

## **INDEX**

<b>SR NO.</b>	<b>CONTENTS</b>	<b>PAGE</b>
1.	INTRODUCTION	03 - 04
2.	SOFTWARE REQUIREMENT AND SPECIFICATIONS	05
3.	SYSTEM DESIGN	06 – 08
4.	IMPLEMENTATION	09- 14
7.	CONCLUSION	15

## INTRODUCTION

The ICC Men's T20 World Cup (earlier known as **ICC World Twenty20**) was first held in 2007. It was first decided that every two years an ICC T20 World Cup tournament is to take place, except in the event of an ICC Cricket World Cup being scheduled in the same year, in which case it will be held the year before. The first tournament was in 2007 in South Africa where India defeated Pakistan in the final. Two Associate teams had played in the first tournament, selected through the 2007 ICC World Cricket League Division One, a 50-over competition. In December 2007 it was decided to hold a qualifying tournament with a 20-over format to better prepare the teams. With six participants, two would qualify for the 2009 World Twenty20 and would each receive \$250,000 in prize money. The second tournament was won by Pakistan who beat Sri Lanka by 8 wickets in England on 21 June 2009. The 2010 ICC World Twenty20 tournament was held in West Indies in May 2010, where England defeated Australia by 7 wickets. The 2012 ICC World Twenty20 was won by the West-Indies, by defeating Sri Lanka at the finals. For the first time, a host nation competed in the final of the ICC World Twenty20. There were 12 participants for the title including Ireland and Afghanistan as 2012 ICC World Twenty20 Qualifier. It was the first time the World Twenty20 tournament took place in an Asian country. Pakistan was the only team to reach the last four in the first four editions of the tournament. 2014 saw the expansion to 16 teams featuring three teams making their debuts. Sri Lanka yet again made it to the Finals this time winning after their two other appearances in previous finals. The ICC Men's T20 World Cup has had five champions from six tournaments.

## REQUIREMENT ANALYSIS

Data Analysis is simply the analysis of various data means cleaning the data, transforming it into understandable form, and then modeling data to extract some useful information for business use or an organizational use. It is mainly used in taking business

decisions. Many libraries are available for doing the analysis. We are using NumPy, Pandas, Seaborn, Matplotlib.

- **Pandas:** Pandas is mainly used for converting data into tabular form and hence, makes the data more structured and easily to read.
- **NumPy:** NumPy is a library written in Python, used for numerical analysis in Python. It stores the data in the form of nd-arrays (n-dimensional arrays).
- **Matplotlib:** Matplotlib is a data visualization and graphical plotting package for Python and its numerical extension NumPy that runs on all platforms.
- **Seaborn:** Seaborn is a Python data visualisation package based on matplotlib that is tightly connected with panda's data structures. The core component of Seaborn is visualization, which aids in data exploration and comprehension.

Data visualization will help the data analysis to make it more understandable and interactive by plotting or displaying the data in pictorial form. Pandas, a Python open-source package that deals with three different data structures: series, data frames, and panels, solves that need of analyzing and visualization of data. Data analysis using Python makes task easier since Python Programming language has many advantages over any other programming language. It has prominent features like being a high-level programming language (the codes are in human readable form) it is easy to understand and use by any programmer or user. Many libraries and functions for statistical, numerical analysis are available in Python.

# SOFTWARE REQUIREMENT AND SPECIFICATIONS

## Purpose:

The purpose of this document is to clearly define the analysis on t20 cricket world cup held in 2022. To get the information about the out comes of the matchs that were played between different counties . This project also provides overall statistical calculations of the total runs scored and wickets taken at the tournament and the total outcome of the matchs

This gives ease to the new users to choose the among the courses based on the statistical data and the graphs showing the count.

## Software Requirements:

- OS Windows 10
- IDE Anaconda (Jupyter Notebook)
- Python 3.10.1

## Hardware Specifications:

- Min 2GB RAM
- Min 2.2Ghz Processor or Higher
- Min 30GB space

## SYSTEM DESIGN

It is a process of planning a new business system or replacing an existing system by defining its components or modules to satisfy the specific requirements. Before planning, you need to understand the old system thoroughly and determine how computers can best be used in order to operate efficiently.

