**Final Project - Problem Statement**

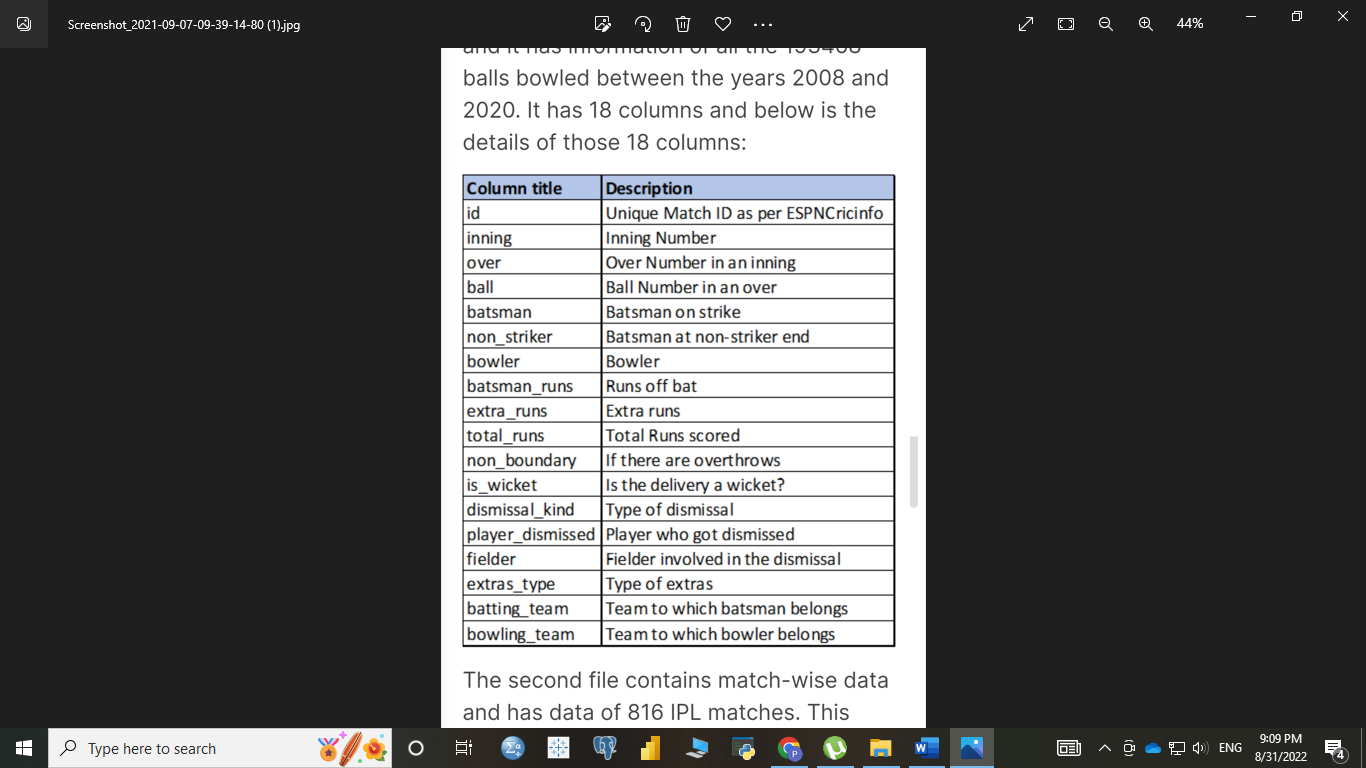
**Final Project - Sports Analytics using SQL**

In this project, you have to perform the job of a sports analyst. You are given two datasets related to IPL (Indian Premier League) cricket matches. One dataset contains ball-by-ball data and the other contains match-wise data. You have to import the datasets into an SQL database and perform the tasks given in this assignment to find important insights from this dataset.

