Assignment 1

Due date: January 25, 2022, 11:55pm IST

General Instructions

1. Please complete this assignment *individually*.

2. You will submit 2 files: EntryNumber.sql and EntryNumber.pdf, corresponding to the dataset and a timings report, respectively.

3. Use PostgreSQL 14 for your homework. See `https://www.postgresql.org/download/` 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 here. To run example.sql in postgres, type the following command in the terminal:

    sudo -u postgres psql dbname

    and then type

    `\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. Errors with respect to equality and inequality conditions will also be evaluated as incorrect and zero marks will be awarded.

7. Assume set semantics, unless stated otherwise.

8. No changes are allowed in the i) data, ii) attribute names, iii) table names

9. 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.

10. The submission will be done on Moodle. Details regarding submission and auto evaluation will be updated on piazza.

1 Dataset

1.1 Description

The dataset in the assignment consists of all information on the Formula 1 races, drivers, constructors, qualifying, circuits, lap times, pit stops, championships from 1950 till the latest 2021 season. The teams in Formula 1 are referred to as constructors. The constructors in Formula 1 are responsible for the research, development, and construction of the Formula 1 car chassis and engine. Two drivers from each constructor take part in a Formula One race. Before each F1 race, qualifying is held. The position at the end of qualifying determines the starting position of the driver in the race. Points are allotted to drivers based on their final positions after the race. For a constructor, points mean the sum of the points of both its drivers. There is one Formula one season held per year. In a year there are multiple races/rounds. The driver with the most points at the end of the season wins the Drivers championship. The constructor with the most points at the end of the season wins the constructor championship.

1.2 Schema

The schema of the dataset can be found here.

1.3 Instructions

1. In this assignment you will analyze Formula 1 data from the years 1950-2021. 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 five tables and you should use only these tables while writing solution of the queries. You can create temporary tables while handling any SQL query but you should include SQL queries for creating and deleting these temporary tables. Note - you don't have to define these tables in the submission file, these will already be present will evaluation.Column name and explanation are only given only for a subset of columns which are ambiguous,other columns are self-explanatory.

    (a) Circuits: This table has details about various circuits/tracks on which the Formula one races takes place. The columns are self-explanatory.

    (b) Drivers: The table has details about the drivers in Formula One. The columns are self-explanatory.

    (c) Constructors: Constructors are people or corporate entities which design key parts of Formula One cars. The Constructor table has details about the constructors in Formula One. The columns are self-explanatory.

    (d) Status: This table has information about the status of the result for drivers in a particular race.

    (e) Qualifying: A qualifying session is held before each race to determine the order cars will be lined up in at the start of the race, with the fastest qualifier starting at the front and the slowest at the back. This table indicates the qualifying position of a driver for a race.

    (f) Pitstops: In F1, a pit stop is a pause for refueling, new tires, repairs, mechanical adjustments, etc. This table has information about the pit stops taken by drivers during a race. Value for the stop column will be 1 for the first stop, 2 for the second stop in a race and so on.

    (g) Laptimes: This table gives information about the lap details for drivers in a particular race.

Column Name	Explanation
stop	It identifies the current stop taken by a driver in a race.
lap	It identifies the lap at which the stop was taken by the driver
milliseconds	Denotes the duration of the stop in ms

Column Name	Explanation
lap	Identifies the current lap in the race.
position	It gives the position held by the driver at the end of the current lap.
milliseconds	Denotes the duration of the lap in ms

Column Name	Explanation
year	Denotes the season of the formula one race.
round	Denotes the round number in a particular year for the current race.

(h) Races: This table has the details about the races held.

(i) Results: This table has information about the results of a particular race. In this table -1 denotes all the invalid/non-applicable values. Note: milliseconds is valid only for those drivers who have completed all the laps in the race, otherwise it is -1. fastestLap is valid only for drivers who have completed at least one lap, otherwise it is -1.

Column Name	Explanation
grid	starting postion of the driver in the race.
positionOrder	final position of the driver after the race.
points	points assigned to the driver for the current race
laps	number of laps completed by the drvier in the race.
milliseconds	total race time for the driver in ms.
fastestLap	The lap number for a driver in which he achieved the personal best lap time (lowest time).
rank	ordering among the drivers based on their fastest lap timing.

(j) Constructor Results: This table has information about the points scored by a constructor for a particular race.

Column Name	Explanation
points	Points scored by the constructor for a particular race.

3. Please stick to the name and the order of the columns mentioned in the order by clause specified for each question.

4. To break ties or to decide the order of the output, use the order specified in "Order By" mentioned for each query. The column name with the lower index in Order By has the higher priority.

5. In case of tie for maximum,minimum queries return all tuples that qualify.

6. Some Definitions that are used in the query.

(a) Career Points : Cumulative sum of points scored over all the races participated in by the entity(driver/constructor).

(b) Career Wins : Number of wins achieved over all the races participated in by the entity(driver/constructor).

(c) Driver Championship: For a season,a driver is said to be a championship winner if he has scored maximum cumulative points for that particular season among drivers.

(d) Constructor Championship: For a season,a constructor is said to be a championship winner if it has scored maximum cumulative points for that particular season among constructors.

1.4 Queries

1. Driver with the slowest lap (largest lap time) in Monaco in 2017. Output format: driverid, forename, surname, nationality, time. (time: duration of the slowest lap in ms) Order by: 1. forename(ascending order) 2.surname(ascending order) 3. nationality(ascending order).

