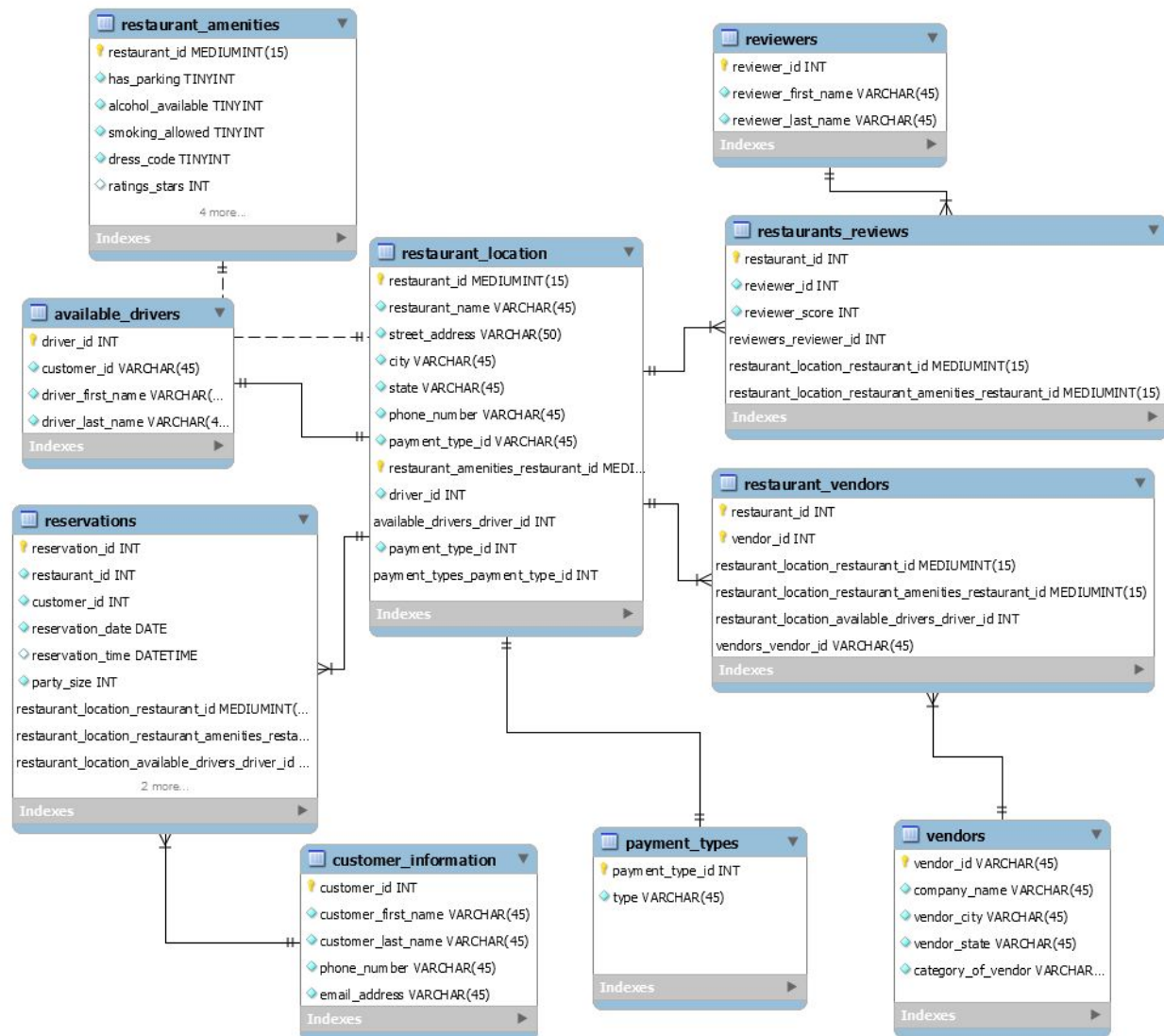


Progress Report

Database Description



Our database will be centered around restaurants in the Montgomery County and Baltimore, Maryland area. The database will include information that varies from details about different restaurants to customer information. The intentions we have for this project are informative, but also indirectly persuasive. Analyzing this database will provide a person with an idea of the variety of restaurants around them to choose from. When using the database users will be able to compare restaurants based on their preferences to determine where they wish to eat. It will also answer many questions like what are the average ratings?

The database will have 10 tables with various columns. Our central table is the restaurant location, which contains all of the restaurants in the database, addresses, and phone numbers. From the restaurant location table we have tables for restaurant amenities, available drivers, reservations, customers, payment types, restaurant vendors, vendors, restaurant reviews, and restaurant reviewers. Restaurant amenities, contains columns for available parking, smoking restrictions, alcohol, dress code, and ratings. Available drivers contains columns based on the available delivery drivers. Reservations contains columns for the date and time of a reservation as well as the party size. Customer information contains columns for a customer's first name, last name, phone number, and email address. Payment types contains columns to keep track of whether a restaurant offers cash, card, or check. Restaurant vendors and Vendors contain columns for the information on all vendors and the restaurants they are linked with. Restaurant reviews and Reviewers contains information on the reviews and the reviewers.