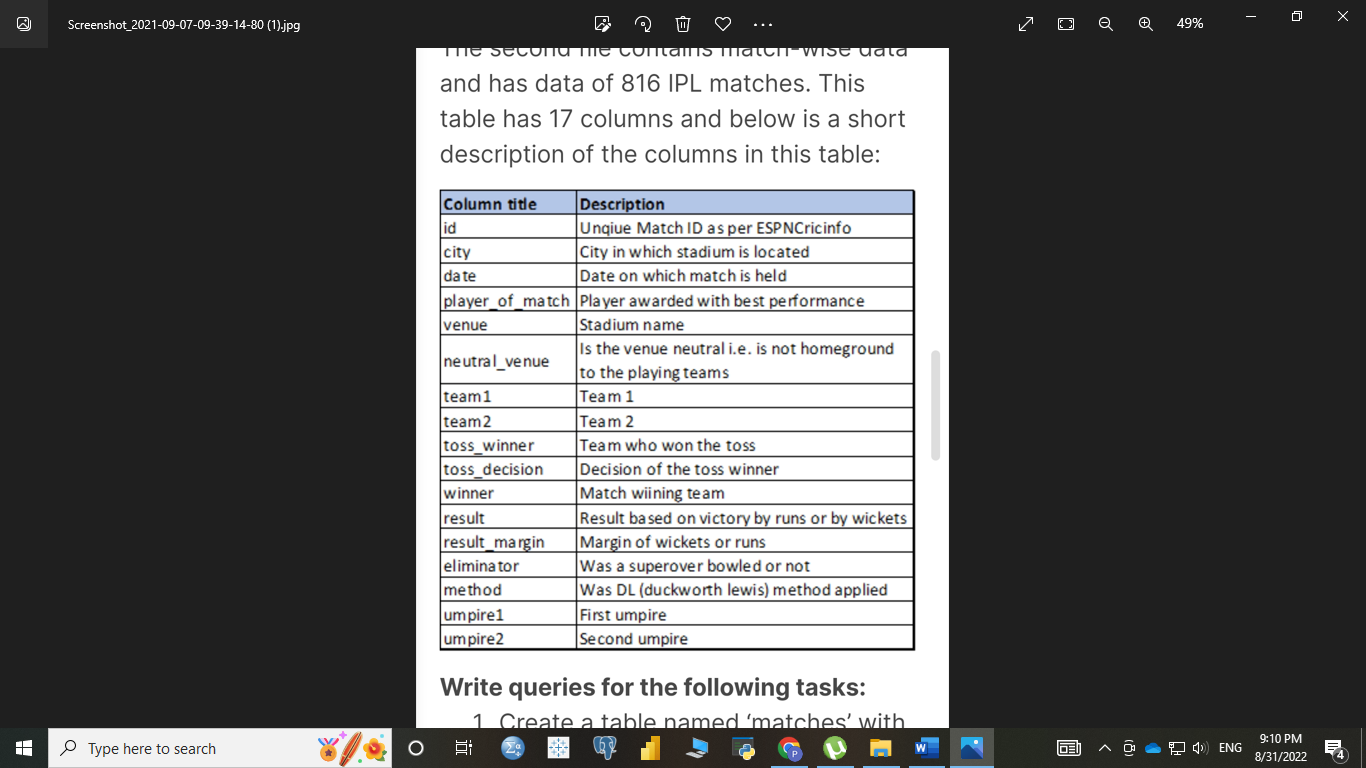
**About the Data**

Please download the datasets by clicking here and have them ready before we get started.

The first CSV file is for ball-by-ball data and it has information of all the 193468 balls bowled between the years 2008 and 2020. It has 18 columns and below are the details of those 18 columns:



The second file contains match-wise data and has data of 816 IPL matches. This table has 17 columns and below is a short description of the columns in this table:



**Write queries for the following tasks:**

1. Create a table named 'matches' with appropriate data types for columns 2. Create a table named 'deliveries' with appropriate data types for columns 3. Import data from csv file 'IPL\_matches.csv'attached in resources to 'matches'

4. Import data from csv file 'IPL\_Ball.csv' attached in resources to 'matches 5. Select the top 20 rows of the deliveries table.

6. Select the top 20 rows of the matches table.

7. Fetch data of all the matches played on 2nd May 2013.

8. Fetch data of all the matches where the margin of victory is more than 100 runs.

9. Fetch data of all the matches where the final scores of both teams tied and order it in descending order of the date.

10. Get the count of cities that have hosted an IPL match.

11. Create table deliveries\_v02 with all the columns of deliveries and an additional column ball\_result containing value boundary, dot or other depending on the total\_run (boundary for >= 4, dot for 0 and other for any other number)

12. Write a query to fetch the total number of boundaries and dot balls

13. Write a query to fetch the total number of boundaries scored by each team

14. Write a query to fetch the total number of dot balls bowled by each team

15. Write a query to fetch the total number of dismissals by dismissal kinds

16. Write a query to get the top 5 bowlers who conceded maximum extra runs

17. Write a query to create a table named deliveries\_v03 with all the columns of deliveries\_v02 table and two additional column (named venue and match\_date) of venue and date from table matches

18. Write a query to fetch the total runs scored for each venue and order it in the descending order of total runs scored.

19. Write a query to fetch the year-wise total runs scored at Eden Gardens and order it in the descending order of total runs scored.

20. Get unique team1 names from the matches table, you will notice that there are two entries for Rising Pune Supergiant one with Rising Pune Supergiant and another one with Rising Pune Supergiants. Your task is to create a matches\_corrected table with two additional columns team1\_corr and team2\_corr containing team names with replacing Rising Pune Supergiants with Rising Pune Supergiant. Now analyse these newly created columns.

21. Create a new table deliveries\_v04 with the first column as ball\_id containing information of match\_id, inning, over and ball separated by'(For ex. 335982-1-0-1 match\_idinning-over-ball) and rest of the columns same as deliveries\_v03)

22. Compare the total count of rows and total count of distinct ball\_id in deliveries\_v04;

23. Create table deliveries\_v05 with all columns of deliveries\_v04 and an additional column for row number partition over ball\_id. (HINT : row\_number() over (partition by ball\_id) as r\_num)

24. Use the r\_num created in deliveries\_v05 to identify instances where ball\_id is repeating. (HINT : select \* from deliveries\_v05 WHERE r\_num=2;)

25. Use subqueries to fetch data of all the ball\_id which are repeating. (HINT: SELECT \* FROM deliveries\_v05 WHERE ball\_id in (select BALL\_ID from deliveries\_v05 WHERE r\_num=2);