

ICC 2023 World Cup Openers Analysis

UCS546 Conversational AI: Accelerated Data Science (Basics)

Project Report

End-Semester Evaluation

Submitted by:

102103671 Anvesha Singh

102103688 Sanskar Sureka

BE Third Year, COE

Group No: 3COE24

Submitted to

Dr. Jasmeet Singh



THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Computer Science and Engineering Department TIET, Patiala

December 2023

Data Collection & Data Description

Project Overview

The project revolves around extracting comprehensive data from **ESPN's** website regarding the **ICC ODI World Cup 2023**. Utilizing web scraping techniques, specifically leveraging **Beautiful Soup**, the objective was to gather detailed statistics on **openers** of different teams throughout the whole tournament and analyze them according to various set parameters using **Power BI** dashboard.

Data Collection Methodology

Employing Beautiful Soup, a Python library for web scraping, we navigated through ESPN's web pages dedicated to the ICC ODI World Cup 2023. The scraper facilitated parsing the HTML content of these pages, allowing us to isolate and extract the desired data points using specific HTML tags and classes and extracted essential information pertaining to participating teams and individual player statistics. The scraped data was stored in a CSV file. The following dataset was collated:

- 1)**Teams:** Capturing team-specific details including team names, player lists, and other associated attributes.
- 2)**Player Statistics:** For each player, data was collected on runs scored, balls faced, strike rates, batting positions, sixes and fours hit, and their respective batting styles.

Website: <https://www.espnricinfo.com/records/tournament/icc-cricket-world-cup-2023-24-15338>

Data Description

The extracted dataset encompasses a comprehensive array of information crucial to understanding openers performances and team dynamics throughout the ICC ODI World Cup 2023:

We collected the data and stored it in following three CSV files. Below mentioned are the main attributes of each file:

1)Player Details:

- Name
- Team
- Batting Style

2)Match Summary:

- Match Date
- Match ID
- Winner

3)Batting Data Summary:

- Match
- Team Innings
- Batting Position
- Batsman
- Out/Not Out
- Runs Scored
- Balls played
- Boundaries
- Strike Rate

Data Preprocessing

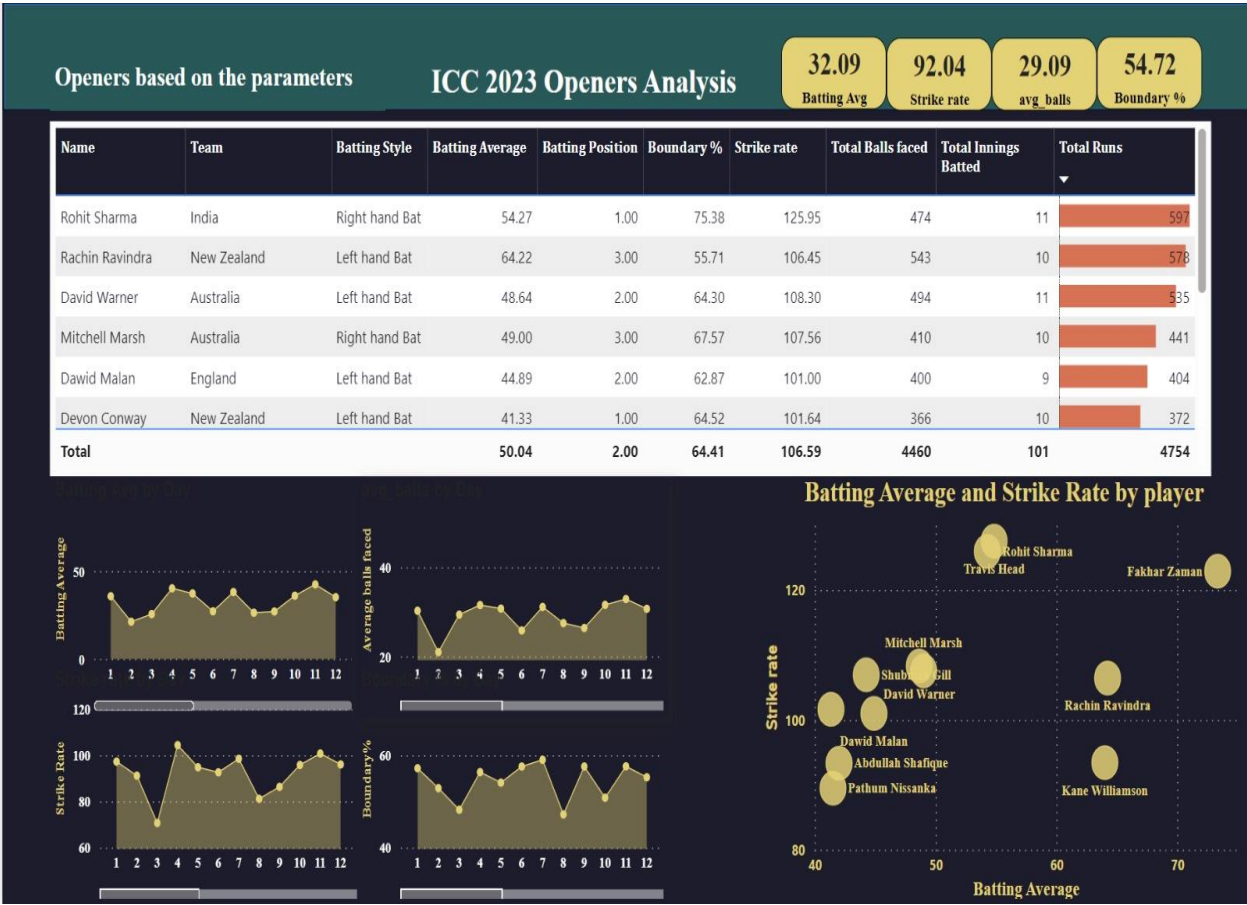
Following are the steps done for data cleaning and preprocessing:

