Lab 5: Aggregations

Due date: Friday, May 19, 10:00am .

Note: Lab 6 will be assigned on May 19 in class.

Lab Assignment

Assignment Preparation

This is an individual lab. Each student has to complete all work required in the lab, and submit all required materials **exactly as specified** in this assignment.

The assignment will involve writing SQL queries for different information needs (quesions asked in English) for each of the five course datasets.

The Task

You are to write and debug (to ensure correct output) the SQL queries that return information as requested for each of the information needs outlined below. The information needs can be addressed with either a simple SELECT statement (i.e., a SELECT statement without grouping, aggregation and nested subqueries) and/or with the UNION operation, or with a SELECT statement that uses some aggregation. Please note, that some queries in this lab may require you to immitate the intersection operation with other available means. Each information need **must be met** with a **single** SQL statement.

For this assignment, you will prepare one SQL script for each database.

NOTE: Please provide a comment in front of each SQL statement in each of your files. The simplest comment can just state the query number (e.g., `"--- Q3."`) for this particular database. This is very useful for the situations when for one reason or another you elected not to implement a query.

NOTE: This assignment does not use the MARATHON and AIRLINES datasets.

STUDENTS dataset

For the STUDENTS dataset, write an SQL script `STUDENTS-lab5.sql` containing SQL statements answering the following information requests.

1. Find all pairs of students with the same first name. Report each pair of students exactly once. Report first and last names of each of the two students, and their grades.
2. Find all first-grade students who are NOT taught by OTHA MOYER. Report their first and last names in alphabetical order by last name.
3. Report the total number of third- and fourth-graders in the school.
4. Find and report the number of students taught by `LORIA ONDERSMA`.

BAKERY dataset

Write an SQL script `BAKERY-lab5.sql` containing SQL statements answering the following information requests.

Note: Your queries must match exactly the wording of the information need. For example, if you are asked to find the price of an `Apricot Tart`, the following query

```
SELECT price
FROM goods
WHERE CODE = '90-APR-PF';
```

is considered to be incorrect because nowhere in the query was the code `'90-APR-PF'` mentioned. (This is especially important when you are expected to produce a join of two or more tables, but instead look up the foreign key value and use it verbatim in the query. Such queries will be marked as incorrect on the spot).

1. Find all customers who purchased, during the same trip to the bakery, two different `Croissant`s. Report first and last names of the customers in alphabetical order by their last name.
2. Find all days on which *either* `ALMETA DOMKOWSKI` made a purchase, *or* someone purchased a `Gongolais Cookie`. Sort dates in chronological order. Each date shall appear exactly once.
3. Report the total amount of money `NATACHA STENZ` spent at the bakery during the month of October, 2007.
4. Report the total amount of money spent by bakery customers in October 2007 on `Chocolate`-flavored items.

CARS dataset

Here are the queries for the CARS dataset. Name the SQL scrips `CARS-lab5.sql`

1. Find all cars made after 1980 with gas mileage better than the 1982 `honda civic`. Report full name of the car, year it was made and the name of the manufacturer. Sort output in descending order by gas mileage.
2. Find the average, maximum and minimum horsepower for 4-cylinder vehicles manufactured by `renault` between 1971 and 1976 inclusively.
3. Find how many cars produced in 1971 had better acceleration than a 1972 `volvo 145e (sw)`. Report just the number.
4. Find how many different car manufacturers produced a vehicle heavier than 4000 lbs.

CSU dataset

Here are the queries for the CSU dataset. Name the SQL scrips `CSU-lab5.sql`

1. Report the total number of degrees granted by `California Polytechnic State University-San Luis Obispo` in the period between 1995 and 2000 (inclusively).
2. Find the largest, the smallest and the average fee on a CSU campus in 2005.
3. Report the average student to faculty (use student FTE to faculty FTE ratio) ratio in 2004 among the campuses where 2004 enrollment (FTE numbers) was greater than 15000.
4. Report all years in which *either* there were more than 17000 students (NOT FTEs) on California Polytechnic State University-San Luis Obispo campus, or California Polytechnic State University-San Luis Obispo graduated (gave degrees) to more than 3500 students. Report years in chronological order, with each year reported once.

