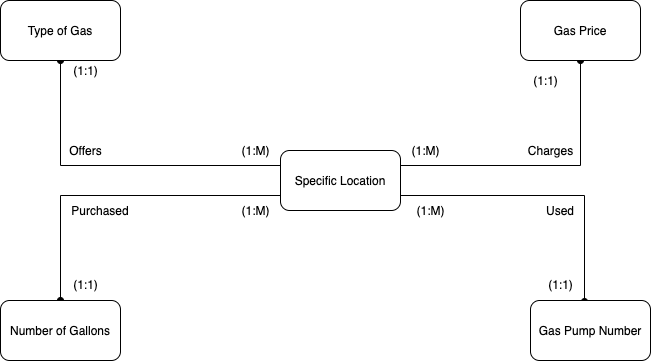
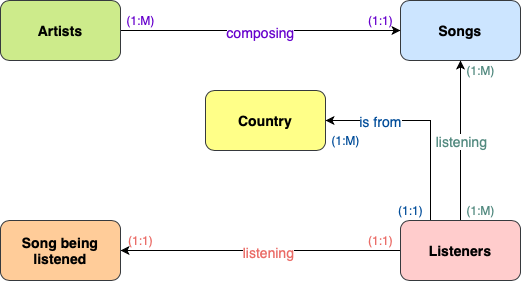
Idea 1: My idea is to compare the data for different aspects of a gas station that is located throughout the United States. In this case, I will use Sheetz. This gas station is unique because it has so many locations with each one differing slightly. In addition, gas prices differ between different locations. I would like to gather data to determine the relationship between which location brings in the most revenue from gas purchases. Since gas is a necessity, there are many interactions that a gas station has throughout the day. A typical interaction that occurs on a daily basis at Sheetz is a customer pulling up to a gas pump, selecting the type of gas they need and purchasing it, and finally pumping the correct amount of gallons into their gas tank. To collect data for this analysis, I will use five tables. These tables are the location of the gas station (different locations throughout the United States), type of gas (regular unleaded, unleaded plus, and premium unleaded), gas price, gas pump number, and the number of gallons purchased. The number of gallons purchased shows real-end time data by continuously being updated to know if the gas station is in need of resupplying. This helps to determine the overall revenue of the gas station. Even though some cars need a certain amount of gallons and others do not, this will be disregarded in this situation because we just want to look at the overall revenue for the gas station. Each gas station has a fair chance of any kind of car coming to get its gas tank filled. Again, we want to look at the overall relationship between the location of Sheetz and the amount of revenue from gas purchases. Each table is connected to the overall revenue in some way and will contribute to determining the relationship.



Idea 2:

Nowadays, people use digital music streaming services to listen to songs and make a playlist of their favorite ones. One of the biggest and most successful ones is Spotify. This application clearly has been using a lot of data analysis techniques to make it such a successful one. However, there are some parts where this function hasn’t been utilized to the maximum, like helping the artists choose the best place to tackle their marketing or in other words, promote themselves. We often see cases where listeners and fans become very disappointed because the artist did not visit their country on tours. Therefore, the goal of the idea presented here is to know the region where an artist is most popular. This idea uniquely functions as a tool to help artists map their tours more precisely and thus satisfying a lot more fans.

The entities that will be presented in this small database are songs, artists, listeners, country, and song being listened to. Artists entity holds the information (attributes) of artist ID, artist name, and year debuted. Songs entity holds information of song name, song ID, artist ID, genre, album, genre, and year composed. Listeners entity holds information of user ID, country ID, and year joined Spotify. The country entity holds information of country ID and country name. The last entity, song being listened, is an end-user entity that records every time a song is being listened to. This entity contains the attributes: song ID, user ID, and date and time listened. The relations between the entity are shown in the table attached below.



Idea 3: Our idea is to replicate an online shopping website such as amazon. We feel this idea is particularly unique because it allows for different types of data to be compiled and compared against each other. Since this is an online shopping network, it creates new challenges that are not present in physical shopping experiences. For example, we can compare the relationship between data such as product type and delivery time. This is a semi-new company since we are centering our business model around a delivery system such as amazon, but we intend to improve the delivery business and streamline the whole process. A typical business interaction at our company is much more complicated than that of a physical store. Interactions at a physical store would consist of the customer taking the item off the shelf and paying for it at the cashier station. For our business, on the other hand, the customer first places an order for an item which is then processed through our server system. This data is recorded, then used to locate the seller and find the closest delivery hub to the customer. The seller then ships out the product and we transfer that item to the customer’s delivery hub. Finally, the package is driven to the customers residence and placed outside their front door. My five tables include customer location, delivery time, shipping options, product type, and time product was ordered. Customer location data includes the exact location that each customer in during the time of purchase. The delivery time tracks the overall time that it took for the package to be delivered. The shipping options include multiple choices such as 1 day, 2 day and regular shipping. The next table is based on the product type. This collects data about the specific size of the packages. The last data type is the time the product was ordered. This includes a specific time of day that the product was ordered on. My overall goal is to compare my other collected data types to the delivery time of each package.

