Description

N1 Health just brought on a new client, and you are tasked with providing some high-level analysis before beginning the engagement. You have been assigned to this client's team, and your first task is to tame the customer's data and to begin to answer several questions.

Challenge

You are being asked to take multiple member rosters and combine them to get a full view of the membership. With the full membership, create summary statistics to be used for further analysis.

First, using Python and SQLite, complete the steps of combining the member rosters and creating the requested tables and summary statistics. Put this script on GitHub or other repository for review.

Next, draft an email to the project manager on the client account explaining the steps you took, any problems you found with the data, and the requested outputs. Finally create a short presentation (only a few slides) to present your work. Expect to present for about 15 minutes which can include introducing yourself, some high-level outputs from your work, and a short walkthrough of your code.

If requirements are not clear, please feel free to make any assumption needed to complete the challenge, but be sure to document that assumption in your email. Specific questions to the team are also welcome.

Data

You have been provided with a SQLite database that contains 5 member roster tables from the client and one table with model scores broken up by zip code.

Expected Final Deliverable

(1-3) should be part of the same python script, using the sqlite3 python package to run sql queries.

Please submit your script and any supporting materials to us at least 1 day prior to your presentation.

1. Create a table called `std_member_info` to aggregate and standardize the five member rosters.
   - Include all members who are eligible during April of 2022
   - Cardinality should be one row per member
2. The columns included in `std_member_info` should be:
   - `member_id`
   - `member_first_name`
   - `member_last_name`
   - `date_of_birth`

- `main_address`
- `city`
- `state`
- `zip_code`
- `payer`

3. Create a script that will calculate and display the answers to the following questions:
   - How many distinct members are eligible in April 2022?
   - How many members were included more than once?
   - What is the breakdown of members by payer?
   - How many members live in a zip code with a food_access_score lower than 2?
   - What is the average social isolation score for the members?
   - Which members live in the zip code with the highest algorex_sdoh_composite_score?

4. Draft an email to the project manager laying out the steps you took, any problems you found with the data, and the requested outputs. Remember that the project manager is non-technical.

5. Create a presentation describing your output and your approach in 5 slides or fewer. Additionally, consider the following question in your presentation:
   - Imagine receiving a roster every week from this client. How would your script need to change?