

# Assignment-1

Due Date: February 26, 2021, 11:55 PM IST

## 1 General Instructions

1. Please complete this assignment individually, on your own.
2. You will submit 2 files: query.sql and EntryNumber.pdf, corresponding to the queries and a timings report, respectively.
3. Use PostgreSQL 13 for your homework. See this link for instructions on how to download and install it on your OS. The .sql files are run automatically using the psql command using the \i option, so please ensure that there are no syntax errors in the file. If we are unable to run your file, you get an automatic reduction to 0 marks. To understand how to run many queries at once from text file, a dummy query file example.sql is available. To run example.sql in PostgreSQL, type the following command in the terminal:  
sudo -u postgres psql dbname  
\i /address/to/example.sql  
This command will run all the queries listed in example.sql at once.
4. The format of the file should be as follows. One line should identify the query number (note the two hyphens before and after the query number), followed by the actual, syntactically correct SQL query. Leave a blank line after each query.  
--1- -  
SQL QUERY  
--2- -  
SQL QUERY  
--3- -  
SQL QUERY
5. All of the queries below require an 'ORDER BY' clause. If you made an error in this clause, your answer will be evaluated as incorrect and zero marks will be awarded.
6. No changes are allowed in the i) data, ii) attribute names, iii) table names
7. The .pdf file should contain a bar graph. The graph should report the timings for each query of the dataset (X-axis legend is the query number, Y-axis legend is the time taken). You will need to figure out how to measure the timings.
8. The submission will be done on Moodle.

9. If unspecified, order the result in ascending order by column 1, then column 2 .. etc. In case of any doubts please ask on Piazza. The instructors ordering will be final and no queries will be entertained on the same.
10. In case any query or sub-query results in NULL values, please discard them before proceeding for any further steps. This is extremely important for some queries which ask us to list TOP n results etc, since you might get erroneous results.
11. For all queries leading to floating point numbers, round off the number to two decimal places right when the number is computed (and not just at the end). Even a 0.01 error will result in a wrong answer.

## 2 Dataset

### 2.1 Instructions

1. In this assignment you will analyze IPL data from the years 2008-2016. The analysis has to be done in postgres. We are providing you cleaned up data and you can download it from this link. The zip file contains a comma- seperated (,) file for each table described below.(Note the order of values in file is same as attributes of table given in next bullet point). You can load the table into database from csv file using the command  
copy Table-Name from '/path/to/file/table-name.csv' DELIMITER ',' CSV HEADER;
2. The database will include following fifteen tables and you should use only these tables while writing solution of the queries. All red coloured values are the primary keys across all tables. Note - you don't have to define these tables in the submission file, these will already be present will evaluation.

(a) Table **player**

<b>player_id:bigint</b>	player_name:text	dob:timestamp
batting_hand:bigint	bowling_skill:bigint	country_id:bigint

(b) Table **match**

<b>match_id:bigint</b>	team_1:bigint	team_2:bigint
match_date:timestamp	season_id:bigint	win_id:bigint
win_margin:bigint	outcome_id:bigint	match_winner:bigint
man_of_the_match:bigint		

(c) Table **player\_match**

<b>match_id:bigint</b>	<b>player_id:bigint</b>	role_id:bigint
team_id:bigint		

(d) Table **ball\_by\_ball**

<b>match_id:bigint</b>	<b>over_id:bigint</b>	<b>ball_id:bigint</b>
<b>innings_no:bigint</b>	team_batting:bigint	team_bowling:bigint
striker_batting_position:bigint	striker:bigint	non_striker:bigint
bowler:bigint		

(e) Table **batsman\_scored**

match_id:bigint	over_id:bigint	ball_id:bigint
runs_scored:bigint	innings_no:bigint	

(f) Table **wicket\_taken**

match_id:bigint	over_id:bigint	ball_id:bigint
player_out:bigint	kind_out:bigint	fielders:bigint
innings_no:bigint		

(g) Table **season**

season_id:bigint	man_of_the_series:bigint	orange_cap:bigint
purple_cap:bigint	season_year:bigint	

(h) Table **win\_by**

win_id:bigint	win_type:text
---------------	---------------

(i) Table **team**

team_id:bigint	team_name:text
----------------	----------------

(j) Table **role**

role_id:bigint	role_desc:text
----------------	----------------

(k) Table **country**

country_id:bigint	country_name:text
-------------------	-------------------

(l) Table **outcome**

outcome_id:bigint	outcome_type:text
-------------------	-------------------

(m) Table **out\_type**

out_id:bigint	out_name:text
---------------	---------------

(n) Table **bowling\_style**

bowling_id:bigint	bowling_skill:text
-------------------	--------------------

(o) Table **batting\_style**

batting_id:bigint	batting_hand:text
-------------------	-------------------

## 2.2 Points for help:

- Every match has two Innings.
- Each team has 11 players in a match.
- In a match, a team is said to be chasing if that team is batting in the second innings.
- A wicket is considered to be taken by the bowler if and only if the out type is one of the following: Caught, Bowled, LBW, Stumped, Caught and Bowled, Hit Wicket.
- Purple cap is awarded to player who took most number of wickets in that season.
- Orange cap is awarded to player who is the leading run-scorer of that season.
- 4, 6 runs are considered as boundaries.
- A fielder will be associated with a fall of wicket when the wicket type is as follows: Caught, Run out, Stumped.
- Batting average of a player is given by total runs scored by the batsman divided by number of matches played.
- A bowler B (say) is said to be conceded X runs (say) if batsmen have scored X runs with B as the bowler.