### **Entity Relationship Diagram:**

- An entity relationship diagram (ERD) shows the relationships of entity sets stored in a database. An entity in this context is an object, a component of data.
- An entity set is a collection of similar entities. These entities can have attributes that define its properties.
- By defining the entities, their attributes, and showing the relationships between them, an ER diagram illustrates the logical structure of databases.
- ER diagrams are used to sketch out the design of a database.

### **Entity Relationship diagram symbols:**

- **Entity:**

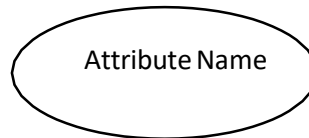
Any real-world object can be represented as an entity about which data can be stored in a database. All the real-world objects like a book, an organization, a product, a car, a person are the examples of an entity. Any living or non-living objects can be represented by an entity. An entity is symbolically represented by a rectangle enclosing its name.



Entity Name

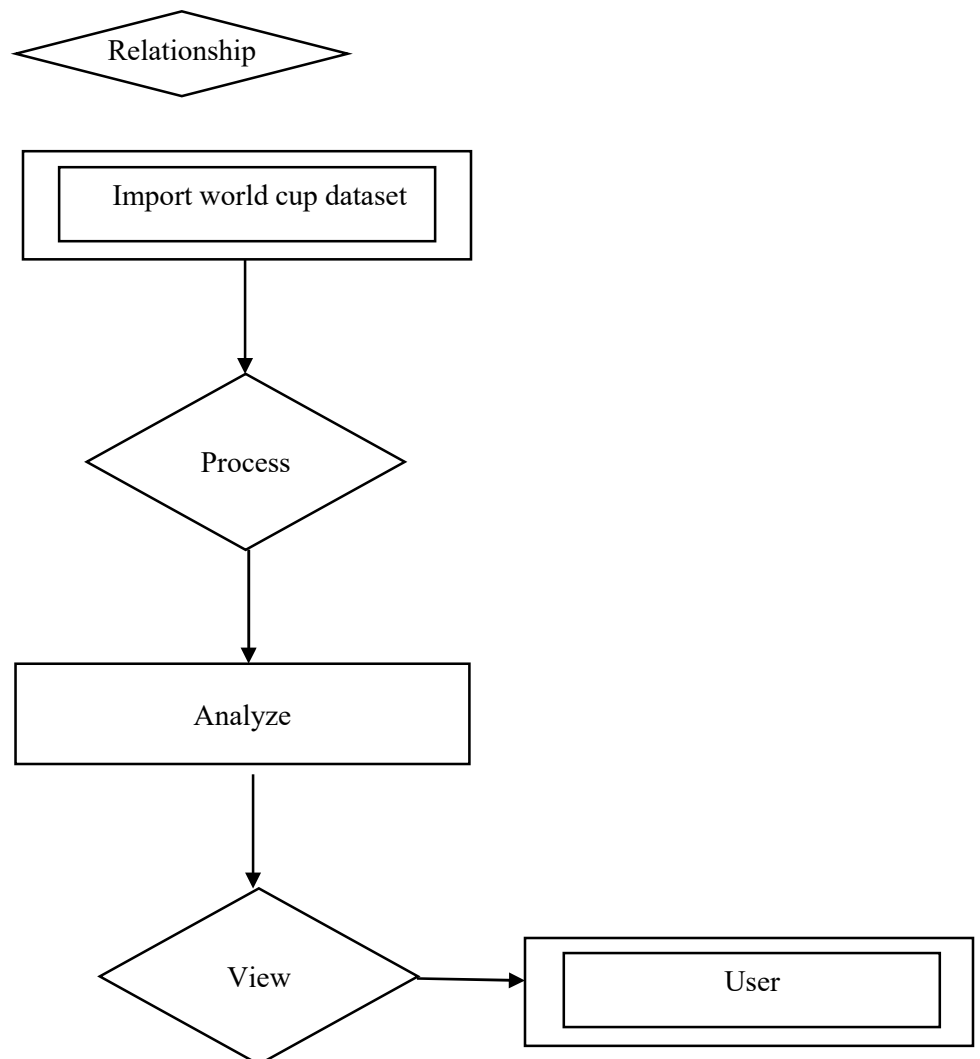
- **Attribute:**

Each entity has a set of properties. These properties of each entity are termed as attributes. For example, a student entity would be described by attributes such as roll no, name, address, etc. Attributes are indicated by ovals in an ER diagram.



