

T20 Analysis

Cricket is undoubtedly one of the most popular sports in the world, and the craze for cricket is not limited to any particular country or region. Cricket has a massive fan base, and its popularity has only grown over the years.

It is a sport that requires a variety of skills, and as a result, there are different types of players required on a team. Here are some of them:

Openers / Power Hitters: An opener is a batsman who opens the batting for their team. Openers need to have a solid technique, as they are the first ones to face the new ball. They need to be able to negotiate the swing and seam movement that the new ball can produce. Criteria considered:

Parameter	Description	Criteria
Batting Average	Average runs scored in an inning	>30
Strike Rate	No. of runs scored per 100 balls	>140
Total no. of innings batted	Total innings batted	>3
Boundary %	% of runs scored in boundaries	>50
Batting Position	Order in which the batsman played	<4

Anchors / Middle Order: Then comes anchor, a batsman who plays a slow and steady innings, with the aim of holding the innings together and keeping the wicket intact. The role of an anchor is to provide stability and support to the team, while allowing other more aggressive batsmen to play around them. They need to have a good technique, as they may face a lot of good deliveries from the opposition as well as should be able to negotiate the bowlers and play defensively when required. Criteria considered:

Parameter	Description	Criteria
Batting Average	Average runs scored in an inning	>40
Strike Rate	No. of runs scored per 100 balls	>125
Total no. of innings batted	Total innings batted	>3

Average ball faced	Average ball faced by a player in an inning	>20
Batting Position	Order in which the batsman played	>2

Finishers / Lower Order Anchors: A finisher is a batsman who is known for his ability to score quick runs towards the end of an innings, usually in the last few overs. They need to have the ability to hit big shots and score quick runs. They often rely on their power-hitting ability to clear the boundary and take the game away from the opposition. Criteria considered:

Parameter	Description	Criteria
Batting Average	Average runs scored in an inning	>25
Strike Rate	No. of runs scored per 100 balls	>130
Total no. of innings batted	Total innings batted	>3
Average ball faced	Average ball faced by a player in an inning	>12
Batting Position	Order in which the batsman played	>4
Innings Bowled	Total innings bowled by a bowler	>1

All-rounders / Lower Order: An all-rounder is a player who can contribute with both bat and ball and is often considered to be one of the most valuable players in a team, as the all-rounder can provide balance and flexibility to the side. Criteria considered:

Parameter	Description	Criteria
Batting Average	Average runs scored in an inning	>15
Strike Rate	No. of runs scored per 100 balls	>140
Total no. of innings batted	Total innings batted	>2
Batting Position	Order in which the batsman played	>4
Innings Bowled	Total innings bowled by a bowler	>2
Bowling Economy	Average runs allowed per over	<7

Bowling Strike Rate	Average no. of balls required to take a wicket	<20
---------------------	--	-----

Fast-bowler: A fast bowler is a player who specializes in bowling at high speeds. Fast bowlers are known for their ability to generate pace and bounce, which can make it difficult for batsmen to score runs and can also result in wickets. Fast bowlers need to have good bowling technique, as they need to be able to maintain their speed and accuracy over a long spell of bowling.

Criteria considered:

Parameter	Description	Criteria
Innings Bowled	Total innings bowled by a bowler	>4
Bowling Economy	Average runs allowed per over	<7
Bowling Strike Rate	Average no. of balls required to take a wicket	<16
Bowling Style	Bowling Style of player	Must contain "Fast"
Bowling Average	No. of runs allowed per wicket	<20
Dot Ball %	% of dot balls allowed	>40

This project aims toward finding out the above players in T20. The process of analysis has following steps included:

- Data Cleaning using Jupyter Notebook
- Data Transformation using Power Query
- Data Modelling and Dashboard Building using Power BI

Data Collection and Preprocessing:

The initial step in our project is data collection. The data needed for this visualization is searched and procured from the kaggle website. We gathered four different files namely, fact_bowling_summary, fact_batting_summary, dim_players and dim_match_summary. These different files are to be connected which will be done in the processing part. Since this acquired data is in its raw form, it requires cleaning and preprocessing before it is brought into Power BI for visualization.

The next step involves Data Cleaning and Preprocessing which is done in Jupyter Notebook.

Firstly, we import all the necessary libraries which are needed throughout the process:

```
In [1]: # Import necessary Libraries
import pandas as pd
import json
```

Now loading and displaying the files which we upload in our Jupyter Notebook:

```
In [2]: with open('t20_json_files/t20_wc_match_results.json') as f:
    data = json.load(f)
    data
```

```
Out[2]: [{'matchSummary': [{team1: 'Namibia',
                           team2: 'Sri Lanka',
                           winner: 'Namibia',
                           margin: '55 runs',
                           ground: 'Geelong',
                           matchDate: 'Oct 16, 2022',
                           scorecard: 'T20I # 1823'},
                           {'team1: 'Netherlands',
                           team2: 'U.A.E.',
                           winner: 'Netherlands',
                           margin: '3 wickets',
                           ground: 'Geelong',
                           matchDate: 'Oct 16, 2022',
                           scorecard: 'T20I # 1825'},
                           {'team1: 'Scotland',
                           team2: 'West Indies',
                           winner: 'Scotland',
                           margin: '42 runs',
                           ground: 'Hobart',
                           matchDate: 'Oct 17, 2022',
                           scorecard: 'T20I # 1826'},
                           {'team1: 'Ireland',
                           team2: 'Zimbabwe',
                           winner: 'Zimbabwe',
                           margin: '31 runs',
                           ground: 'Hobart',
                           matchDate: 'Oct 17, 2022',
                           scorecard: 'T20I # 1828'},
                           {'team1: 'Namibia',
                           team2: 'Netherlands',
                           winner: 'Netherlands',
                           margin: '5 wickets',
                           ground: 'Geelong',
                           matchDate: 'Oct 18, 2022',
                           scorecard: 'T20I # 1830'}]]
```

As visible, the data is in a dictionary format. So now we need to convert this data into a table which will be easier to access and understand. For this purpose, we create a dataframe:

```
# creating dataframe
df_match = pd.DataFrame(data[0]['matchSummary'])
df_match.head()
```

	team1	team2	winner	margin	ground	matchDate	scorecard
0	Namibia	Sri Lanka	Namibia	55 runs	Geelong	Oct 16, 2022	T20I # 1823
1	Netherlands	U.A.E.	Netherlands	3 wickets	Geelong	Oct 16, 2022	T20I # 1825
2	Scotland	West Indies	Scotland	42 runs	Hobart	Oct 17, 2022	T20I # 1826
3	Ireland	Zimbabwe	Zimbabwe	31 runs	Hobart	Oct 17, 2022	T20I # 1828
4	Namibia	Netherlands	Netherlands	5 wickets	Geelong	Oct 18, 2022	T20I # 1830

Now checking the dimension of the newly created table:

```
# checking dimension
df_match.shape
# 45 rows, 7 columns
```

(45, 7)

Further, we wish to use the column ‘scorecard’ as the key of this particular dataframe. Meaning, this column is to be used as Match ID for connecting with other tables (as a primary key).

Renaming ‘scorecard’ as ‘match_id’:

```
# Renaming scorecard
# to treat scorecard as a primary key in order to connect to other tables in powerbi

df_match.rename({'scorecard' : 'match_id'}, axis = 1, inplace = True)
df_match.head()
```

	team1	team2	winner	margin	ground	matchDate	match_id
0	Namibia	Sri Lanka	Namibia	55 runs	Geelong	Oct 16, 2022	T20I # 1823
1	Netherlands	U.A.E.	Netherlands	3 wickets	Geelong	Oct 16, 2022	T20I # 1825
2	Scotland	West Indies	Scotland	42 runs	Hobart	Oct 17, 2022	T20I # 1826
3	Ireland	Zimbabwe	Zimbabwe	31 runs	Hobart	Oct 17, 2022	T20I # 1828
4	Namibia	Netherlands	Netherlands	5 wickets	Geelong	Oct 18, 2022	T20I # 1830

Creating a Dictionary for all Teams and their corresponding Match Ids:

```
# Creating a dictionary to give match IDs to the corresponding teams

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"]
```

Saving this data into a csv file:

```
df_match.to_csv('t20_csv_files/dim_match_summary.csv', index = False)
```

Now, open another file. In this we create a list ‘all_records’ and then use a for loop for appending all the records of ‘battingsummary’ in that list. Then using this newly created list to make and display a dataframe.

