NUMPY AND PANDAS AND VISUALIZATION

TOTAL MARKS:30 **DURATION: 2 HOURS**

DATA SET:

- Scorecard: Contains an index
- Team 1: Contains name of the host Team
- Team 2: Contains name of the visiting Team
- Winner: Contains name of the winning team
- Margin: Contains margin by which a team won. It is either in number of wickets or number of runs
- Ground: Contains name of the ground on which the game was played
- Match Date: Contains date on which the match was played

SECTION A: 5 MARKS

1. Read the dataset and find the total number of ODI Matches in the dataset (1 mark)

A. How many teams are unique and not repeating in the "Team 1" column. Make a list to display the same (1 mark)
B. How many teams are unique and not repeating in the "Team 2" column. Make a list to display the same (1 mark)
C. Find the Total Number of Unique Teams (from both Team 1 and Team 2 combined) using Set Operations and display the same in a list (1 mark)
D. Remove all the team names ending with 'XI' and display the filtered output to get the final list of Teams (1 mark)

SECTION B: 10 MARKS

2.1 A) Display the Records of every ODI_Matches played after removing those team names ending with "XI" which was done in the previous question – (3 marks)
2.1 B) Rename "Scorecard" to "Odi_No" – (1 mark)
2.1 C) Set "Scorecard" as index for these records – (1 mark)

2.2 A) Find the total number of matches won by each team in ascending order and make it in a data frame – (2 marks)
2.2 B) Make a new column "Team" to capture the different teams in the dataset – (1 mark)
2.2 C) Rename the "Odi_No" column "Number of Wins" – (1 mark)
2.2 D) Drop the rows - "Result" and "Not Tied" – (1 mark)

SECTION C: 15 MARKS

3.1 Find the total number of matches played by each Country as both Team 1 and Team 2. After that find the total number of matches played by each country as a whole – (5 marks)

3.2 :-
A) With the given hint, first replace all the NAN values with 0 using the below code.

--your merged dataset here--.fillna(0,inplace=True)

B) Convert the data type of the Team 2 column from float64 to int64 – (1 mark)
C) Add the values of Columns "Team 2" and "Team 1" – (1 mark)
D) Make a new column "Country" – (1 mark)
E) Sort and display the total number of matches played by each country in descending order and state your insights – (2 marks)

3.3 Find out the win percentage of each country and also make a bar plot on the same. Provide your insights on the same – (5 marks)