# Project Report

**Student Name: Harshit Kumar Jaiswal**

**Branch: MCA-General**

**Semester: 1st**

**Subject: Python Programming Lab**

**UID: 24MCA20148**

**Section: 24MCA-3A**

**Date of Performance: 05/11/2024**

**Subject Code: 24CAH-606**

## Aim of the Project:

1. Cricket Playing XI Dashboard using python.

## Hardware and Software Requirements:

- **Software:**
  - **Download:** I have downloaded the latest version (3.12.5) of Python from the official website: python.org.
  - **Installation:** During installation, I have checked the option to "Add Python to PATH" on Windows to make running Python from the command line easier.
  - **Jupyter Notebook:** An interactive environment that allows us to write code, display outputs, and include markdown notes in a notebook format. Ideal for data science and experimentation. Install Jupyter Notebook.
  - **Anaconda**: A distribution that includes Python and many scientific libraries, along with a package manager (conda) and the Jupyter Notebook. It's especially useful for data science. Downloaded Anaconda version (2.6.2).
  - **Power BI:** Microsoft Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence. It is part of the Microsoft Power Platform. Power BI is a collection of software services, apps, and connectors that work together to turn various sources of data into static and interactive data visualizations. Data may be input by reading directly from a database, webpage, PDF, or structured files such as spreadsheets, CSV, XML, JSON, XLSX, and SharePoint

- **Hardware:**
  - I have used my personal laptop here are the following specification:
  - **Processor:** Intel Core i5 8th Generation.
  - **RAM:** 8GB DDR4.
  - **Storage:** 1TB and 128GB SSD.
  - **Operating System:** Windows 11 Pro.
  - **GPU:** Nvidia GeForce GTX 1050(4GB).

**Program Logic:**

1. **Dataset Description**: The research utilizes a comprehensive dataset from the ESPN Cric-info website, focusing on the T20 World Cup. This dataset includes a wide range of variables such as match details, player statistics, scores, and more, providing a rich source of information for in-depth analysis of games, strategies, and player performances.

2. **Web Scraping Technique**: To extract the data, web scraping techniques were implemented. Bright Data's services were employed to navigate the ESPN Cric-info website and systematically gather the required data. Web scraping was carried out with respect to the website's terms of service, ensuring ethical data collection practices.

3. **Data Transformation and Cleaning with Python**: Once the raw data was collected, Python’s data transformation capabilities were harnessed to clean and structure the dataset. This involved removing irrelevant information, handling missing values, normalizing data formats, and correcting inconsistencies, ensuring the dataset was of high quality and ready for detailed analysis.

4. **Data Manipulation with Pandas**: Pandas, a powerful Python data analysis toolkit, was used to manipulate the data effectively. With Pandas, the data was organized, sorted, and grouped to facilitate comprehensive analysis. Computational operations were performed with ease, allowing for the aggregation and summarization necessary to derive meaningful patterns and insights from the dataset.

5. **Data Visualization with Power BI:** The final step involved translating the processed data into visual form using Power BI. This tool allowed the creation of interactive dashboards that provide at-a-glance insights through charts, graphs, and tables. Power BI's dynamic visualizations enabled stakeholders to interpret complex data intuitively and make informed decisions supported by the data analysis.

