

# Data Visualization - DS304.3

# **Final Report**

# 2022 T20 World Cup Cricket Dashboard

Group - B

# **Group Members**

Student Name	Student Number		
M K I M Rohana	24572		
M R K Karunathilaka	24490		

Module Lecturer: Ms. Lakni Peiris

# **Table of Contents**

1. Introduction	3
1.1 Background Study of the Selected Sector	3
1.2 Problems Related to the Selected Area	4
1.3 Objectives of the Project	4
1.4 Expected Limitations on the Visualization Process of the Selected Area	6
2. Literature Review	7
2.1 Introduction to the dataset selected.	7
2.2 Variables and Definitions	7
3. Data Preprocessing and Data Wrangling	9
3.1 Data Blending & Integration	
3.2 Data Cleaning	9
3.3 Data Transformation	10
4. Methodology	11
4.1 Introduction	11
4.2 Type of Data to be collected and data sources.	
4.3 Data collection tools and plan	
4.4 Methods, techniques and tools used to visualize the data	
5. Data Analysis	14
Descriptive Statistics	14
Correlation Analysis	30
6. Dashboard	
7. Conclusion	
8. References	
9. Attachments	56

### 1. Introduction

# 1.1 Background Study of the Selected Sector

Our project centres around cricket, a globally popular sport. We specifically focus on analysing the T20 cricket format, emphasizing the 2022 T20 World Cup. This choice stems from a comprehensive study of cricket's landscape and the distinctive features of T20 tournaments.

# Rationale for Selecting T20 Cricket Analysis:

### • Influence on Cricketing Strategies:

T20 cricket significantly shapes strategies in other formats, setting trends and highlighting its broader strategic importance.

## • Global Appeal:

T20 cricket, especially in the T20 World Cup, is globally renowned, attracting diverse audiences with top teams and players.

# • Strategic Insights:

The T20 format demands strategic brilliance, providing a unique platform to assess individual performances and team strategies, making it ideal for comprehensive data analysis.

## 1.2 Problems Related to the Selected Area

While the T20 cricket format and the 2022 T20 World Cup provide a rich landscape for analysis, several challenges and problems need consideration:

## • Privacy and Ethical Considerations:

Handling player-related data requires strict adherence to privacy and ethical standards. Ensuring that our analysis respects the confidentiality of sensitive information while providing meaningful insights poses a challenge.

## • Complex Team and Player Dynamics:

In T20 cricket, intricate team strategies and individual player dynamics pose a challenge in concise analysis and representation. Balancing depth with user-friendly visualizations is crucial.

• Finding Data set resources is limited.

### 1.3 Objectives of the Project

#### • Dashboard Creation:

It develops an interactive and user-friendly dashboard that comprehensively overviews the 2022 T20 World Cup.

Thereby, implementing intuitive visualizations to highlight key performance metrics, such as runs, wickets, and strike rates, allowing users to quickly grasp the tournament's unfolding narrative.

## • Match Insights:

This provides in-depth insights into each match, including team compositions, player statistics, and match commentary.

It incorporates dynamic features allowing users to explore detailed statistics of individual players and compare their performances across matches.

### • Team Performance Analysis:

Analyse and visualize team performance trends throughout the tournament, showcasing run rates, wicket-taking abilities, and overall team dynamics.

Implement features for users to compare team statistics, enabling a deeper understanding of strategic strengths and weaknesses.

## • Player Profiles:

It creates individual player profiles featuring comprehensive statistics, highlighting standout performances and contributions.

Thereby incorporating filters and sorting options for users to explore player data based on various criteria such as runs scored, strike rates, and wickets taken.

### • User Interaction and Customization:

It provides interactive elements, such as dropdowns, and checkboxes, enabling users to customize their viewing experience.

Ensure responsiveness across different devices, making the dashboard accessible to a broad audience.

# 1.4 Expected Limitations on the Visualization Process of the Selected Area

# • Privacy Constraints:

Respecting players' privacy constraints may limit the depth of personal insights. Visualizations may need to prioritize aggregated team-level data to navigate privacy concerns

## • Device Compatibility:

Ensuring that visualizations are accessible across various devices and screen sizes may pose challenges. Optimizing for both desktop and mobile platforms requires careful consideration.

## • Resource Intensity:

Resource-intensive visualizations, particularly those involving complex animations or intricate graphics, may pose limitations on the responsiveness of the dashboard. Optimization is key to maintaining user experience.