Batting Summary

```
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()
```

	match	teamInnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c Mendis b Karunaratne	20	12	1	2	166.66
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan	26	24	2	0	108.33
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	c Gunathilaka b PWH de Silva	20	24	0	0	83.33

Instead of having that ‘dismissal’ column we can rather have a column named ‘out/not out’ which will tell us whether any player was out or not. Therefore, we created a new column based on the dismissal column and removed the previous column as it was not needed anymore.

```
# Creating new column 'out/not out' based on dismissal
df_batting["out/not_out"] = df_batting.dismissal.apply(lambda x : "out" if len(x) > 0 else "not out")
df_batting.head(10)
```

	match	teamInnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	out/not_out
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	out
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00	out
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c Mendis b Karunaratne	20	12	1	2	166.66	out
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan	26	24	2	0	108.33	out
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	c Gunathilaka b PWH de Silva	20	24	0	0	83.33	out
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	run out (Gunathilaka/Mendis)	44	28	4	0	157.14	out
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	c Mendis b Theekshana	0	1	0	0	0.00	out
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit		31	16	2	2	193.75	not out
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	c Smit b Shikongo	9	10	1	0	90.00	out
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendis	c Green b Wiese	6	6	0	0	100.00	out

Sakshi Zagade, Shruti Sakpal

```
# Removing 'dismissal' column  
df_batting.drop(columns=["dismissal"], inplace = True)  
df_batting.head(10)
```

	match	teamInnings	battingPos	batsmanName	runs	balls	4s	6s	SR	out/not_out
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	6	0	0	50.00	out
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	9	1	0	100.00	out
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	12	1	2	166.66	out
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	24	2	0	108.33	out
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	24	0	0	83.33	out
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	44	28	4	0	157.14	out
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	0	1	0	0	0.00	out
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit	31	16	2	2	193.75	not out
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	9	10	1	0	90.00	out
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendisâ€	6	6	0	0	100.00	out

The column ‘batsmanName’ has certain unnecessary or illegible characters which don’t make any sense. So our next step includes removing those unnecessary characters.

```
# Removing unnecessary characters from 'batsmanName'  
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(11)
```

	match	teamInnings	battingPos	batsmanName	runs	balls	4s	6s	SR	out/not_out
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	6	0	0	50.00	out
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	9	1	0	100.00	out
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	12	1	2	166.66	out
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	24	2	0	108.33	out
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	24	0	0	83.33	out
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	44	28	4	0	157.14	out
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	0	1	0	0	0.00	out
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit	31	16	2	2	193.75	not out
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	9	10	1	0	90.00	out
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendis	6	6	0	0	100.00	out
10	Namibia Vs Sri Lanka	Sri Lanka	3	Dhananjaya de Silva	12	11	1	0	109.09	out

To bring the previously created ‘match_id’ dictionary, we create a new column and give the dictionary to it. By doing this, we have created a link between the two tables.

```
# Creating new column 'Match ID' and giving the previously created dictionary
df_batting["match_id"] = df_batting["match"].map(match_ids_dict)
df_batting.head()

# Now we have linked the two tables using match_id
```

	match	teamInnings	battingPos	batsmanName	runs	balls	4s	6s	SR	out/not_out	match_id
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	6	0	0	50.00	out	T20I # 1823
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	9	1	0	100.00	out	T20I # 1823
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	12	1	2	166.66	out	T20I # 1823
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	24	2	0	108.33	out	T20I # 1823
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	24	0	0	83.33	out	T20I # 1823

Now checking the shape of the dataframe and exporting it into csv file.

```
df_batting.shape
```

```
(699, 11)
```

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

Similarly, doing this same procedure with another table ‘fact_bowling_summary’:

Bowling Summary

```
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]

[{'match': 'Namibia Vs Sri Lanka',
 'bowlingTeam': 'Sri Lanka',
 'bowlerName': 'Maheesh Theekshana',
 'overs': '4',
 'maiden': '0',
 'runs': '23',
 'wickets': '1',
 'economy': '5.75',
 '0s': '7',
 '4s': '0',
 '6s': '0',
 'wides': '2',
 'noBalls': '0'},
 {'match': 'Namibia Vs Sri Lanka',
 'bowlingTeam': 'Sri Lanka',
 'bowlerName': 'Dushmantha Chameera',
 'overs': '4',
 'maiden': '0',
 'runs': '39',
 'wickets': '1',
 'economy': '9.75',
 '0s': '6',
 '4s': '3',
 '6s': '1',
 'wides': '2',
 'noBalls': '0'}]
```

Sakshi Zagade, Shruti Sakpal

```
df_bowling = pd.DataFrame(all_records)
print(df_bowling.shape)
df_bowling.head()
```

```
(500, 13)
```

	match	bowlingTeam	bowlerName	overs	maiden	runs	wickets	economy	0s	4s	6s	wides	noBalls
0	Namibia Vs Sri Lanka	Sri Lanka	Maheesh Theekshana	4	0	23	1	5.75	7	0	0	2	0
1	Namibia Vs Sri Lanka	Sri Lanka	Dushmantha Chameera	4	0	39	1	9.75	6	3	1	2	0
2	Namibia Vs Sri Lanka	Sri Lanka	Pramod Madushan	4	0	37	2	9.25	6	3	1	0	0
3	Namibia Vs Sri Lanka	Sri Lanka	Chamika Karunaratne	4	0	36	1	9.00	7	3	1	1	0
4	Namibia Vs Sri Lanka	Sri Lanka	Wanindu Hasaranga de Silva	4	0	27	1	6.75	8	1	1	0	0

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

	match	bowlingTeam	bowlerName	overs	maiden	runs	wickets	economy	0s	4s	6s	wides	noBalls	match_id
0	Namibia Vs Sri Lanka	Sri Lanka	Maheesh Theekshana	4	0	23	1	5.75	7	0	0	2	0	T20I # 1823
1	Namibia Vs Sri Lanka	Sri Lanka	Dushmantha Chameera	4	0	39	1	9.75	6	3	1	2	0	T20I # 1823
2	Namibia Vs Sri Lanka	Sri Lanka	Pramod Madushan	4	0	37	2	9.25	6	3	1	0	0	T20I # 1823
3	Namibia Vs Sri Lanka	Sri Lanka	Chamika Karunaratne	4	0	36	1	9.00	7	3	1	1	0	T20I # 1823
4	Namibia Vs Sri Lanka	Sri Lanka	Wanindu Hasaranga de Silva	4	0	27	1	6.75	8	1	1	0	0	T20I # 1823

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

Opening another file containing the players information:

Process Players Information

```
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)
```

```
(219, 6)
```

	name	team	battingStyle	bowlingStyle	playingRole	description
0	Michael van Lingen	Namibia	Left hand Bat	Left arm Medium	Bowling Allrounder	
1	Divan la Cock	Namibia	Right hand Bat	Legbreak	Opening Batter	
2	Jan Nicol Loftie-Eaton	Namibia	Left hand Bat	Right arm Medium, Legbreak	Batter	
3	Stephan Baard	Namibia	Right hand Bat	Right arm Medium fast	Batter	
4	Gerhard Erasmus(c)	Namibia	Right hand Bat	Right arm Offbreak	Allrounder	
5	Jan Frylinck	Namibia	Left hand Bat	Left arm Fast medium	Allrounder	
6	David Wiese	Namibia	Right hand Bat	Right arm Medium fast	Allrounder	David Wiese joined a marked outflow of South Africa's top-order batsmen.
7	JJ Smit	Namibia	Right hand Bat	Left arm Medium fast	Bowling Allrounder	
8	Pathum Nissanka	Sri Lanka	Right hand Bat		Top order Batter	
9	Kusal Mendis	Sri Lanka	Right hand Bat	Legbreak	Wicketkeeper Batter	Blessed with a compact technique, an aggressive and accurate shot selection.

