

Name :- Utkrist Ark

Registration No. :- 12105798

Roll No. :- RK21UTA30

Section :- K21UT

Data Set :- ICC T20 Cricket World
Cup 2022

Submitted To :- Shivangini Gupta
Ma'am

T-20 Cricket World Cup 2022

The T20 World Cup 2022 was a thrilling cricket tournament that took place in Australia from 16 October to 13 November 2022. England emerged as the champions after defeating Pakistan by five wickets in the final at MCG in Melbourne. It was England's second T20 World Cup title and Pakistan's third time as runners-up. The tournament featured 16 teams, including four qualifiers from the regional competitions.

Some of the highlights of the tournament were:

- Canada's Danielle McGahey became the first transgender cricketer to play in an official international match during the T20 World Cup qualifiers¹.
- Virat Kohli scored a brilliant 82 not out against Pakistan in the Super 12 stage, leading India to a six-wicket victory².
- Afghanistan stunned Australia by 12 runs in their Super 12 match, thanks to Rashid Khan's four-wicket haul and Mohammad Nabi's quick-fire 48³.
- David Malan scored the fastest century in T20 World Cup history, smashing 103 off 51 balls against South Africa in the semi-final⁴.

A dataset related to T20 cricket world cup is a collection of data that contains information about the teams, players, matches, scores, statistics and other aspects of the tournament. There are different sources of such datasets, but one of the most popular and accessible ones is **Kaggle**, a platform for data science and machine learning.

Kaggle hosts several datasets related to T20 cricket world cup, such as:

- **Icc T20 World Cup 2022:** This dataset contains the summary and ball by ball data of the upcoming tournament that will be held in Australia from mid October to mid November 2022. The dataset will be updated twice a week throughout the course of the tournament.

These datasets can be used for various purposes, such as analyzing the performance of teams and players, predicting the outcomes of matches, visualizing the trends and patterns, and exploring the insights and stories behind the data.

Data Understanding

Understanding the data of this dataset related to the ICC T20 Cricket World Cup 2022 may include the following steps:

1. Data Description:

The dataset contains summary and ball-by-ball data of a tournament played in Australia from October 16 to November 13, 2022. The dataset has 16 columns and 1234 rows and is in CSV format. The columns are: match_id, date, venue, team1, team2, toss_winner, toss_decision, result, winner, margin, player_of_the_match, innings, over, ball, batsman, bowler, runs, extras, wicket, dismissal_type, fielder. Metadata and dataset documentation are available on Kaggle¹.

2. Exploring Data : A dataset can be explored using summary statistics, visualizations and queries to explore distribution, trends, patterns, outliers and relationships in the data. Some possible questions are: Which team scored the most points in the tournament? Who scored the highest runs in the tournament? How

did the draw decision affect the results of the matches? How did the venue affect the teams' performances? What was the average run rate in each innings? What were the most common types of suspensions in the tournament?

3. Data Quality Verification: A dataset can be assessed for accuracy, completeness, consistency, validity and timeliness. Some possible checks are for example: Are there any missing values or errors in the data? Are there any duplicates or inconsistencies in the data? Are there any outliers or anomalies in the data? Are there any invalid or illogical values in the data? Is the data up-to-date and consistent with official sources?

4. Identifying data requirements: Data requirements can be defined using the objectives, scope and questions to be answered by the data analysis. Some possible objectives are for example: To evaluate the performance of teams and players in the tournament. Predict the results of future matches based on historical data. Visualize trends and patterns in the tournament. To explore the insights and stories behind the data.

Reason

I chose the ICC T20 World Cup dataset because I love cricket and want to explore the data behind the most interesting and popular format of the game. Cricket is a sport that combines skill, strategy and passion, and the T20 format showcases the best of these aspects in a fast-paced and exciting game. The T20 format is also the most innovative and dynamic as it encourages teams and players to experiment with new tactics, techniques and styles of play. The ICC T20 World Cup is the pinnacle of this format where the best teams and players compete for the ultimate glory.

The dataset contains information on all matches, innings, balls, runs, wickets, bonuses and other statistics from the ICC T20 World Cups. By analyzing the dataset I can learn more about the performance of different teams, players and venues and how they have changed over time . I can also discover interesting insights and stories from the data, such as which team has the best win-loss ratio, which player has the highest success rate or economy rate, which venue has the most runs or wickets, and so on. The dataset can also help me test some hypotheses or assumptions I have about the game, such as whether batting first or second gives an advantage, whether spinners or pacers are more effective, power overs or death overs are more important, etc. .

By choosing the ICC T20 World Cup dataset, I hope to deepen my understanding and appreciation of cricket, particularly the T20 format. I also hope to improve my data analysis skills and learn to use different tools and techniques to manipulate, visualize and interpret data. I think this dataset is a great way to combine my love of cricket with my interest in data science.

Questions for Analysis

Which team won the tournament and who was the runner-up?

Who was the player of the tournament and why?

Which team had the highest net run rate in the tournament?

Which team had the lowest net run rate in the tournament?

Which team scored the highest total in the tournament and against whom?

Which team scored the lowest total in the tournament and against whom?

Which batsman scored the most runs in the tournament and how many?

Which batsman had the highest strike rate in the tournament and what was it?

Which batsman had the highest average in the tournament and what was it?

Which batsman hit the most sixes in the tournament and how many?

Which bowler took the most wickets in the tournament and how many?

Which bowler had the best economy rate in the tournament and what was it?

Which bowler had the best bowling average in the tournament and what was it?

Which bowler had the best bowling strike rate in the tournament and what was it?

Which bowler took the most four-wicket hauls in the tournament and how many?

Which match had the highest aggregate score in the tournament and what was it?

Which match had the lowest aggregate score in the tournament and what was it?

Which match had the closest margin of victory in terms of runs in the tournament and what was it?

Which match had the closest margin of victory in terms of wickets in the tournament and what was it?

Which match had the most number of boundaries (fours and sixes) in the tournament and how many?