Exploratory Data Analysis on Cricket T20 Internationals From 2005-2024

Author: Irfan Ullah Khan

GITHUB PROFILE KAGGLE PROFILE LINKEDIN PROFILE

YOUTUBE PROFILE FACEBOOK PROFILE TIKTOK PROFILE

TWITTER PROFILE INSTAGRAM PROFILE EMAIL CONTACT ME

About Dataset

This extensive dataset serves as a comprehensive repository of historical information regarding T20 International (T20I) cricket matches dating back to the inception of the format. T20I cricket is renowned for its thrilling encounters, and this dataset meticulously documents the particulars of these matches. It stands as a valuable resource for cricket enthusiasts, statisticians, and analysts eager to delve into and dissect T20I cricket data.

Key Features:

- Match Details: Thorough information pertaining to each T20I match, encompassing match date, location, and format.
- Teams and Players: In-depth details about the participating teams, encompassing player names, roles, and batting/bowling statistics).
- Match Outcomes: Insights into match results, encompassing the victorious team and the margin of victory.
- Player of the Match: Recognition of the standout player in each T20I match.
- Umpires and Match Referees: Particulars of the officials responsible for overseeing the match.
- Toss Details: Revelations about the toss winner's decisions, which can significantly influence the game's trajectory.
- Venue Information: Location specifics, including stadium name, city, and country.

Importing Liabraries

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

Loading datasets

```
# All matches records
matches_data = pd.read_csv("/content/t20i_Matches_Data.csv")

# Players information
players_info = pd.read_csv("/content/players_info.csv")

#Batting Stats
batting_data = pd.read_csv("/content/t20i_Batting_Card.csv")

#Bowling Stats
bowling_data = pd.read_csv("/content/t20i_Bowling_Card.csv")
```

Understanding and cleaning matches data

```
matches_data.head()
```

	T20I Match	Match ID	Match Name	Series ID	Series Name	Match Date	Match Format	Team1 ID	Team1 Name	Ca

	No									
0	52	291356	Australia Vs India Only T20I	291355	India tour of Australia - 2007 (2007/08)	2008-02-01	T20	6	India	
1	54	300436	New Zealand Vs England 2Nd T20I	300418	England tour of New Zealand - 2008 (2007/08)	2008-02-07	T20	1	England	
2	65	361531	Netherlands Vs Scotland 2Nd Semi Final	353665	ICC World Twenty20 Qualifier Bermuda, Canada, ...	2008-08-04	T20	30	Scotland	4
3	66	354459	Kenya Vs Scotland 3Rd Place Playoff	353665	ICC World Twenty20 Qualifier Bermuda, Canada,	2008-08-04	T20	26	Kenya	

```
matches_data.columns
```

```
    Index(['T20I Match No', 'Match ID', 'Match Name', 'Series ID', 'Series Name',
           'Match Date', 'Match Format', 'Team1 ID', 'Team1 Name', 'Team1 Captain',
           'Team1 Runs Scored', 'Team1 Wickets Fell', 'Team1 Extras Rec',
           'Team2 ID', 'Team2 Name', 'Team2 Captain', 'Team2 Runs Scored',
           'Team2 Wickets Fell', 'Team2 Extras Rec', 'Match Venue (Stadium)',
           'Match Venue (City)', 'Match Venue (Country)', 'Umpire 1', 'Umpire 2',
           'Match Referee', 'Toss Winner', 'Toss Winner Choice', 'Match Winner',
           'Match Result Text', 'MOM Player', 'Team1 Playing 11',
           'Team2 Playing 11', 'Debut Players'],
          dtype='object')
```

```
matches_data.shape
```

```
    (2411, 33)
```

```
# Renaming columns
new_col = []
for col_name in matches_data.columns.to_list():
    new_col.append(col_name.replace(" ","_").lower())

matches_data.columns = new_col
```

```
matches_data.columns
```

```
    Index(['t20i_match_no', 'match_id', 'match_name', 'series_id', 'series_name',
           'match_date', 'match_format', 'team1_id', 'team1_name', 'team1_captain',
           'team1_runs_scored', 'team1_wickets_fell', 'team1_extras_rec',
           'team2_id', 'team2_name', 'team2_captain', 'team2_runs_scored',
           'team2_wickets_fell', 'team2_extras_rec', 'match_venue_(stadium)',
           'match_venue_(city)', 'match_venue_(country)', 'umpire_1', 'umpire_2',
           'match_referee', 'toss_winner', 'toss_winner_choice', 'match_winner',
           'match_result_text', 'mom_player', 'team1_playing_11',
           'team2_playing_11', 'debut_players'],
          dtype='object')
```

```
matches_data.head()
```

	t20i_match_no	match_id	match_name	series_id	series_name	match_date	match_
0	52	291356	Australia Vs India Only T20I	291355	India tour of Australia - 2007 (2007/08)	2008-02-01	
1	54	300436	New Zealand Vs England 2Nd T20I	300418	England tour of New Zealand - 2008 (2007/08)	2008-02-07	
2	65	361531	Netherlands Vs Scotland 2Nd Semi Final	353665	ICC World Twenty20 Qualifier Bermuda, Canada, ...	2008-08-04	
3	66	354459	Kenya Vs Scotland 3Rd Place Playoff	353665	ICC World Twenty20 Qualifier Bermuda, Canada, ...	2008-08-04	
4	69	361653	Sri Lanka Vs Zimbabwe 1St Match	361644	T20 Canada in Canada - 2008 (2008/09)	2008-10-10	