- **Relationships:**

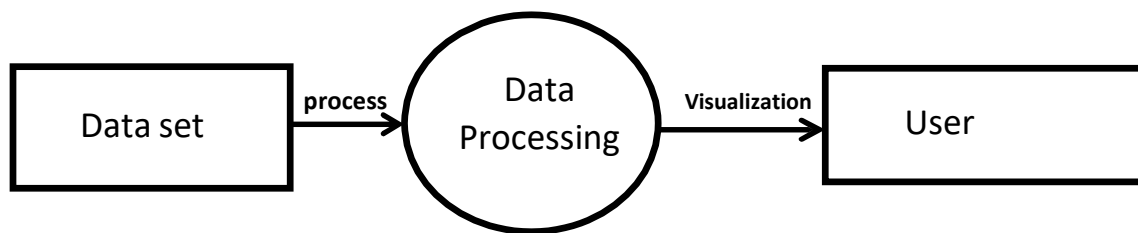
A relationship is defined as bond or attachment between 2 or more entities. Normally, a verb in a sentence signifies a relationship. A diamond is used to symbolically represent a relationship in the ER diagram.



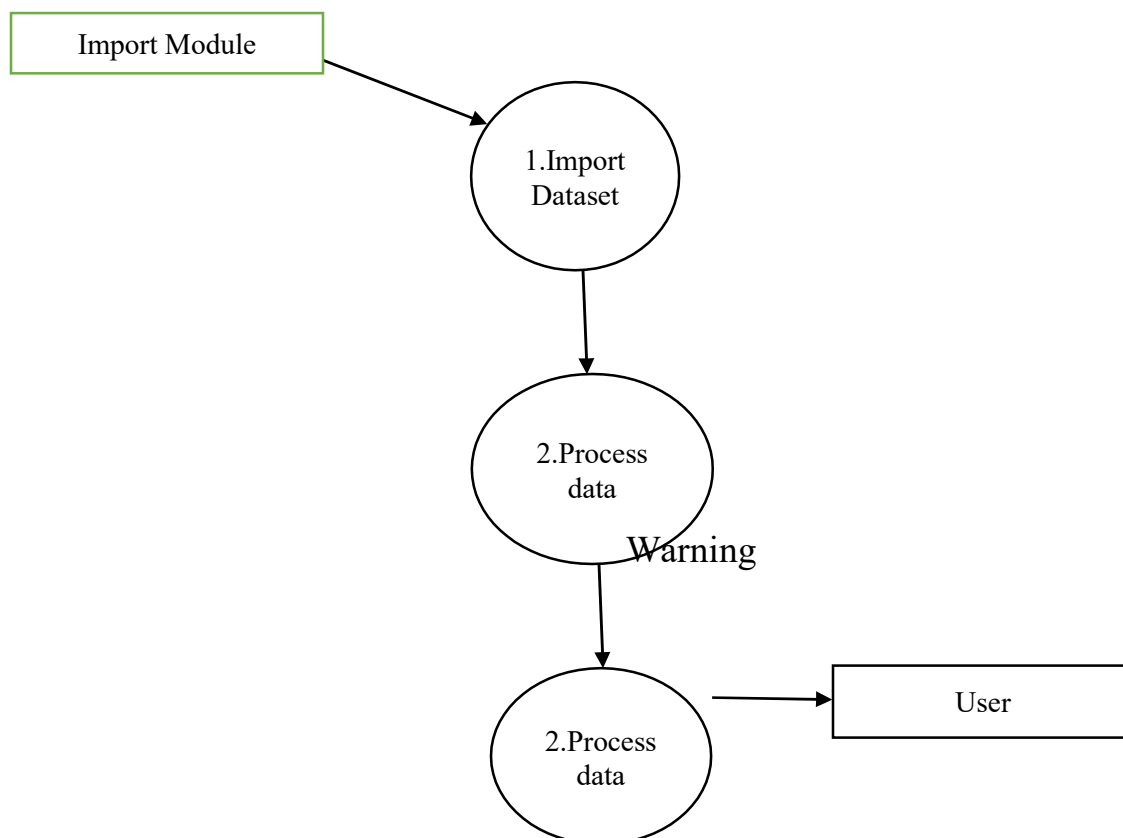
### Data Flow Diagram

- A data flow diagram is a way of representing a flow of data of a process or system. It includes data inputs and outputs, data stores, and the various sub processes the data moves through. DFDs are built using standardized symbols and notation to describe various entities and their relationships.
- Data flow diagrams visually represent systems and processes that would be hard to describe in a chunk of text.