Sakshi Zagade, Shruti Sakpal

As did earlier, removing unnecessary characters from players name:

```
# Cleaning unnecessary characters  
df_players['name'] = df_players['name'].apply(lambda x: x.replace('ä', ''))  
df_players['name'] = df_players['name'].apply(lambda x: x.replace('ü', ''))  
df_players['name'] = df_players['name'].apply(lambda x: x.replace('\xa0', ''))  
df_players.head(10)
```

	name	team	battingStyle	bowlingStyle	playingRole	description
0	Michael van Lingen	Namibia	Left hand Bat	Left arm Medium	Bowling Allrounder	
1	Divan la Cock	Namibia	Right hand Bat	Legbreak	Opening Batter	
2	Jan Nicol Loftie-Eaton	Namibia	Left hand Bat	Right arm Medium, Legbreak	Batter	
3	Stephan Baard	Namibia	Right hand Bat	Right arm Medium fast	Batter	
4	Gerhard Erasmus(c)	Namibia	Right hand Bat	Right arm Offbreak	Allrounder	
5	Jan Frylinck	Namibia	Left hand Bat	Left arm Fast medium	Allrounder	
6	David Wiese	Namibia	Right hand Bat	Right arm Medium fast	Allrounder	David Wiese joined a marked outflow of South Africa's...
7	JJ Smit	Namibia	Right hand Bat	Left arm Medium fast	Bowling Allrounder	
8	Pathum Nissanka	Sri Lanka	Right hand Bat		Top order Batter	
9	Kusal Mendis	Sri Lanka	Right hand Bat	Legbreak	Wicketkeeper Batter	Blessed with a compact technique, an aggressive...

Display players information having team as India:

	name	team	battingStyle	bowlingStyle	playingRole	description
127	KL Rahul	India	Right hand Bat		Opening Batter	A tall, elegant right-hand batsman who can keep his head in the game...
128	Rohit Sharma(c)	India	Right hand Bat	Right arm Offbreak	Top order Batter	Languid and easy on the eye, Rohit Sharma owned the crease...
129	Virat Kohli	India	Right hand Bat	Right arm Medium	Top order Batter	India has given to the world many a great cricketer, Virat Kohli...
130	Suryakumar Yadav	India	Right hand Bat	Right arm Medium, Right arm Offbreak	Batter	Hard-hitting 360-degree batter Suryakumar Yadav is a force to be reckoned with...
131	Axar Patel	India	Left hand Bat	Slow Left arm Orthodox	Bowling Allrounder	Left-arm spinner Axar Patel has been instrumental in India's success...
132	Hardik Pandya	India	Right hand Bat	Right arm Medium fast	Allrounder	Hardik Pandya swears by living life king size...
133	Dinesh Karthik	India	Right hand Bat	Right arm Offbreak	Wicketkeeper Batter	Not many would forget the sight of Dinesh Kartik behind the stumps...
134	Ravichandran Ashwin	India	Right hand Bat	Right arm Offbreak	Bowling Allrounder	R Ashwin took the tricks and skills he learned from his father and...
135	Bhuvneshwar Kumar	India	Right hand Bat	Right arm Medium	Bowler	At the time of his India debut in 2012, Bhuvneshwar Kumar was a...
136	Arshdeep Singh	India	Left hand Bat	Left arm Medium fast	Bowler	
137	Mohammed Shami	India	Right hand Bat	Right arm Fast	Bowler	Mohammed Shami was India's leading fast bowler...
192	Deepak Hooda	India	Right hand Bat	Right arm Offbreak	Allrounder	An allrounder who can bat in any position, Deepak Hooda...
211	Rishabh Pant	India	Left hand Bat		Wicketkeeper Batter	A match-turning, swashbuckling batter-keeper in the making...

Exporting this file into csv:

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

Here we complete the preprocessing part. The next step consists of bringing these csv files into PowerBI for further transformation.

Sakshi Zagade, Shruti Sakpal

Import all the four csv files into PowerBI for transformation.

In dim_match_summary we create a new column 'stage' which tells us if the match was before 22nd october 2022 it was a qualifier else it was Super 12.

The screenshot shows the 'Add Conditional Column' dialog in PowerBI. The 'New column name' is 'stage'. The 'If' condition is 'matchDate is before or equal... 10/22/2022 Then Qualifier'. The 'Else' condition is 'Super 12'. Below the dialog is the 'dim_match_summary' table view, showing columns: team1, team2, winner, margin, ground, matchDate, match_id, stage. The data includes various international teams and their match details, with the 'stage' column indicating whether the match was a 'Qualifier' or 'Super 12'.

team1	team2	winner	margin	ground	matchDate	match_id	stage
Namibia	Sri Lanka	Namibia	55 runs	Geelong	16 October 2022	T20I # 1823	Qualifier
Netherlands	U.A.E.	Netherlands	3 wickets	Geelong	16 October 2022	T20I # 1825	Qualifier
Scotland	West Indies	Scotland	42 runs	Hobart	17 October 2022	T20I # 1826	Qualifier
Ireland	Zimbabwe	Zimbabwe	31 runs	Hobart	17 October 2022	T20I # 1828	Qualifier
Namibia	Netherlands	Netherlands	5 wickets	Geelong	18 October 2022	T20I # 1830	Qualifier
Sri Lanka	U.A.E.	Sri Lanka	79 runs	Geelong	18 October 2022	T20I # 1832	Qualifier
Ireland	Scotland	Ireland	6 wickets	Hobart	19 October 2022	T20I # 1833	Qualifier
West Indies	Zimbabwe	West Indies	31 runs	Hobart	19 October 2022	T20I # 1834	Qualifier
Netherlands	Sri Lanka	Sri Lanka	16 runs	Geelong	20 October 2022	T20I # 1835	Qualifier
Namibia	U.A.E.	U.A.E.	7 runs	Geelong	20 October 2022	T20I # 1836	Qualifier
Ireland	West Indies	Ireland	9 wickets	Hobart	21 October 2022	T20I # 1837	Qualifier
Scotland	Zimbabwe	Zimbabwe	5 wickets	Hobart	21 October 2022	T20I # 1838	Qualifier
Australia	New Zealand	New Zealand	89 runs	Sydney	22 October 2022	T20I # 1839	Super 12
Afghanistan	England	England	5 wickets	Perth	22 October 2022	T20I # 1840	Super 12
Ireland	Sri Lanka	Sri Lanka	9 wickets	Hobart	23 October 2022	T20I # 1841	Super 12
India	Pakistan	India	4 wickets	Melbourne	23 October 2022	T20I # 1842	Super 12
Bangladesh	Netherlands	Bangladesh	9 runs	Hobart	24 October 2022	T20I # 1843	Super 12
South Africa	Zimbabwe	no result		Hobart	24 October 2022	T20I # 1844	Super 12
Australia	Sri Lanka	Australia	7 wickets	Perth	25 October 2022	T20I # 1845	Super 12
England	Ireland	Ireland	5 runs	Melbourne	26 October 2022	T20I # 1846	Super 12

In fact_bowling_summary, we want to create a new column 'balls'.

For that we split the column 'overs' into two 'over1' and 'over2' by using '.' as a delimiter.

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter
--Custom--
.

- Split at
 Left-most delimiter
 Right-most delimiter
 Each occurrence of the delimiter

Advanced options

Quote Character
"
 Split using special characters
Insert special character



Since the new column now has null values at some places, we replace it with 0:

Replace Values

Replace one value with another in the selected columns.

Value To Find

null

Replace With

0

OK

Cancel

Now we create our new column 'balls' based on over 1 and over 2:

Custom Column

Add a column that is computed from the other columns.

New column name

balls

Custom column formula

= [overs.1]*6 + [overs.2]

Available columns

match

team

bowlerName

overs.1

overs.2

maiden

runs

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.