# 2. Literature Review

### 2.1 Introduction to the dataset selected.

The dataset obtained from Kaggle is a comprehensive collection of cricket statistics, capturing individual player performances during the ICC Men's T20 World Cup 2022, hosted in Australia. Comprising two distinct files, "batting\_card.csv" and "bowling\_card.csv," the dataset offers a granular insight into the dynamic world of T20 cricket.

### 2.2 Variables and Definitions

Variables in both batting and bowling data sets.

Variables	Definition
match_id	Match Id number (Primary key)
match_name	Home team name vs Away team name
home_team	Home Team name
away_team	Away team name
bowling_team	Bowling Team name
venue	Ground (stadium) Name
city	Match city of Australia
country	The country where the competition was held. (Australia)
current_innings	Current innings team name
innings_id	Inning ID 1 or 2 (These matches have two innings)
fullName	Player's full name
minutes	How many minutes did the player play in the match?
runs	Player's runs
ballsFaced	How many balls did the player face?
fours	Number of fours taken by the player.
sixes	Number of six taken by the player.
strikeRate	Indicating how quickly a player scores runs during their innings.
captain	Whether the player is captain or not.
isNotOut	Whether the player is out or not.
runningScore	Score of the inning when the player comes.

runningOver	The number of overs when the player comes.
shortText	Out details of the players.
overs	Number of overs bowled by the bowler
maidens	Number of maiden overs bowled by the bowler.
conceded	Scores of the taken from bowler.
wickets	Number of wickets taken by bowler.
economyRate	Average scores of the taken for an over from bowler.
dots	Number of dot balls bowled by the bowler.
foursConceded	Numbers of the fours taken from the bowler.
sixesConceded	Numbers of the sixes taken from the bowler.
wides	Numbers of the Wides given by bowler.
noballs	Numbers of the no balls taken from bowler.

# 3. Data Preprocessing and Data Wrangling

## 3.1 Data Blending & Integration

Under the data blending and integration process we used two data sets,

- First data set: Batters data set (batting\_card.csv): batting\_card.csv
- Second data set: Bowler's data set (bowling\_card.csv): bowling\_card.csv

Also, we created new tables using inner join with above mention tables to understand correlation ships.

## 3.2 Data Cleaning

- Under the data cleansing process firstly check whether the data set has a Noice variable (unwanted columns).
- Secondly, check whether the data set has duplicated values.
- Thirdly, check whether the data set has missing values.
- After checking outliers visualizing data.
- By taking the above steps we tried to enhance the quality and accuracy of the data set.
- ➤ Removed "season" Columns for both Batting and Bowling data sets.

#### Reason:

- Because in season column only contains 2022 which is the year of the T20 World Cup.
- ➤ Removed "name" Columns for both Batting and Bowling data sets.

### Reason:

• Because the name column contains the short Name of the player. Also, both data sets contained the full names of the payers.

> Removed "href" Columns in Batting data sets.

#### Reason:

- In "href" column contains HTML content; due to the primary data set extracted from the web scraping process.
- > Removed "commentary" and "link" Columns in the Batting data set.

#### Reason:

- In "commentary" and "link" Columns contained HTML format data and those data contained huge descriptions.
- ➤ By the way, also we have to remove other columns such as {"running Score", "short Text", and "running Over"} from the batting data set.

#### 3.3 Data Transformation

Converted the "strike Rate" variable (column) into Integer format (Numerical) in the batting data set.

#### Reason:

- Because "strike Rate" is not a String (Categorical) variable and represents the Indicating how quickly a player scores runs during their innings.
- ➤ Clean Batting dataset: <u>Clean\_batting\_dataset.xlsx</u>
- ➤ Clean Bowling dataset: <u>Clean\_bowling\_dataset.xlsx</u>

# 4. Methodology

### 4.1 Introduction

- > Data collection
  - In cricket, the T20 World Cup data set includes both Quantitative variables and Qualitative variables.

# > Analytical strategies

- Univariate analysis understands signal variables.
- Bivariate analysis understands the relationship between two variables.
- Statistical analysis is used to understand data distribution, especially numerical (Quantitative) variables.

## 4.2 Type of Data to be collected and data sources.

- The data set was gathered from a web open Data Source called Kaggle: https://www.kaggle.com/datasets
- The data collection strategy is gathering data from secondary sources which have already been created by some other party.
- Also, the primary dataset was created from a web mining technique called data scraping.