INN dataset

For the INN dataset, create a SQL script file `INN-lab5.sql` with SQL queries for the following information needs. (When no year is supplied in the query descriptions below, assume 2010).

1. Find all rooms that were occupied on all three of the following dates: May 15, 2010, August 18, 2010 and December 12, 2010. Report just the full name of the room and the room code. Sort output in alpabetical order by room name.

2. Find the names of all people[1] staying at the inn at the same time as `HERBERT FRYDAY`. Sort the output in alphabetical order by last name.

3. Find the number of August reservations (both checkin and checkout dates are in August) where two adults are staying with two children.

4. Find the average number of nights of stay in the `'Interim but salutary'` room for all reservations that commenced May 1, 2010 or later and ended before August 31, 2010.

5. Find how many different durations of stay for trips that commenced and ended in July of 2010 were in `Interim but salutary` room.

WINE dataset

Create a SQL script `WINE-lab5.sql` containing SQL statements representing the following information needs.

1. List all 2006 vintage wines from Napa (any appellation within the county) whose total revenue exceeds that of the 2006 `'Appelation Series'`[2] Paso Robles Zinfandel from `'Rosenblum'` winery. For each wine report grape, winery and name, score and revenue. Order by revenue.

2. Find the average score of a `Paso Robles Zinfandel`. ;

3. Find the total revenue from all red wines made by `Kosta Browne`.

4. Find the average number of cases of a `Pinor Noir` produced from grapes sourced from the `Central Coast`.

5. Report the overall number of different red wines produced in `Russian River Valley` during the year when this AVA had a wine with a score of 98.

KATZENJAMMER dataset

Note: I made a small change in two records in the `Instruments` table to fix a factual error[3] in the original dataset. This affects the results of one of the queries below. The exact commands updating the database were:

```
mysql> update Instruments
       set Instrument = 'banjo'
       where song=3 and bandmate=3;
```

[1] We only know the names of the people who made the reservations, so only those names are subject to the query.

[2] There is a typo there. Let it be for now.

[3] In case you are wondering, on `Demon Kitty Rag` Anne-Marit plays banjo, while Marianne plays bass balalaika.

```
mysql> update Instruments
       set Instrument = 'bass balalaika'
       where song=3 and bandmate=2;
```

Please be aware of this.

Create a SQL script `KATZENJAMMER-lab5.sql` containing SQL statements representing the following information needs.

1. Find the number of times Turid played bass balalaika on Katzenjammer songs.
2. Find the number of times Solveig was positioned center stage while Marianne was playing drums.
3. Find the number of times Solveig sang lead while Marianne was performing out front (left, right or center stage).
4. Find the total number of different instruments Anne-Marit played on Katzenjammer songs.
5. List all the instruments that both Marianne and Turid played on (possibly different) Katzenjammer songs.
6. Find how many different performers played guitar.
7. Find on how many songs at least two performers played the same instrument.
8. Find how many times the lead vocalist on a song played bass balalaika.

Submission Instructions

You must submit nine `<DATASET>-info.sql` files. In addition, submit a simple `README` file with your name and email address.

There is no need to submit any other files.

You must submit all your files in a single archive. Accepted formats are `gzipped tar` (`.tar.gz`) or `zip` (`.zip`). The file you are submitting must be named `lab4.ext` where `ext` is one of the extensions above.

When unpacked, your archive must place the seven `<DATASET>-lab5.sql` files into the current directory (i.e., in the root of your `handin` directory for Lab 4 submission). No other subdirectories are needed.

Submit using handin:

```
$ handin dekhtyar lab05 <file>
```