### Zero Level Diagram:



### 1<sup>st</sup> Level Diagram:



## IMPLEMENTATION

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
data = pd.read_csv("t20-world-cup-22.csv")
```

### Display Top 5 Records

```
: data.head(5)#display the top 5 rows of thr dataset
```

	venue	team1	team2	stage	toss winner	toss decision	first innings score	first innings wickets	second innings score	second innings wickets	winner	won by	player of the match	top scorer	highest score	best bowler	best bowling figure
0	SCG	New Zealand	Australia	Super 12	Australia	Field	200.0	3.0	111.0	10.0	New Zealand	Runs	Devon Conway	Devon Conway	92.0	Tim Southee	3-6
1	Optus Stadium	Afghanistan	England	Super 12	England	Field	112.0	10.0	113.0	5.0	England	Wickets	Sam Curran	Ibrahim Zadran	32.0	Sam Curran	5-10
2	Blundstone Arena	Ireland	Sri Lanka	Super 12	Ireland	Bat	128.0	8.0	133.0	1.0	Sri Lanka	Wickets	Kusal Mendis	Kusal Mendis	68.0	Maheesh Theekshana	2-19
3	MCG	Pakistan	India	Super 12	India	Field	159.0	8.0	160.0	6.0	India	Wickets	Virat Kohli	Virat Kohli	82.0	Hardik Pandya	3-30
4	Blundstone Arena	Bangladesh	Netherlands	Super 12	Netherlands	Field	144.0	8.0	135.0	10.0	Bangladesh	Runs	Taskin Ahmed	Colin Ackermann	62.0	Taskin Ahmed	4-25

### Display bottom 5 records

```
data.tail(5)#used to display the last 5 rows of the dataset
```

	venue	team1	team2	stage	toss winner	toss decision	first innings score	first innings wickets	second innings score	second innings wickets	winner	won by	player of the match	top scorer	highest score	best bowler	best bowling figure
28	Adelaide Oval	Bangladesh	Pakistan	Super 12	Bangladesh	Bat	127.0	8.0	128.0	5.0	Pakistan	Wickets	Shaheen Afridi	Najmul Hossain	54.0	Shaheen Afridi	4-22
29	MCG	India	Zimbabwe	Super 12	India	Bat	186.0	5.0	115.0	10.0	India	Runs	Suryakumar Yadav	Suryakumar Yadav	61.0	Ravichandran Ashwin	3-22
30	SCG	New Zealand	Pakistan	Semi-final	New Zealand	Bat	152.0	4.0	153.0	3.0	Pakistan	Wickets	Mohammad Rizwan	Mohammad Rizwan	57.0	Shaheen Afridi	2-24
31	Adelaide Oval	India	England	Semi-final	England	Field	168.0	6.0	170.0	0.0	England	Wickets	Alex Hales	Alex Hales	86.0	Chris Jordan	3-43
32	MCG	Pakistan	England	Final	England	Field	137.0	8.0	138.0	5.0	England	Wickets	Sam Curran	Ben Stokes	52.0	Sam Curran	3-12



## Check for null values in each column

```
data.isnull().any()#identifying the null values in the columns
```

```
venue                False
team1                False
team2                False
stage                False
toss winner          True
toss decision         True
first innings score   True
first innings wickets True
second innings score  True
second innings wickets True
winner               True
won by               True
player of the match   True
top scorer            True
highest score         True
best bowler           True
best bowling figure   True
dtype: bool
```

## Checking for number of null values