**Code:**

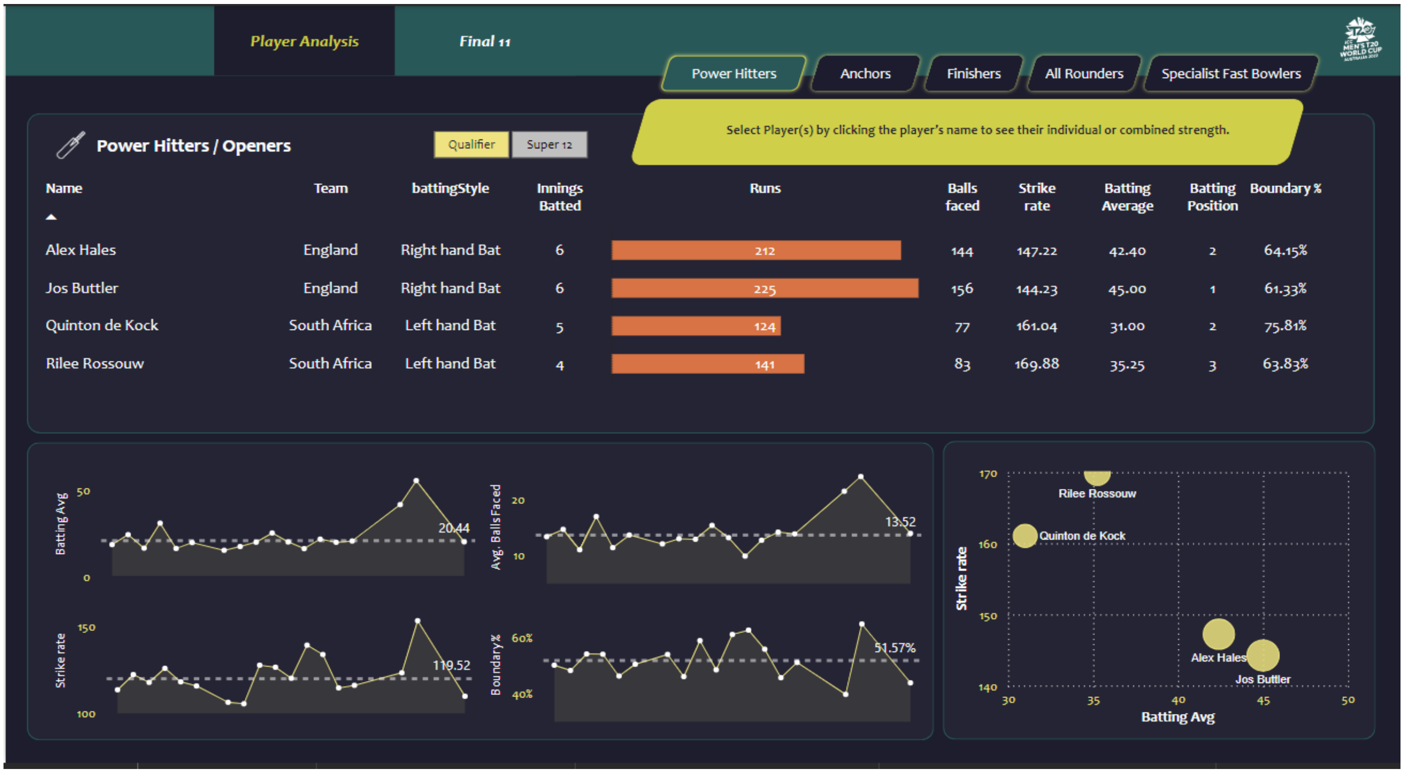
**Data Pre-processing:**

```
#import necessary libraries
import pandas as pd
import json
with open('t20_json_files/t20_wc_match_results.json') as f:
    data = json.load(f)
df_match = pd.DataFrame(data[0]['matchSummary'])
df_match.head()
df_match.shape
df_match.rename({'scorecard': 'match_id'}, axis = 1, inplace = True)
df_match.head()
match_ids_dict = {}
for index, row in df_match.iterrows():
    key1 = row['team1'] + ' Vs ' + row['team2']
    key2 = row['team2'] + ' Vs ' + row['team1']
    match_ids_dict[key1] = row['match_id']
    match_ids_dict[key2] = row['match_id']
df_match.to_csv('t20_csv_files/dim_match_summary.csv', index = False)
with open('t20_json_files/t20_wc_batting_summary.json') as f:
    data = json.load(f)
    all_records = []
    for rec in data:
        all_records.extend(rec['battingSummary'])
```

```
df_batting = pd.DataFrame(all_records)

df_batting.head(11)

df_batting['out/not_out'] = df_batting.dismissal.apply(lambda x: "out" if len(x)>0 else "not_out")

df_batting.head(11)

df_batting['match_id'] = df_batting['match'].map(match_ids_dict)

df_batting.head()

df_batting.drop(columns=["dismissal"], inplace=True)

df_batting.head(10)

df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace('â€', ''))

df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace('\xa0', ''))

df_batting.head()

df_batting.shape

df_batting.to_csv('t20_csv_files/fact_bating_summary.csv', index = False)

with open('t20_json_files/t20_wc_bowling_summary.json') as f:

    data = json.load(f)

    all_records = []

    for rec in data:

        all_records.extend(rec['bowlingSummary'])

all_records[:2]

df_bowling = pd.DataFrame(all_records)

print(df_bowling.shape)

df_bowling.head()

df_bowling['match_id'] = df_bowling['match'].map(match_ids_dict)
```

