Capstone Project

Cricket Scorecard

Let’s make a score card

1. Defining the problem statement
2. T20 data analysis
3. Preparing data set for modelling
4. Applying model
5. Model validation and selection
6. Creating the score cards

The International Cricket Council (ICC) is the world governing body of cricket. It was founded as the Imperial Cricket Conference in 1909 by representatives from Australia, England and South Africa. It was renamed as the International Cricket Conference in 1965, and took up its current name in 1989. It organizes world championship events such as Cricket World Cup, Women's Cricket World Cup, ICC T20 World Cup, ICC Women's T20 World Cup, ICC Champions Trophy and Under-19 Cricket World Cup.

The ICC has 106 member nations currently:
12 Full Members that play Test matches and 94 Associate Members.
The ICC is responsible for the organization and governance of cricket's major international tournaments, most notably the Cricket World Cup and the T20 World Cup. It also appoints the umpires and referees that officiate at all sanctioned Test matches, One Day Internationals and Twenty20 Internationals.

The data consists of 1417 T20 matches. Each row corresponds to a single T20 match.

- Rename all the column names to their appropriate names, for example meta.created should be renamed as created_date
- Find out the top three venues which hosted the greatest number of matches.
- Find out the pair of cricket teams who played the most number of T20 matches against each other.
- Print the top five teams by their win percentages. Win percentage is defined as the number of matches won divided by the number of matches played and then multiplied by 100.
- Write a function to get the scorecard of each match. This function would take the innings value as argument and return two scorecard dataframes each for one team as shown below. So the first dataframe would contain the top 4 scorers of the team who batted first and the top 4 bowlers of the opponent team. And the second dataframe would contain the top 4 scorers of the team who batted second and the top 4 bowlers of the opponent team.

#	Column	Non-Null Count	Dtype
0	innings	1417 non-null	object
1	meta.data_version	1417 non-null	float64
2	meta.created	1417 non-null	object
3	meta.revision	1417 non-null	int64
4	info.dates	1417 non-null	object
5	info.gender	1417 non-null	object
6	info.match_type	1417 non-null	object
7	info.outcome.by.wickets	651 non-null	float64
8	info.outcome.winner	1372 non-null	object
9	info.overs	1417 non-null	int64
10	info.player_of_match	1255 non-null	object
11	info.teams	1417 non-null	object
12	info.toss.decision	1417 non-null	object
13	info.toss.winner	1417 non-null	object
14	info.umpires	1395 non-null	object
15	info.venue	1417 non-null	object
16	info.city	1228 non-null	object
17	info.outcome.by.runs	721 non-null	float64
18	info.match_type_number	684 non-null	float64
19	info.neutral_venue	328 non-null	float64
20	info.outcome.method	51 non-null	object
21	info.outcome.result	45 non-null	object
22	info.outcome.eliminator	16 non-null	object
23	info.supersubs.New Zealand	1 non-null	object
24	info.supersubs.South Africa	1 non-null	object
25	info.bowl_out	2 non-null	object
26	info.outcome.bowl_out	2 non-null	object

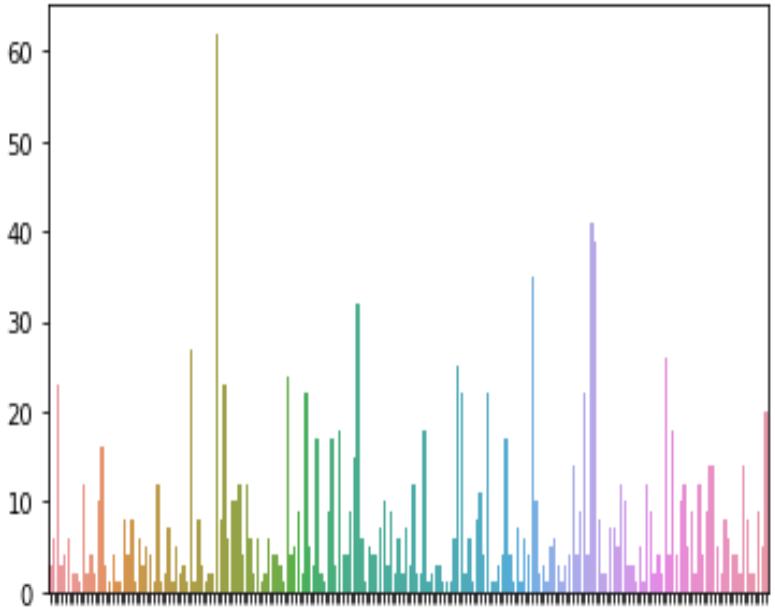
Using replace method

```
[10] df.columns

     Index(['innings', 'data_version', 'created', 'revision', 'dates', 'gender',
            'match_type', 'outcome_by_wickets', 'outcome_winner', 'overs',
            'player_of_match', 'teams', 'toss_decision', 'toss_winner', 'umpires',
            'venue', 'city', 'outcome_by_runs', 'match_type_number',
            'neutral_venue', 'outcome_method', 'outcome_result',
            'outcome_eliminator', 'supersubs_New Zealand', 'supersubs_South Africa',
            'bowl_out', 'outcome_bowl_out'],
           dtype='object')
```