## 4.3 Data collection tools and plan

• Also, the primary dataset was created from a web mining technique called data scraping. We used the data obtained in this way as a secondary source.

### 4.4 Methods, techniques and tools used to visualize the data.

Our proposal outlines a comprehensive approach to visualize data from the 2022 T20 World Cup dataset, employing a range of methods, techniques, and tools to ensure meaningful representation. The visualization process will be facilitated using the following:

### ➤ Power BI:

Tool:

• Leveraged Power BI as the primary visualization tool due to its robust features, ease of use, and seamless integration with various data sources.

## ➤ Charts and Graphs:

## Method/Technique:

 Implemented various charts and graphs, including bar charts, line charts, pie charts, and scatter plots, to represent diverse aspects of match statistics, player performances, and team dynamics.

# > Interactive Dashboards:

# Method/Technique:

Designed interactive dashboards in Power BI to allow users to dynamically
explore the dataset, apply filters, and drill down into specific details based
on their interests.

# Responsive Design:

# Method/Technique:

• Ensured that visualizations were designed with a responsive layout, optimizing for both desktop and mobile platforms to enhance accessibility across various devices.

# 5. Data Analysis

### **Descriptive Statistics**

Descriptive statistics is a branch of statistics which is a process of data collecting, organizing, presenting, analysing, and interpreting data.

## 1. Central tendency

Mean, Median, Mode

# 2. Dispersion

variance, standard deviation, IQR

## 3. Graphical charts

Bar chart, scatter plot, line chart, pie chart

## Summary statistic for the Numerical variable of the Batting data set

	runs	ballsFaced	minutes	fours	sixes	strikeRate
count	699.000000	699.000000	699.000000	699.000000	699.000000	699.000000
mean	15.978541	13.648069	22.808298	1.300429	0.473534	95.882232
std	18.709462	12.716854	21.054790	1.834448	0.978465	61.428293
min	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000
25%	2.000000	4.000000	8.000000	0.000000	0.000000	50.000000
50%	9.000000	10.000000	16.000000	1.000000	0.000000	100.000000
75%	22.500000	19.000000	31.000000	2.000000	1.000000	133.330000
max	109.000000	64.000000	106.000000	10.000000	8.000000	400.000000

According to the above chart 2022 T20 World Cup

- ➤ Batsmen's average runs is approximately 16. That means we can **predict average one-inning score goes up to 160 (16 \*10)**
- ➤ Also, Batsmen's highest number of runs is 109 which means any player couldn't take one hundred and half centuries.
- > All 50 % of the batsmen in that tournament inning run is less than or equal to 9.

- ➤ Batsmen's average no of faced balls is equal to approximately 14 (overs 2.2).
- In that tournament batsmen faced a maximum no of balls is 64 (overs 10.4)
- Also, in that tournament average one player at least has taken one four runs.
- In an inning batsman's highest taken no of sixes is 8.
- ➤ In the tournament, 75% of batsmen's strike rate is less than or equal to (133.33). Hence, we can say playgrounds are suitable for batters.
- ➤ Also, the Average batsmen strike rate is approximately equal to 96.0.

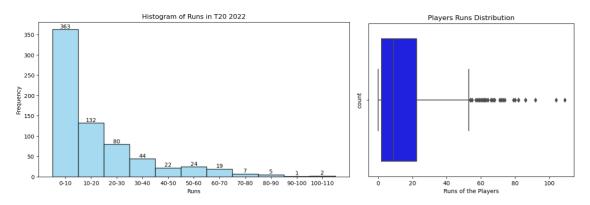
# Summary statistic for the Categorical variable of the Batting data set

	match_name	home_team	away_team	venue	city	current_innings	name	fullName
count	699	699	699	699	699	699	699	699
unique	42	16	15	7	7	16	198	198
top	PAK v SA	INDIA	NED	Bellerive Oval, Hobart	Hobart	ZIM	BFW de Leede	Bas de Leede
freq	21	76	107	136	136	74	8	8

## According to the above chart 2022 T20 World Cup

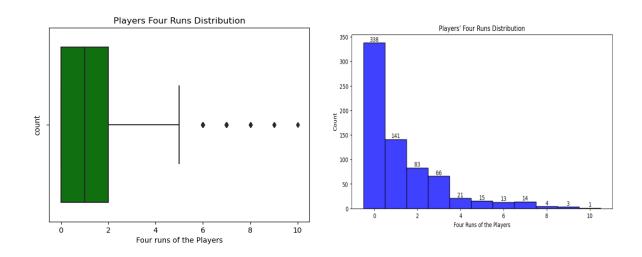
- > Seven different playgrounds have been used.
- > The home team has played in 16 countries.
- While as an away team has been played in 15 countries.