- 1) **Insertion of Foreign Key:** Match ID from the Match Summary CSV was inserted as a foreign key into the Batting Summary CSV to establish a relationship between match details and batting statistics.
- 2) **Strike Rate Replacement:** For missing strike rate values, replaced '-' entries with zeros to ensure uniformity in the dataset.
- 3) **Captain Suffix Removal:** Removed the suffix '(c)' from captain names to standardize player names across the dataset.
- 4) **Conversion of Out/Not-Out to Binary:** Converted 'Out' and 'Not Out' strings into binary values, assigning 'Out' as 1 and 'Not Out' as 0 to facilitate numerical analysis.
- 5) **Calculation of Balls Faced:** Computed the number of balls faced by players based on the overs column, applying a formula to derive the count of balls played.
- 6) **Boundary Percentage Creation:** Constructed a new feature 'Boundary Percentage' by calculating the percentage of boundaries (four or six) scored by a player out of the total balls faced, utilizing the calculated balls faced from the overs column.
- 7) **Data Aggregation:** Aggregated the data of each players from different matches and modelled it into a single row.
- 8) **Data Validation:** Checked for inconsistencies or errors in data types, ranges, and formats.

Following are the parameters that we used to filter the openers in our dashboard:

- 1) **Batting average > 40**
- 2) **Batting Position < 4**
- 3) **Boundary Percentage > 50**
- 4) **Strike Rate > 80**