2. Top 5 constructors for the 2012 season, based on the total points scored in the 2012 season. Output format: constructor_name, constructorid, nationality, points. (points: points scored by the constructor in the 2012 season) Order by: 1. Points (descending order), 2.constructor name(ascending order) 4.nationality(ascending order). 5.constructorid(ascending order)

3. Output the Driver with the maximum points scored totally for the duration 2001-2020. (including 2001 and 2020). output format: driverid, forename, surname, points. Order by: 1. forename(ascending order) 2.surname(ascending order), 3.driverid(ascending order)

4. Output the Constructor with the maximum points scored totally for the duration 2010-2020. (including 2010 and 2020). output format: constructorid, name, nationality, points. Order by: 1. name(ascending order) 2.nationality (ascending order) 3.constructorid(ascending order)

5. Output the Driver with the maximum number of career race wins. Output format: driverid, forename, surname, race_wins (number of wins). Order by: 1. Forename(ascending order) 2.surname (ascending order) 3.driverid(ascending order)

6. Output the Constructor with the all-time maximum number of race wins. Output format: constructorid, name, num_wins. Order by: 1. name (ascending order) 2.constructorid(ascending order)

7. Consider all the drivers who have not won a championship, output the top 3 drivers with the most career points. Output Format: driverid, name, points. Order by: 1. Points (descending order), 2.forename(ascending order) 3.surname (ascending order) 4.driverid(ascending order)

8. Output the driver who has won a race in the maximum number of countries. output format: driverid, forename, surname, num_countries. Order by: 1. Forename(ascending order) 2.surname (ascending order) 3.driverid(ascending order)

9. Overall (including all seasons), output the top 3 drivers who started at the 1st position on the grid (1st place in qualifying) and went on to win the race, the maximum number of times. output format: driverid, forename, surname,num_wins. Order by: 1. num_wins (descending order), 2. Forename(ascending order) 3.surname(ascending order) 4.driverid(ascending order)

10. Consider all race winners across all the seasons. Output the maximum number of pitstops taken in a race. Output format: raceid, num_stops, driverid, forename, surname, circuitId, name. (name: name of the circuit) Order by: 1. Forename(ascending order) 2.surname(ascending order), 3. Circuit_name (ascending order) 4.circuitId(ascending order) 5.driverid(ascending order)

11. Output the race that had the maximum number of collisions. Output format: raceid, name, location, num_collisions. (name: name of the circuit) Order by: 1.circuit_name (ascending order), 2.location(ascending order) 3.raceid(ascending order)

12. Output the driver who was the race winner and also had the fastest lap in that particular race for the maximum number of times over his career. Output format: driverid, forename, surname, count. (Count should denote the number of times the driver had the fastest lap in a race and was also the race winner) Order by: 1. Forename(ascending order) 2.surname (ascending order). 3.driverid(ascending order)

13. Output the maximum point difference achieved in the constructor's championship between the first and the second-placed constructors at the end of a season. Output: year, point_diff (constructor1 points - constructor2 points),constructor1_id (constructor who finished first), constructor1_name, constructor2_id (constructor who finished second), constructor2_name. Order by: 1.constructor1_name (ascending order) 2.Constructor2_name (ascending order). 3.constructor1_id(ascending order) 4.constructor2_id(ascending order)

14. Output the largest starting position (started at the farthest position on the grid) for a race winner in the 2018 season. Output format: driverid, forename, surname, ciruitId, country, pos. (position on the grid). Order by: 1.forename(descending order) 2.surname(ascending order) 3.country(ascending order). 4.driverid(ascending order) 5.ciruitId(ascending order)

15. Output the constructor who has faced the maximum number of Engine Failures (status ID: 5) since 2000(including 2000 and including 2021 season). Output format: constructorid, name, num. (num: number of engine failures) Order by: 1. name: ascending order. 3.constructorid(ascending order)

16. Output 5 american racers (in lexicographical order) who have won a race in a circuit located in the USA. Output format: driverid,forename,surname Order by: 1.forename (ascending order), 2.surname(ascending order) 3.driverid(ascending order)

17. Output the maximum number of 1-2 finishes in a race by a constructor since 2014 (including 2014). (A constructor is said to have a 1-2 finish in a race if drivers of the constructor finish at the first and second position). Output format: constructorid, name, count. (count: number of 1-2 finishes by the constructor) Order by: 1.name (ascending order) 2.constructorid(ascending order)

18. Output the driver who is the race leader in the maximum number of laps throughout his career. (race leader: has position 1 in the lap). Output: driverid, forename, surname,num_laps (number of laps as the race leader) Order by: 1.Forename(ascending order) 2.surname(ascending order) 3.driverid(ascending order)

19. Output the name of the driver who has finished the maximum number of times on the podium. (Podium finish: finished the race at 1st, 2nd or 3rd position). Output Format: driverid, forename, surname, count. (count: number of times the driver has finished on the podium). Order by: 1. Forename (ascending order), 2.suranme (descending order) 3.driverid(ascending order)

20. Output the all-time top-5 drivers with the most number of driver championship wins. Output format : driverid,forename,surname,num_champs. Order by: 1.num_champs (descending order) 2. Forename (ascending order) 3.suranme (descending order) 4.driverid(ascending order).