INST 327
Section 0201
Project Final Report
12/09/2024
Team 1
Kim Mai, Ruth Fekadu, Nell Yonkos, Thanh Hoang, Sebastian Boyer

Project Final Report

1. Introduction:

The world is in the midst of a transformative shift toward sustainable energy, and at the heart of this movement is the rise of electric vehicles (EVs). Once a niche product, EVs are now emerging as a mainstream alternative to traditional internal combustion engine vehicles. As concerns about climate change, air quality, and the depletion of non-renewable resources continue to grow, EVs provide a viable solution to reduce carbon emissions and reliance on fossil fuels. In the United States alone, the EV population has surpassed two million as of 2021, a number that is expected to increase significantly due to advancements in technology, government incentives, and the rapid expansion of charging infrastructure.

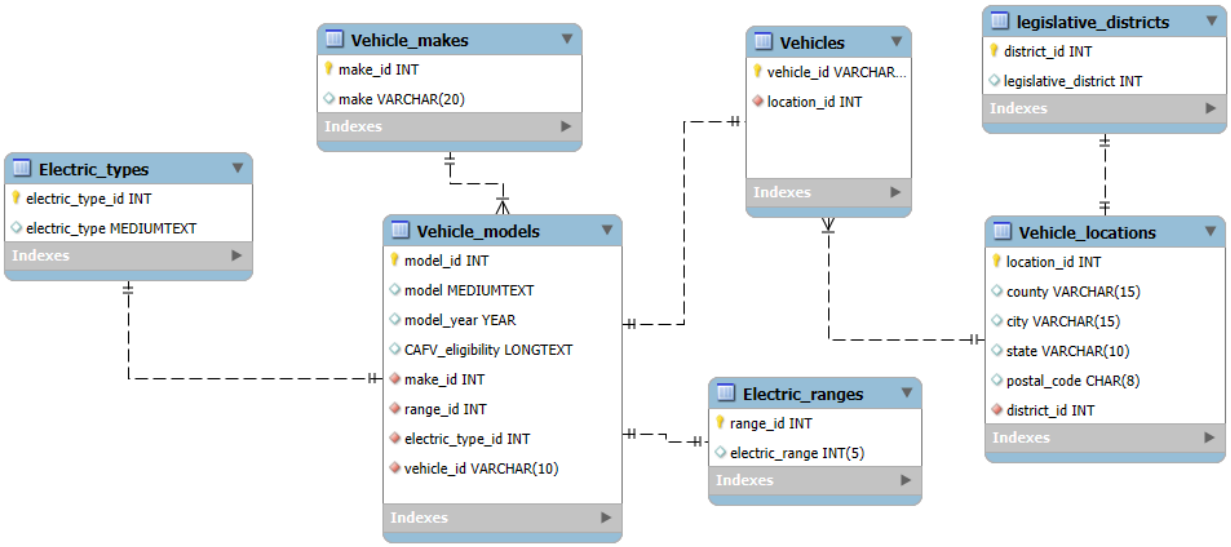
For our project, we will utilize a comprehensive dataset provided by the Washington State Department of Licensing to track the electric vehicle population within the state. This database contains detailed information such as vehicle make, model, type (Battery Electric Vehicle or Plug-in Hybrid Electric Vehicle), year of manufacture, year of registration, and eligibility for government incentives like tax exemptions under the Clean Alternative Fuel Vehicle program. The dataset also includes the geographic distribution of registered vehicles, enabling analysis of trends in EV adoption across different regions of the state. By leveraging this existing data, we can explore market dominance by specific manufacturers, assess the effectiveness of policies, and provide insights into the continued growth of electric transportation.

The database will be valuable for a broad range of stakeholders, including policy makers, automotive manufacturers, environmental organizations, and consumers, all of whom have a vested interest in the future of sustainable transportation. Ultimately, this project aims to provide

actionable insights that can help shape the policies and innovations needed to accelerate the shift towards clean energy vehicles.

2. Database Description:

- **Logical Design:**



The final logical design of our team is somewhat different from the one that we designed in the project progress report. Only the legislative_districts table and electric_ranges table remained the same. Since we moved the columns between tables, we redesigned the relationships to make our database more logical. The electric_types, vehicle_makes, electric_ranges, and vehicles tables are linked (JOIN) by the vehicle_models table. It included make_id, range_id, electric_type_id, and vehicle_id as the foreign keys. Additionally, the vehicle_locations table serves as a link between the legislative_district table and the vehicles.

- **Physical Database:**

Our database included 7 tables that had two linked tables, the vehicle_models table, and the vehicle_locations table. The electric_types, electric_models, vehicles, and electric_ranges table are one-to-one relationships. It is a one-to-many relationship between vehicle_makes and vehicle_models; one vehicle make can have many models. Legislative_districts and vehicle_locations table are one-to-one relationships. Vehicle_locations and vehicles table are one-to-many relationships that one location can have many vehicle registrations.

- **Sample data:**

Our database will public data from the Washington DOL on the registered electric vehicles in Washington State. The tables we will be adding are vehicles, locations, electric types, electric ranges, vehicle models, vehicle makes, and legislative districts. Since the dataset includes over 50,000 entries, we decided to only use a smaller dataset, in order to be manageable. The data will come from the publicly available sources in the Washington DOL website and other sources. In order to make sure our dataset is not too long, we will remove the incomplete records and normalize the data to avoid adding excess info. We did not include the Base MSRP, DOL_id or Electric Utility because we felt as though it was a little unnecessary for the goals of our project.

