Name \_\_\_\_\_\_\_Mao, Keying Linxia Li\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_Nov, 26, 2022\_\_\_\_\_\_\_

(last name, first name)

NYU ID: \_\_\_\_\_\_ km5580 ll4764\_\_\_\_\_\_\_\_

Course Section Number: \_\_DS-GA-2433-001\_\_\_\_\_

**Project Part 2**

**Total in points** (100 points total): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Professor’s Comments:**

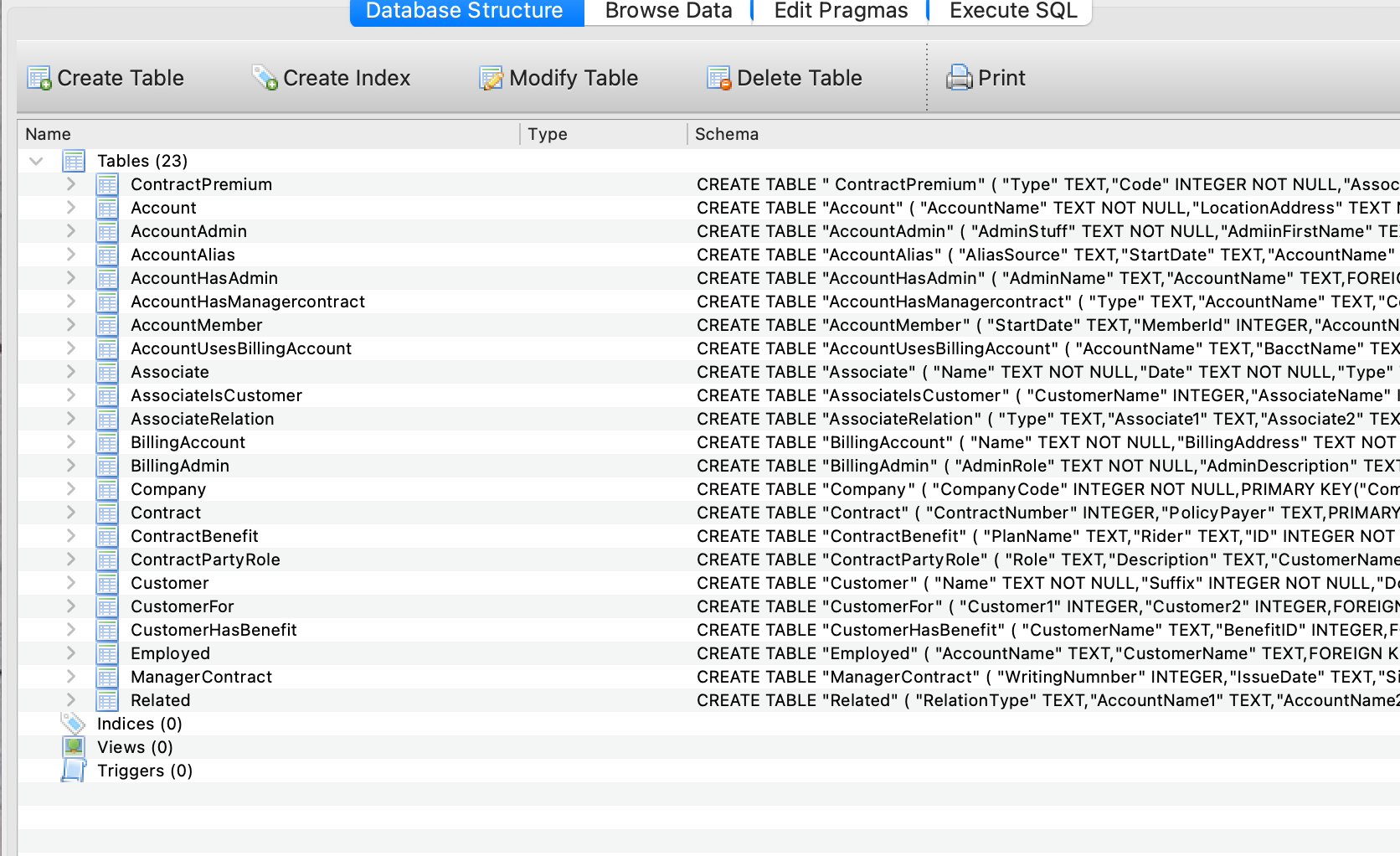
Affirmation of my independent effort:

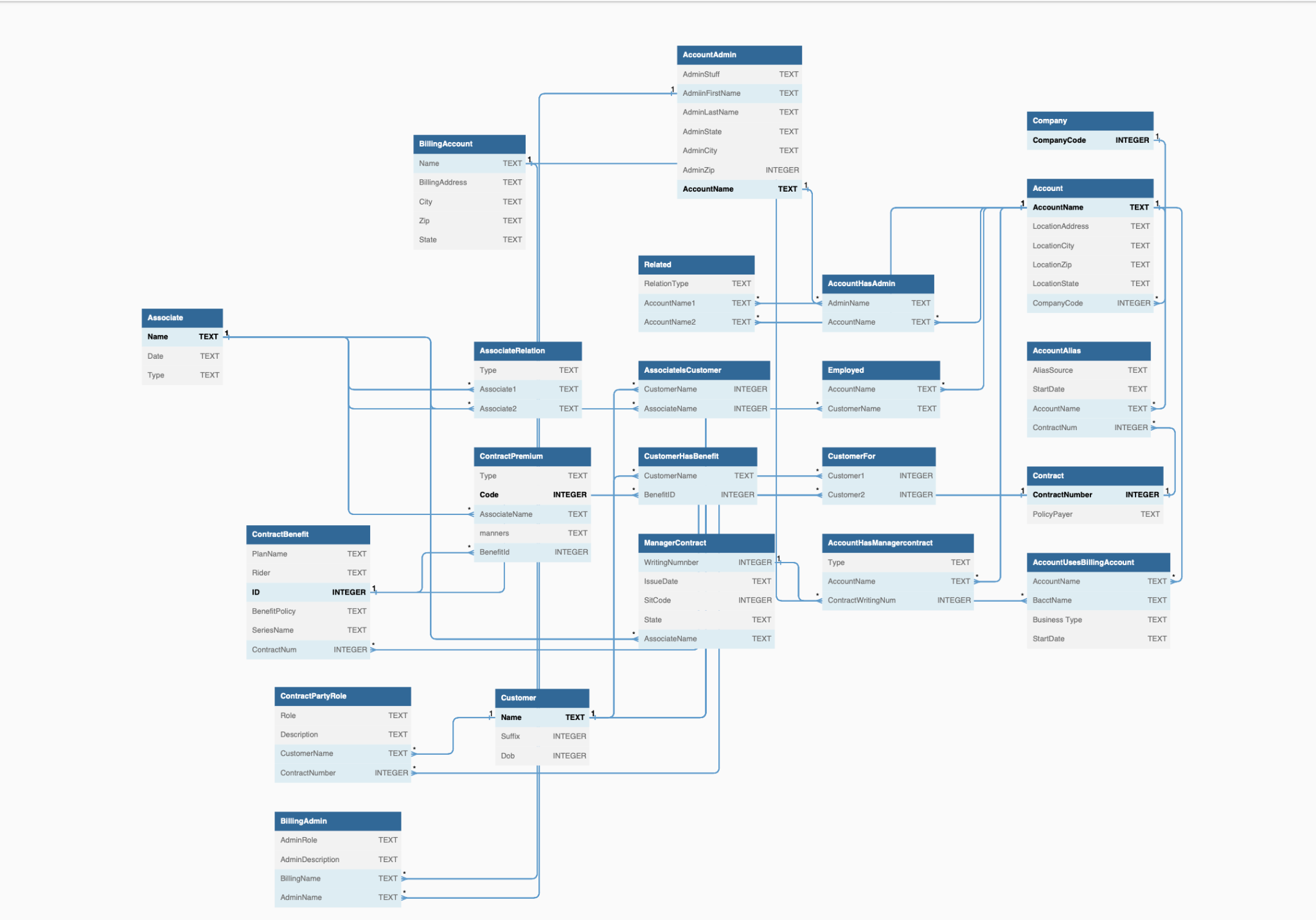
Keying Mao

Linxia Li

We found the cardiovascular dataset from Kaggle <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>. It has factors such as age, height, blood pressure, smoking habits etc.. The data corresponds to the diagram of factors to chronic disease. Based on those factors, we could predict the possibilities that clients have