5 rows × 33 columns

```
#Looking for null values
matches_data.isna().sum()
```

```
t20i_match_no              0
match_id                   0
match_name                 0
series_id                  0
series_name                0
match_date                11
match_format               0
team1_id                   0
team1_name                 0
team1_captain              0
team1_runs_scored         13
team1_wickets_fell        13
team1_extras_rec          13
team2_id                   0
team2_name                 0
team2_captain              0
team2_runs_scored         46
team2_wickets_fell        46
team2_extras_rec          46
match_venue_(stadium)      0
match_venue_(city)         0
match_venue_(country)      0
umpire_1                   2
umpire_2                   2
match_referee            441
toss_winner                1
toss_winner_choice        10
match_winner              91
match_result_text          0
mom_player               297
team1_playing_11           0
team2_playing_11           0
debut_players              0
dtype: int64
```

```
# Dropping Redundant columns
matches_data.drop(columns=['match_referee', 'umpire_1', 'umpire_2'], inplace=True)
```

```
matches_data.head(2)
```

	t20i_match_no	match_id	match_name	series_id	series_name	match_date	match_
0	52	291356	Australia Vs India Only T20I	291355	India tour of Australia - 2007 (2007/08)	2008-02-01	
1	54	300436	New Zealand Vs England 2Nd T20I	300418	England tour of New Zealand - 2008 (2007/08)	2008-02-07	

2 rows × 30 columns

```
matches_data.sort_values('t20i_match_no', inplace=True)
```

Double-click (or enter) to edit

```
matches_data
```

	t20i_match_no	match_id	match_name	series_id	series_name	match_date	mat
10	1	211048	New Zealand Vs Australia Only T20I	238218	Australia tour of New Zealand - 2005 (2004/05)	NaN	
11	2	211028	England Vs Australia Only T20I	238219	Australia tour of England and Scotland - 2005...	NaN	
12	3	222678	South Africa Vs New Zealand Only T20I	238166	New Zealand tour of South Africa - 2005 (2005...	NaN	
13	4	226374	Australia Vs South Africa Only T20I	226335	South Africa tour of Australia - 2005 (2005/06)	NaN	
14	5	237242	New Zealand Vs West Indies Only T20I	237244	West Indies tour of New Zealand - 2006 (2005/06)	NaN	
...	...	...	...	...	...	...	
2406	2407	1373581	West Indies Vs England 3Rd T20I	1373560	England tour of West Indies - 2023 (2023/24)	2023-12-16	
2407	2408	1412221	Mozambique Vs Rwanda 11Th Match Group A	1412206	Africa Cricket Association Cup in South Africa...	2023-12-17	
2408	2409	1412222	Ghana Vs Sierra Leone 12Th Match Group B	1412206	Africa Cricket Association Cup in South Africa...	2023-12-17	
2409	2410	1412223	Botswana Vs Uganda 1St Semi Final	1412206	Africa Cricket Association Cup in South Africa...	2023-12-18	
2410	2411	1412224	Kenya Vs Malawi 2Nd Semi Final	1412206	Africa Cricket Association Cup in South Africa...	2023-12-18	

2411 rows × 30 columns

```
# Converting date values to Date Time format i.e Timestamp

matches_data['match_date'] = pd.to_datetime(matches_data['match_date'])
```

```
# Finding out time range of data
print(matches_data['match_date'].min())
print(matches_data['match_date'].max())
```

```
2006-06-15 00:00:00
2023-12-18 00:00:00
```

Understanding & Cleaning player information data

```
players_info.head()
```

	player_id	player_object_id	player_name	dob	dod	gender	batting_style	bo
0	93957	1046619	Avinash Pai	1982-01-24	NaN	M	right-hand bat	
1	2182	51462	Sherwin Campbell	1970-11-01	NaN	M	right-hand bat	
2	48391	56194	Tamim Iqbal	1989-03-20	NaN	M	left-hand bat	

```
players_info.shape
```

```
(6697, 11)
```

```
players_info.isna().sum()
```

```
player_id              0
player_object_id       0
player_name            0
dob                   26
dod                 5439
gender                 0
batting_style         97
bowling_style       1160
country_id             2
image_url           2957
image_metadata      2957
dtype: int64
```

```
players_info['gender'].value_counts()
```

```
M    6696
F       1
Name: gender, dtype: int64
```

```
# Dropping Redundant Columns
players_info.drop(columns=['gender', 'image_url', 'image_metadata'], inplace=True)
```

```
players_info.head()
```

	player_id	player_object_id	player_name	dob	dod	batting_style	bowling_st
0	93957	1046619	Avinash Pai	1982-01-24	NaN	right-hand bat	right-offbı
1	2182	51462	Sherwin Campbell	1970-11-01	NaN	right-hand bat	right-mec
2	48391	56194	Tamim Iqbal	1989-03-20	NaN	left-hand bat	N

Understanding & Cleaning batting stats data

```
batting_data.shape
```

```
(52419, 13)
```

```
batting_data.head(20)
```