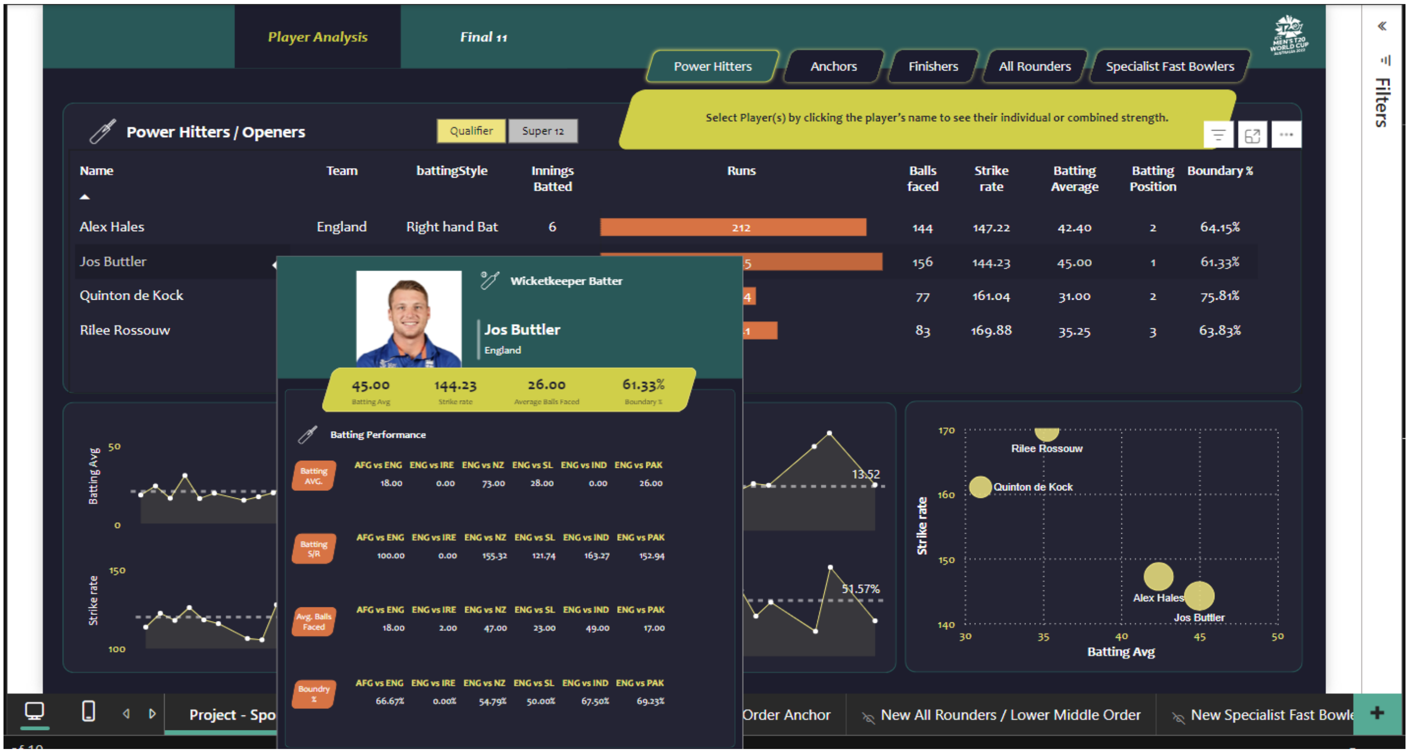
```
df_bowling.head()

df_bowling.to_csv('t20_csv_files/fact_bowling_summary.csv', index = False)

with open('t20_json_files/t20_wc_player_info.json') as f:

    data = json.load(f)

df_players = pd.DataFrame(data)

print(df_players.shape)

df_players.head(10)

df_players['name'] = df_players['name'].apply(lambda x: x.replace('â€', ''))

df_players['name'] = df_players['name'].apply(lambda x: x.replace('†', ''))

df_players['name'] = df_players['name'].apply(lambda x: x.replace('\xa0', ''))

df_players.head(10)

df_players[df_players['team'] == 'India']

df_players.to_csv('t20_csv_files/dim_players_no_images.csv', index = False)
```

## Result:



| Name | Team | battingStyle | Innings Batted | Runs | Balls faced | Strike rate | Batting Average | Batting Position | Boundary % |
|---|---|---|---|---|---|---|---|---|---|
| Alex Hales | England | Right hand Bat | 6 | 212 | 144 | 147.22 | 42.40 | 2 | 64.15% |
| Jos Buttler | England | Right hand Bat | 6 | 225 | 156 | 144.23 | 45.00 | 1 | 61.33% |
| Quinton de Kock | South Africa | Left hand Bat | 5 | 124 | 77 | 161.04 | 31.00 | 2 | 75.81% |
| Rilee Rossouw | South Africa | Left hand Bat | 4 | 141 | 83 | 169.88 | 35.25 | 3 | 63.83% |

