Names: Eugene Chang, Sanho Lee

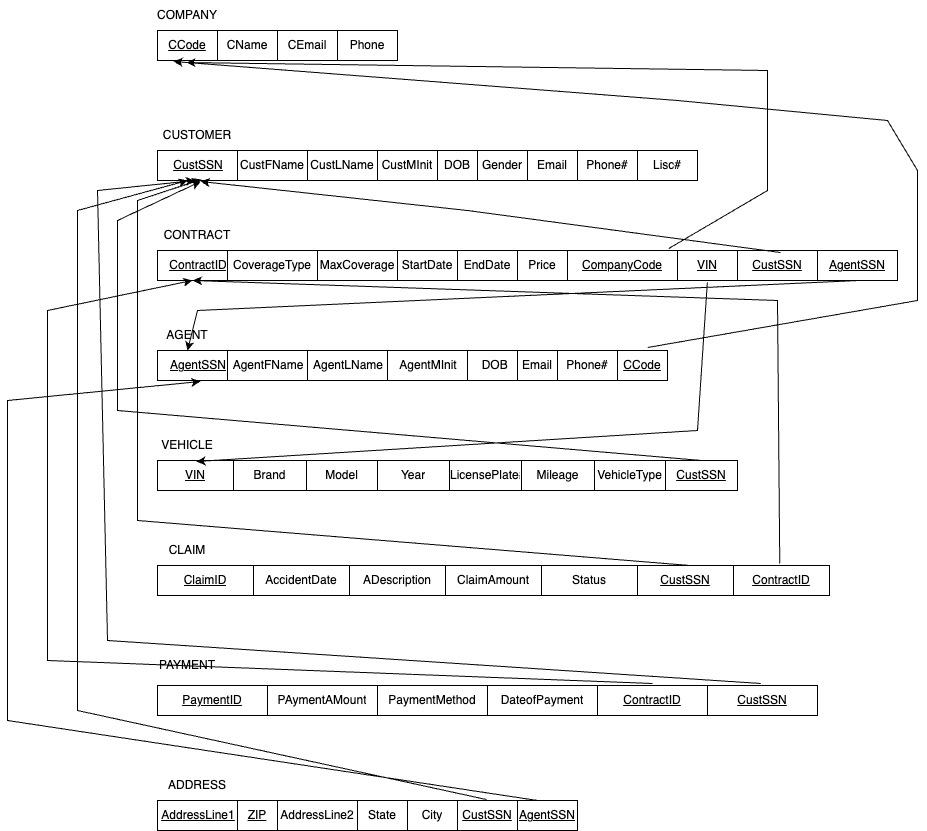
NYU IDs: N174044284, N15250101

Course Section Number: CSCI-GA.2433-001 - Fall 2023

Date: 11/16/2023

**EDA Project Part 2**

**Relational Schema Mapping**



**Physical Model Design**

TABLE

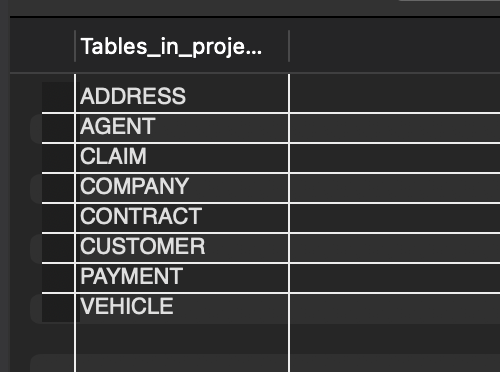


TABLE ADDRESS