### • Batsmen's runs in the tournament.



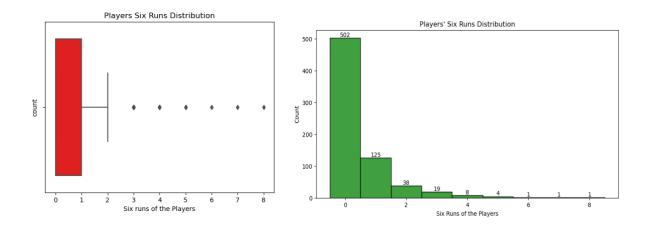
According to the above chart 2022 T20 World Cup

- ➤ Only got two centuries throughout the tournament.
- Also have been taking 56 half-centuries throughout the tournament.
- ➤ Most of the players were not able to play long innings in that tournament.
- ➤ Overall, most of the batsmen couldn't come good runs in the tournament.
- Number of fours hit by players in the 2022 T20 tournament.

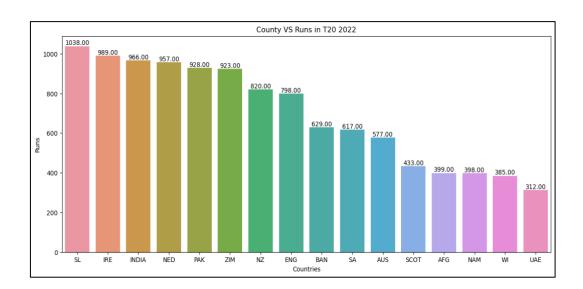


- ➤ Throughout the tournament, only one player takes 10 four runs, and it is the maximum no of fours taken in an inning by one player.
- Also, in the tournament, at least one-four has been taken by players on 361 instances.
- In the tournament, at least eight four runs have been taken by players in 8 instances.

• Number of sixes hit by players in the 2022 T20 tournament.

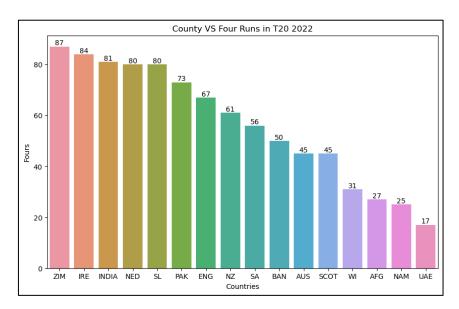


- Throughout the tournament, only one player takes 8 sixes, and it is the maximum no of sixes taken in an inning by one player.
- ➤ In the tournament at least one has been taken one six by players on 197 instances.
- Also, in the tournament, at least 5 sixes have been taken by players in 7 instances.
- Total runs scored by countries in T20 2022.



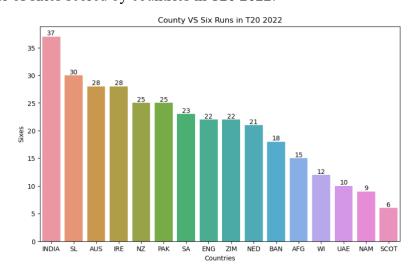
- According to the above chart, Sri Lanka is the maximum total run's country.
- > UAE has the lowest total runs.
- ➤ India's total runs are higher than Pakistan's total runs in the tournament.
- Also, Australia's total runs are less than New Zealand's total runs.

• Number of fours scored by countries in T20 2022.



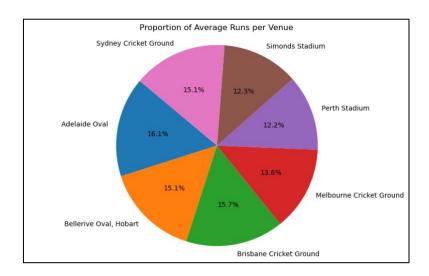
- According to the above chart, the highest fours taken country is Zimbabwe which is 87.
- ➤ While the lowest fours taken country is UAE which is 17.
- > India taken four runs is higher than Sri Lanka.
- ➤ Also, England taken fours is higher than South Africa.

• Number of sixes scored by countries in T20 2022.