	Match ID	innings	team	batsman	runs	balls	fours	sixes	strikeRate	isO
0	361657	1	Zimbabwe	10423.0	53.0	38.0	5.0	3.0	139.47	Tr
1	361657	1	Zimbabwe	49282.0	12.0	22.0	2.0	0.0	54.54	Tr
2	361657	1	Zimbabwe	47619.0	8.0	26.0	0.0	0.0	30.76	Tr
3	361657	1	Zimbabwe	10421.0	4.0	6.0	1.0	0.0	66.66	Tr
4	361660	1	Pakistan	11647.0	44.0	41.0	4.0	1.0	107.31	Tr
5	361657	1	Zimbabwe	45252.0	4.0	8.0	0.0	0.0	50.00	Tr
6	361660	1	Pakistan	47738.0	0.0	5.0	0.0	0.0	0.00	Tr
7	361657	1	Zimbabwe	10639.0	0.0	1.0	0.0	0.0	0.00	Tr
8	361660	1	Pakistan	4169.0	14.0	13.0	1.0	0.0	107.69	Tr
9	361531	1	Scotland	45548.0	22.0	27.0	1.0	0.0	81.48	Tr
10	361660	1	Pakistan	10439.0	0.0	4.0	0.0	0.0	0.00	Tr
11	361657	1	Zimbabwe	49274.0	7.0	9.0	1.0	0.0	77.77	Tr
12	361531	1	Scotland	46048.0	40.0	44.0	4.0	2.0	90.90	Tr
13	361660	1	Pakistan	8270.0	19.0	19.0	1.0	1.0	100.00	Tr
14	361657	1	Zimbabwe	51554.0	1.0	6.0	0.0	0.0	16.66	Tr
15	361531	1	Scotland	46142.0	25.0	27.0	1.0	0.0	92.59	Tr
16	361660	1	Pakistan	19596.0	23.0	25.0	2.0	0.0	92.00	Fal
17	361657	1	Zimbabwe	45326.0	0.0	2.0	0.0	0.0	0.00	Fal
18	361531	1	Scotland	8221.0	0.0	1.0	0.0	0.0	0.00	Tr

Next steps: Generate code with `batting_data` View recommended plots

```
batting_data.tail(20)
```

	Match ID	innings	team	batsman	runs	balls	fours	sixes	strikeRate
52399	1412220	2	Botswana	53245.0	12.0	11.0	1.0	0.0	109.09
52400	1412224	2	Malawi	103952.0	5.0	5.0	0.0	0.0	100.00
52401	1412220	2	Botswana	103191.0	19.0	39.0	1.0	0.0	48.71
52402	1412224	2	Malawi	104842.0	4.0	3.0	0.0	0.0	133.33
52403	1412220	2	Botswana	108745.0	13.0	13.0	1.0	0.0	100.00
52404	1412224	2	Malawi	65386.0	NaN	NaN	NaN	NaN	NaN
52405	1412220	2	Botswana	103973.0	25.0	25.0	0.0	1.0	100.00
52406	1412224	2	Malawi	113234.0	NaN	NaN	NaN	NaN	NaN
52407	1412220	2	Botswana	98611.0	3.0	7.0	0.0	0.0	42.85
52408	1412224	2	Malawi	113229.0	NaN	NaN	NaN	NaN	NaN
52409	1412220	2	Botswana	55147.0	0.0	1.0	0.0	0.0	0.00
52410	1412224	2	Malawi	114011.0	NaN	NaN	NaN	NaN	NaN
52411	1412220	2	Botswana	68480.0	7.0	16.0	0.0	0.0	43.75
52412	1412224	2	Malawi	110420.0	NaN	NaN	NaN	NaN	NaN
52413	1412220	2	Botswana	112124.0	1.0	1.0	0.0	0.0	100.00
52414	1412224	2	Malawi	110423.0	NaN	NaN	NaN	NaN	NaN
52415	1412220	2	Botswana	71147.0	1.0	5.0	0.0	0.0	20.00
52416	1412224	2	Malawi	114966.0	NaN	NaN	NaN	NaN	NaN
52417	1412220	2	Botswana	98613.0	1.0	2.0	0.0	0.0	50.00

```
batting_data.sample(10)
```

	Match ID	innings	team	batsman	runs	balls	fours	sixes	strikeRat
4901	527013	2	Zimbabwe	47618.0	22.0	17.0	2.0	0.0	129.4
26924	1263472	2	Sri Lanka	78229.0	0.0	5.0	0.0	0.0	0.0
21918	1201670	2	Spain	93983.0	NaN	NaN	NaN	NaN	Na
2632	403375	1	Pakistan	19930.0	NaN	NaN	NaN	NaN	Na
4115	446961	2	England	8107.0	1.0	2.0	0.0	0.0	50.0
8722	690351	2	Bangladesh	52364.0	NaN	NaN	NaN	NaN	Na
21213	1199536	2	Kenya	98584.0	3.0	14.0	0.0	0.0	21.4
51751	1411434	1	Cameroon	110563.0	0.0	3.0	0.0	0.0	0.0
[illegible]	[illegible]	[illegible]	New	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

```
#Checking null values
batting_data.isna().sum()
```

```
Match ID          0
innings           0
team             26
batsman          26
runs          13542
balls         13542
fours         13542
sixes         13542
strikeRate    13542
isOut            26
wicketType       26
fielders      22560
bowler        26771
dtype: int64
```

```
# Dropping rows with null values in runs, balls, fours, sixes, strikeRate
batting_data.dropna(subset=['runs','balls','fours','sixes', 'strikeRate'], inplace=True)
```

```
batting_data.isna().sum()
```

```
Match ID          0
innings           0
team              0
batsman           0
runs              0
balls             0
fours             0
sixes             0
strikeRate        0
isOut             0
wicketType        0
fielders       9018
bowler        13229
dtype: int64
```

```
batting_data
```