Cancel

Structure		Mark as date table	Manage relationships	New measure column	New table	Data																
						match	team	bowlerName	overs.1	overs.2	maiden	runs	wickets	economy	zeros	fours	sixes	wides	noBalls	match_id	balls	Bou
a Vs Sri Lanka	Namibia	David Wiese	4	0	0	16	2	4	13	1	0	0	0	0	0	0	0	0	0	T20I # 1823	24	
a Vs Sri Lanka	Namibia	Bernard Scholtz	4	0	0	18	2	4.5	10	1	0	0	0	0	0	0	0	0	0	T20I # 1823	24	
s Netherlands	Netherlands	Tim Pringle	4	0	0	13	1	3.25	11	0	0	0	0	0	0	0	0	0	0	T20I # 1825	24	
s Netherlands	Netherlands	Paul van Meekeren	4	0	0	21	0	5.25	11	1	0	0	0	0	0	0	0	0	0	T20I # 1825	24	
s Netherlands	UAE	Karthik Meyappan	4	0	0	22	1	5.5	11	2	0	0	0	0	0	0	0	0	0	T20I # 1825	24	
s Netherlands	UAE	Zahoor Khan	4	0	0	11	1	2.75	14	0	0	0	0	0	0	0	0	0	0	T20I # 1825	24	
we Vs Ireland	Ireland	Josh Little	4	0	0	24	3	6	11	3	0	0	0	0	0	0	0	0	0	T20I # 1828	24	
we Vs Ireland	Zimbabwe	Richard Ngarava	4	0	0	22	2	5.5	13	3	0	0	0	0	0	0	0	0	0	T20I # 1828	24	
a Vs Netherlands	Namibia	JJ Smit	4	0	0	24	2	6	9	2	0	0	0	0	0	0	0	0	0	T20I # 1830	24	
a Vs U.A.E.	UAE	Zahoor Khan	4	0	0	26	2	6.5	10	3	0	0	0	0	0	0	0	0	0	T20I # 1832	24	
dies Vs Zimbabwe	Zimbabwe	Sikandar Raza	4	0	0	19	3	4.75	7	0	0	0	0	0	0	0	0	0	0	T20I # 1834	24	
a Vs Netherlands	Sri Lanka	Wanindu Hasaranga de Silva	4	0	0	28	3	7	11	3	0	0	0	0	0	0	0	0	0	T20I # 1835	24	
dies Vs Ireland	West Indies	Jason Holder	4	0	0	23	0	5.75	7	1	0	0	0	0	0	0	0	0	0	T20I # 1837	24	
d Vs Zimbabwe	Scotland	Brad Wheal	4	0	0	26	1	6.5	12	4	0	0	0	0	0	0	0	0	0	T20I # 1838	24	
d Vs Zimbabwe	Scotland	Mark Watt	4	0	0	19	1	4.75	9	1	0	0	0	0	0	0	0	0	0	T20I # 1838	24	
istan Vs England	Afghanistan	Mujeeb Ur Rahman	4	0	0	22	1	5.5	10	1	0	0	0	0	0	0	0	0	0	T20I # 1840	24	
istan Vs England	Afghanistan	Rashid Khan	4	0	0	17	1	4.25	10	0	0	0	0	0	0	0	0	0	0	T20I # 1840	24	
Vs Sri Lanka	Sri Lanka	Maheesh Theekshana	4	0	0	19	2	4.75	8	1	0	0	0	0	0	0	0	0	0	T20I # 1841	24	
i Vs India	India	Hardik Pandya	4	0	0	30	3	7.5	10	4	0	0	0	0	0	0	0	0	0	T20I # 1842	24	
i Vs India	Pakistan	Naseem Shah	4	0	0	23	1	5.75	10	1	0	0	0	0	0	0	0	0	0	T20I # 1842	24	
i Vs India	Pakistan	Shadab Khan	4	0	0	21	0	5.25	10	2	0	0	0	0	0	0	0	0	0	T20I # 1842	24	
Jesh Vs Netherlands	Netherlands	Paul van Meekeren	4	0	0	21	2	5.25	10	1	0	0	0	0	0	0	0	0	0	T20I # 1843	24	
Jesh Vs Netherlands	Bangladesh	Taskin Ahmed	4	0	0	25	4	6.25	16	5	0	0	0	0	0	0	0	0	0	T20I # 1843	24	
Vs England	England	Adil Rashid	4	0	0	24	0	6	6	1	0	0	0	0	0	0	0	0	0	T20I # 1846	24	
ifrica Vs Bangladesh	South Africa	Keshav Maharaj	4	0	0	24	1	6	6	1	0	0	0	0	0	0	0	0	0	T20I # 1847	24	
Netherlands	Netherlands	Tim Pringle	4	0	0	30	0	7.5	6	3	0	0	0	0	0	0	0	0	0	T20I # 1848	24	
Mathurandu	India	Mohammed Shami	4	0	0	27	1	6.75	10	2	0	0	0	0	0	0	0	0	0	T20I # 1849	24	

Sakshi Zagade, Shruti Sakpal

In fact_batting_summary, transforming column 'out/not out' by renaming it and inserting only 0 and 1 meaning not out and out respectively.

The screenshot shows the Power BI Data Editor interface. At the top, there's a 'Replace Values' dialog box with fields for 'Value To Find' (containing 'out') and 'Replace With' (containing '0'). Below the dialog is a table titled 'fact_batting_summary'. The table has columns: match, team, battingPos, batsmanName, runs, balls, fours, sixes, SR, out, match_id, boundary runs. The 'out' column contains binary values (0 or 1). On the right side of the editor, there's a 'Data' pane showing a tree view of tables: Key_Measures, dim_match_summary, dim_players, fact_batting_summary (which is selected), and fact_bowling_summary.

match	team	battingPos	batsmanName	runs	balls	fours	sixes	SR	out	match_id	boundary runs
Namibia Vs Sri Lanka	Namibia	7	David Wiese	0	1	0	0	0.00	1	T20I # 1823	0
Namibia Vs Sri Lanka	Sri Lanka	4	Darushka Gunathilaka	0	1	0	0	0.00	1	T20I # 1823	0
U.A.E. Vs Netherlands	U.A.E.	10	Junaid Siddique	0	1	0	0	0.00	0	T20I # 1825	0
U.A.E. Vs Netherlands	Netherlands	7	Roelof van der Merwe	0	2	0	0	0.00	1	T20I # 1825	0
Scotland Vs West Indies	West Indies	9	Altzam Joseph	0	1	0	0	0.00	1	T20I # 1826	0
Zimbabwe Vs Ireland	Zimbabwe	7	Regis Chakabva	0	2	0	0	0.00	1	T20I # 1828	0
Zimbabwe Vs Ireland	Ireland	1	Paul Stirling	0	2	0	0	0.00	1	T20I # 1828	0
Zimbabwe Vs Ireland	Ireland	9	Simi Singh	0	1	0	0	0.00	1	T20I # 1828	0
Namibia Vs Netherlands	Namibia	2	Divan la Cock	0	2	0	0	0.00	1	T20I # 1830	0
Namibia Vs Netherlands	Namibia	4	Jan Nicot Loftie-Eaton	0	2	0	0	0.00	1	T20I # 1830	0
Namibia Vs Netherlands	Netherlands	5	Colin Ackermann	0	2	0	0	0.00	1	T20I # 1830	0
Sri Lanka Vs U.A.E.	Sri Lanka	5	Charith Asalanka	0	1	0	0	0.00	1	T20I # 1832	0
Sri Lanka Vs U.A.E.	Sri Lanka	6	Dasun Shanaka	0	1	0	0	0.00	1	T20I # 1832	0
Sri Lanka Vs U.A.E.	U.A.E.	8	Kashif Daud	0	4	0	0	0.00	1	T20I # 1832	0
Scotland Vs Ireland	Scotland	6	Calum MacLeod	0	3	0	0	0.00	1	T20I # 1833	0
West Indies Vs Zimbabwe	West Indies	6	Shamarth Brooks	0	3	0	0	0.00	1	T20I # 1834	0
Sri Lanka Vs Netherlands	Sri Lanka	3	Dhananjaya de Silva	0	1	0	0	0.00	1	T20I # 1835	0
Sri Lanka Vs Netherlands	Netherlands	4	Colin Ackermann	0	1	0	0	0.00	1	T20I # 1835	0
Sri Lanka Vs Netherlands	Netherlands	8	Timm van der Gugten	0	1	0	0	0.00	1	T20I # 1835	0
Sri Lanka Vs Netherlands	Netherlands	11	Roelof van der Merwe	0	1	0	0	0.00	0	T20I # 1835	0
Scotland Vs Zimbabwe	Zimbabwe	3	Wessly Madhevere	0	5	0	0	0.00	1	T20I # 1838	0
New Zealand Vs Australia	Australia	10	Adam Zampa	0	2	0	0	0.00	1	T20I # 1839	0
Afghanistan Vs England	Afghanistan	8	Rashid Khan	0	1	0	0	0.00	1	T20I # 1840	0
Afghanistan Vs England	Afghanistan	9	Mujeeb Ur Rahman	0	1	0	0	0.00	1	T20I # 1840	0
Afghanistan Vs England	Afghanistan	11	Fazalhaq Farooqi	0	2	0	0	0.00	1	T20I # 1840	0
Ireland Vs Sri Lanka	Ireland	8	Mark Adair	0	1	0	0	0.00	1	T20I # 1841	0
Pakistan Vs India	Pakistan	2	Babar Azam	0	1	0	0	0.00	1	T20I # 1842	0
Pakistan Vs India	Pakistan	0	Tahir Ahmad	0	1	0	0	0.00	1	T20I # 1843	0

In this step, we create some DAX measures which we will be using for building the actual visuals.