Table 1: Vehicles

Vehicle_id (PK)	**Location_id (FK)**
1C4JJXP60M	1
1C4JJXP60N	2

Table 2: Electric_ranges

Range_id (PK)	**electric_range**
1	6
2	12

Table 3: Vehicle_makes

Make_id (PK)	**make**
1	JEEP
2	FORD

Table 4: Vehicle_models

Model_id (PK)	**model**	**model_year**	**CAFV_eligibility**	**Make_id (FK)**	**Range_id (FK)**	**electric_type_id**	**vehicle_id**
1	WRANG	202	Not eligible due to	1	10	1	1C4JJXP

	LER	1	low battery range				60M
2	WRANGLER	2022	Not eligible due to low battery range	1	8	1	1C4JJXP60N

Table 5: Vehicle_locations

Location_id (PK)	**county**	**city**	**state**	**postal_code**	**District_id (FK)**
1	Spokane	Spokane	WA	99217	5
2	Spokane	Spokane	WA	99204	2

Table 6: Legislative_districts

District_id (PK)	**legislative_district**
1	2
2	3

Table 7: Electric_types

Electric_type_id (PK)	**electric_type**
1	Plug-in Hybrid Electric Vehicle (PHEV)
2	Battery Electric Vehicle (BEV)

- **CRUD (Views/Queries):**

View Name	Req. A	Req. B	Req. C	Req. D	Req.E
vehicle_info	x	x		x	
counts_by_make	x	x	x		
vehicle_locations	x			x	
range_and_type	x	x		x	

make_model_counts	x	x	x		x
Avg_Electric_Range	x	x	x	x	
Vehicle_Count_By_Location	x		x		

3. **Change From Original Design:**

Throughout the project, we made significant changes to our original design to improve the database's functionality and alignment with our objectives. Initially, our design combined too many attributes in large, redundant tables such as Vehicles and Vehicle Locations, which complicated querying and analysis. To address this, we normalized the database to 3NF, creating smaller, specialized tables like Vehicle_Types, Legislative_Districts, and Location_Points. This separation reduced redundancy and clarief relationship between electric vehicles and their CAFV eligibility, leading to creation of Electric_Types and CAFV_Eligibility tables. These adjustments not only streamlined our design but also made the database more scalable and efficient for generating insights, such as geographic and legislative trends, which were key objectives.

4. **Database Ethics Considerations:**

As there are different barriers that can prevent people from accessible purchasing and equitable charging due to price and location, these are things that need to be considered when monitoring the data. With the rise in popularity over the last couple of years comes a rise in price when compared to traditional gas powered vehicles. While there are certain tax breaks in states that can help reduce the price of electric vehicles in the long term, the prices do end up expensive and not available to everyone at a great level. We also need to keep in mind the consideration of where the charging ports are in greatest frequency. There needs to be an open mind and make sure that the charging stations are spread equally throughout the cities no matter the specific income level by area. We think one of the last things that we need to keep in main consideration is that since we are targeting government agencies, we need to make sure that the information that we are spreading is available to be accessed by people of all cultures and origins.

Since we are creating a database that will track the electric vehicle populations in Washington State, we need to take privacy and ethical concerns into account. The Washington

State Department of Licensing (DOL) states they will notify you if there is a data breach that includes any of your personal information, such as social security, driver's license number, or any biometric data, unless told by law enforcement to be delayed. By using these public datasets we need to ensure that any sensitive data is properly anonymized by restricting location data that are saved and we will remove unnecessary characteristics like exact tracking coordinates or demographic details. We will try to use enough data just so we can be able to study the trends while also protecting each individual's confidentiality. Our project will be using an open dataset that comes from the Washington State Department of Licensing and we are not using any "closed" data sources, so we will lower the possibility of being exposed to legal liabilities. While recognizing fair use, we do not think we will have any copyright issues because the dataset is non-proprietary and made available for public use by the government. However, it is still crucial that we follow the appropriate usage guidelines while using the dataset. Finally, while working on this project we will ensure that we give credit to all sources and make sure that everything is done ethically.

5. **Lessons Learned:**

This project provided valuable lessons in database design, normalization, and query writing. One of the main challenges we faced was understanding and implementing normalization to eliminate redundancy while preserving data integrity. We learned the importance of organizing data into logical entities and ensuring each attribute depends solely on its table's primary key. Additionally, crafting complex SQL queries, particularly those involving multiple JOIN, filtering, and aggregations, helped us improve our technical skills and structured queries. Collaboration was also a key takeaway, as dividing tasks among team members and integrating feedback ensured our design met the objectives while adhering to best practice. Finally, we gained insight into the ethical considerations of data management, particularly around ensuring privacy when dealing with location-based data.

6. **Potential Future Work:**

While our database effectively captures and analyzes electric vehicle data, there are several areas for improvement. One priority is incorporating additional datasets, such as charging station availability or population density, to provide deeper insights into EV adoption patterns. Enhancing geographic data granularity by integrating census tract-level details or advanced spatial coordinates could also help stakeholders make more localized decisions. Additionally,

implementing automated reporting tools or dashboards could make query results more accessible to non-technical users. Another area of improvement is optimizing query performance for larger datasets, as scalability will become increasingly important with real-world usage. Lastly, expanding the database to track trends over time, such as year on year EV adoption growth or impact of changing CAFV policies, could provide valuable longitudinal insights for policymakers and manufacturers.

Citations:

Washington State Department of Licensing. (October 16, 2024). *Electric Vehicle Population Data*. [Dataset]. Retrieved November 17, 2024, from https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2

Washington State Department of Licensing. (n.d.). *Data transparency*. Retrieved November 17, 2024, from https://dol.wa.gov/about/privacy-center/data-transparency#:~=We'll%20notify%20you%20any,or%20Washington%20ID%20card%20number.