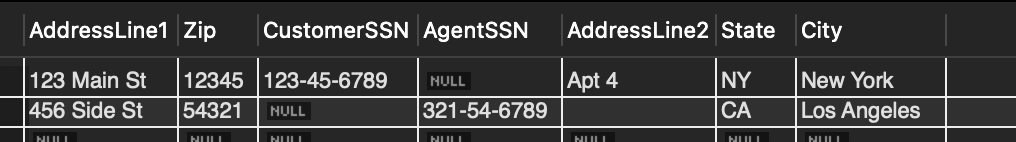


TABLE AGENT



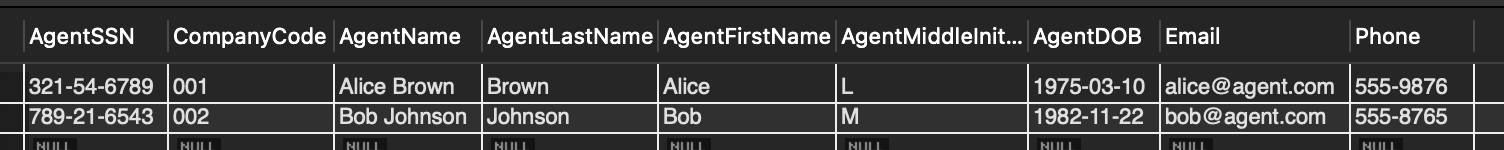
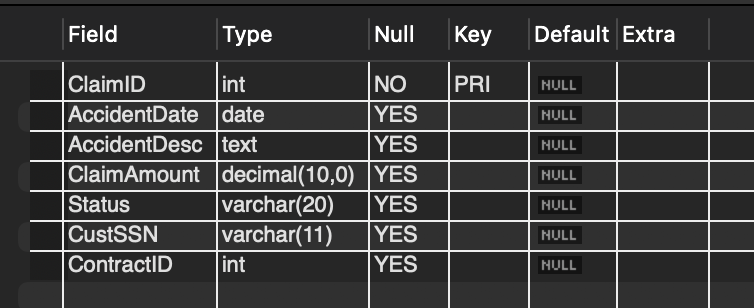


TABLE CLAIM



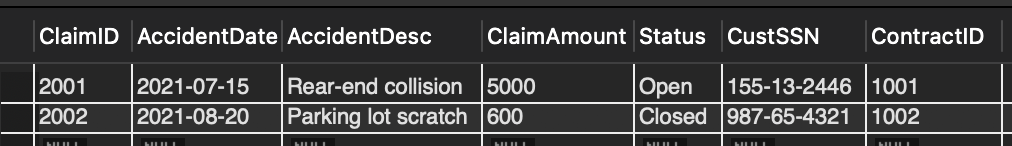
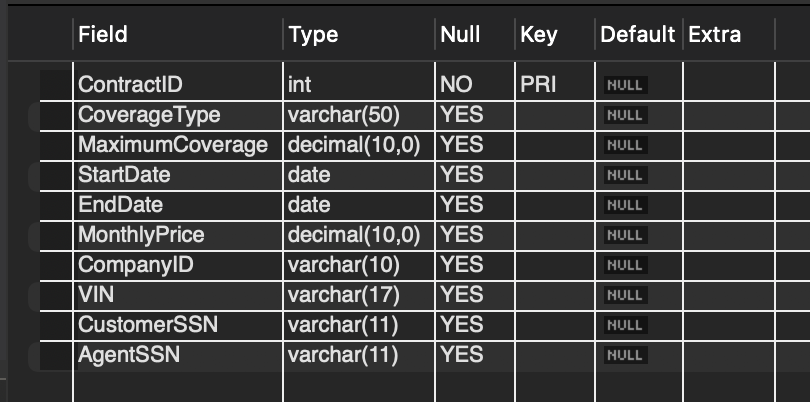


TABLE CONTRACT



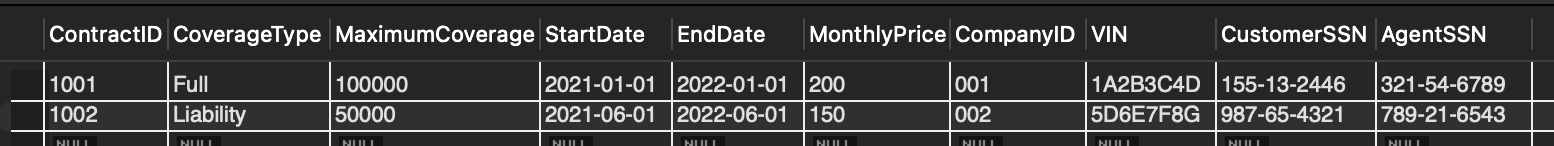
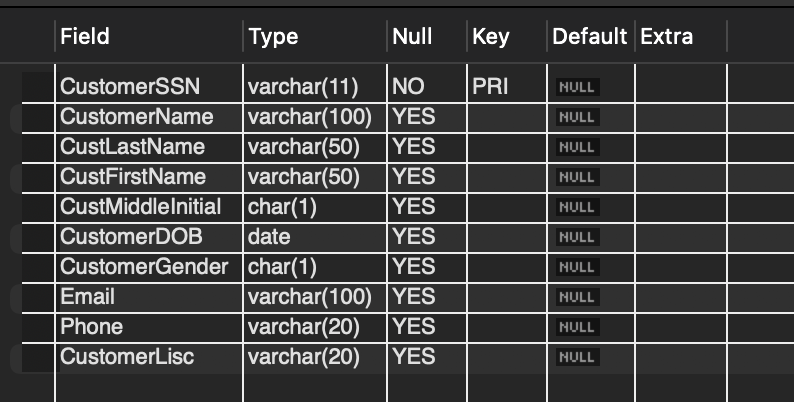