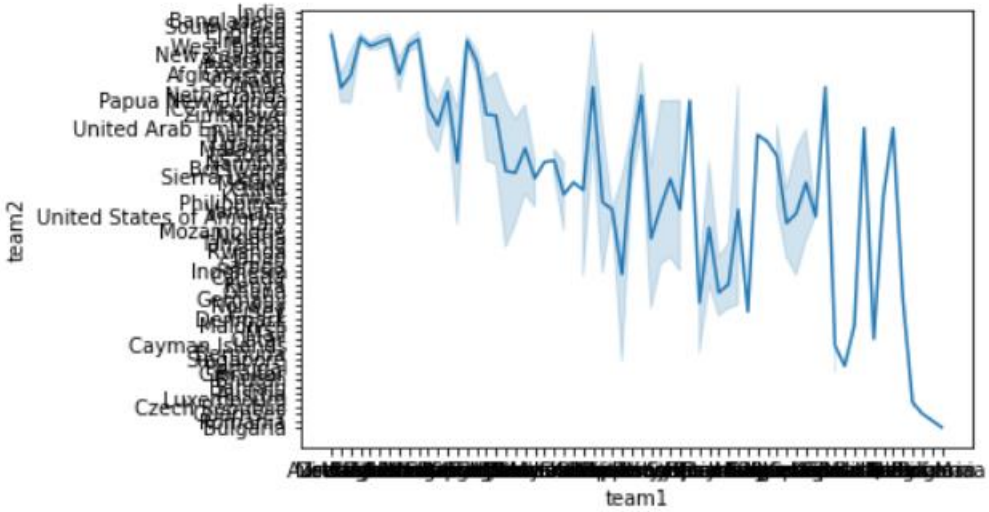
Find out the top three venues which hosted the greatest number of matches.

```
import seaborn as sns
x,y=np.unique(df['venue'],return_counts=True)
sns.barplot(x,y)
```




```
the top 3 venues are:
venue
Dubai International Cricket Stadium    62
Sheikh Zayed Stadium                   41
Shere Bangla National Stadium          39
Name: venue, dtype: int64
```

Find out the pair of cricket teams who played the most number of T20 matches against each other.



Teams that played the most matches together are Australia and England and they played a total of 45 games

Print the top five teams by their win percentages. Win percentage is defined as the number of matches won divided by the number of matches played and then multiplied by 100.

	Team	Played	Wins
0	Australia	214	132
1	Sri Lanka	182	69
2	Ireland	126	44
3	Hong Kong	49	17
4	Zimbabwe	75	19

Top Five Teams with the highest Win Percentage

	Team	Played	Wins	Win_Percentage
65	Belgium	2	2	100.00
57	Spain	6	5	83.33
38	Tanzania	3	2	66.67
40	Indonesia	3	2	66.67
70	Romania	3	2	66.67

The top 4 scorers of the team who batted first and the top 4 bowlers of the opponent team.

The top Players Scorecard 1st Inning

	Player	Runs	Balls_Played	Balls_Bowled	Runs_given	wicket	Strike_Rate	Economy
0	HM Amla	62	44	0	0	0	141.0	NaN
1	F du Plessis	36	26	0	0	0	138.0	NaN
2	JP Duminy	29	16	0	0	0	181.0	NaN
3	AB de Villiers	26	17	0	0	0	153.0	NaN
4	TA Boult	0	0	24	12	2	NaN	3.00
5	C de Grandhomme	0	0	19	22	2	NaN	6.95
6	MJ Santner	0	0	24	44	0	NaN	11.00
7	C Munro	0	0	6	11	0	NaN	11.00

The second dataframe contains the top 4 scorers of the team who batted second and the top 4 bowlers of the opponent team.

	Player	Runs	Balls_Played	Balls_Bowled	Runs_given	wicket	Strike_Rate	Economy
0	TC Bruce	33	27	0	0	0	122.0	NaN
1	TG Southee	20	6	0	0	0	333.0	NaN
2	C de Grandhomme	15	8	0	0	0	188.0	NaN
3	KS Williamson	13	14	0	0	0	93.0	NaN
4	CH Morris	0	0	18	10	2	NaN	3.33
5	AL Phehlukwayo	0	0	19	19	3	NaN	6.00
6	Imran Tahir	0	0	23	24	5	NaN	6.26
7	D Paterson	0	0	12	14	0	NaN	7.00

Thank you