cardiovascular. So that we can use that information to adjust the insurance benefit. That’s also why we can feed the data into EDA.

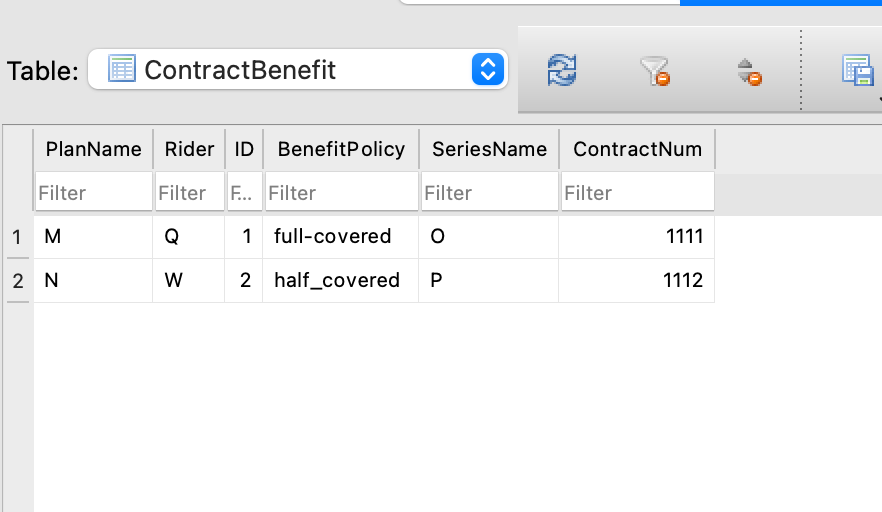
We built our insurance schema in a DB browser and below is our database structure. We chose to convert into the SQL file and then import into DB Diagram. Then we generated the logical database diagram as the second picture shows.

