Hive SQL Queries

Question 8: Calculate the average score of the away team, and the home team for every season. Show results by ordering by the season.

Hive Query

-- Seasonwise average score for both, home and away teams

select season,

avg(home_team_id) Average_score_away,

avg(home_goals) Average_score_home

from game

group by season

sort by season;

-- End of the query

Result

season	average_score_away	average_score_home
20122013	16.588089330024815	2.8647642679900074
20132014	16.804232804232804	2.88435374149965987
20142015	17.589082638362395	2.83775587566633814
20152016	17.649507948523844	2.805450416351249
20162017	17.538344722854973	2.915717539863326
20172018	19.206642066420663	3.1202952029520294

Question 9: For every season, find the highest scoring home team. Include the score in your answer.

Hive Query

-- List all the team ids which scored highest goals in a season

create temporary table if not exists season_go_max as

select season,

max(home_goals) goals_max

from game

group by season;

select game.season,game.home_team_id,game.home_goals

from game

join season_go_max

on game.season=season_go_max.season and game.home_goals=season_go_max.goals_max

-- End of the query

Result

game.season	game.home_team_id	game.home_goals
20122013	5	8
20122013	14	8
20132014	24	9
20132014	28	9
20122013	17	8
20142015	52	8
20142015	26	8
20142015	5	8
20152016	2	8
20172018	8	10
20162017	8	10
20152016	24	8
20162017	29	10
20172018	16	10

Question 10: Provide team_id and name for the team that played as home team at TD Garden.

Hive Query

-- Team id and Team name which played at TD Garden location

select team_id,teamname from team_info

where team_info.team_id in (select distinct home_team_id from game

                                                    where venue='TD Garden')

-- End of the query

Result

team_id	teamname
6	Bruins

Question 11: Create a new table that lists all games that happened at TD Garden or Madison Square Garden. Add a column that summarizes away and home goals.

Hive Query

-- Summarize goals of home and away team which played at TD Garden or Medison Square Garden

create table if not exists mie_mtrivedi_goals as

select *, concat('Away Goals + Home Goals: ',away_goals+home_goals) as goal_summary

from game

where venue='TD Garden' or venue='Madison Square Garden';

select * from mie_mtrivedi_goals

-- End of the query

Result

mie_mtrivedi_goals.venue_time_zone_id	mie_mtrivedi_goals.venue_time_zone_offset	mie_mtrivedi_goals.venue_time_zone_tz	mie_mtrivedi_goals.goal_summary
America/New_York	-4	EDT	Away Goals + Home Goals: 5
America/New_York	-4	EDT	Away Goals + Home Goals: 7
America/New_York	-4	EDT	Away Goals + Home Goals: 3
America/New_York	-4	EDT	Away Goals + Home Goals: 7
America/New_York	-4	EDT	Away Goals + Home Goals: 4
America/New_York	-4	EDT	Away Goals + Home Goals: 3
America/New_York	-4	EDT	Away Goals + Home Goals: 1
America/New_York	-4	EDT	Away Goals + Home Goals: 7
America/New_York	-4	EDT	Away Goals + Home Goals: 7
America/New_York	-4	EDT	Away Goals + Home Goals: 1
America/New_York	-4	EDT	Away Goals + Home Goals: 4

Question 12: Use subquery and count unique team names of teams that played as away team at TD Garden center and scored > 6. Provide the code and the answer.

Hive Query

-- Use of subquery and counting unique team names of teams that played as away teams at TD Garden center and scored >6

select count(distinct(team_id)) as Distinct_count_teams from team_info

where team_id in (select away_team_id from game

                  where away_goals>6 and venue='TD Garden')

-- End of the query

Result

distinct_count_teams
2