INST327 - Database Design and Modeling

Section 0202

Final Project Report

12/09/2024

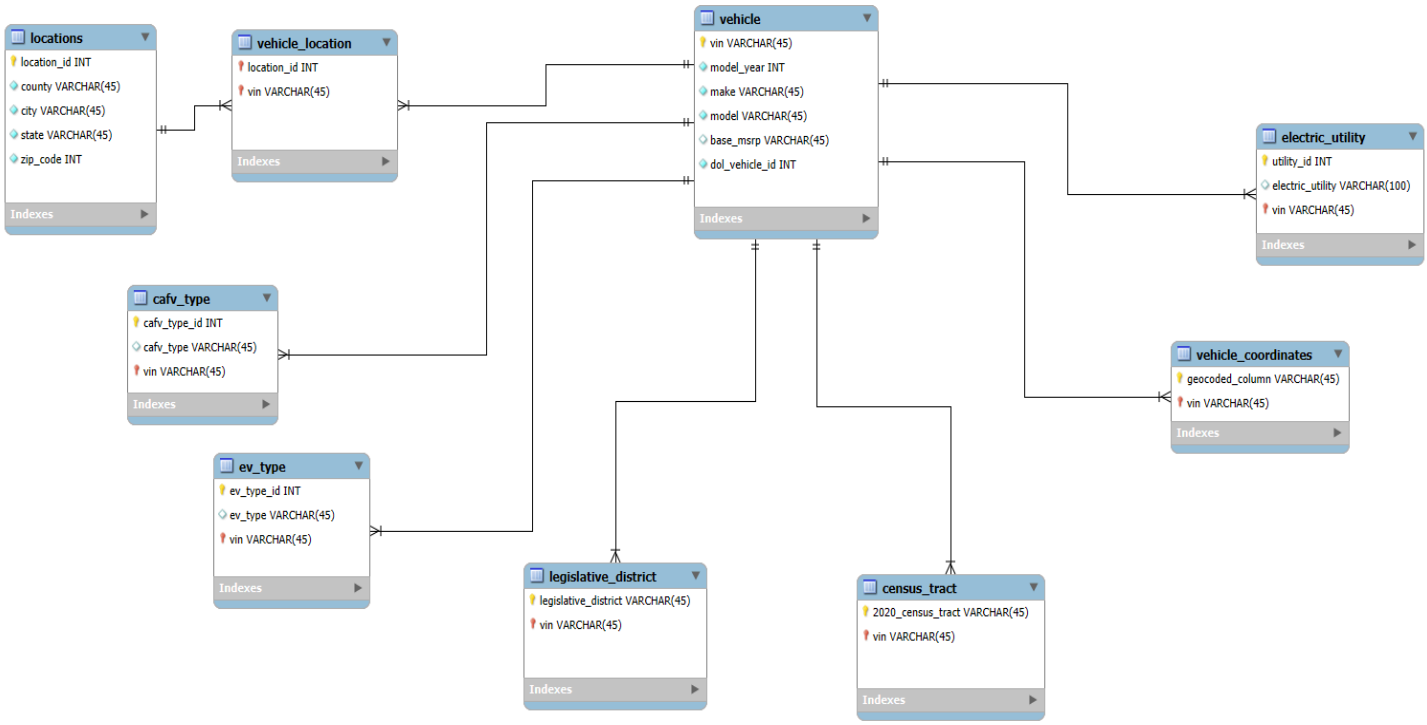
Team 2

Jack Blessing, Don Divin Simbakira Duhimbaze, Ashton Kabou, Ethan Tran, Chris Xu

Introduction

Our focus for this project was to develop and refine an electric vehicle (EV) database. The goal of our database was to serve as a comprehensive repository for detailed information about various electric vehicles. Key attributes of interest include the *Year*, *Make*, *Model*, *Vehicle Type*, and *Clean Alternative Fuel Vehicle (CAFV) Eligibility*. We chose to focus on electric vehicles as they are becoming prevalent nationwide. It is crucial for various reasons that information relating to existing EVs is stored in a well-organized database. For example, our database could potentially help the state improve electric vehicle infrastructure in various regions of the state as the number of EVs registered in each electric utility zone will be easily accessible. We believe that this database can be utilized by multiple groups of people. The first is administrator users, such as government officials or law enforcement agencies. These users would have access to sensitive information such as the “vehicle” and “location” tables. Regarding the example given previously, electric utility officials in Washington could have access to the “electric_utility” vehicle so they are aware of the number of vehicles registered in their service areas without having access to other sensitive information.

Database Description



Logical Design

Our main goal was to design the tables in a way where information is easily accessible, and organized well, in a way where sensitive information is kept in separate tables to protect the privacy of vehicle owners. We wanted to make sure that views could be easily designed to protect the more sensitive tables, such as vehicle_coordinates. We chose to have the majority of tables contain the VIN as a foreign key, linking them back to the vehicle table.

Physical Design

Our database includes nine tables with our vehicle table being the main one. It has a one-to-many relationship with vehicle_location, cafv_type, ev_type, legislative_district, census_tract, vehicle_coordinates, and electric_utility. Vehicle_location is a linking table for locations and has a many-to-one relationship.

