bowlers:

Bowlers in IPL T20 have a crucial role in taking wickets, containing runs, and exerting control over the game. To identify effective bowlers, the following criteria are considered:

- The economy rate should be below 9 runs per over.
- A strike rate of less than 23, indicating the ability to take wickets efficiently.
- Preference is given to bowlers who have secured four-wicket hauls, serving as a hallmark of excellence in a batsman-dominated game.

Name: Kagiso Rabada			**Bowler**		**South Africa**	
Matches	Wickets	Econ.	SR	W/Mat	4W	$ Value
134	104	8.35	14.65	1.55	12	1.11M

Name: Adam Zampa			**Bowler**		**Australia**	
Matches	Wickets	Econ.	SR	W/Mat	4W	$ Value
19	28	8.08	14.11	1.47	0	180K

Name: Mohsin Khan			**Bowler**		**India**	
Matches	Wickets	Econ.	SR	W/Mat	4W	$ Value
11	15	6.64	14.4	1.36	2	24K

Name: Harshal Patel			**Bowler**		**India**	
Matches	Wickets	Econ.	SR	W/Mat	4W	$ Value
89	109	8.6	16.69	1.23	4	1.29M

Name: Khaleel Ahmed			**Bowler**		**India**	
Matches	Wickets	Econ.	SR	W/Mat	4W	$ Value
41	55	8.59	16.98	1.34	0	630K

Name: Mohit Sharma			**Bowler**		**India**	
Matches	Wickets	Econ.	SR	W/Mat	4W	$ Value
95	105	8.37	18.24	1.11	4	60K

C. Wicketkeeper:

In the IPL T20, wicketkeepers play a vital role as key players contributing to both fielding and batting capabilities. The following criteria are considered when selecting a wicketkeeper:

- A minimum strike rate of 130 with the bat.
- Preference is given to wicketkeepers with high dismissal rates per match and better averages.

Name: Sanju Samson		**Wicketkeeper**		**India**	
Matches	Runs	SR	Avg	Dismissals/Mat	$ Value
151	3888	137.27	24.74	.06	1.68M

Name: Jitesh Sharma		**Wicketkeeper**		**India**	
Matches	Runs	SR	Avg	Dismissals/Mat	$ Value
24	534	160.97	20.79	.67	24K

D. All-Rounder:

An all-rounder in cricket is a player who excels in both batting and bowling, making significant contributions to the team's performance. In the IPL T20, the selection of all-rounders is based on the following criteria:

- A contribution score of not less than -10, contribution is the difference between the batting average and number of runs conceded per wicket. Thus, indicating their positive impact on the team.
- A minimum strike rate of 130 with the bat.
- Averaging at least one wicket per game, highlighting their effectiveness in bowling.

These selection criteria ensure the identification of skilled and impactful players in their respective roles, contributing to a well-balanced and competitive team in the IPL T20.

Name: Andre Russel		**All-Rounder**		**West Indies**	
Mat	Runs	Batting SR	Wkts/Mat	Contribution	$ Value
111	2255	174.67	0.87	+ 4.93	1.92M

Name: Mark Wood		**All-Rounder**		**England**	
Mat	Runs	Batting SR	Wkts/Mat	Contribution	$ Value
5	12	150	2.2	- 10.27	900K

E. Starting XI Players

S.No	Name	Specialization
1	David Warner (Captain) ✈	Batsman
2	Rajat Patidar	Batsman
3	Yashasvi Jaiswal	Batsman
4	Rituraj Gaikwad	Batsman
5	Suryakumar Yadav	Batsman
6	Sanju Samson	Wicketkeeper
7	Andre Russel ✈	All-Rounder
8	Kagiso Rabada ✈	Bowler
9	Adam Zampa ✈	Bowler
10	Mohsin Khan	Bowler
11	Harshal Patel	Bowler

F. Substitutes

S.No	Name	Specialization
1	Aaron Finch ✈	Batsman
2	David Miller ✈	Batsman
3	Abdul Samad	Batsman
4	Khaleel Ahmed	Bowler
5	Mohit Sharma	Bowler
6	Jitesh Sharma	Wicketkeeper
7	Mark Wood ✈	All-Rounder