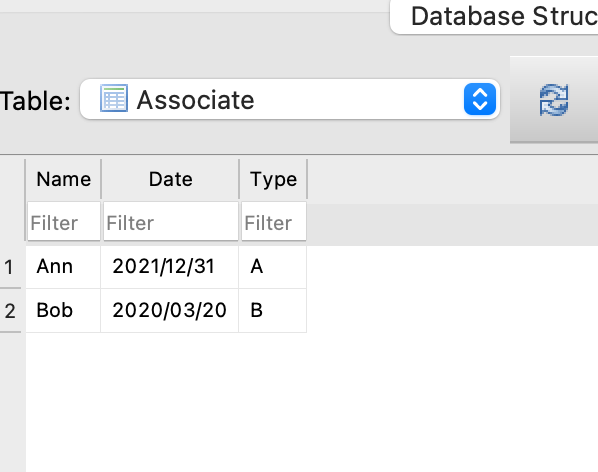


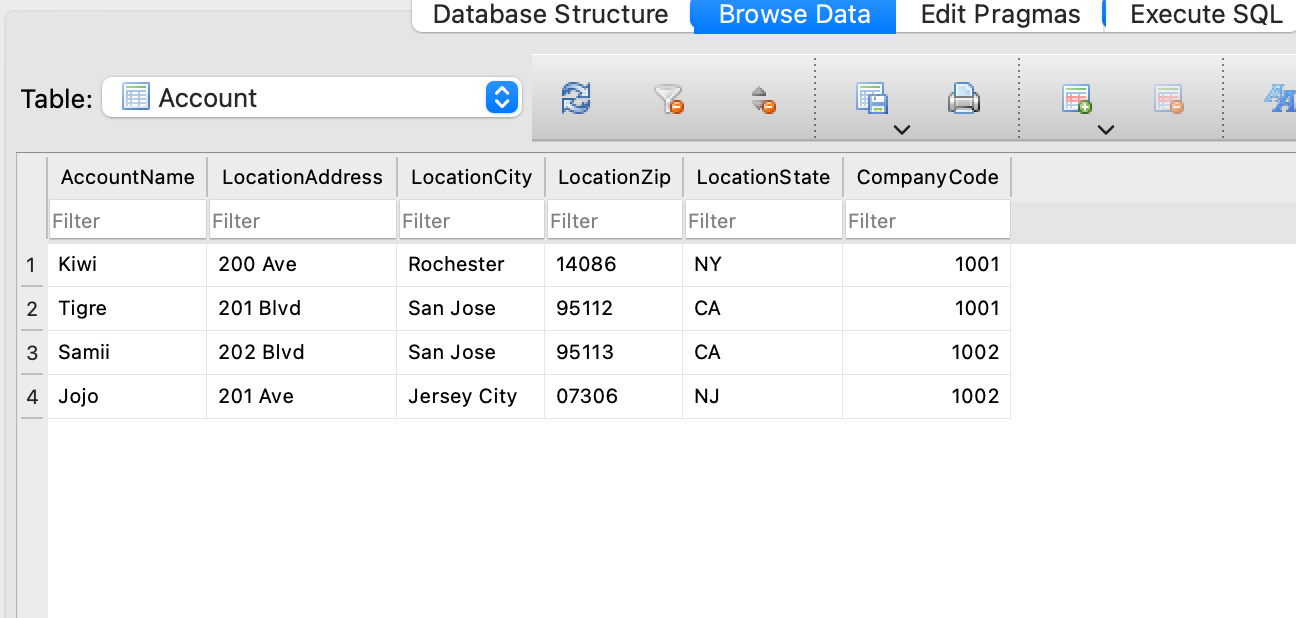


We optimized the logical schema by normalizing the schema and relationship. Our database schema can achieve 3NF. Relation should not have a non-key attribute that is functionally determined by another non-key attribute (or by a set of non-key attributes). That is, there should be no transitive dependency of a non-key attribute on the primary key. For example, in the Account entity, Name is the primary key. Other entities only use “accountName” as the foreign key to refer to the Account entity.

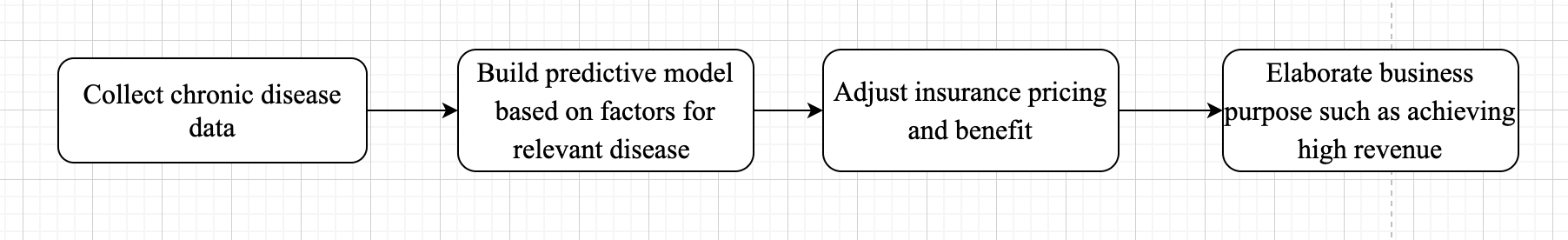
We also fake some data for the insurance schema to check and see how the relation of the database works. Several sample screenshots are attached below.







In order to leverage hybrid data as part of business, we design a simple reference architecture workflow.



Finally, we leveraged our data from kaggle, the insurance fake data, the logical database diagram into the cloud platform (Google Cloud).