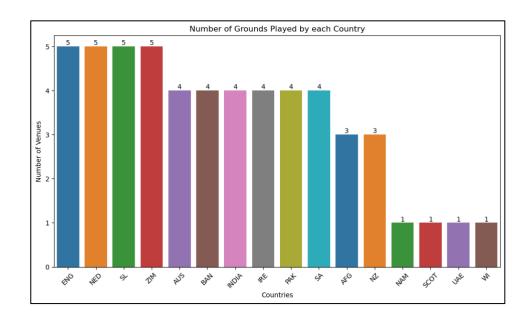
- According to the above chart, India is the Highest six runs taken country which is 37.
- ➤ While Scotland is the lowest six taken country and which is 6.
- Sri Lanka no of six taken is higher than Australia no of six taken in the tournament.
- Also, Ireland no of six taken is higher than no of six taken by Bangladesh.

• Proportion of the Average runs in the playgrounds 2022 T20 tournament.



- According to the above Pie chart highest average runs taken on the playground is Adelaide Oval which is 16.1%.
- ➤ The lowest average runs taken playground is Perth Stadium which is 12.2%.
- ➤ Brisbane Cricket Ground average taken runs is higher than Melbourne Cricket Ground.
- Also, the Sydney Cricket Ground average taken runs is approximately equal to the Bellerive Oval, Hobart average taken runs.

• The number of the Grounds Played by each country in the 2022 T20 tournament.



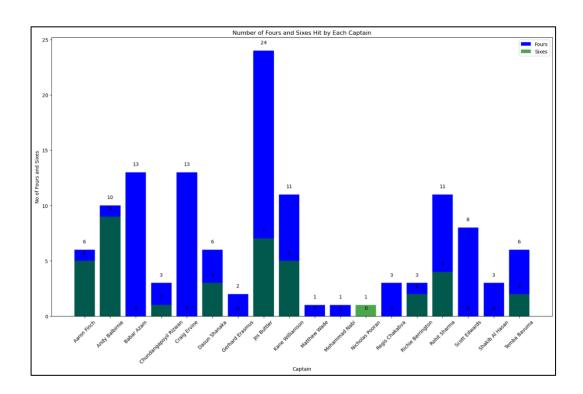
- According to the above graph, four countries have played five different playgrounds, and those countries are (England, Netherlands, Sri Lanka, and Zimbabwe).
- Four countries have played only one playground throughout the tournament and those countries are (Namibia, Scotland, UAE, and Western Indies).
- Six countries have played four different playgrounds throughout the tournament and those countries are (Australia, Bangladesh, India, Ireland, Pakistan, and South Africa).
- Afghanistan and New Zealand have been played on only three different grounds in the tournament.

• All the captains' summary descriptive statistical data

	match_id	innings_id	runs	ballsFaced	minutes	fours	sixes	strikeRate
count	8.300000e+01	83.000000	83.000000	83.000000	83.000000	83.000000	83.000000	83.000000
mean	1.298157e+06	1.493976	18.337349	16.783133	28.530120	1.493976	0.469880	89.504819
std	1.356194e+01	0.503003	18.783281	13.311510	21.643642	1.934119	0.860195	46.080327
min	1.298135e+06	1.000000	0.000000	0.000000	2.000000	0.000000	0.000000	0.000000
25%	1.298145e+06	1.000000	4.000000	7.000000	11.500000	0.000000	0.000000	60.000000
50%	1.298157e+06	1.000000	13.000000	13.000000	22.000000	1.000000	0.000000	88.880000
75%	1.298169e+06	2.000000	26.500000	23.000000	37.000000	2.000000	1.000000	118.180000
max	1.298179e+06	2.000000	80.000000	54.000000	89.000000	9.000000	3.000000	189.470000

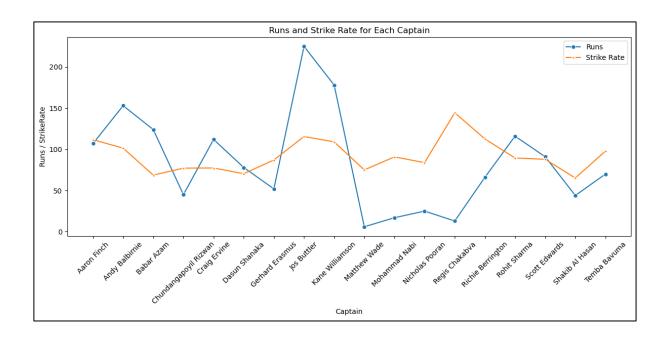
- > Captain's average taken runs is approximately equal to 18.
- Also, A captain taken highest run is 80.
- A captain-faced average balls are approximately equal to 17 (Overs 2.5).
- A captain's average no of fours is equal to 1.
- Also, A captain's average strike rate is 89.50.
- > 25% of the captains' strike rate is lower than or equal to 60.