Dashboard

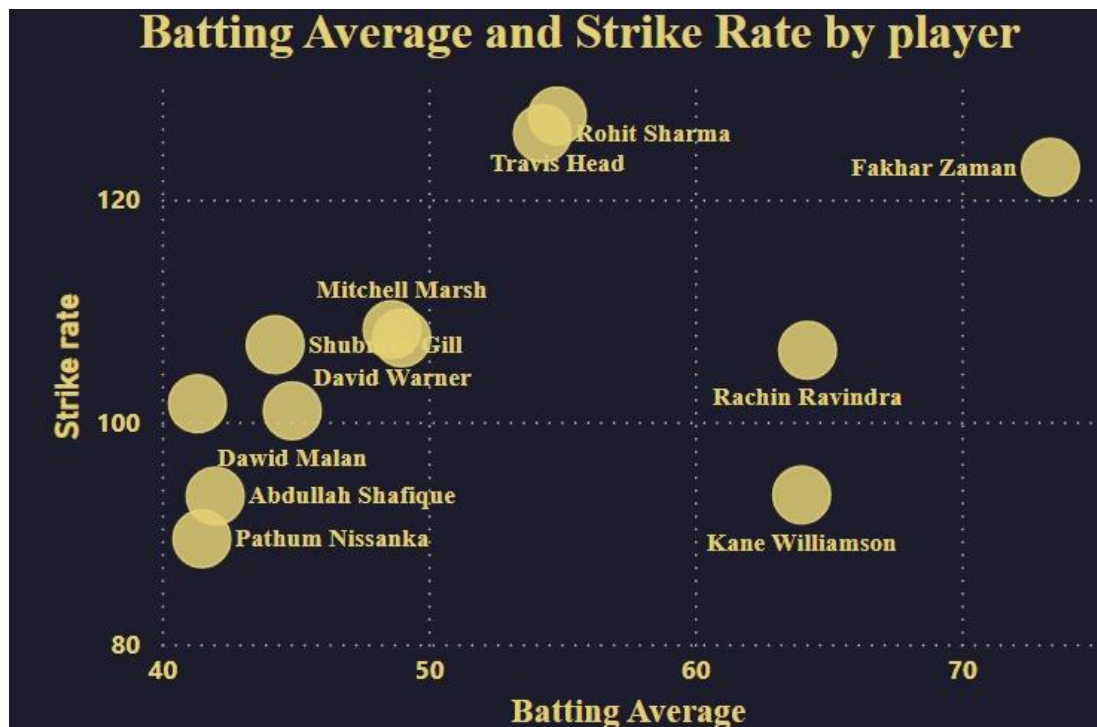


Components

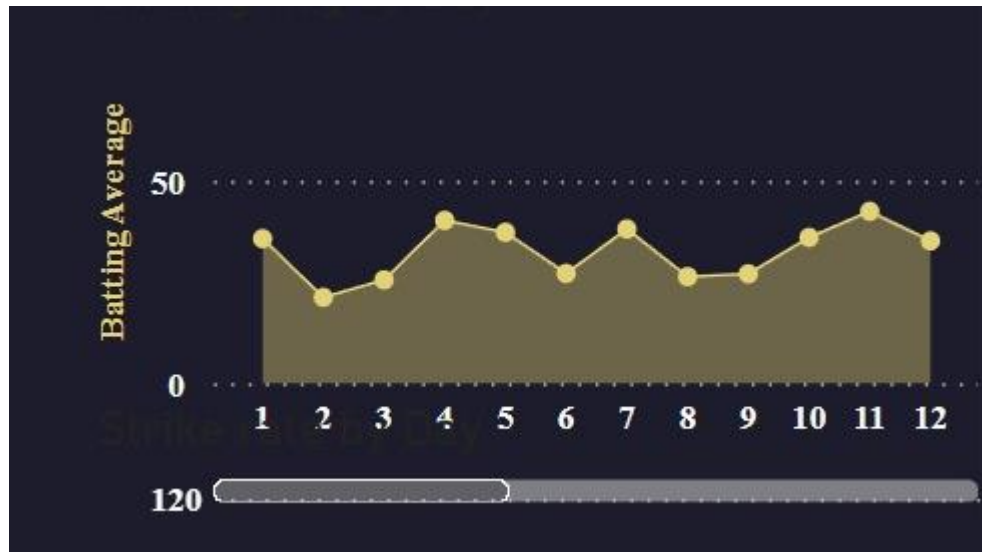
1)Table

Name	Team	Batting Style	Batting Average	Batting Position	Boundary %	Strike rate	Total Balls faced	Total Innings Batted	Total Runs
Rohit Sharma	India	Right hand Bat	54.27	1.00	75.38	125.95	474	11	597
Rachin Ravindra	New Zealand	Left hand Bat	64.22	3.00	55.71	106.45	543	10	578
David Warner	Australia	Left hand Bat	48.64	2.00	64.30	108.30	494	11	535
Mitchell Marsh	Australia	Right hand Bat	49.00	3.00	67.57	107.56	410	10	441
Dawid Malan	England	Left hand Bat	44.89	2.00	62.87	101.00	400	9	404
Devon Conway	New Zealand	Left hand Bat	41.33	1.00	64.52	101.64	366	10	372
Total			50.04	2.00	64.41	106.59	4460	101	4754

2)Scatter Plot (Batting Average vs Strike Rate by Player)



3)Graph (Batting Average)



4)Graph (Average Balls Faced)



5)Graph (Strike Rate)



6)Graph (Boundary Percentage)



7)Consolidated Parameters

32.09	92.04	29.09	54.72
Batting Avg	Strike rate	avg_balls	Boundary %

Link to Dashboard

Note: The Power BI dashboard file is attached with the submission of the report in .pbix format