(a) Power Hitter/ Openers Dashboard



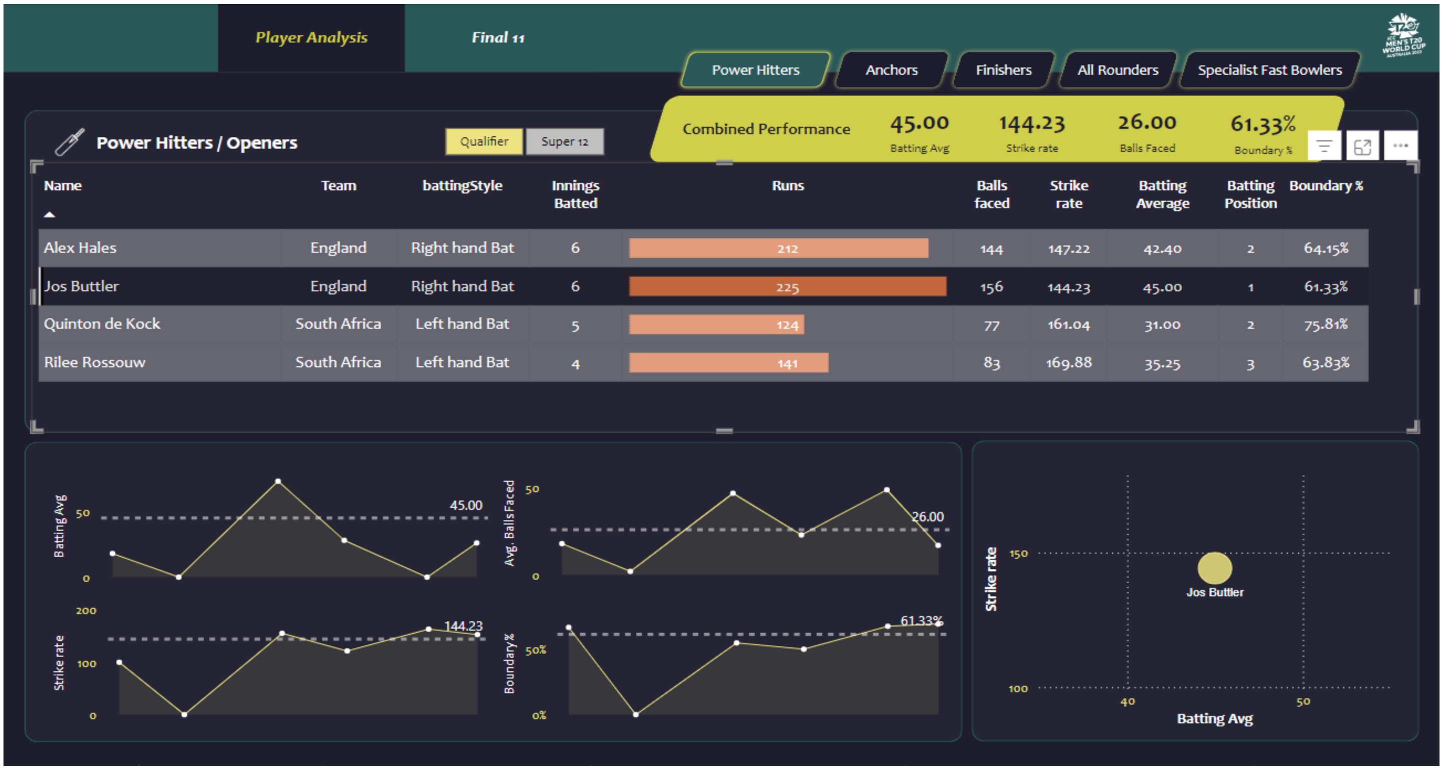
| Name | Team | battingStyle | Innings Batted | Runs | Balls faced | Strike rate | Batting Average | Batting Position | Boundary % |
|---|---|---|---|---|---|---|---|---|---|
| Alex Hales | England | Right hand Bat | 6 | 212 | 144 | 147.22 | 42.40 | 2 | 64.15% |
| Jos Buttler | | | | 5 | 156 | 144.23 | 45.00 | 1 | 61.33% |
| Quinton de Kock | | | | 4 | 77 | 161.04 | 31.00 | 2 | 75.81% |
| Rilee Rossouw | | | | 1 | 83 | 169.88 | 35.25 | 3 | 63.83% |

Jos Buttler
England

| Batting Avg | Strike rate | Average Balls Faced | Boundary % |
|---|---|---|---|
| 45.00 | 144.23 | 26.00 | 61.33% |

| | AFG vs ENG | ENG vs IRE | ENG vs NZ | ENG vs SL | ENG vs IND | ENG vs PAK |
|---|---|---|---|---|---|---|
| Batting AVG. | 18.00 | 0.00 | 73.00 | 28.00 | 0.00 | 26.00 |
| Batting S/R | 100.00 | 0.00 | 155.32 | 121.74 | 163.27 | 152.94 |
| Avg. Balls Faced | 18.00 | 2.00 | 47.00 | 23.00 | 49.00 | 17.00 |
| Boundry % | 66.67% | 0.00% | 54.79% | 50.00% | 67.50% | 69.23% |

