"Predicting Best Cricket XI"

Project By

Prasad Tak

Introduction

This is a solo project done by me in order to apply my knowledge of data analysis to create the best IPL T20 XI using the recent IPL 2023 data.

The motive is to analyse the recent IPL data and to get the best Xi players according to their performance.

In this project I have a built a scrapper using JavaScript and J Query. I have using python to format the data and then used Power BI to create dashboard for visualization.

Methodology

The project includes 4 steps:

1. Scrapping Data- I have used Bright Data proxy API's to scrape data from ESPN Cric Info. The data is exported as JSON files.
2. Processing data – I have used Python for data cleaning and preprocessing. The JSON files are imported, cleaned and then exported as 'csv' files.
3. Modelling Data and Creating Dashboard – The csv files are imported into Power BI and then the relation in identified between the attributes and interactive dashboard is created.
4. Studying Data – The dashboard is then studied to create a playing XI.

Implementation

1. Scraping Data

- I've created scrappers using JavaScript in Bright Data Web scrapper.
- I've created four scrappers for scrapping four tables
    - Match table containing playing teams and winner team's name.
    - Batting summary table containing batting data from each match of every batsman
    - Bowling summary table containing bowling data from each match of every bowler
    - Players data containing player's batting and bowling style

Note: I have uploaded Scraper's code.

2. Preprocessing Data

- The JSON files are imported into the Jupyter notebook
- The data is cleaned, duplicated removed and then the tables are imported as csv files

Note: I have uploaded the Python file

3. Building Dashboards

- The csv files are imported into Power BI
- Connection is established between the attributes
- Key measures are created to show statistics
- Dashboards are created
- Data is filtered according to the constraints given in the problem scoping file.

Note – I have uploaded the filter criteria and the problem statement file. Also I have uploaded Power Bi file

4. Studying Data

- The dashboard is studied and the best XI is selected.

Results

After analysing the dashboard, following Players can be identified as the Best XI.

Shubman Gill
Yashasvi Jaiswal
Suryakumar Yadav
Cameron Green
Heinrich Klaasen
Rinku Singh
Ravindra Jadeja
Rashid Khan
Matheesha Pathirana
Mohammed Shami
Mohit Sharma

References

- https://brightdata.com/
- Youtube Tutorials