TABLE CUSTOMER



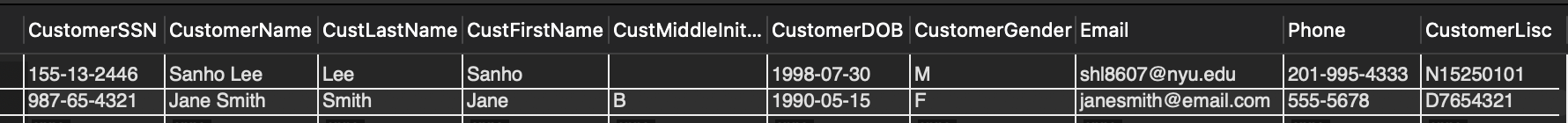


TABLE COMPANY

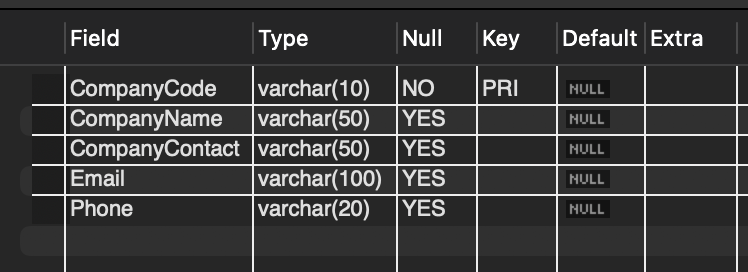
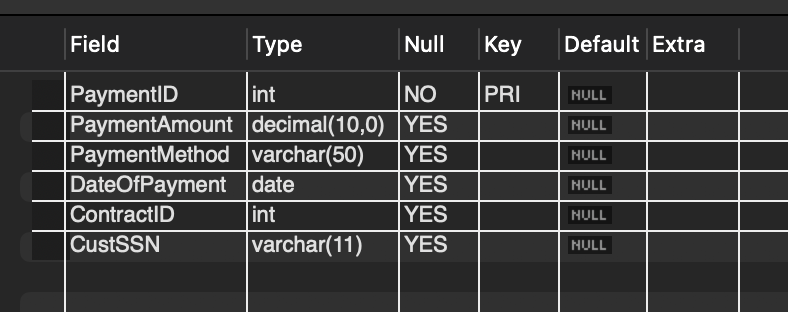




TABLE PAYMENT



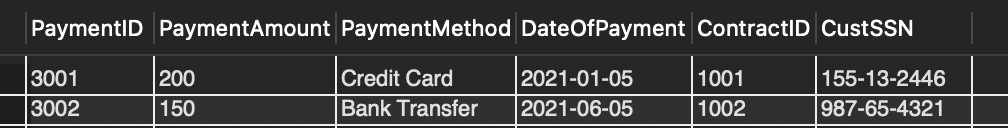
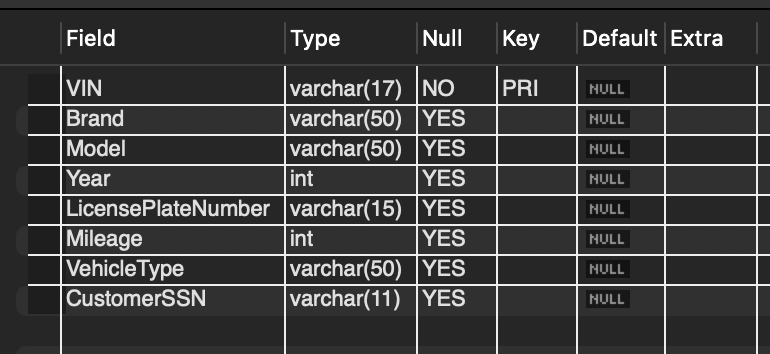
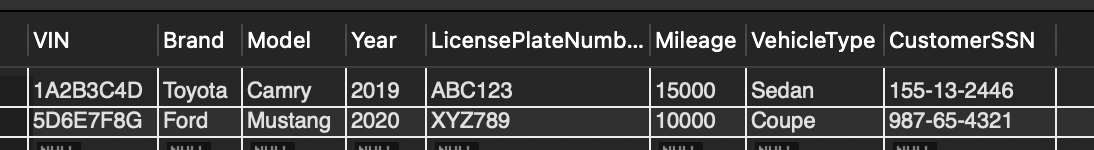


TABLE VEHICLE





Normalization

* Our database model already satisfied the Third Normal Form (3NF)
* We had already removed transitive dependencies

Machine Learning Idea:

**Driver Segmentation and Clustering Method to Determine Insurance Premiums**

* Use Clustering Method Such as K-means Clustering to segment customers based on many factors such as age, gender, vehicle type, driving history, etc.
* By using this algorithm, we can train a program to assign risk factor ratings associated with each driver type/group (eg. A group consisting of very young, inexperienced drivers will be given a high risk factor rating)
* Based on the risk factor, we can determine the opimal insurance premium (fee) rates to offer to minimize risk on the end of the insurance company and boost revenue.
* Users can enter a website and input their information (age, gender, driving history, vehicle id number, location, etc.) in order to recieve a car insurance quote (estimated cost of a policy) that is generated by the trained learning algorithm.

Future plans

* Use a cloud-based service to manage data