```
data.isnull().sum()#identifying the sum of null values in the dataset
```

```
venue                0
team1                0
team2                0
stage                0
toss winner          3
toss decision         3
first innings score   3
first innings wickets 3
second innings score  3
second innings wickets 3
winner               4
won by               4
player of the match   4
top scorer            3
highest score         3
best bowler           3
best bowling figure   3
dtype: int64
```

## Provides info on the dataset

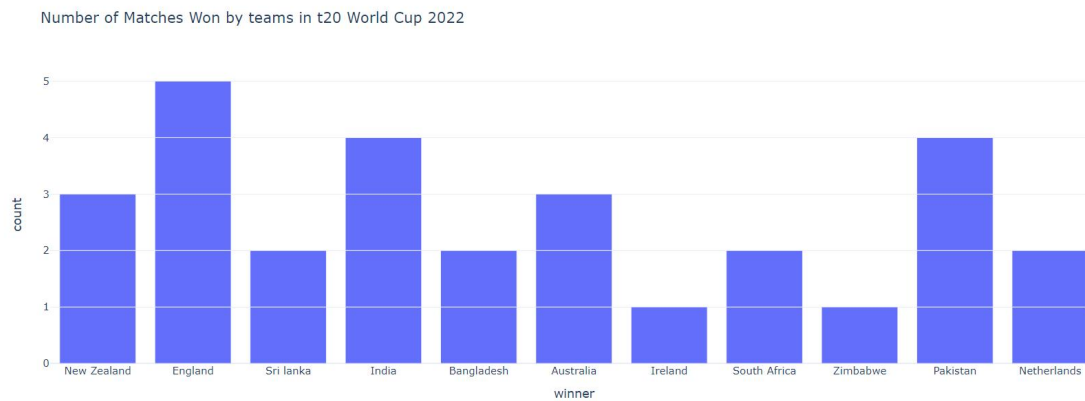
```
data.info()#details of dataset
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33 entries, 0 to 32
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   venue                 33 non-null    object
1   team1                 33 non-null    object
2   team2                 33 non-null    object
3   stage                 33 non-null    object
4   toss winner           30 non-null    object
5   toss decision         30 non-null    object
6   first innings score   30 non-null    float64
7   first innings wickets 30 non-null    float64
8   second innings score  30 non-null    float64
9   second innings wickets 30 non-null    float64
10  winner                29 non-null    object
11  won by                29 non-null    object
12  player of the match    29 non-null    object
13  top scorer            30 non-null    object
14  highest score         30 non-null    float64
15  best bowler           30 non-null    object
16  best bowling figure    30 non-null    object
dtypes: float64(5), object(12)
memory usage: 4.5+ KB
```

