

Term Project

Phase 3: Queries and Results (due Fri Dec 3rd, by 8pm)

Now comes the payoff for all your hard work! Write SQL queries to find the answers to your questions. This will likely be an exploratory process. As you discover things, you will have follow-up questions that may take you in new directions. You don't have to stick to your original plan – this exploration is what I'd like to see.

Here's what we will be looking for from your queries:

- The queries should be relevant to your investigative questions and should at least move towards addressing them.
- I'd like to see that to some degree the result of one query led to a new question and a new query. So for 3 investigative questions, there should certainly be substantially more than 3 queries.
- The queries should make connections across tables and dig deeply into the data.
- The queries do not have to be as complex as on A2. Some might be quite simple, in fact. SQL can dig deeply without you having to do much work.

Hand in the following:

- One or more files containing the SQL queries you have written. You can decide if you want to have one big file, one file per investigative question (each with a series of queries used to explore that investigative question), or some other organization. The name of each file must begin with the string “queries”.
- A file called `demo-queries.txt` containing an example interaction with the PostgreSQL shell where you (a) load the schema and data successfully, then (b) run each query and show its result. I am providing a “runner” file that you can customize and use to quickly generate your queries demo file.
- A file called `discussion.pdf` containing a written discussion of what you learned about your domain. We are not looking for earth-shattering results; we expect your results to be modest but your discussion to be of high quality. Conclusions should be supported by the results. The discussion should be clear and well organized. Use proper grammar, spelling and punctuation, and write in the language of a formal report. Use full sentences organized into paragraphs. Make it something you'd be proud to show in an interview. Refer back to the queries using labels that you add to the queries demo file. Consider plotting your results where appropriate; you can use any tool you like for this.

You will be graded on content (how accurately and insightfully you interpret your results, and the depth of your analysis), as well as the professionalism of your writing.

Your discussion can be up to two pages. If you include figures, you may go to three pages, but don't exceed 1,000 words.

- You will also need to resubmit several files from earlier phases so that the TAs have them for reference
 - `questions.pdf`: These can be the same investigative questions that you had in phase 1 or, if your questions have evolved, describe what they are now.

-
- **schema.ddl**: Most of you have a stable schema at this point, in which case you should hand in the file from Phase 2. But if you needed to make changes to your schema, hand in your revised DDL and include comments at the top summarizing and justifying your changes.. In either case, it must be the exact schema you used to run your queries. Note: We won't be grading the design of your schema again, just that you submit it and include appropriate comments.
 - **demo-data.txt**: Your demo from phase 2 (but note the slightly different file name this time). If anything in your schema or dataset has changed, regenerate the demo – it needs to show the same database that the queries are run on.

Enjoy exploring your data!