## 2.3 Queries

1. Return the bowlers that took 5 or more wickets in a single match. sort in descending order of num\_wickets. Break ties by ascending order by player\_name and then team\_name. Columns: match\_id, player\_name, team\_name, num\_wickets.
2. Top 3 (all of them if count is less than 3) players who won most Man-of-the-matches being in a losing team. Sort in descending order of num\_matches. Break ties with ascending (lexicographical) order of player names. Columns: Player\_name, num\_matches.
3. Return the player who took most number of catches (as a fielder) in the year 2012. Break ties by ascending (lexicographical) order of player names. Columns: player\_name.
4. Return number of matches played by the purple cap player in that respective season. Sort in ascending order of season year. Columns: season\_year, player\_name, num\_matches.
5. Return the players who scored more than 50 runs in the matches their team lost. Sort in ascending (lexicographical) of player names. Columns: player\_name.
6. Return top 5 teams with most left handed foreign batsmen in each season. Sort in ascending order of season year. Break ties of number of such batsmen by ascending (lexicographical) order of team names in a season. Columns: season\_year, team\_name, rank.
7. Return the teams in the order of maximum match wins in the season 2009. Break ties by ascending (lexicographical) order of team names. Columns: team\_name  
Note: consider the match\_winner column of table match to decide the winner and take care of null values.

8. Return the top run scorer of each team in the year, 2010. Sort in the ascending order of team names. Break ties between players in the ascending order of player names. Columns: team\_name, player\_name, runs.
9. Return top 3 teams with maximum number of sixes (runs\_scored is 6 in batsman\_scored) in a single innings in the season 2008. Break ties by ascending order of team names. Columns: team\_name, opponent\_team\_name, number of sixes
10. Return players with maximum batting average who took more wickets than an average bowler in each bowling category in all the seasons (Consider all matches of all seasons). Sort in ascending order of bowling\_skill. Break ties in ascending order of player\_name. Columns: bowling\_category, player\_name, batting\_average.
11. Return all the left handed batsmen who scored 150 or more runs and took 5 or more wickets and played 10 or more matches in a season. Sort in descending order of number of wickets, descending order of runs in the season. Break ties by ascending order of player names. Columns: season\_year, player\_name, num\_wickets, runs.
12. Find the season, match id, player name, team name and number of wickets where the highest number of wickets taken by a player in a match. Sort in descending order of number of wickets. Break ties by ascending order of player name, ascending order of match\_id. Columns: match\_id, player\_name, team\_name, num\_wickets, season\_year
13. Return all the players who played in all the seasons. Sort in order of ascending order of player names. Columns: player\_name.
14. Return top 3 teams for each season based on number of batsmen with a score of 50 or more in a match they won. Sort in ascending order of season year and rank the teams in descending order of number of batsmen with a score of 50+. Break ties by ascending order of team names. Columns: season\_year, match\_id, team\_name
15. Return Players with second highest runs, second highest wickets along with the number of runs and wickets for each season. Sort in ascending order of season year. Break ties by ascending order of player names for both batsmen and bowlers while ranking. Columns: season\_year, top\_batsman, max\_runs, top\_bowler, max\_wickets
16. Find all teams against which 'Royal Challengers Bangalore' lost a match in 2008. Sort in descending order of number of matches won against 'Royal Challengers Bangalore' in 2008. Break ties by ascending order of team name. Columns: team\_name.
17. For each team, return the player who has been awarded man of the match maximum number of times. Sort in order of ascending order of team names. Break ties by ascending order of player name while ranking. Columns: team\_name, player\_name, count.
18. Return top 5 players who played in 3 or more teams and have conceded more than 20 runs in an over for the most number of times. Break ties by ascending order of player names. Columns: player\_name.
19. Return average runs of each team in season 2010, rounded off to 2 decimals. (Ignore the extras). Sort in ascending order of team name. Columns: team\_name, avg\_runs.
20. Return top 10 players who got out in the first over (of the match) for most number of times. Break ties by ascending order of player names. Columns: player\_names.

21. Return top 3 matches where team wins by chasing with least number of boundaries. Sort in ascending order of number of boundaries. Break ties by ascending order of match\_winner team name, team\_1 name, team\_2 name. Columns: Match\_id, team\_1\_name, team\_2\_name, match\_winner\_name, number\_of\_boundaries.
22. Return the countries of top 3 players with lowest average runs conceded per number of wickets taken over all matches. Break ties by ascending order of player name. Discard players with 0 total wickets. Columns: country\_name.