```
import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio
pio.templates.default = "plotly_white"
```

### Number of matches won by each team

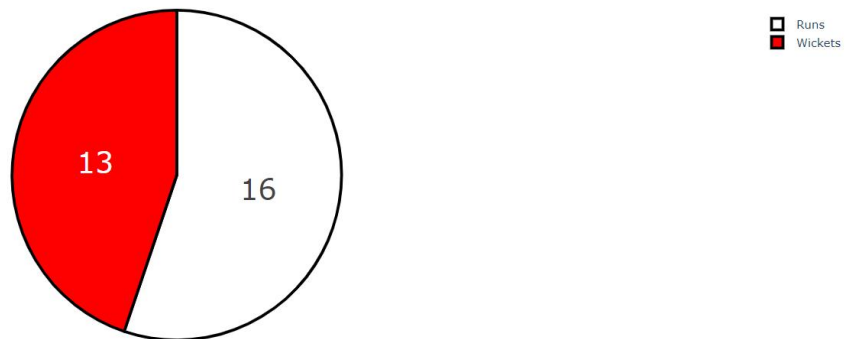
```
figure = px.bar(data,
                x=data["winner"],
                title="Number of Matches Won by teams in t20 World Cup 2022")
figure.show()
#Number of matches won by each teams
```



### Number of Matches Won By Batting Or Chasing

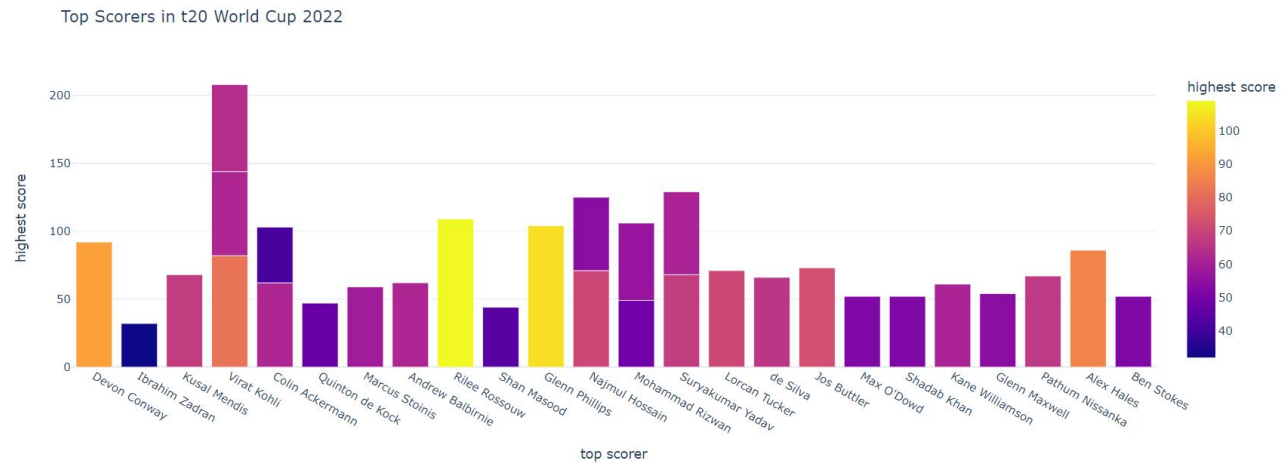
```
won_by = data["won by"].value_counts()
label = won_by.index
counts = won_by.values
colors = ['white', 'red']
fig = go.Figure(data=[go.Pie(labels=label, values=counts)])
fig.update_layout(title_text='Number of Matches Won By Batting Or Chasing')
fig.update_traces(hoverinfo='label+percent', textinfo='value', textfont_size=30,
                  marker=dict(colors=colors, linedict(color='black', width=3)))
fig.show()
#the number of matches won by batting first or second in the t20 world cup 2022
```

Number of Matches Won By Batting Or Chasing



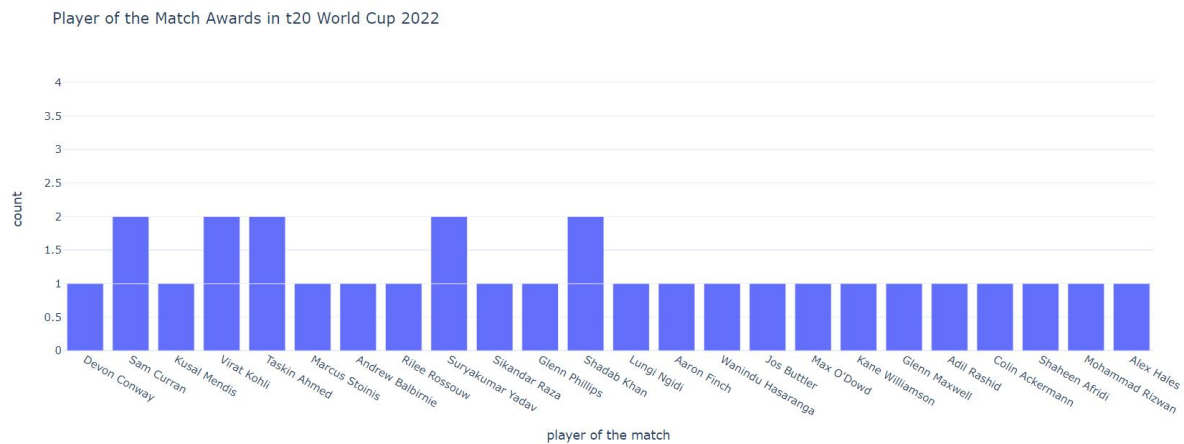
## Top scores at the world cup 2022