• All the captains' no of Fours and Sixes in the 2022 T20 Tournament



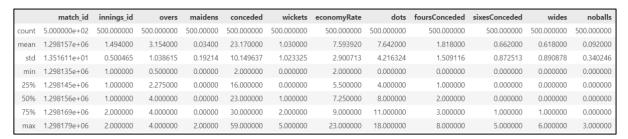
- > Jos Buttler is the captain who has taken the highest no of fours which is 24.
- Also, Andy Balbirnie is the captain who has taken the highest no of sixes which is 9.
- > Sri Lankan captain Dasun Shanaka has taken 6 fours and 3 sixes throughout the tournament.
- Also, Babar Azam and Craig Ervine captains no of fours and sixes are the same.

• All the captains' Runs and strike Rates in the 2022 T20 Tournament

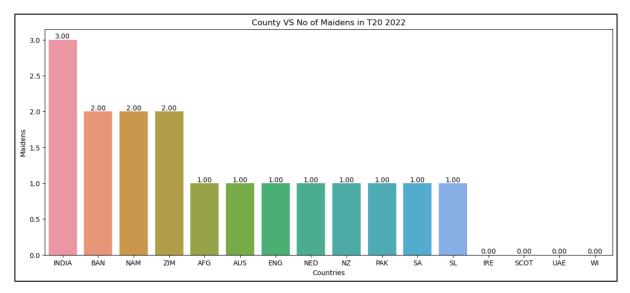


- According to the above line chart, Jos Buttler is the captain of the highest total runs.
- Regis Chakbva is the captain of the highest strike Rate in the tournament.
- Dasun Shanaka's tournament average strike rate is lower than Kane Williamson's.
- > Rohit Sharma's average strike rate in the tournament is higher than Shakib Al Hasan's.

## Summary statistic for the Numerical variable of the Bowling data set

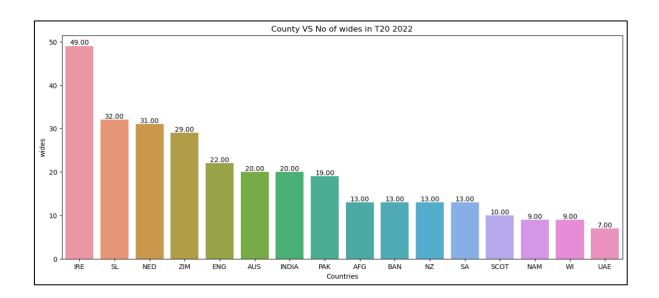


- According to the above table, a bowler's average no of bowled overs is 3.
- Also, throughout the tournament has taken 2 maiden overs.
- The bowler average taken wickets is 1 in the tournament.
- In an inning, a bowler taken maximum no of wickets is five.
- ➤ A bowler's average Economy Rate is 7.59.
- In the tournament, 50% of bowlers taken dots is equal to or less than 8.
- In an inning, a bowler given a maximum number of no balls is three.
- Total no of maidens' overs obtained by countries in the 2022 T20 tournament.



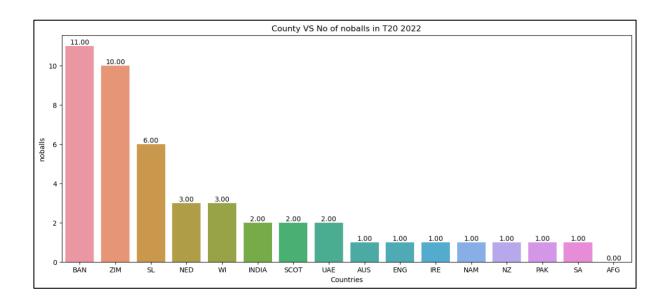
- According to the above chart India has the highest taken maidens overs in the tournament which is 3.
- > Ireland, Scotland, UAE, and Western Indies have not taken any maidens overs in the tournament.
- Bangladesh, Namibia, and Zimbabwe have taken two maiden overs in the tournament.
- Afghanistan, Australia, England, Netherlands, New Zealand, Pakistan, South Africa, and Sri Lanka have taken only one maiden over in the tournament.