There is a one to one relationship between the restaurant location table and restaurant amenities, available drivers, and payment types. There is a one to many relationship between the restaurant location table and reservations, restaurant vendors, and restaurant reviews. Reservations contains a link to customer information. Also, restaurant vendors has a vendors linking table along with restaurant reviews having a link to reviewers.

Sample Data Plan

In our initial proposal, we anticipated using datasets from Kaggle, however our approach to collecting data has changed a bit. We will be using two datasets: one includes information

about restaurants in Montgomery County, while the other holds information about restaurants in Baltimore. These datasets only satisfy the table which will be labeled, restaurant_information. To get information for the remaining tables, we will have to do more research, or create our own data for tables that describe individuals.

In an effort to maximize every group member, we will most likely split up the research amongst five of us. Some will be responsible for completing the research for a particular area, while others will be tasked with imagining what a diverse spread of customers looks like.

Data Sources:

<https://data.montgomerycountymd.gov/Health-and-Human-Services/Food-Inspection/5pue-gfbc>

<https://data.baltimorecity.gov/Culture-Arts/Restaurants/k5ry-ef3g>

Progress Report

We have made considerable progress since we last submitted our project project. For our group proposal, the instructor appreciated the structure of our tables and advised that we plan ahead of time to collect sortable data. We welcomed this feedback and have since then worked on making the latter tables more feasible with meaningful data.

We started by designing ER diagrams for our restaurants and their different entities. In the process of doing this, we realized that a lot of our tables could have been combined into more comprehensive tables thus reducing our total number of tables. We also decided that some tables needed to be renamed in order for it to make more logical sense. For example, the table previously called 'General information' was broken down and renamed 'restaurant location'.

Structuring the table to make logical sense was just the starting point to create a good restaurant database. We eventually came up with new and attribute names to comprehensively complement our database. Establishing relationships between these tables was another interesting milestone that we were able to accomplish. We put to practice our lessons on primary and foreign keys. Normalization was an inevitable step in the process of establishing our database. With the help of Rob, we normalized our tables to 1st and 2nd normal forms. He

suggested some edits to our records' data types and also helped us set up linking tables to link tables for tables with many-to-many relationships.

Another progress worth mentioning is that while we searched for sortable data to include in our database, we came across large datasets that did not necessarily meet all levels of our needs. Although we are still going to create a database of local restaurants in the DMV, each team member will be assigned a well-defined geographical location in the DMV where she/he would gather restaurant information as it pertains to the completion of our database. This means that we would use data sets from verified sources as well as we will complete necessary gaps with information gathered by individual research.

In order for us to make so much progress on our project, we stay in constant communication and have met remotely 3 times since the submission of our project proposal.

Changes from Initial Proposal

The biggest changes from our initial proposal were the tables. Initially, we had eight tables all with varying degrees of relatedness to restaurants. However, when we were building the ERD, a lot of the entities could not stand alone and instead could just be combined into the restaurant information table. We had to think outside of the box to generate relationships between more tables.

We were able to maintain the general information table, now known as *restaurant_location*. It really serves as the starting point for the database. Our next table was going to be about additional features the restaurant has to offer such as if there is parking, whether or not customers are allowed to smoke, and more. However, we decided to make this table about amenities that each customer would prefer so it could be a table on its own. (*restaurant_amenities*).

Two additional tables closely connected to the restaurant information are *available_drivers* and *payment_types*. *Available_drivers* connects drivers to a particular restaurant, as well as to a customer. *Payment_types* assigns identification numbers to the most common form of payment used at a restaurant.

Because our database will now feature some sample customers, we also needed to create a table with customer contact information (*customer_information*). This table is important because, like *restaurant_location*, it will serve as the basis for a lot of other tables in the database. With the addition of customer information to our database, we decided to create a table for storing reservations (*reservations*). The reservations will link back to the customers, as well as specific restaurants. Each reservation will also include the date, time, and size of the party.

The ratings table still exists (*restaurant_reviews*) and now has a connecting table for information on the reviewer (*reviewers*). Our last table is about the restaurant vendors (*vendors*). Within this table we will include basic information about the vending company.

Another major change that will affect the rest of the project is our sample data. As mentioned before, we will most likely end up collecting much of the restaurant data ourselves. We don't anticipate it will be too much of a challenge since we can split it up. The primary concern is ensuring that we don't choose the same restaurants. The best way to avoid that dilemma would be to not use chain restaurants. Additionally, since we are now using customers and drivers in the table, we will be creating characters to fill those tables.

Given this change we have also altered our scope to just Maryland and not Washington, D.C. and Virginia.

Plans for Remaining Work

There is still some normalization left to be performed on the tables. This will ensure that our database will be more organized. We also have a large amount of data mining to be done based on the entities we defined in our models. We have to decide how the data will be collected, and what locations each group member will focus on. This includes determining how we will obtain our information and how we will upload it to our database.