	Match ID	innings	team	batsman	runs	balls	fours	sixes	strikeRate
0	361657	1	Zimbabwe	10423.0	53.0	38.0	5.0	3.0	139.47
1	361657	1	Zimbabwe	49282.0	12.0	22.0	2.0	0.0	54.54
2	361657	1	Zimbabwe	47619.0	8.0	26.0	0.0	0.0	30.76
3	361657	1	Zimbabwe	10421.0	4.0	6.0	1.0	0.0	66.66
4	361660	1	Pakistan	11647.0	44.0	41.0	4.0	1.0	107.31
...	...	...	...	...	...	...	...	...	...
52409	1412220	2	Botswana	55147.0	0.0	1.0	0.0	0.0	0.00
52411	1412220	2	Botswana	68480.0	7.0	16.0	0.0	0.0	43.75
52413	1412220	2	Botswana	112124.0	1.0	1.0	0.0	0.0	100.00
52415	1412220	2	Botswana	71147.0	1.0	5.0	0.0	0.0	20.00
52417	1412220	2	Botswana	98613.0	1.0	2.0	0.0	0.0	50.00

Next steps: Generate code with batting_data View recommended plots

```
batting_data.rename(columns= {'Match ID' : 'match_id'}, inplace=True)
batting_data
```

	match_id	innings	team	batsman	runs	balls	fours	sixes	strikeRat
0	361657	1	Zimbabwe	10423.0	53.0	38.0	5.0	3.0	139.4
1	361657	1	Zimbabwe	49282.0	12.0	22.0	2.0	0.0	54.5
2	361657	1	Zimbabwe	47619.0	8.0	26.0	0.0	0.0	30.7
3	361657	1	Zimbabwe	10421.0	4.0	6.0	1.0	0.0	66.6
4	361660	1	Pakistan	11647.0	44.0	41.0	4.0	1.0	107.3
...	...	...	...	...	...	...	...	...	..
52409	1412220	2	Botswana	55147.0	0.0	1.0	0.0	0.0	0.0
52411	1412220	2	Botswana	68480.0	7.0	16.0	0.0	0.0	43.7
52413	1412220	2	Botswana	112124.0	1.0	1.0	0.0	0.0	100.0
52415	1412220	2	Botswana	71147.0	1.0	5.0	0.0	0.0	20.0
52417	1412220	2	Botswana	98613.0	1.0	2.0	0.0	0.0	50.0

38877 rows × 13 columns

Next steps: Generate code with batting_data View recommended plots

Understanding & Cleaning bowling stats data

```
bowling_data
```

	Match ID	innings	team	opposition	bowler id	overs	balls	maidens	conc
0	300436	1	New Zealand	England	10303.0	4.0	24.0	0.0	
1	300436	1	New Zealand	England	9711.0	4.0	24.0	0.0	
2	300436	1	New Zealand	England	10325.0	3.0	18.0	0.0	
3	300436	1	New Zealand	England	49108.0	4.0	24.0	0.0	
4	300436	1	New Zealand	England	9570.0	3.0	18.0	0.0	
...	...	...	...	...	...	...	...	...	
28476	1412220	2	Kenya	Botswana	63485.0	3.0	18.0	1.0	
28477	1412220	2	Kenya	Botswana	68249.0	4.0	24.0	0.0	
28478	1412220	2	Kenya	Botswana	53203.0	4.0	24.0	0.0	

Next steps: Generate code with `bowling_data` View recommended plots

```
bowling_data.sample(20)
```

	Match ID	innings	team	opposition	bowler id	overs	balls	maidens
2518	523736	2	Pakistan	Zimbabwe	45170.0	1.0	6.0	0.0
27253	1400984	1	Nigeria	Ghana	102141.0	2.0	12.0	0.0
21152	1326826	1	Mozambique	Eswatini	103111.0	1.3	9.0	0.0
12018	1202010	2	Malaysia	Vanuatu	71052.0	3.1	19.0	0.0
19196	1317138	2	Serbia	Bulgaria	112445.0	1.5	11.0	0.0
12780	1217741	1	Malaysia	Thailand	71041.0	4.0	24.0	0.0
19656	1320975	2	Israel	Spain	112439.0	2.0	12.0	0.0
9849	1187681	1	New Zealand	India	48928.0	4.0	24.0	0.0
27524	1404392	2	Namibia	Zimbabwe	63683.0	3.0	18.0	0.0
20105	1321281	2	Austria	Luxembourg	109665.0	1.0	6.0	0.0
5579	919603	2	Pakistan	Zimbabwe	24723.0	4.0	24.0	0.0
7387	1115807	1	New Zealand	Pakistan	67586.0	4.0	24.0	0.0
23624	1349388	1	Singapore	Malaysia	94203.0	4.0	24.0	0.0
19755	1320985	2	Israel	Hungary	93967.0	1.0	6.0	0.0
14554	1272096	1	Bangladesh	New Zealand	54674.0	3.3	21.0	0.0
10193	1192816	1	Samoa	Vanuatu	105861.0	4.0	24.0	0.0
5188	875521	2	Hong Kong	Ireland	66839.0	4.0	24.0	0.0
17821	1298174	2	Netherlands	South Africa	97878.0	4.0	24.0	0.0

```
# Checking null values
bowling_data.isna().sum()
```

```
Match ID        0
innings         0
team           26
opposition     26
bowler id      26
```

```
    overs           26
    balls           26
    maidens         26
    conceded        26
    wickets         26
    economy         26
    dots          2277
    fours         2277
    sixes         2277
    wides           26
    noballs         26
    dtype: int64
```

```
# Verying Null rows
bowling_data[bowling_data['team'].isna()]
```