```
figure = px.bar(data,
                x=data["top scorer"],
                y = data["highest score"],
                color = data["highest score"],
                title="Top Scorers in t20 World Cup 2022")
figure.show()
#top scorers at the world cup
```



## Man of the match awards won

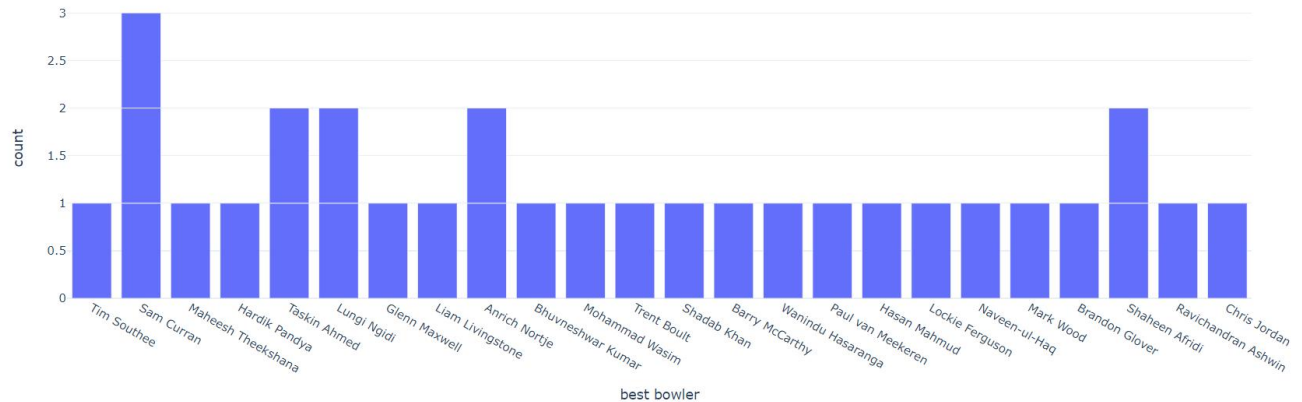
```
figure = px.bar(data,
                x = data["player of the match"],
                title="Player of the Match Awards in t20 World Cup 2022")
figure.show()
#players with most man of the matches
```



## All the bowling figures

```
figure = px.bar(data,
                xdata["best bowler"],
                title="Best Bowlers in t20 World Cup 2022")
figure.show()
#best bowling figures
```

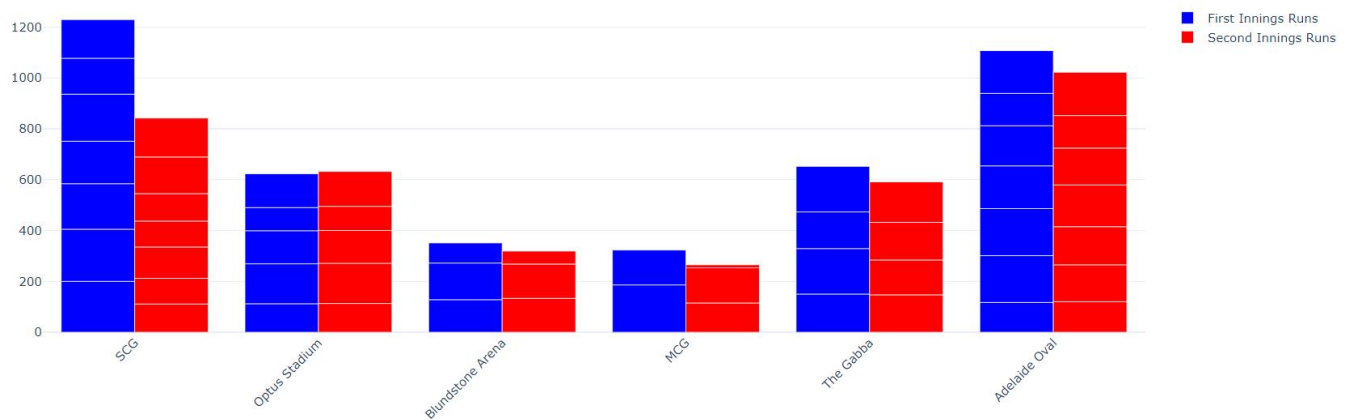
Best Bowlers in t20 World Cup 2022



## All matches that are won by batting or second in each stadium

```
fig = go.Figure()
fig.add_trace(go.Bar(
    xdata["venue"],
    ydata["first innings score"],
    name='First Innings Runs',
    marker_color='blue'
))
fig.add_trace(go.Bar(
    xdata["venue"],
    ydata["second innings score"],
    name='Second Innings Runs',
    marker_color='red'
))
fig.update_layout(barmodes='group',
                  xaxis_tickangle=-45,
                  title="Best Stadiums to Bat First or Chase")
fig.show()
#number of runs scored in first and second innings in each stadium
```

Best Stadiums to Bat First or Chase



## Games won in each stadium by chasing