LEGENDS:

✈ - International Players

G. Future Research

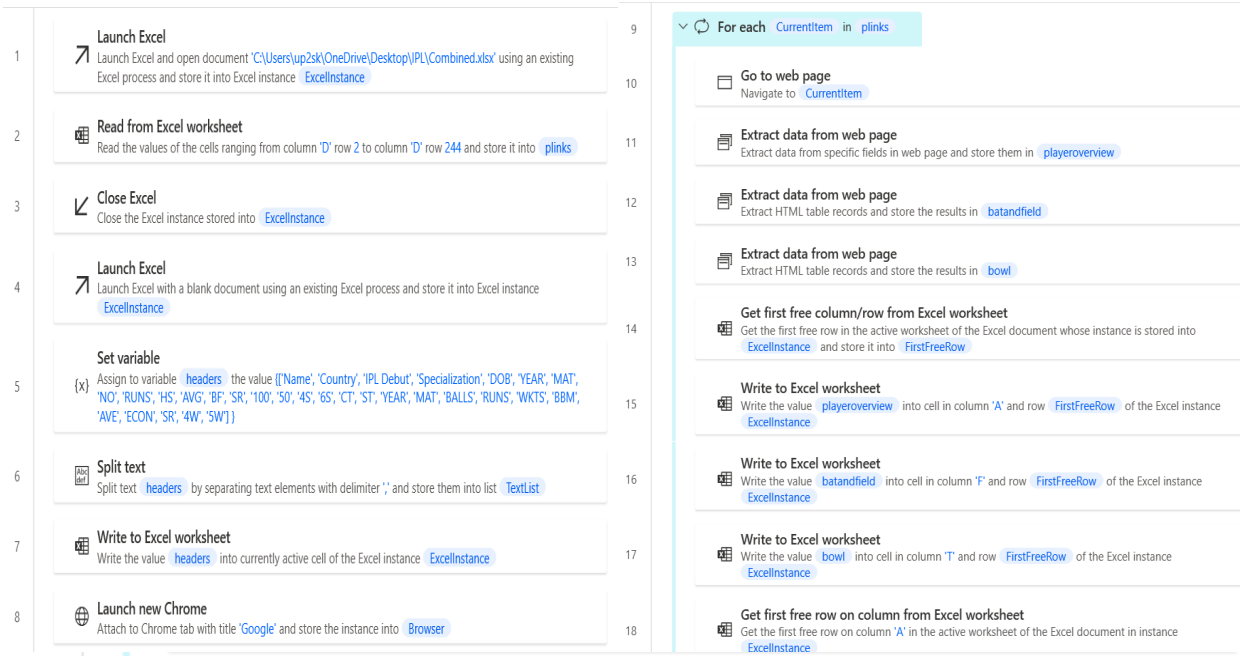
The team aimed to incorporate object-oriented programming to enable the tool to suggest players in their ranked order based on their specialization. However, due to the project's time constraints, further research is necessary for its implementation. Furthermore, the analysis conducted solely relies on data obtained from the official IPL website, limiting its scope. Professional companies that collect ball-by-ball data could greatly enhance the analysis by providing more comprehensive insights. Additionally, the current data collection process fails to capture other crucial aspects of the game, such as the actual auctioning of players, which varies for each team based on their individual strategies. Nevertheless, the team firmly believes that the assembled roster possesses the necessary qualities to become championship winners, showcasing a well-balanced composition of specialists with suitable backups.