The screenshot shows the Power BI Data Visualizations pane. It displays a list of DAX measures categorized into groups:

- Bowling** (selected):
 - balls Bowled
 - Bowling Average
 - Bowling Strike Rate
 - Dot ball %
 - Economy
 - Runs Conceded
 - Total Innings Bowled
 - wickets
- Others**:
 - Color Callout Value
 - Display Text
 - Player Selection

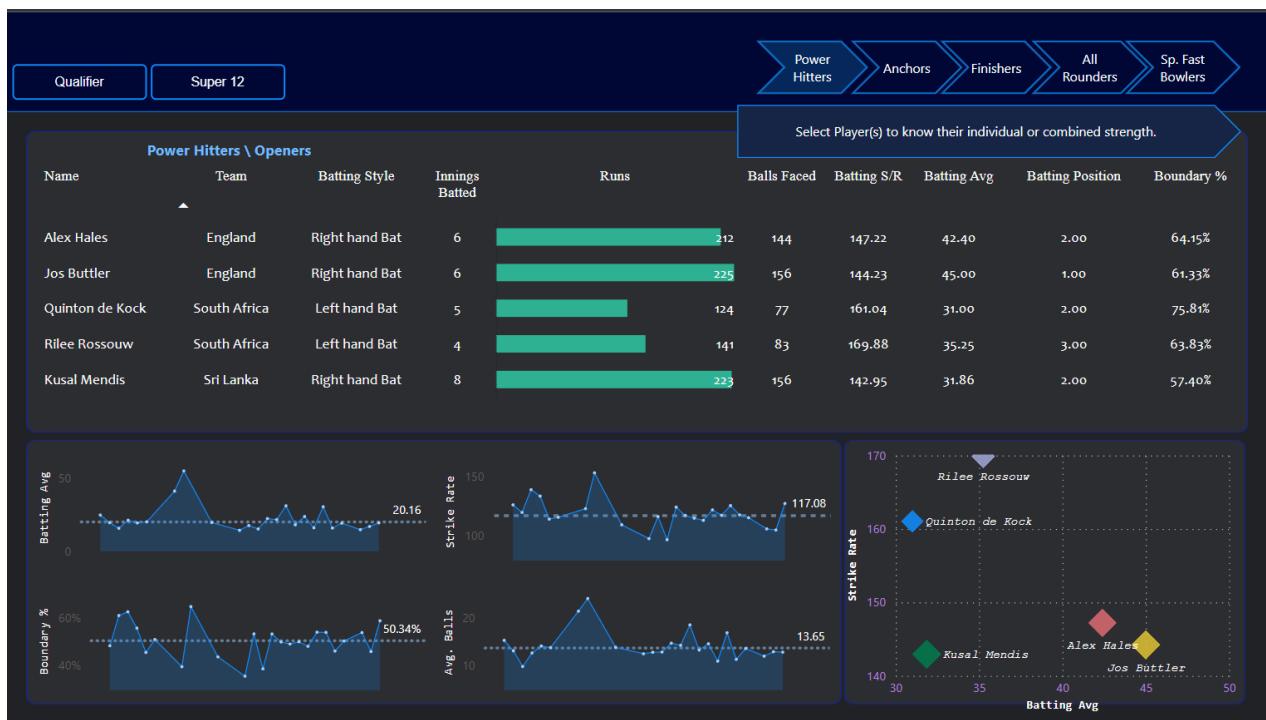
Creating three calculated columns in three different tables:

dim_players	fact_batting_summary	fact_bowling_summary
<input type="checkbox"/> battingStyle	<input type="checkbox"/> \sum balls	<input type="checkbox"/> \sum balls
<input type="checkbox"/> bowlingStyle	<input type="checkbox"/> batsmanName	<input type="checkbox"/> Boundary runs bowling
<input checked="" type="checkbox"/> Custom Batting Order	<input type="checkbox"/> \sum battingPos	<input type="checkbox"/> bowlerName
<input type="checkbox"/> description	<input type="checkbox"/> boundary runs	<input type="checkbox"/> \sum economy
<input type="checkbox"/> image	<input type="checkbox"/> \sum fours	<input type="checkbox"/> \sum fours
<input type="checkbox"/> name	<input type="checkbox"/> match	<input type="checkbox"/> \sum maiden
<input type="checkbox"/> playingRole	<input type="checkbox"/> match_id	<input type="checkbox"/> match
<input type="checkbox"/> team	<input type="checkbox"/> \sum out	<input type="checkbox"/> match_id
	<input type="checkbox"/> \sum runs	<input type="checkbox"/> \sum noBalls
	<input type="checkbox"/> \sum sixes	<input type="checkbox"/> \sum overs.1
	<input type="checkbox"/> SR	<input type="checkbox"/> \sum overs.2
	<input type="checkbox"/> team	<input type="checkbox"/> \sum runs
		<input type="checkbox"/> \sum sixes

Data Visualization

The final step of this project is building an interactive dashboard on the basis of all these prior preprocessing.

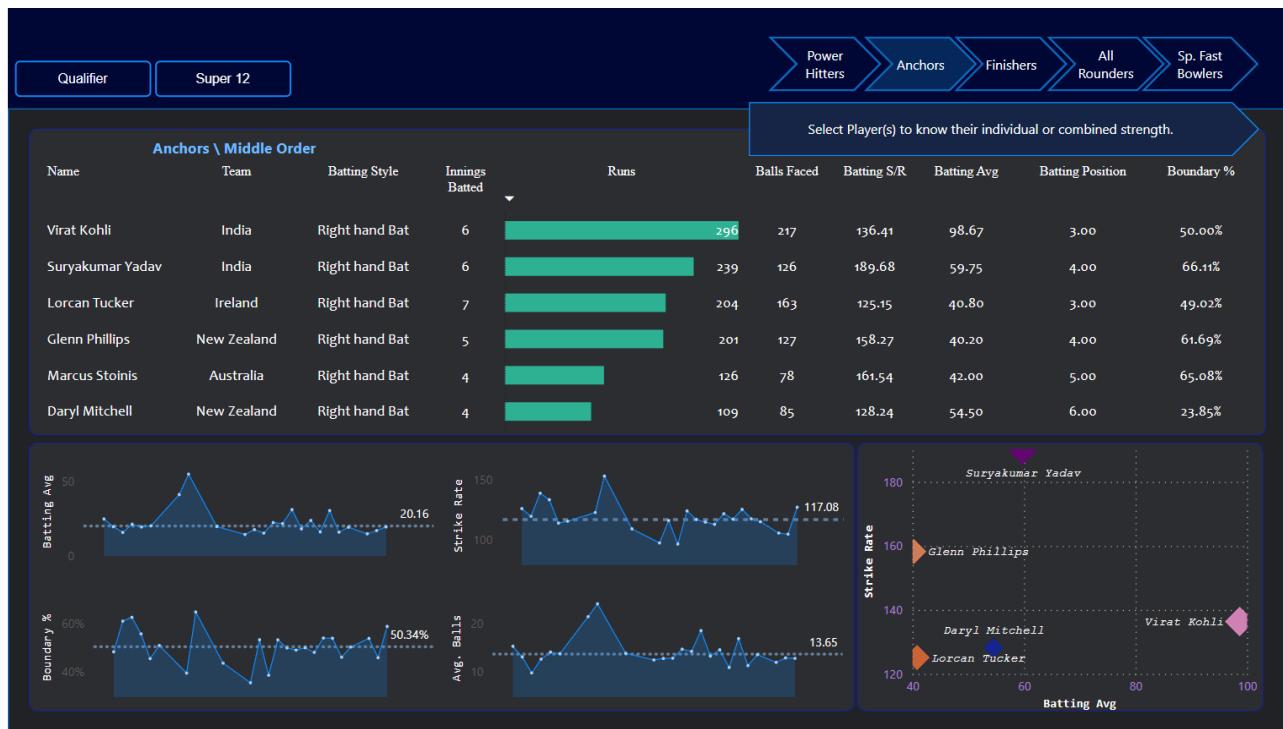
Our final dashboard looks like this:



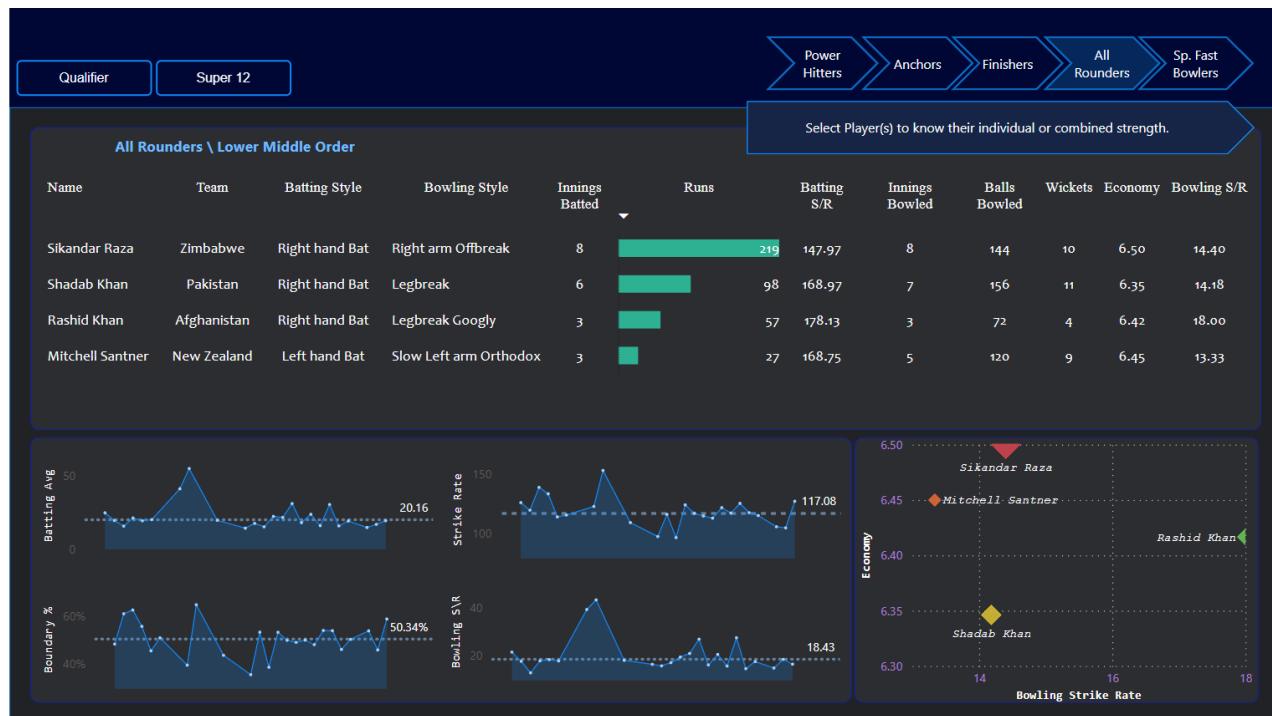
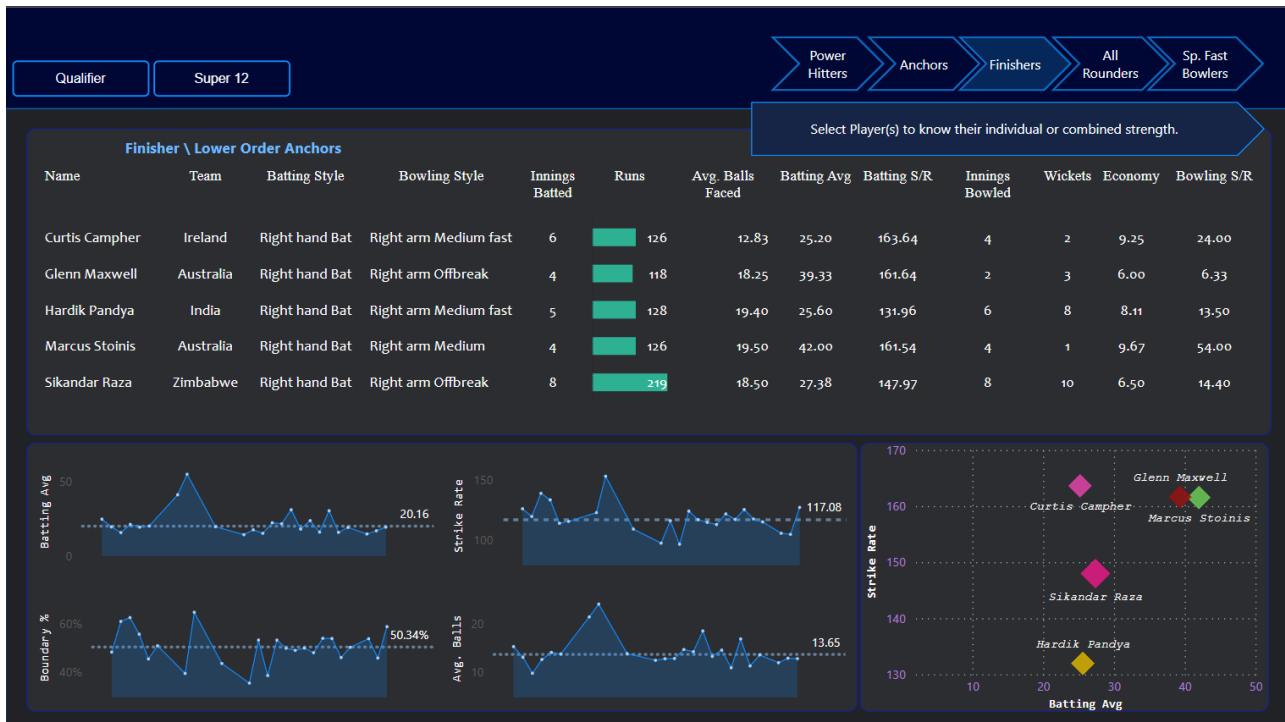
There are 5 different main tabs (categories in the problem statement) namely:

- Power hitters,
- Anchors,
- Finishers,
- All rounders,
- Special Fast Bowlers.

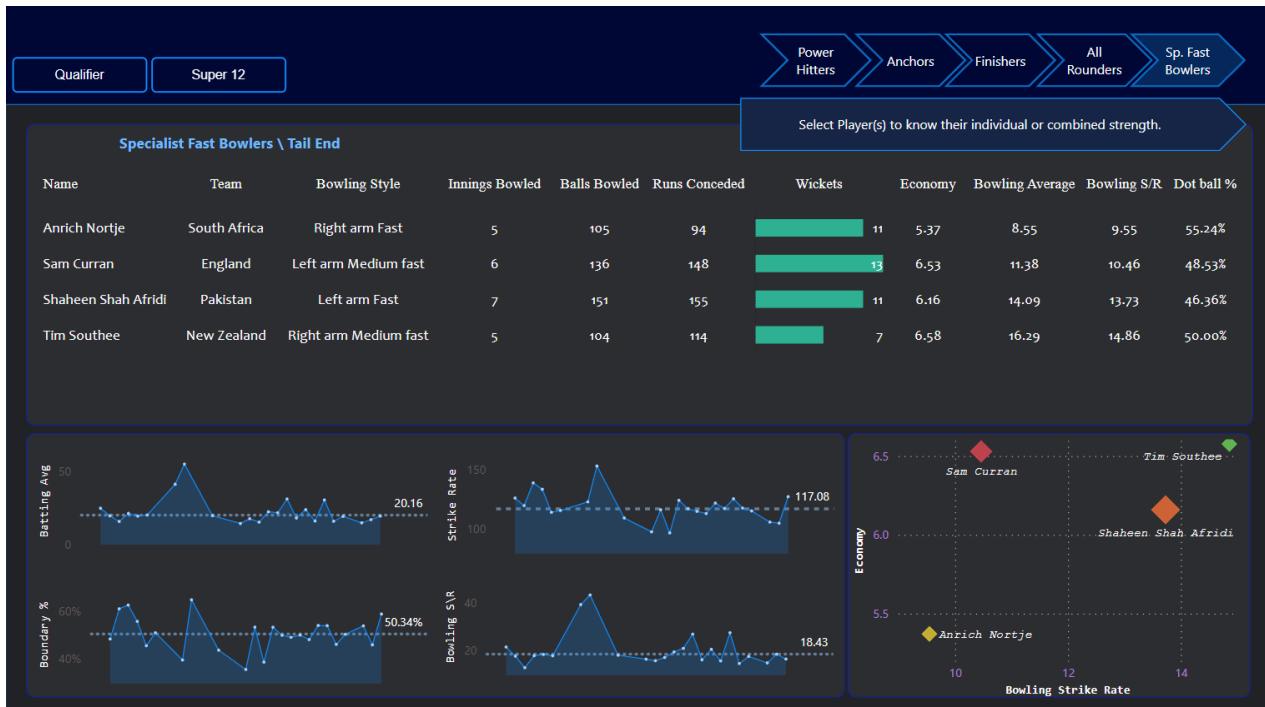
Whenever any particular tab is selected, the players in that category along with other statistics such as runs, strike rate, batting average etc. will be displayed.