	Match ID	innings	team	opposition	bowler id	overs	balls	maidens	conceded
369	287859	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
370	287859	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
881	354461	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
882	354461	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5238	876465	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5239	876465	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5262	876467	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5263	876467	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5845	951317	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5846	951317	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9433	1185189	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9434	1185189	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12083	1202242	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12084	1202242	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15406	1275041	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15407	1275041	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22636	1343759	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22637	1343759	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23475	1349141	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23476	1349141	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25207	1381459	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25208	1381459	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25222	1381460	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25224	1381460	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
26055	1391336	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
# Dropping these rows
bowling_data.dropna(subset="team", inplace=True)
```

```
bowling_data.isna().sum()
```

```
    Match ID         0
    innings          0
    team             0
    opposition       0
    bowler id        0
    overs            0
    balls            0
    maidens          0
    conceded         0
    wickets          0
```

```
economy          0
dots          2251
fours         2251
sixes         2251
wides            0
noballs          0
dtype: int64
```

```
bowling_data[bowling_data['dots'].isna()]

# No need to drop these
```

	Match ID	innings	team	opposition	bowler id	overs	balls	maidens	cor
239	264065	1	Zimbabwe	Bangladesh	4177.0	4.0	24.0	0.0	
240	264065	1	Zimbabwe	Bangladesh	46653.0	3.0	18.0	0.0	
241	264065	1	Zimbabwe	Bangladesh	45252.0	2.0	12.0	0.0	
242	264065	1	Zimbabwe	Bangladesh	45326.0	4.0	24.0	0.0	
243	264065	1	Zimbabwe	Bangladesh	47206.0	4.0	24.0	0.0	
...	...	...	...	...	...	...	...	...	
28476	1412220	2	Kenya	Botswana	63485.0	3.0	18.0	1.0	
28477	1412220	2	Kenya	Botswana	68249.0	4.0	24.0	0.0	
28478	1412220	2	Kenya	Botswana	53203.0	4.0	24.0	0.0	
28479	1412220	2	Kenya	Botswana	113058.0	4.0	24.0	0.0	
28480	1412220	2	Kenya	Botswana	21463.0	4.0	24.0	0.0	

```
bowling_data.rename(columns= {'Match ID' : 'match_id', 'bowler id' : 'bowler_id'}, inplace=True)
bowling_data
```

	match_id	innings	team	opposition	bowler_id	overs	balls	maidens	co
0	300436	1	New Zealand	England	10303.0	4.0	24.0	0.0	
1	300436	1	New Zealand	England	9711.0	4.0	24.0	0.0	
2	300436	1	New Zealand	England	10325.0	3.0	18.0	0.0	
3	300436	1	New Zealand	England	49108.0	4.0	24.0	0.0	
4	300436	1	New Zealand	England	9570.0	3.0	18.0	0.0	
...	...	...	...	...	...	...	...	...	
28476	1412220	2	Kenya	Botswana	63485.0	3.0	18.0	1.0	
28477	1412220	2	Kenya	Botswana	68249.0	4.0	24.0	0.0	
28478	1412220	2	Kenya	Botswana	53203.0	4.0	24.0	0.0	

Next steps: Generate code with bowling_data   View recommended plots

Performing Exploratory Data Analysis(EDA) for finding useful insights and answering analytical questions.

Which country has hosted the most number of matches?

```
matches_data.head(2)
```

	t20i_match_no	match_id	match_name	series_id	series_name	match_date	match_
10	1	211048	New Zealand Vs Australia Only T20I	238218	Australia tour of New Zealand - 2005 (2004/05)	NaT	
11	2	211028	England Vs Australia Only T20I	238219	Australia tour of England and Scotland - 2005...	NaT	

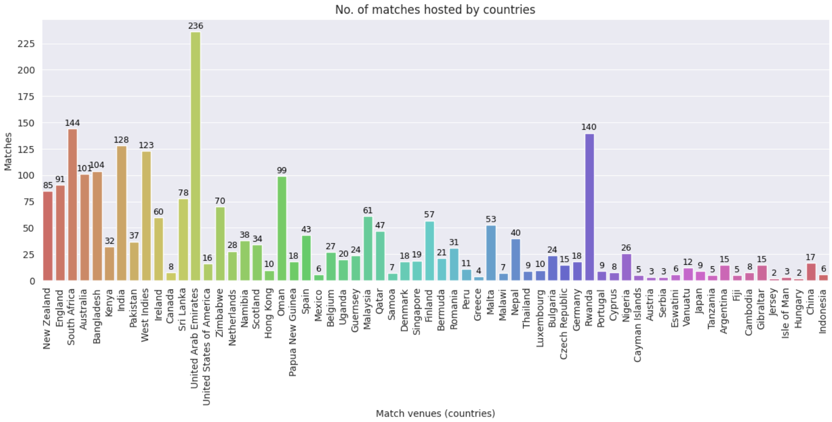
2 rows × 30 columns