(b) Individual Stats Dashboard

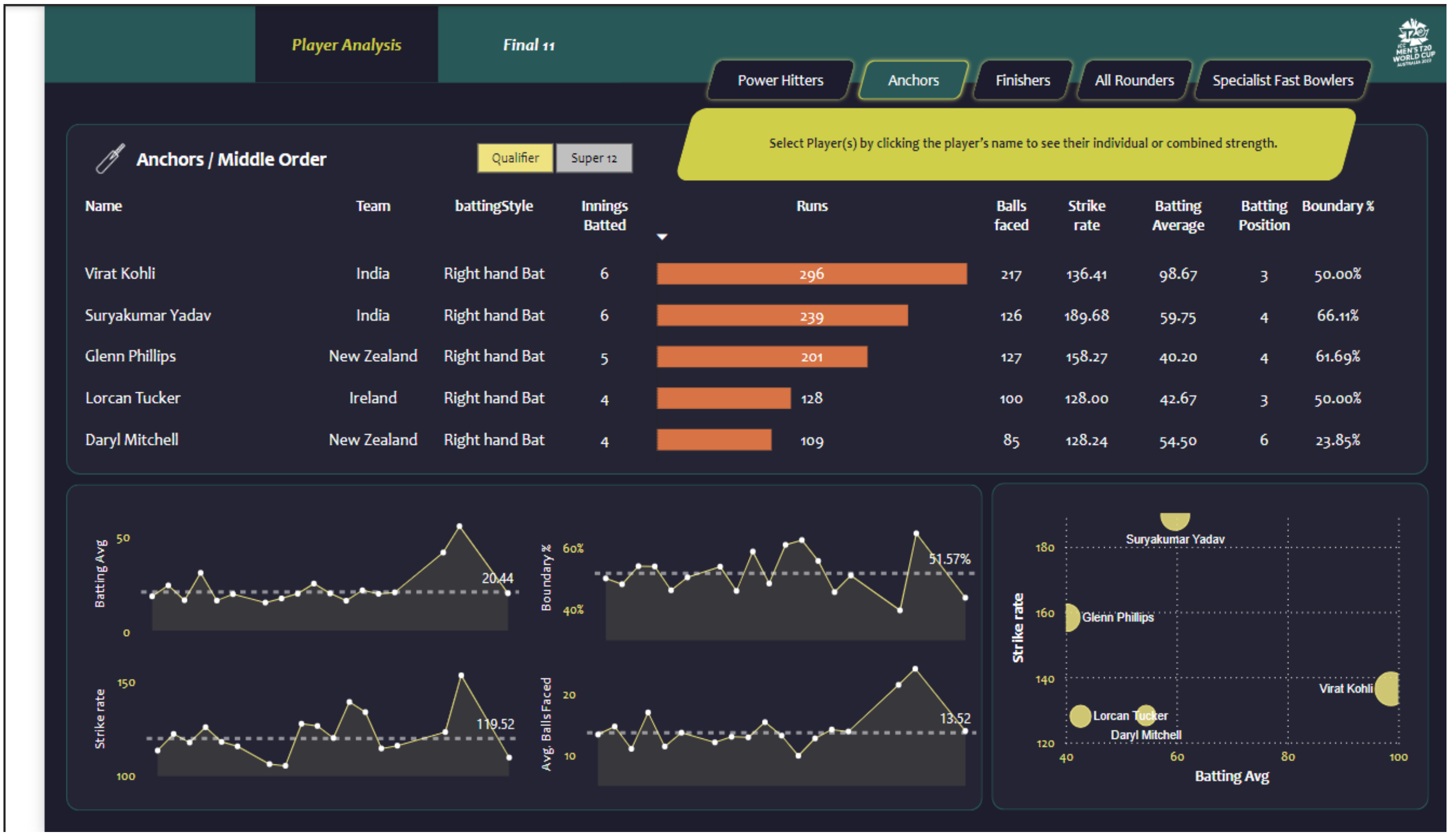


Power Hitters / Openers

Combined Performance 45.00 Batting Avg 144.23 Strike rate 26.00 Balls Faced 61.33% Boundary %

| Name | Team | battingStyle | Innings Batted | Runs | Balls faced | Strike rate | Batting Average | Batting Position | Boundary % |
|---|---|---|---|---|---|---|---|---|---|
| Alex Hales | England | Right hand Bat | 6 | 212 | 144 | 147.22 | 42.40 | 2 | 64.15% |
| Jos Buttler | England | Right hand Bat | 6 | 225 | 156 | 144.23 | 45.00 | 1 | 61.33% |
| Quinton de Kock | South Africa | Left hand Bat | 5 | 124 | 77 | 161.04 | 31.00 | 2 | 75.81% |
| Rilee Rossouw | South Africa | Left hand Bat | 4 | 141 | 83 | 169.88 | 35.25 | 3 | 63.83% |