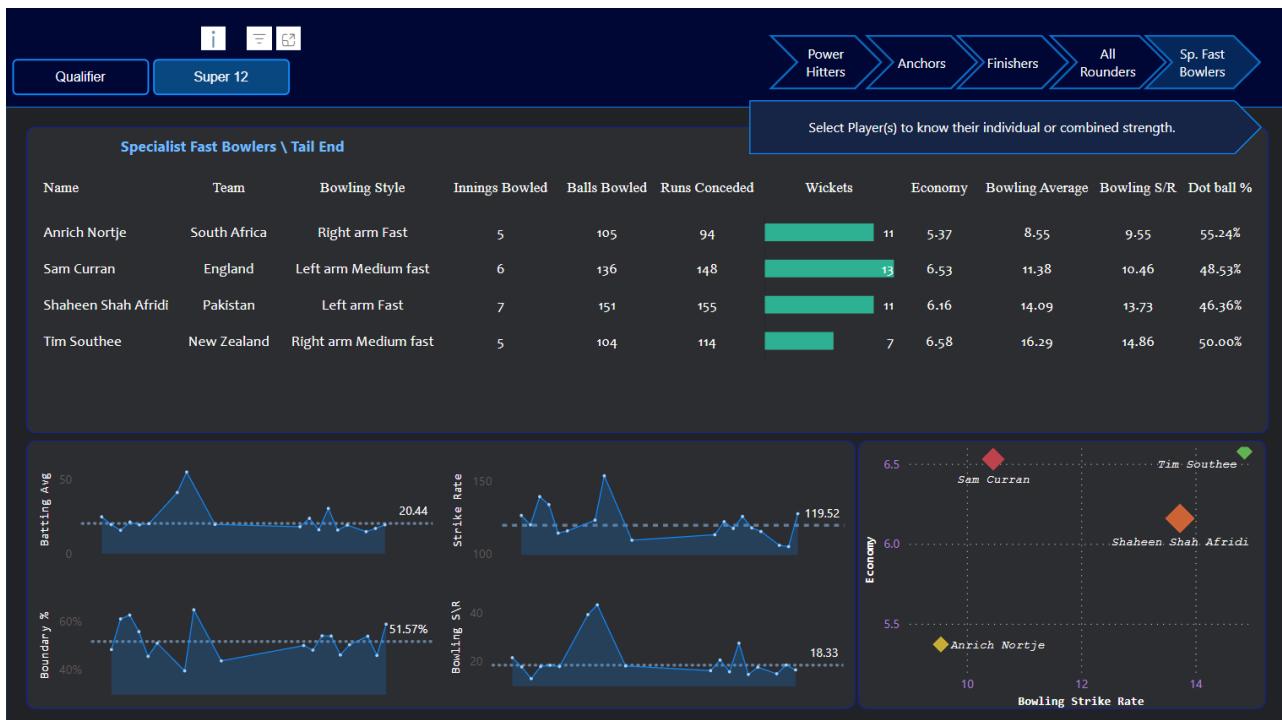
Sakshi Zagade, Shruti Sakpal



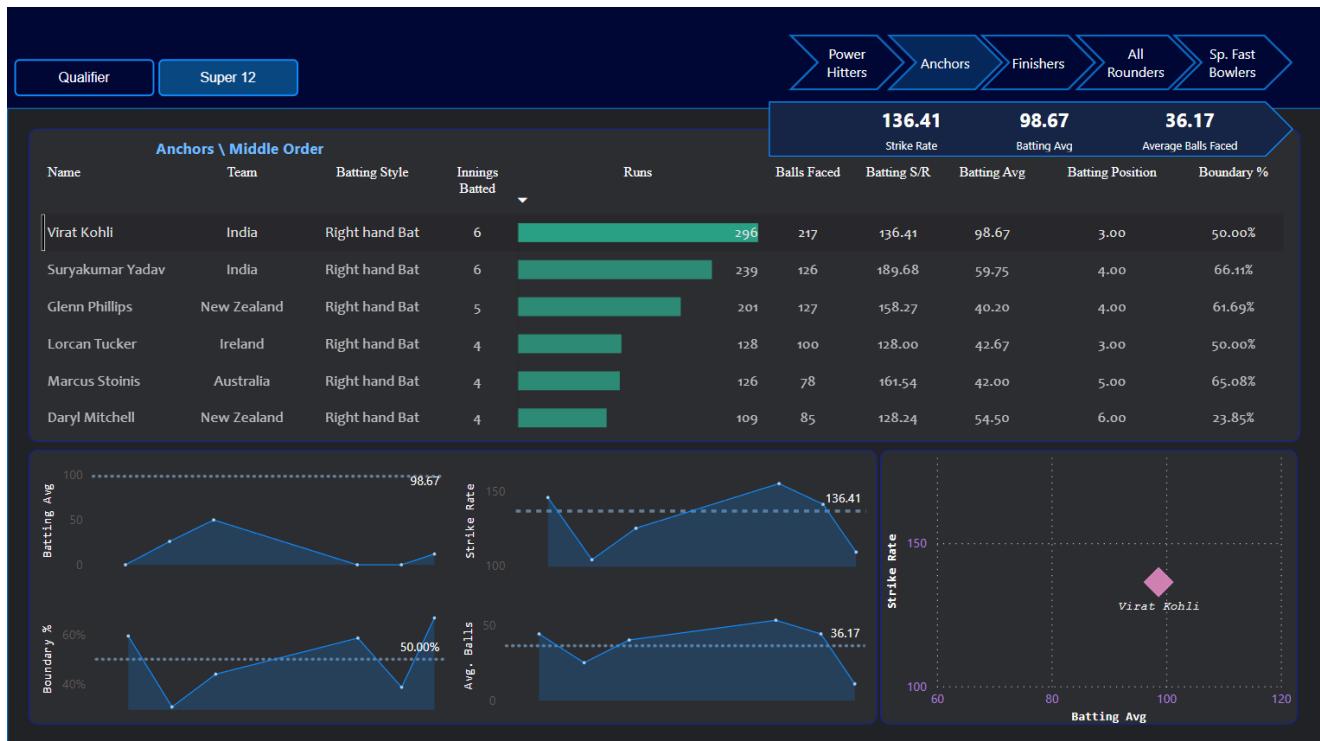
Sakshi Zagade, Shruti Sakpal



There are also two sub tabs, Qualifier and Super 12 which show different information when clicked.