• Total number of wides obtained by countries in the 2022 T20 tournament.



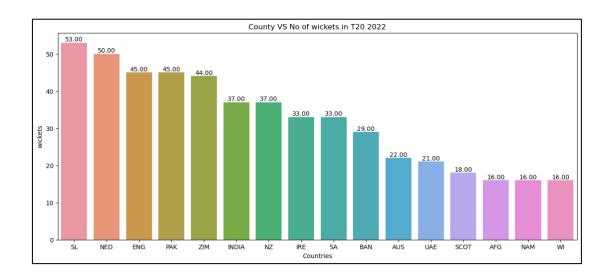
- > Ireland is the most no wides given country in the tournament, which is 49.
- ➤ UAE has the lowest no-wides given in the tournament, which is 7.
- > Sri Lankan given no of wides is 32.
- Australia and India both countries given no of wides are 20.

• Total number of no balls obtained by countries in the 2022 T20 tournament.



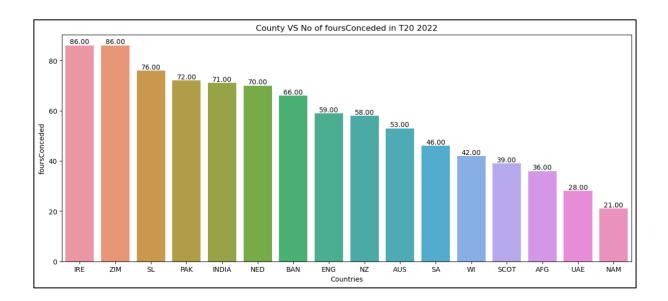
- According to the above chart, Bangladesh is the most no-ball given country in the tournament, which is 11.
- Afghanistan is the lowest no-ball given country, which is 0.
- The number of no-balls given by Sri Lanka is 6.
- ➤ India, Scotland, and UAE are given only two no-balls in the tournament.
- > Zimbabwe is the second highest no-ball given country, and which is 10.

• Total number of wickets obtained by countries in the 2022 T20 tournament.



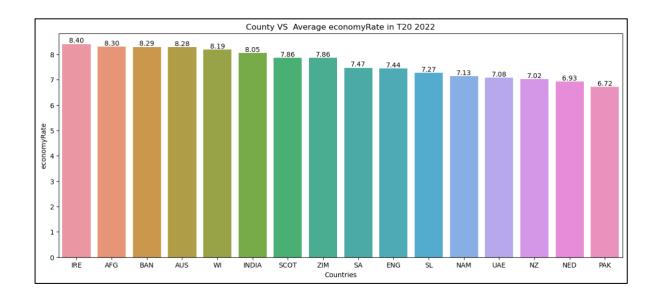
- > Sri Lanka has the highest total no of wickets country in the tournament.
- ➤ Western Indies, Namibia, and Afghanistan are the lowest wicket-taken countries, and which is 16.
- ➤ Pakistan total no of wickets taken is higher than India total no of wickets taken in the tournament.

• The total number of fours conceded was obtained by countries in the 2022 T20 tournament.



- According to the above chart Ireland was given the maximum four runs in the tournament which is 86.
- Namibia is the lowest four runs given country in the tournament which is 21.
- > Sri Lanka's given number of fours is higher than India's given number of fours.
- > South Africa's given number of fours is less than Australia's given number of fours.

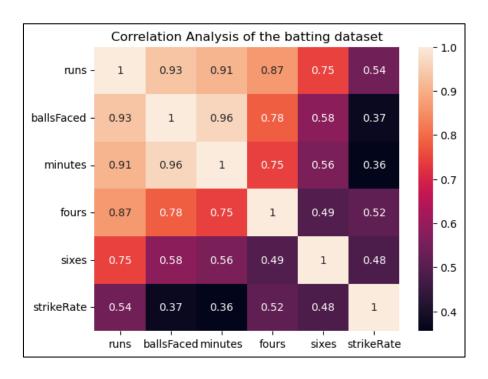
• Country Average Bowling Economy Rate in the 2022 T20 Tournament.



- > The highest average Bowling Economy rate country is Ireland, which is 8.4.
- ➤ The lowest average Bowling Economy rate country is Pakistan, which is 6.72.
- > Sri Lanka's average Bowling Economy rate is Less than India's average Bowling Economy rate.
- Australia's average Bowling Economy rate is higher than New Zealand's average Bowling Economy rate.

