

IPL-T20 Web Scraping and Data Analysis

By Akash Kandarkar

Data Source :

1) <https://www.iplt20.com/> (official IPL website)

The Indian Premier League Twenty20 is known as IPL T20. The Board of Control for Cricket in India (BCCI) established a professional Twenty20 cricket league in India in 2007. Each year, eight teams from various Indian cities compete in the competition. Top cricket talent from all around the world is drawn to the IPL T20, one of the most well-known and lucrative cricket competitions in the world. Every year, the event, which features action-packed, high-scoring games that are aired worldwide, is typically held between March and May.

Twenty20 cricket, usually referred to as T20 cricket, was first played in England in 2003. Each team plays a single inning, which is limited to 20 overs (120 deliveries), in this sport's structure.

Compared to the classic forms of cricket, such as Test and One Day International (ODI), T20 cricket is meant to be faster and more thrilling. Fans who do not have the time or tolerance for lengthy bouts may find the game more appealing due to its shorter duration.

Since its inception, T20 cricket has gained immense popularity around the world and has led to the development of various domestic and international T20 leagues, including the Indian Premier League (IPL), Big Bash League (BBL), Caribbean Premier League (CPL), and United States League aka [Major League Cricket](#) (MLC) which saw recent investment in MI New York. T20 cricket has also played a significant role in bringing in new audiences to the sport and has attracted a number of top players from around the world.

Goal:

This analysis aims to analyze IPL-T20 by examining historical data on each player performance based on batting and bowling metrics. With the goal of understanding what sets the best and worst performers apart considering rising auction prices of each player in 2023 and the revenue generated by this championship brings lots of sponsors to invest in the game. This analysis will give both bidders and sponsor companies.

Web Scraping:

Official website of IPL-T20 provides the batting stats of each player in each season. Using this data we can create a dataset which will help us analyze the stats to our need and may reveal surprising facts.

To get all data, Selenium and Request packages we used as main scraping tools along with python.

Each year stats for batting can be navigated through website using selenium and same used for scraping all data and storing it in each year CSV file.

1. Batting_df_{year}.csv
2. Team_wins.csv

Data Processing:

Bowling and Batting data had mix values like team name attached to player's name, no year mentioned for each season, I did process to make sure all data is aligned and have no null values.

Batting data file:

```
POS,Player,Mat,Inns,NO,Runs,HS,Avg,BF,SR,100,50,4s,6s,season_year,team
1,Shaun Marsh,11,11,2,616,115,68.44,441,139.68,1,5,59,26,2008,PBKS
2,Gautam Gambhir,14,14,1,534,86,41.07,379,140.89,0,5,68,8,2008,DC
3,Sanath Jayasuriya,14,14,2,518,114*,43.16,309,167.63,1,2,58,31,2008,MI
4,Shane Watson,15,15,5,472,76*,47.2,311,151.76,0,4,47,19,2008,RR
5,Graeme Smith,11,11,2,441,91,49.0,362,121.82,0,3,54,8,2008,RR
6,Adam Gilchrist,14,14,1,436,109*,33.53,318,137.1,1,3,51,19,2008,PBKS
7,Yusuf Pathan,16,15,1,435,68,31.07,243,179.01,0,4,43,25,2008,RR
8,Suresh Raina,16,14,3,421,55*,38.27,296,142.22,0,3,35,18,2008,CSK
```

Column Description:

POS: Ranking of player for batting

Player: Name of player

Mat: Matches played

Inns: Innings played (got chance to play)

NO: Not out total number of times.

Runs: Runs made by player

HS: Highest score by player

Avg: Average Score

BF: Balls faced by player

SR: Strike rate (Runs made by player / Number of balled played)

100s: number of run century by player

50s: number of run half-century by player

4s: number of boundaries hit by player

6s: number of sixes hit by player

Season_year: year of IPL season

Team: team name

Team Wins data file:

winning_team: team name

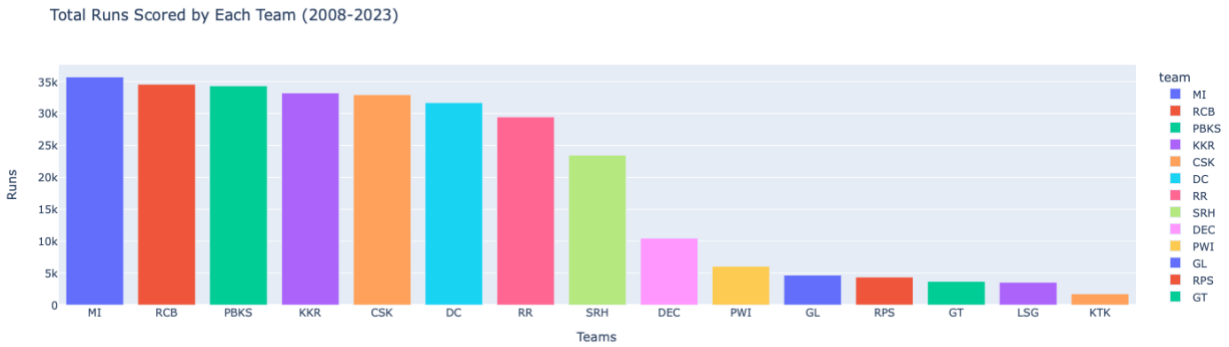
winning_year: year in which the team won

Joining all data sets with no redundant data was goal, added column Season_year and Team, made it unique data. To get all insights from the data we need to combine the related files.