```
matches_data['match_venue_(country)'].value_counts().nlargest(10)
```

```
United Arab Emirates    236
South Africa            144
Rwanda                  140
India                   128
West Indies             123
Bangladesh              104
Australia               101
Oman                     99
England                  91
New Zealand              85
Name: match_venue_(country), dtype: int64
```

```
# Setting figure size
plt.figure( figsize = (15,5) )

# Setting figure style or theme
sns.set_style('darkgrid')

# Plotting chart
ax = sns.countplot(data=matches_data, x='match_venue_(country)', palette='hls')

# Setting labels with each bar count
for container in ax.containers:
    ax.bar_label(container, label_type="edge", padding=1, size=9, color="black")

# Customizations
plt.tick_params('x', rotation=90)
plt.xlabel("Match venues (countries)")
plt.ylabel("Matches")
plt.title("No. of matches hosted by countries")
plt.yticks([0,25,50,75,100,125,150,175,200,225])

# Show chart
plt.show()
```

```
<ipython-input-47-563f56ba76a3>:8: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.

  ax = sns.countplot(data=matches_data, x='match_venue_(country)', palette='hls')
```



From the chart, we can see that United Arab Emirates has hosted the most number of T20I matches till now i.e 236

Top 10 players with most man-of-the-match awards

```
matches_data.head(2)
```

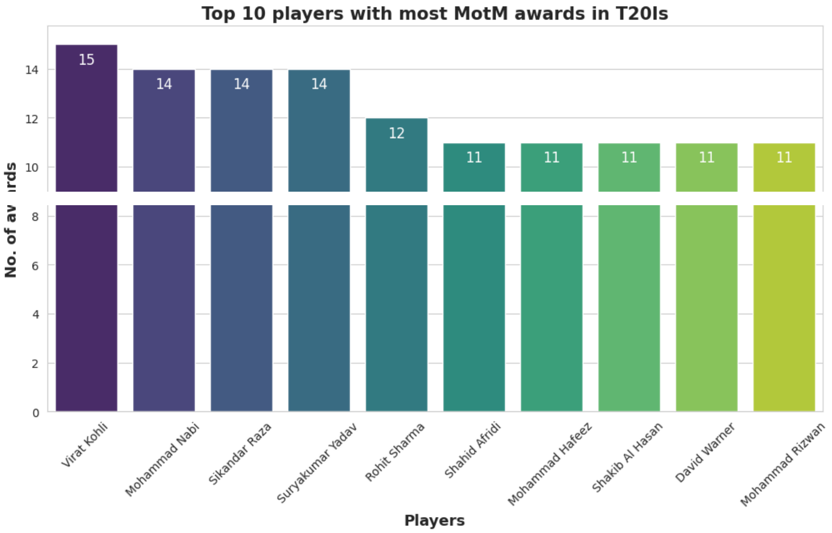
	t20i_match_no	match_id	match_name	series_id	series_name	match_date	match
10	1	211048	New Zealand Vs Australia Only T20I	238218	Australia tour of New Zealand - 2005 (2004/05)	NaT	
11	2	211028	England Vs Australia Only T20I	238219	Australia tour of England and Scotland - 2005...	NaT	

2 rows × 30 columns

```
# Grouping the dataframe by 'mom_player' columns and aggregating the rows by counting the no. of awards
motm_players = matches_data.groupby('mom_player')[['match_id']].count().rename(columns={'match_id' : 'awards'}

# Merging the result dataframe with player info on each unique player id
motm_with_names = motm_players.merge(players_info, left_index=True, right_on='player_id')

# Setting size and theme
plt.figure(figsize=(12,6))
sns.set_style('whitegrid')

# Plotting by 10 largest no. of awards
ax = sns.barplot(x='player_name', y='awards', data=motm_with_names.nlargest(10,'awards'), palette='viridis')
```

```
# Customizations
ax.tick_params('x', rotation=45)
ax.set_title("Top 10 players with most MotM awards in T20Is", fontweight = 'bold', fontsize = 15)
ax.set_xlabel('Players', fontweight = 'bold', fontsize = 13)
ax.set_ylabel('No. of awards', fontweight = 'bold', fontsize = 13)

# Setting bar label on each container
for container in ax.containers:
```

```
<ipython-input-49-53cd16561b5c>:12: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.

  ax = sns.barplot(x='player_name', y='awards', data=motm_with_names.nlargest(10,'
```



Virat Kholi has won the most Man-of-the-Match awards followed by Muhammad Nabi and Sikandar Raza

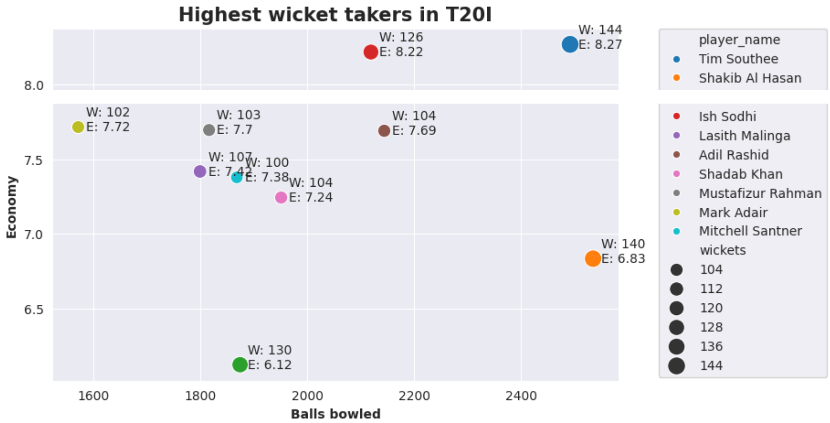
Highest wicket takers in T20 Internationals