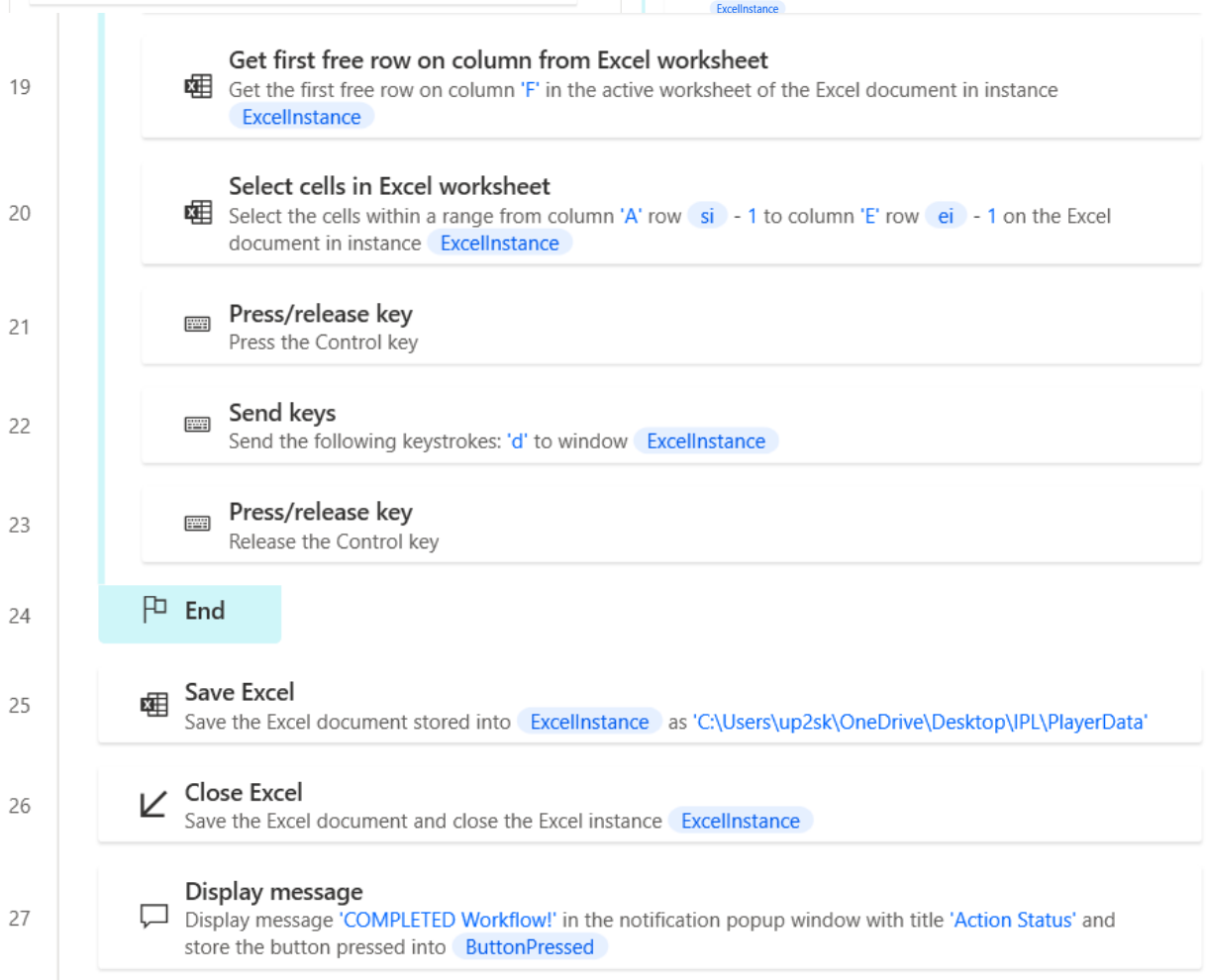
H. References

- Indian Premier League. (2023). *TATA IPL 2023*. Retrieved from Indian Premier League: https://www.iplt20.com/stats/2022
- Microsoft. (2023). *Microsoft Learn*. Retrieved from Microsoft Learn: https://learn.microsoft.com/en-in/training/browse/?products=power-automate&WT.mc_id=webupdates_GEP_PowerAutomate-web-wwl
- Richardson, L. (2023). *Beautiful Soup Documentation*. Retrieved from Beautiful Soup 4.12.0 documentation: https://www.crummy.com/software/BeautifulSoup/bs4/doc/

I. Appendix

➢ Web-Scraping (MS Power Automate)





➢ Data Cleaning and Visualization

a. Python

All the cleaning and visualization contain in the file:

Data Cleaning.ipynb

strikerate_matplotlib.ipynb

b. Tableau

https://public.tableau.com/views/IPLAnalytics_16854181898210/Batsman?:language=en-US&:display_count=n&:origin=viz_share_link