Sample Data

Our database is populated with information from an existing database provided by the Washington State Department of Licensing. It holds extensive information regarding electric vehicles registered in Washington. The original database includes a CSV file with all points of data, however since there were over 200,000 entries we decided to make our own CSV files for each table to be imported. We chose 21 sample data points for our database. Below is an example of some of the data points included in the "vehicle" table:

	vin	model_year	make	model	base_msrp	dol_vehicle_id
	5YJ3E1EB5J	2018	TESLA	MODEL 3		474363746
	5YJ3E1EC0N	2022	TESLA	MODEL 3		258924354
	5YJSA1DN2D	2013	TESLA	MODEL 3	69900	180374904
	5YJSA1DPXC	2012	TESLA	MODEL S	59900	188634442
	5YJSA1E20H	2017	TESLA	MODEL S		348605603
	5YJSA1E51N	2022	TESLA	MODEL S		187584183
	5YJYGDEEXL	2020	TESLA	MODEL Y		123289442
	7SAXCAE59P	2023	TESLA	MODEL X		245689888
	JF2GTDNC8K	2019	SUBARU	CROSST...	34995	175410441
	JTDKARFP8H	2017	TOYOTA	PRIUS P...		181646636
	JTMAB3FV5R	2024	TOYOTA	RAV4 P...		271853067

CRUD - Views / Queries

We created five queries for our database. They are as follows:

Query 1: Creates a view that counts the number of vehicles registered in each legislative district.

Query 2: Creates a view that displays the VIN number of vehicles that have the CAFV type "Clean Alternative".

Query 3: Creates a view that displays locations where one or more vehicles with EV type "Plug-in Hybrid" are registered.

Query 4: Creates a view that displays the location and coordinates of each registered vehicle.

Query 5: Creates a view that counts the number of vehicles registered in each city.

	Join clauses	Filtering	Aggregate functions	Join and source tables	Subquery
count_of_vehicles_in_district	X		X		
vehicles_cafv_clean_alternative	X	X			X
location_of_hybrid_vehicles	X	X		X	
precise_vehicle_location	X				

electric_vehicle_count_by_city	X	X	X		
Total	5	3	2	1	1

Changes From Original Design

Since the creation of our original ERD, mostly minor changes have taken place. The most noteworthy change is the creation of a linking table called vehicle_location, connecting the locations table to the vehicle table. The creation of this table makes it much easier to connect multiple vehicles to one location using a unique location ID. Another recent change is the creation of the cafv_type table. It was originally mixed together with the ev_type table and had one ev_type ID representing both ev_type and cafv_type. We realized that this structure did not work according to the normalization process and after consulting our TA, we decided to separate the tables.

Database Ethics Considerations

The ethical considerations and data privacy concerns have stayed the same from our initial proposal. Our database will include every hybrid and electric vehicle registered within the state of Washington. As long as an electric vehicle was registered properly, it will be included. This helps to eliminate selection bias as no demographic within Washington is excluded from this dataset. However, it is still important to consider that there may be electric vehicles within Washington that haven't been registered properly with the DOL or are registered within another US state. It is also important to consider that many of the vehicles registered within Washington have owners residing in other US states, meaning that these vehicles may not be present in Washington. With our design, these vehicles will appear in the database regardless as they once were in Washington.

A primary concern with our database is that it contains sensitive information about the vehicle such as GPS location which includes ZIP code, state, city, and county. This information could potentially be exploited if data were to be breached or leaked. Although government officials and law enforcement would have access to the location data, there would need to be strict access policies in place to protect the user's privacy. Another possible data concern is the use of the VIN, if exposed this could trace back to the vehicle owner and can potentially expose

their information. The vehicle location can also prove to be a potential problem as it stores precise coordinates which can be exploited if improperly accessed. In order to protect user privacy and unethical use of data, researchers and students should only have access to non-identifiable data so they would only be able to track things like vehicle stats. We will implement separate VIEWS that expose only non-sensitive information for non-admin positions and we will grant access to the raw tables to those who are administrators. For potential copyright and fair use, we will have to ensure that the database follows guidelines and is structured to comply with any restrictions to prevent any legal problems. Lastly, as long as our data is only gathered from official government sources, we shouldn't face any legal liability in terms of copyright or any breach of terms and services.

Lessons Learned

One of the first big challenges we faced as a group came following the Project Logical Design assignment. Following the normalization process, we only had five tables, which didn't meet project requirements. We struggled to determine how the tables should be broken up. After discussing our situation with our lab TA during the Project Status 2 assignment, she suggested ways we could reach seven tables, such as breaking up the vehicle table which had fourteen proposed columns at the time. This helped us to reach seven tables at the time, which is now nine. The second big challenge we faced was working with the ERD and forward engineering it into our database. The ERD had to be revisited multiple times due to several errors, such as the VIN being an integer when it contained letters. We had also set certain ID columns to be auto-incrementing when they shouldn't have been, causing the CSV import process to fail for certain tables. Besides problems relating directly to our database, we also faced challenges as a group when it came to communication. It was difficult to find times for all group members to meet and work together as we were all super busy this semester. We made sure to communicate often in our Discord server regarding project status so that we were all on the same page.

Potential Future Work

Currently, our database only contains 21 sample data entries. As mentioned earlier, the CSV file provided from the original database contains over 200,000 entries. We can incorporate all of the entries into our database so that it is a true representation of every electric vehicle in the

state of Washington. We could design more views tailored to specific positions that may benefit from the utilization of our database, such as automakers and Washington state government officials. We could also expand the scope of our database and seek out electric vehicle data from other US states. By doing this, we could make our database represent the entire nation.

Citations

"Electric Vehicle Population Data." *Washington State Open Data Portal*, 16 Sept. 2024, https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2/about_data.