Elliot Topper, Spandana Bondalapati, Shannon Joseph, Natalia Silva, Rudy Wadle

CSC 315, BUS 385 / MGT 385

Phase III

Relational databases are the most popular form of management of data applications, and this is for good reason. Relational databases are databases that primarily rely on the concepts of *relations* and *relationships* to store data in a way that is efficient to update and query. Relational databases are formed using rigid schemas, meaning the data at any given point in time is required to follow a set of constraints defined by this schema, and must be organized in a predictable way for the guarantee of this ease of access. Relational databases are made up of relations, which, through relationships, can be connected to other relations for complex queries on the stored data.

Relational databases are valuable because they allow dynamic complex queries on parts of a given state, as opposed to only allowing queries on the entire state, as some other database models do. The flexibility of relational databases makes them very versatile, hence their common use despite the age of the technology.

The first relation in the schema is Municipality. This relation contains information to uniquely identify each municipality in the state. To uniquely identify this, both the name of the municipality and the county in which it resides are necessary, as some town names are used in multiple counties. This is represented by the uppermost box in the ER diagram, with its two attributes branching off.

Branching off of Municipality is the weak relationship DESCRIBED_BY. The weak relationship refers to Municipality being necessary to uniquely identify Datum. The year attribute uniquely identifies a given datum of a given type in the set of all data present for that municipality of that type. Additionally, each Municipality can have any number of pieces of data, but each datum is linked to only one municipality.

Datum describes any piece of information stored about a municipality, and can be specialized into EV_Datum, Generalized_GHG_Datum, and Vehicle_GHG_Datum. Any piece of data must fall into exactly one of these three categories. Each piece of data contains information about a different metric measured in some municipality, to which it is related using DESCRIBED_BY, where the year measured is specified on the Datum.

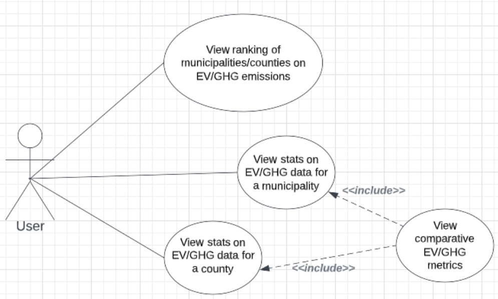
The EV_Datum represents the number of electric vehicles and the number of emissions-based vehicles within a given municipality. Furthermore, the inclusion of EV_Datum would help to reduce redundancy within our dataset. This would be the case because it would be redundant to store the average in our database when it would be simpler to calculate it through the attributes already present, hence the dotted circle around the attribute in the ER diagram.

The greenhouse gas emissions by vehicle type dataset consists of 18 columns with municipality name, county, MPO, year, as well as greenhouse gas (GHG) emissions (metric tons carbon dioxide equivalent) for 13 vehicles. For instance, passenger cars, light commercial trucks, motorcycles, and school buses are only a few examples. Each row holds data for the various municipalities for the 21 counties in New

Jersey. The Vehicle_GHG_Datum also shows the total emissions that correlate with electric vehicles.

Each Generalized_GHG_Datum consists of 4 different types of emission metrics, stored in General_ems. These represent the residential, industrial, commercial, and streetlight metrics found in the data provided by Sustainable Jersey. Each type of emission keeps track of the natural gas emissions and electrical emissions related to the specified part of infrastructure. Additionally, each Generalized_GHG_Datum keeps track of a municipality's carbon emissions related to vehicle use, as well as the total MTCO2e produced.

In regards to our database, the Generalized_GHG_Datum would show specific data such as showing a correlation between either the increase/decrease of electric vehicle usage with the percentage of greenhouse gas emissions in a given year. Essentially, the GHG datum will help conceptualize the data and create comparative metrics. Furthermore, you can view the data values included in the vehicle greenhouse gas data.



Use case: Select region for data display

1. System displays county selection options to User.
2. System displays municipality selection options to User.
3. System validates municipality selection.
4. System retrieves relevant data from the database.
5. System calculates comparative metrics based on aggregated data across the state, including Greenhouse Gas emission data and Elevtric Vehicle ownership data.
6. System displays requested information and comparative metrics to User.

Use case: Select data for visual display

1. System allows User to view either the entire state or an individual county.
2. System displays selection options for comparative display to User, specifying Greenhouse Gas emission data or Electric Vehicle ownership data.
3. System retrieves data from the database based on the user’s selection.
4. System calculates a normalization value for each region included in the selected area (either counties or municipalities in a county) based on the selected data and the maximum datum displayed in the interface.
5. System calculates the ranking of each county and municipality compared to all selected regions.
6. System displays the retrieved data to the User as a list of ranked percentile scores.