(c) Batting Performance Dashboard

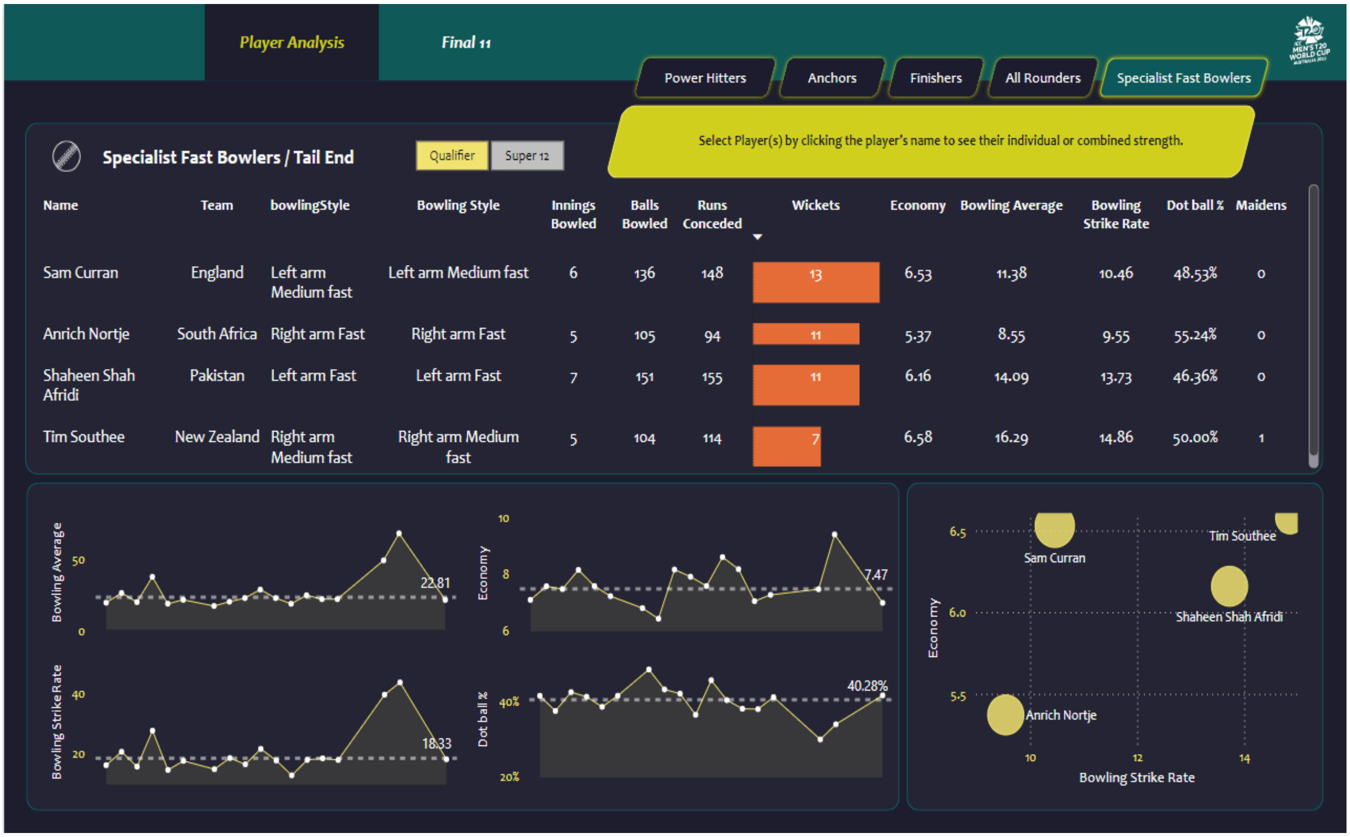


Anchors / Middle Order

Select Player(s) by clicking the player's name to see their individual or combined strength.

| Name | Team | battingStyle | Innings Batted | Runs | Balls faced | Strike rate | Batting Average | Batting Position | Boundary % |
|---|---|---|---|---|---|---|---|---|---|
| Virat Kohli | India | Right hand Bat | 6 | 296 | 217 | 136.41 | 98.67 | 3 | 50.00% |
| Suryakumar Yadav | India | Right hand Bat | 6 | 239 | 126 | 189.68 | 59.75 | 4 | 66.11% |
| Glenn Phillips | New Zealand | Right hand Bat | 5 | 201 | 127 | 158.27 | 40.20 | 4 | 61.69% |
| Lorcan Tucker | Ireland | Right hand Bat | 4 | 128 | 100 | 128.00 | 42.67 | 3 | 50.00% |
| Daryl Mitchell | New Zealand | Right hand Bat | 4 | 109 | 85 | 128.24 | 54.50 | 6 | 23.85% |