```
new=data[data['won by']=='Wickets']['venue']
new
#games won in stadiums by wickets
```

```
1      Optus Stadium
2      Blundstone Arena
3      MCG
6      Optus Stadium
16     Optus Stadium
17     Optus Stadium
19     The Gabba
21     Adelaide Oval
26     SCG
28     Adelaide Oval
30     SCG
31     Adelaide Oval
32     MCG
Name: venue, dtype: object
```

## Match data on team India

```
data[data['winner']=='India']
#data set in with team india has won the matches
```

	venue	team1	team2	stage	toss winner	toss decision	first innings score	first innings wickets	second innings score	second innings wickets	winner	won by	player of the match	top scorer	highest score	best bowler	best bowling figure
3	MCG	Pakistan	India	Super 12	India	Field	159.0	8.0	160.0	6.0	India	Wickets	Virat Kohli	Virat Kohli	82.0	Hardik Pandya	3-30
10	SCG	India	Netherlands	Super 12	India	Bat	179.0	2.0	123.0	9.0	India	Runs	Suryakumar Yadav	Virat Kohli	62.0	Bhuvneshwar Kumar	2-19
22	Adelaide Oval	India	Bangladesh	Super 12	Bangladesh	Field	184.0	6.0	145.0	6.0	India	Runs	Virat Kohli	Virat Kohli	64.0	Hasan Mahmud	3-47
29	MCG	India	Zimbabwe	Super 12	India	Bat	186.0	5.0	115.0	10.0	India	Runs	Suryakumar Yadav	Suryakumar Yadav	61.0	Ravichandran Ashwin	3-22

## Matches in which virat kohli was the top scorer

```
data[data['top scorer']=='Virat Kohli']
#matches where virat kohli was the top scorer
```

	venue	team1	team2	stage	toss winner	toss decision	first innings score	first innings wickets	second innings score	second innings wickets	winner	won by	player of the match	top scorer	highest score	best bowler	best bowling figure
3	MCG	Pakistan	India	Super 12	India	Field	159.0	8.0	160.0	6.0	India	Wickets	Virat Kohli	Virat Kohli	82.0	Hardik Pandya	3-30
10	SCG	India	Netherlands	Super 12	India	Bat	179.0	2.0	123.0	9.0	India	Runs	Suryakumar Yadav	Virat Kohli	62.0	Bhuvneshwar Kumar	2-19
22	Adelaide Oval	India	Bangladesh	Super 12	Bangladesh	Field	184.0	6.0	145.0	6.0	India	Runs	Virat Kohli	Virat Kohli	64.0	Hasan Mahmud	3-47

## Highest runs scored by batting first

```
data[data['first innings score'].max()==data['first innings score']]
#highest runs scored in first innings
```

	venue	team1	team2	stage	toss winner	toss decision	first innings score	first innings wickets	second innings score	second innings wickets	winner	won by	player of the match	top scorer	highest score	best bowler	best bowling figure
9	SCG	South Africa	Bangladesh	Super 12	South Africa	Bat	205.0	5.0	101.0	10.0	South Africa	Runs	Rilee Rossouw	Rilee Rossouw	109.0	Anrich Nortje	4-10

## Lowest runs scored by batting first

```
data[data['first innings score'].min()==data['first innings score']]
#lowest runs scored in first innings
```

	venue	team1	team2	stage	toss winner	toss decision	first innings score	first innings wickets	second innings score	second innings wickets	winner	won by	player of the match	top scorer	highest score	best bowler	best bowling figure
5	Blundstone Arena	Zimbabwe	South Africa	Super 12	Zimbabwe	Bat	79.0	5.0	51.0	0.0	NaN	NaN	NaN	Quinton de Kock	47.0	Lungi Ngidi	2-20

## **CONCLUSION**

We did Prediction data Analysis on the features of this dataset and saw how each feature is analyzed and predicted.

We did data cleaning with various technique involved by visualizing with various plots including count plot, bar plot, his plot, line graph and pie plot.

We analyzed each variable to check if data is cleaned and normally distributed. We cleaned the data and removed NA values.