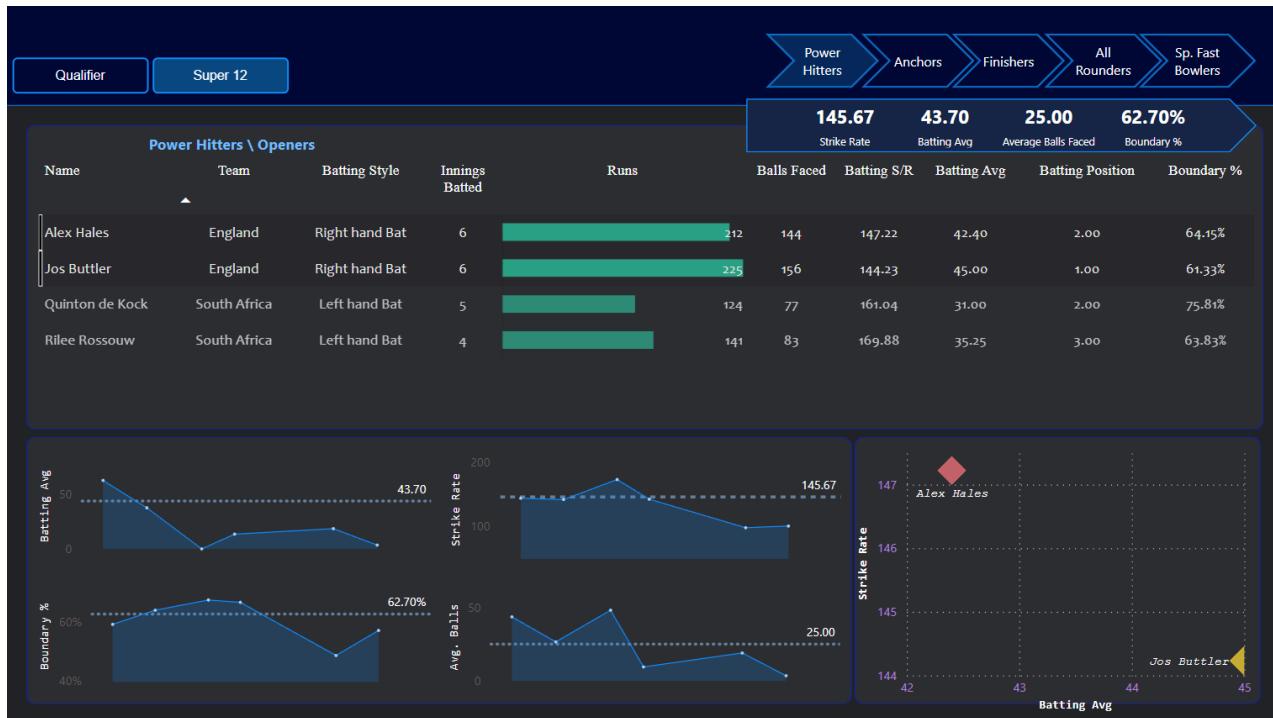


By clicking on any particular player, we get statistics and graphs associated with that player.

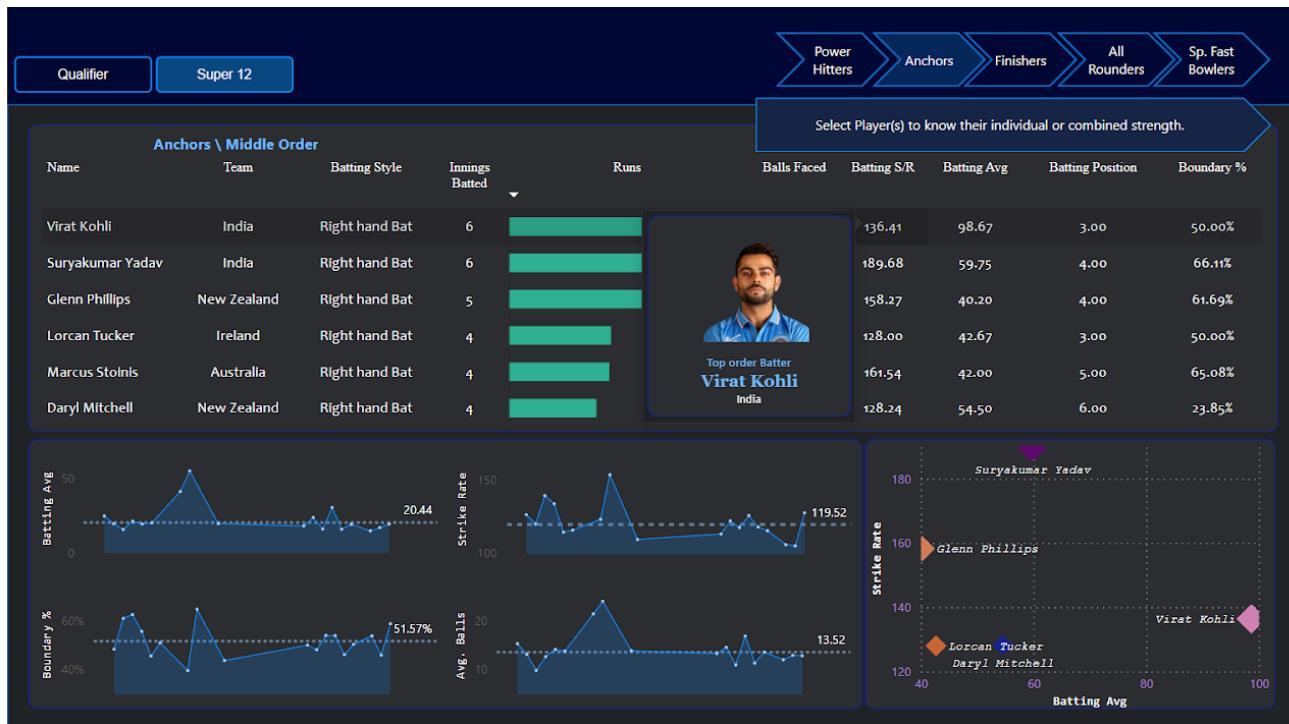


We can also get combined statistics of any players by Clicking Shift and then selecting:

Sakshi Zagade, Shruti Sakpal

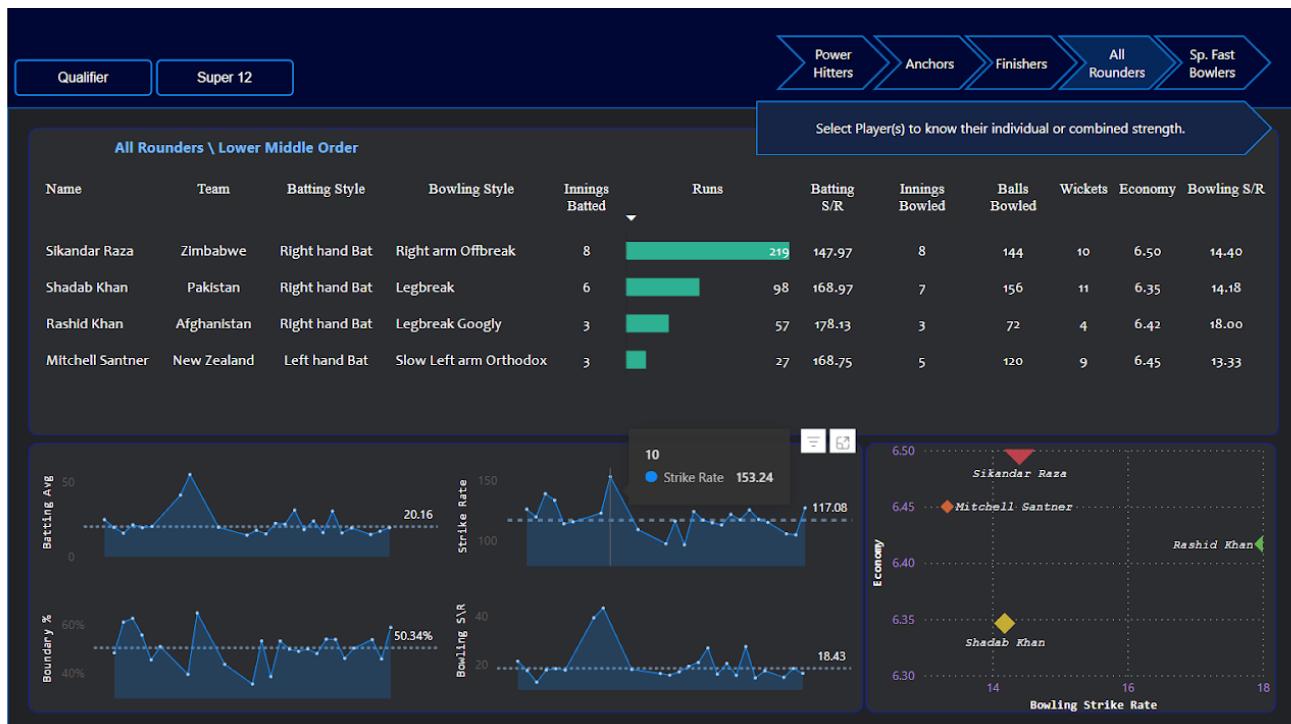


By hovering on any of the parameter of any player, it displays a picture of that particular player.



Sakshi Zagade, Shruti Sakpal

At the bottom we have some trends based on various statistics represented in a line chart . By hovering on the line charts we get the actual value of that specific parameter at that point.



In the bottom right corner we have a scatter chart between Strike Rate and Batting Average. By hovering on any of the point on the scatter chart we get the information of that player with respect to certain parameters.

Sakshi Zagade, Shruti Sakpal



Dashboard Link:

<https://app.powerbi.com/view?r=eyJrljoiNGM5NDg4NWItNzhiOC00M2U0LWI2NjQtNzMwODdkYmYWQ2liwidCI6ImNjNTcwZWQwLTlyN2ItNGU2YS1iODQ5LTMwNzYzM2UxMTNkYSJ9>