(d) Middle Order Batsmen Dashboard

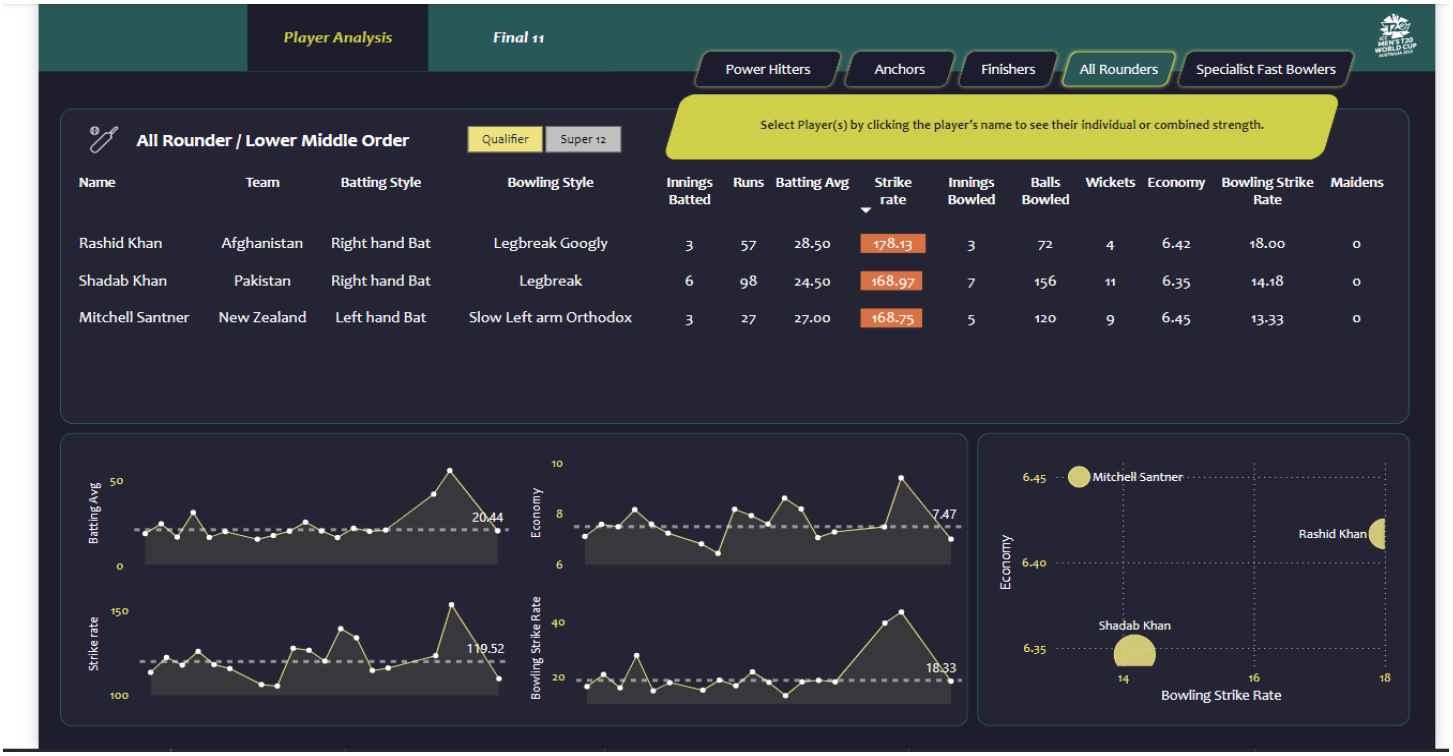


Select Player(s) by clicking the player's name to see their individual or combined strength.

Specialist Fast Bowlers / Tail End

| Name | Team | bowlingStyle | Bowling Style | Innings Bowled | Balls Bowled | Runs Conceded | Wickets | Economy | Bowling Average | Bowling Strike Rate | Dot ball % | Maidens |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sam Curran | England | Left arm Medium fast | Left arm Medium fast | 6 | 136 | 148 | 13 | 6.53 | 11.38 | 10.46 | 48.53% | 0 |
| Anrich Nortje | South Africa | Right arm Fast | Right arm Fast | 5 | 105 | 94 | 11 | 5.37 | 8.55 | 9.55 | 55.24% | 0 |
| Shaheen Shah Afridi | Pakistan | Left arm Fast | Left arm Fast | 7 | 151 | 155 | 11 | 6.16 | 14.09 | 13.73 | 46.36% | 0 |
| Tim Southee | New Zealand | Right arm Medium fast | Right arm Medium fast | 5 | 104 | 114 | 7 | 6.58 | 16.29 | 14.86 | 50.00% | 1 |