## **Correlation Analysis**

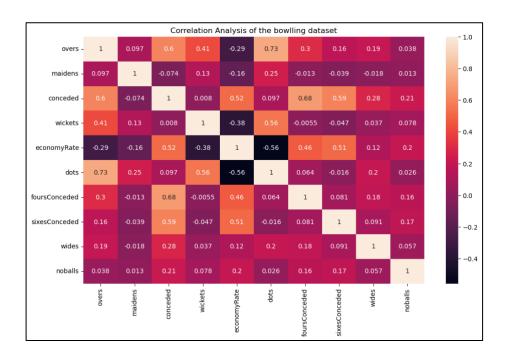
# Batting Dataset



- According to the above heatmap, we can say that batsman runs affect more highly variables such as,
  - o Number of the balls faces which is 0.93 relation.
  - o Number of minutes spent on the ground. which is 0.91 relation.
  - o Number of the fours taken which is 0.87 relation.
  - Number of the sixes taken which is 0.75 relation.
    - Also, taken runs depend on the strike rate. which is 0.54 relation.
- The number of the balls faced highly depends on variables such as,
  - o Number of minutes on the ground as a batsman. which is 0.96 relation.
  - Taken runs which is 0.93 relation.

- A Batsman was taken four (4) runs highly depending on variables such as,
  - Taken runs which is 0.87 relation.
  - o Number of the balls faces which is 0.78 relation.
  - Number of minutes on the ground as a batsman. which is 0.75 relation.

## Bowling Dataset



- According to the above heatmap, we can say the number of the taken wickets by the bowlers highly depends on variables such as,
  - o Number of the dots taken by the player which is 0.56 relation.
  - Number of the overs played by player which is 0.41 relation.
- Bowlers increasing the Economy rates depends on variables such as,
  - o Summation of the taken runs for him. which is 0.52 relation.
  - o Number of the taken six (6) runs. which is 0.51 relation.
  - Number of the taken four (4) runs. which is 0.46 relation.

- > Bowlers decreasing the Economy rates depends on variables such as,
  - o Number of the taken dot balls which is 0.56 relation.
  - o Number of the taken wickets which is 0.38 relation.

# 6. Dashboard

The 2022 T20 Dashboard has 5 main sections.

- HOME
- BATTING
- BOWLING
- BATTING DETAILS
- BOWLING DETAILS

### 1. HOME



#### 2. BATTING



### 3. BOWLING



#### 4. BATTING DETAILS



## 5. BOWLING DETAILS



## 7. Conclusion

The report delves into an in-depth analysis of the 2022 T20 World Cup, focusing on various aspects such as team performances, player statistics, and match dynamics. It outlines the rationale behind choosing T20 cricket as the subject of analysis, emphasizing its global appeal and strategic significance within the cricketing landscape. However, it also highlights challenges such as privacy considerations and data limitations.

The objectives of the project are clearly defined, aiming to create an interactive dashboard, provide match insights, analyze team performances, and offer detailed player profiles. These objectives are aimed at providing a comprehensive understanding of the tournament while ensuring user engagement and customization options.

Methodologies for data collection, visualization, and analysis are outlined, with a focus on leveraging tools like Power BI for visualization and employing various analytical strategies such as univariate and bivariate analysis. Using descriptive statistics and graphical charts provides insights into batting and bowling performances, and team dynamics.

Key findings from the data analysis include average runs scored by batsmen, distribution of fours and sixes, total runs scored by countries, and bowling economy rates. These findings offer valuable insights into individual and team performances throughout the tournament.

Overall, the report presents a comprehensive and insightful analysis of the 2022 T20 World Cup, highlighting key trends, challenges, and opportunities within the cricketing landscape.

# 8. References

- Kaggle. (2022). ICC Men's T20 World Cup 2022 Cricket Statistics. Kaggle. [https://www.kaggle.com/datasets]
- Microsoft. (2022). Power BI. Microsoft. [https://powerbi.microsoft.com/]
- International Cricket Council. (2022). ICC Men's T20 World Cup 2022. International Cricket Council. [https://www.icc-cricket.com]

# 9. Attachments

- EDA part for the Batting Card Dataset <u>Batting. ipynb</u>
- EDA part for the Bowling Card Dataset <u>Bowling. ipynb</u>
- Power BI Dashboard 2022 T20 Dashboard. pbix