

How did Sunrisers Hyderabad reach from bottom to top in Indian Premier League(IPL) 2024?

Bhavya Bhargava

An Enthusiast's Perspective

Dated: 2024-08-31

Data Details

Blog post

Behind the Scenes

Catching Data in the Wild

The study of the Sunrisers Hyderabad (SRH) cricket team was hard in many ways, and careful steps had to be taken to make sure the data was correct. One big job was to get information about players from the team’s website. This meant changing code for each of the three years (2022, 2023, and 2024) because the website was being updated regularly. This step was important, but it wasn’t very interesting.

Once all the information was gathered, it was carefully checked against the official website for the Indian Premier League (IPL). In this step, mistakes like misspelt player names and wrong job assignments were found and fixed for each year’s data. This method was necessary to keep the quality and reliability of the data, even though it was time-consuming. Along with that, data transformation was needed to get the data ready for visualisations. For example, the points table data had to be rearranged in order to make a stacked bar chart. Here each row showed the team, the outcome (win, loss, or no result), and the value that went with it. This process wasn’t very pretty, but it was necessary for the final presentation.

It is important to remember that the risk of website changes continuously persists, leaving the existing data gathering code ineffective. Thus, continuous monitoring and code updates may be necessary to maintain the accuracy of data collection. Moving to the analysis process, it brings to light the important but often ignored work that goes on behind the scenes of data analysis. The end charts and insights get a lot of attention, but the careful steps of data preparation, validation, and transformation are just as important for making sure the results are reliable and correct. Even though they aren’t very exciting, these steps are very important to the general quality and reliability of the analysis.

Challenges of Analysis

During my analysis of the Sunrisers Hyderabad (SRH) cricket team, I did face some unexpected challenges that I had not expected initially. One issue was the need to clean up and standardize the player names. When scraping data from the website, I found player names were written differently, with variations in capitalization, abbreviations, and the order of first and last names. This required me to write extra code functions to make all the names consistent across the dataset, which was an added task I didn’t foresee.

Another surprise was dealing with special cases related to player roles and statistics. Sometimes, a player’s main role was different from the role recorded for a specific match. I had to add more instructions to my code to properly categorize players based on the actual roles they played, rather than relying only on the listed roles. At first, my plan was to only analyze SRH’s performance metrics and trends. However, as I progressed, I realized it was important to compare their performance with other teams in the league to provide better context. This led me to expand the scope of the project to include gathering and visualizing data for all teams’ performances, which required more data processing and transformation work than I had initially planned for.

Additionally, I didn’t expect there to be differences between the data on SRH’s website and the official IPL website. This meant I had to cross-check and verify the data from multiple sources, which was while being time-consuming, necessary to ensure accuracy and reliability.

While the core coding tasks like handling dates or doing calculations weren’t too difficult, the unexpected need for cleaning up names, dealing with special cases, expanding the scope, and reconciling data from different sources added complexity and extra effort to the analysis.

Further Scope of Improvement

While the analysis conducted provided me with some valuable insights, I feel there are still some areas for improvement.

1. The data collection process from web scraping could be made more efficient and reliable. Perhaps using RSelenium could be the way to get the data from dynamic websites.
2. The data cleaning and transformation steps could be further automated to reduce the manual work.
3. Expanding the scope to include more in-depth analysis of specific aspects of the like per match run rate, player performances in different situations and venues, and local vs international training metrics could help in getting additional insights.
4. Additionally, incorporating data from future IPL seasons would keep the analysis relevant and help us in getting long-term trends.