(e) Fast Bowlers Dashboard

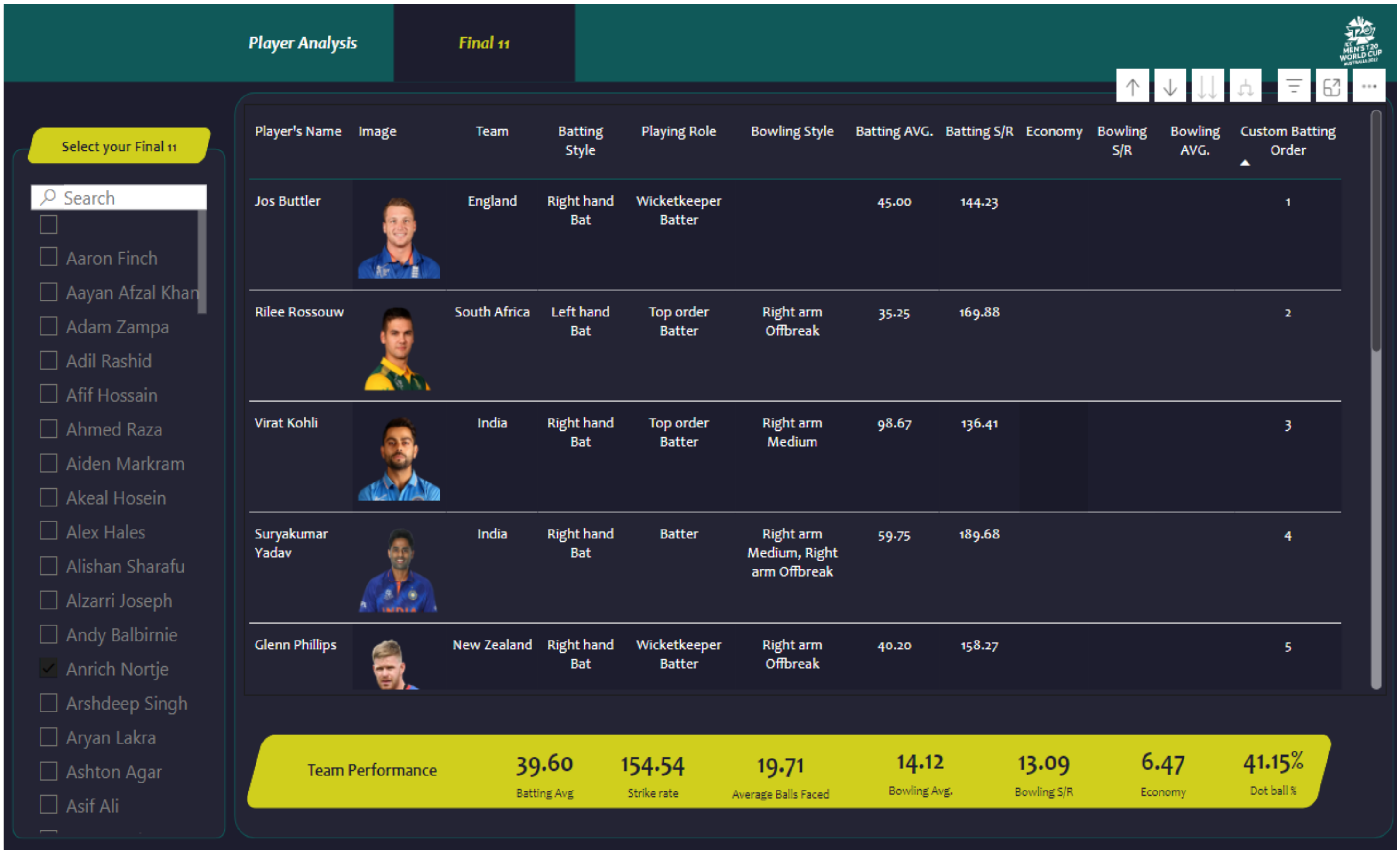


Select Player(s) by clicking the player's name to see their individual or combined strength.

All Rounder / Lower Middle Order

| Name | Team | Batting Style | Bowling Style | Innings Batted | Runs | Batting Avg | Strike rate | Innings Bowled | Balls Bowled | Wickets | Economy | Bowling Strike Rate | Maidens |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rashid Khan | Afghanistan | Right hand Bat | Legbreak Googly | 3 | 57 | 28.50 | 178.13 | 3 | 72 | 4 | 6.42 | 18.00 | 0 |
| Shadab Khan | Pakistan | Right hand Bat | Legbreak | 6 | 98 | 24.50 | 168.97 | 7 | 156 | 11 | 6.35 | 14.18 | 0 |
| Mitchell Santner | New Zealand | Left hand Bat | Slow Left arm Orthodox | 3 | 27 | 27.00 | 168.75 | 5 | 120 | 9 | 6.45 | 13.33 | 0 |

(f) Limited Brawler Insights



☐
☐ Aaron Finch
☐ Aayan Afzal Khan
☐ Adam Zampa
☐ Adil Rashid
☐ Afif Hossain
☐ Ahmed Raza
☐ Aiden Markram
☐ Akeal Hosein
☐ Alex Hales
☐ Alishan Sharafu
☐ Alzarri Joseph
☐ Andy Balbirnie
☑ Anrich Nortje
☐ Arshdeep Singh
☐ Aryan Lakra
☐ Ashton Agar
☐ Asif Ali

| Player's Name | Image | Team | Batting Style | Playing Role | Bowling Style | Batting AVG. | Batting S/R | Economy | Bowling S/R | Bowling AVG. | Custom Batting Order |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jos Buttler | | England | Right hand Bat | Wicketkeeper Batter | | 45.00 | 144.23 | | | | 1 |
| Rilee Rossouw | | South Africa | Left hand Bat | Top order Batter | Right arm Offbreak | 35.25 | 169.88 | | | | 2 |
| Virat Kohli | | India | Right hand Bat | Top order Batter | Right arm Medium | 98.67 | 136.41 | | | | 3 |
| Suryakumar Yadav | | India | Right hand Bat | Batter | Right arm Medium, Right arm Offbreak | 59.75 | 189.68 | | | | 4 |
| Glenn Phillips | | New Zealand | Right hand Bat | Wicketkeeper Batter | Right arm Offbreak | 40.20 | 158.27 | | | | 5 |

| Team Performance | Batting Avg | Strike rate | Average Balls Faced | Bowling Avg. | Bowling S/R | Economy | Dot ball % |
|---|---|---|---|---|---|---|---|
| | 39.60 | 154.54 | 19.71 | 14.12 | 13.09 | 6.47 | 41.15% |

(g) Final XI Dashboard

## Learning outcomes (What I have learnt):

a) Learned about data analysis and visualization libraries like pandas, NumPy, Matplotlib and Seaborn.
b) Learned about web scrapping.
c) Leaned about creating dataset using pandas library.
d) Learned about converting the dataset to DataFrame using pd.DataFrame.
e) Power BI dashboard creation.
f) Data Modeling and visualization techniques.

## Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

| Sr. No. | Parameters | Marks Obtained | Maximum Marks |
|---|---|---|---|
| 1. | Worksheet | | 08 |
| 2. | Viva | | 10 |
| 3. | Simulation | | 12 |
| 4. | Total Marks | | 30 |

Teacher Signature