Top Scoring teams

We may learn some things about each team's performance in the IPL T20 by looking at their top runs scored. We can determine the teams' strengths and weaknesses by examining the top run scorers for each club. A powerful batting lineup is indicated, for instance, if a club has a top run scorer who consistently scores a high number of runs throughout the season. On the other hand, it may be a sign that the club is having trouble at the plate if it has a low-scoring leading run scorer.

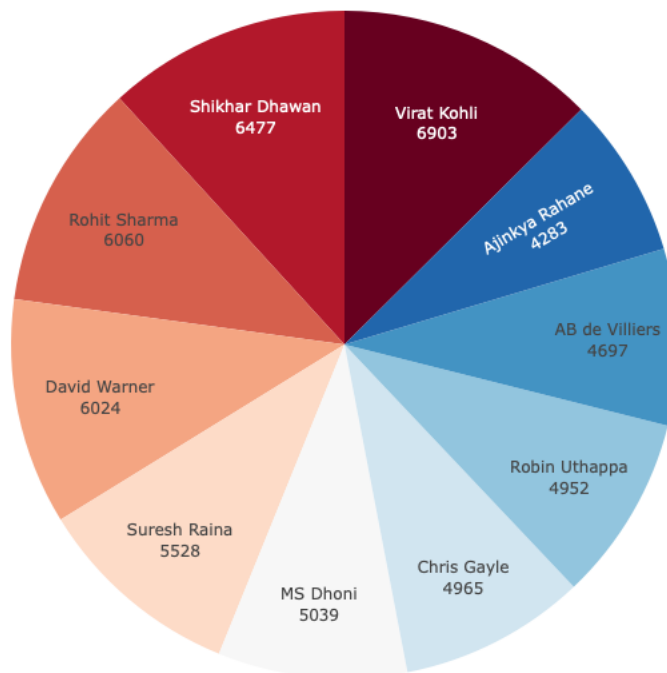
Similar to this, analyzing the top run scorers from other teams might reveal which groups have the best batting lineups overall. Predicting the results of games and competitions can be done with the help of this information. Overall, looking at each team's top runs can give us useful information about how they performed in the IPL T20, which can then be used to make decisions about the teams and players.



Data shows that MI is highest run scorer till date, which shows potential of the Batsman in the Team, where was the Team which never won IPL trophy RCB also shows almost same potential.

Top 10 Batsman

Top 10 Batsmen (All)



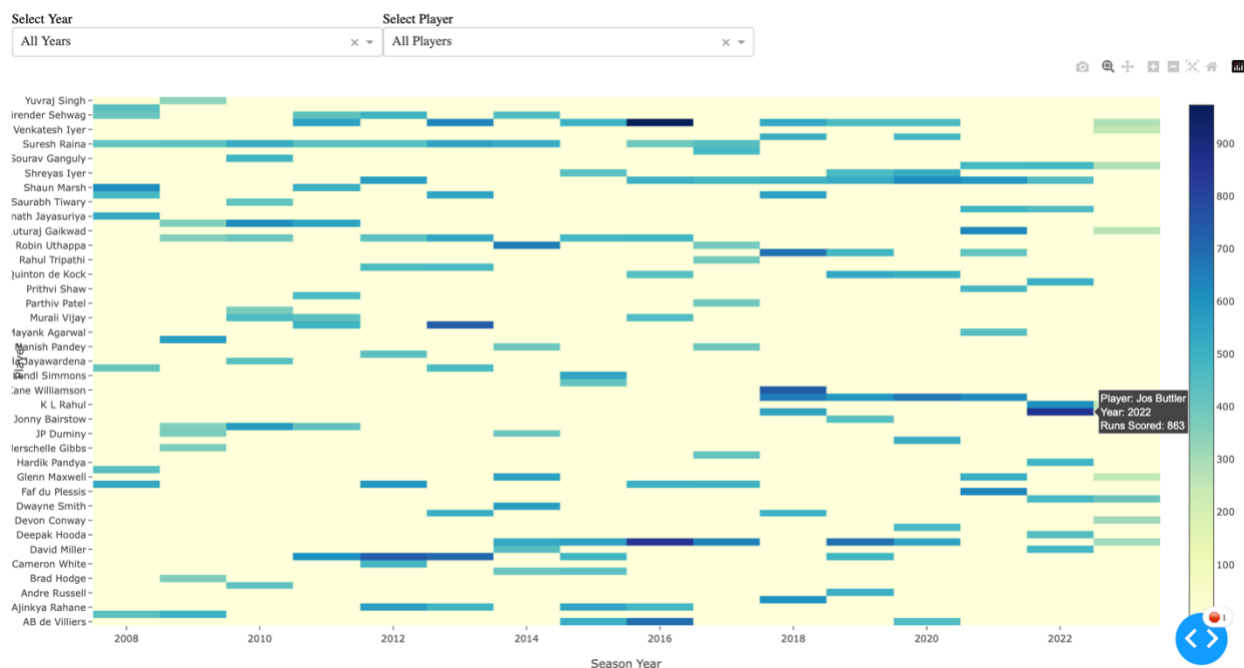
We may learn following things by looking at the IPLt20 player who scored the most runs overall:

- **Consistency:** It can provide information about a player's continuity across time. A player has been performing well throughout the seasons if they are continuously listed among the leading run-scorers.

- **Impact:** A player's contribution to the team's performance can be judged by how many runs they score in a game. A player has contributed significantly to the success of his side if he has scored a significant number of runs.
- **Comparison:** The best batsman in IPLt20 history can be determined by evaluating the performances of various players. This analysis can aid in our understanding of the abilities needed to win the competition and assist teams in locating suitable recruits.
- **Engagement with the audience:** IPL viewers are always curious to learn about the league's best players. Fans' engagement and interest in the league can be increased by using this analysis.

Runs Scored by Top 10 IPL Batsman in Each Season

Heatmap:



It can quickly compare their performances over a number of seasons and reveal patterns or trends in the way they score. For instance, we can check to see if a certain batsman consistently performs well over the course of several seasons or if their performance drastically differs from one season to the next. We can also find any connections or correlations between several batsmen's performances during the same season.

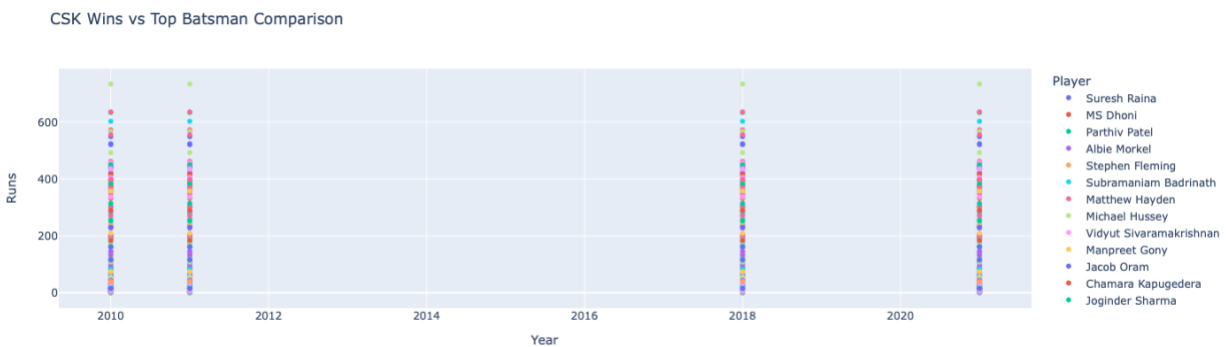
The heat map may also be used to analyze the performance of batsmen from various teams and pinpoint any organizations that regularly generate the best batsmen. Overall, the heat map offers a potent visual tool for exploring and analyzing the top IPL batsmen's historical performance.

Team Wins and Top Batsman Comparison

Each year's IPL champion teams are included along with the name of the best run-scoring batsman in a bar graph called the Team Wins and best Batsman Comparison. The teams that win the IPL championship and the best batter of the same year can be compared using this graph to look for any similarities or trends.

For instance, it may be inferred that having a high-performing batsman is a crucial element in winning the tournament if we note that the team that wins the IPL title most years also has the best runs-scoring batter of the year. On the other hand, if we see that there is no connection between the team that wins and the best batter of the year, it may indicate that other elements, such as the bowlers' effectiveness or the club's general plan, are involved in the tournament's outcome.

Overall, this chart can assist us understand the relationship between individual player performance and IPL team success.



Conclusion:

The analysis of IPL T20 data revealed a substantial association between the performance of the best batsmen and their teams' success in winning titles. The heatmap visualization of the top 10 batsmen's runs scored in each season revealed that some players consistently performed well across numerous seasons, while others had spectacular performances in a single season. The parallel coordinates graphic made it possible to compare numerous performance parameters for separate players. Furthermore, a bar chart comparing the top batsmen for each year with the winning teams revealed that on many occasions, the team with the top batsman of the year also won the championship. Overall, the data emphasizes the significance of excellent batting performances in IPL T20 success.

Files :

Project folder –

csv_data/ batting_df_{year2008-2023}.csv

scraping/ batting_data.py

dashboard.py

--all csv scraped from website

--csv scraping code

-- dash, plotly dashboard