```
bowling_data.head()
```

	match_id	innings	team	opposition	bowler_id	overs	balls	maidens	conced
0	300436	1	New Zealand	England	10303.0	4.0	24.0	0.0	3
1	300436	1	New Zealand	England	9711.0	4.0	24.0	0.0	4
2	300436	1	New Zealand	England	10325.0	3.0	18.0	0.0	4

Next steps: Generate code with bowling_data   View recommended plots

```
# Grouping data by bowler ids and aggregating balls, wickets and economy
bowlers_record = bowling_data.groupby('bowler_id')[['balls', 'wickets', 'economy']].agg(
    {'balls' : 'sum',
     'wickets' : 'sum',
     'economy' : 'mean'})

# Extracting top 10 bowlers by wickets
top_10_bowlers = bowlers_record.nlargest(10, 'wickets')

# Merging dataframe with players info
top_10_bowlers_names = top_10_bowlers.merge(players_info, left_index=True, right_on='player_id')

# Setting size and theme
plt.figure(figsize=(8,5))
sns.set_style('darkgrid')

# Plotting with balls bowled on x-axis and economy on y-axis
ax = sns.scatterplot(x='balls', y='economy', data=top_10_bowlers_names, size='wickets', hue='player_name', sizes=(100,200)

# Customizing legend
plt.legend(bbox_to_anchor=(1.07, 1), loc='upper left', borderaxespad=0)

# Annotating points on the plot with wickets and economy
for lab,row in top_10_bowlers_names.iterrows():
    ax.annotate(f"W: { int(row['wickets']) }\nE: { round(row['economy'],2) }", xy=(row['balls']+15, row['economy']-0.025))

# Customizing the chart
ax.set_title("Highest wicket takers in T20I", fontweight = 'bold', fontsize=15)
ax.set_xlabel("Balls bowled", fontweight = 'bold')
ax.set_ylabel("Economy", fontweight = 'bold')

plt.show()
```



- **Rashid Khan is the most economical bowler with 130 wickets**
- **Tim Southee has the highest wickets uptill now but also with a high economy**

Highest run scorers in T20 Internationals

```
batting_data.head()
```

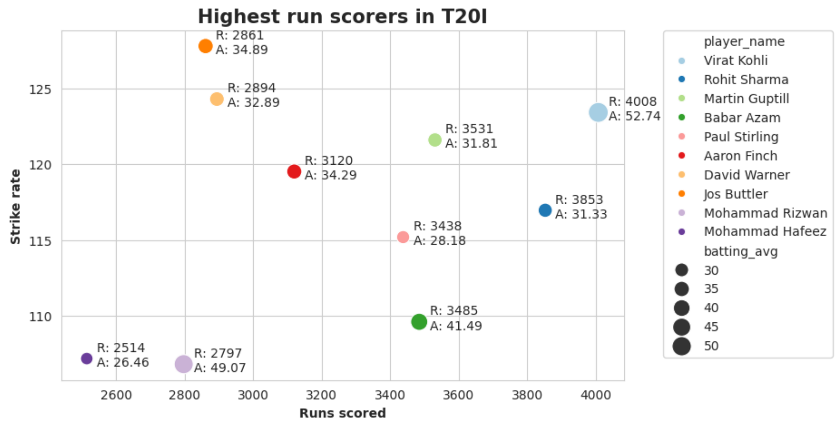
	match_id	innings	team	batsman	runs	balls	fours	sixes	strikeRate	is
0	361657	1	Zimbabwe	10423.0	53.0	38.0	5.0	3.0	139.47	T
1	361657	1	Zimbabwe	49282.0	12.0	22.0	2.0	0.0	54.54	T
2	361657	1	Zimbabwe	47619.0	8.0	26.0	0.0	0.0	30.76	T
3	361657	1	Zimbabwe	10421.0	4.0	6.0	1.0	0.0	66.66	T
4	361660	1	Pakistan	11647.0	44.0	41.0	4.0	1.0	107.31	T

Next steps: Generate code with batting_data View recommended plots

```
batsman_records = batting_data.pivot_table(values=['runs', 'balls', 'fours', 'sixes', 'strikeRate', 'isOut'], index='batsman',
                         aggfunc={'runs': 'sum', 'balls' : 'sum', 'fours' : 'sum', 'sixes' : 'sum', 'strikeRate' : 'mean', 'isOut' : 'sum'})

top_10_batsman = batsman_records.nlargest(10,'runs').reset_index()

top_10_batsman['batting_avg'] = top_10_batsman['runs'] / top_10_batsman['isOut']

top_10_batsman_names = top_10_batsman.merge(players_info, left_on='batsman', right_on='player_id')

top_10_batsman_names
```

	batsman	balls	fours	isOut	runs	sixes	strikeRate	batting_avg	player_id
0	49752.0	2905.0	356.0	76	4008.0	117.0	123.425047	52.736842	49752
1	48405.0	2767.0	348.0	123	3853.0	182.0	116.962286	31.325203	48405
2	48927.0	2602.0	309.0	111	3531.0	173.0	121.606356	31.810811	48927
3	56880.0	2714.0	371.0	84	3485.0	53.0	109.595408	41.488095	56880
4	52631.0	2533.0	399.0	122	3438.0	123.0	115.188647	28.180328	52631
5	35812.0	2189.0	309.0	91	3120.0	125.0	119.522136	34.285714	35812

```
# Setting size and theme
plt.figure(figsize=(8,5))
sns.set_style('whitegrid')

# Plotting with balls bowled on x-axis and economy on y-axis
ax = sns.scatterplot(x='runs', y='strikeRate', data=top_10_batsman_names, size='batting_avg',
                     hue='player_name', sizes=(100,250), palette='Paired')

# Customizing legend
plt.legend(bbox_to_anchor=(1.07, 1), loc='upper left', borderaxespad=0)

# Annotating points on the plot with wickets and economy
for lab,row in top_10_batsman_names.iterrows():
    ax.annotate(f"R: { int(row['runs']) }\nA: {round(row['batting_avg'],2)}", xy=(row['runs']+30, row['strikeRate']-0.5))

# Customizing the chart
ax.set_title("Highest run scorers in T20I", fontweight = 'bold', fontsize=15)
ax.set_xlabel("Runs scored", fontweight = 'bold')
ax.set_ylabel("Strike rate", fontweight = 'bold')

plt.show()
```



- **Highest Run scorer is Virat Kohli with 4008 runs with average of 52.74**
- **Jos Buttler has the highest strike rate.**
- **Babar Azam has scored 3485 runs with average of 41.49**

Who hit most sixes?

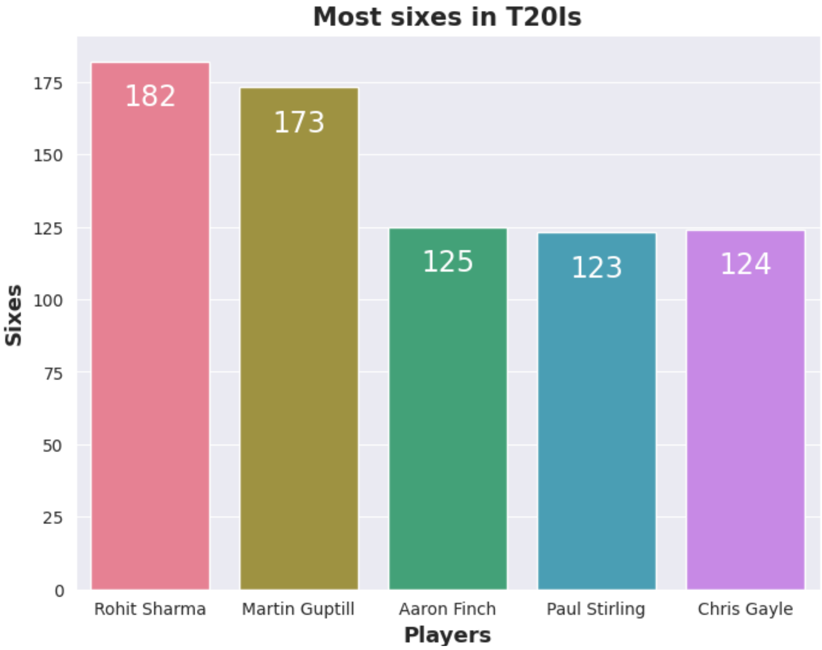
```
batsman_records.head()
```

batsman	balls	fours	isOut	runs	sixes	strikeRate
1934.0	12.0	2.0	True	10.0	0.0	83.330
1939.0	13.0	0.0	2	3.0	0.0	23.610
1979.0	8.0	0.0	True	5.0	0.0	62.500
1988.0	487.0	76.0	27	629.0	23.0	96.322
2034.0	15.0	0.0	False	11.0	1.0	73.330

Next steps: Generate code with batsman_records   View recommended plots

```
# Filter top 5 players with most sixes and merging it with player info dataframe
top_5_sixes = batsman_records.nlargest(5,'sixes')
top_5_sixes_names = top_5_sixes.merge(players_info, left_index=True, right_on='player_id')

# Setting size and theme
plt.figure(figsize=(8,6))
sns.set_style('darkgrid')

# Plotting values and setting title and labels
ax = sns.barplot(x='player_name', y='sixes', data=top_5_sixes_names.sample(5), palette='husl')
ax.set_title("Most sixes in T20Is", fontweight = 'bold', fontsize=15)
ax.set_xlabel("Players", fontweight = 'bold', fontsize=13)
ax.set_ylabel("Sixes", fontweight = 'bold', fontsize=13)

# Setting label on each bar
for container in ax.containers:
    ax.bar_label(container, padding=-30, fontsize = 17, color='white')
```

```
<ipython-input-56-c3b9fd4a1234>:10: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.

  ax = sns.barplot(x='player_name', y='sixes', data=top_5_sixes_names.sample(5), p
```



- **Rohit Sharma has hitted most sixes till now i.e 182**

Most successful teams in T20Is

```
matches_data.head(2)
```

	t20i_match_no	match_id	match_name	series_id	series_name	match_date	match