Final Report

Team 6

Maria Nedumthuruthil, Sakshi Shetty Santhosh Kumar,
Lauren Reyes, John Michael Papagiannopoulos, Andrew Khmaladze
University of Maryland, College Park

INST 327

Section 0101

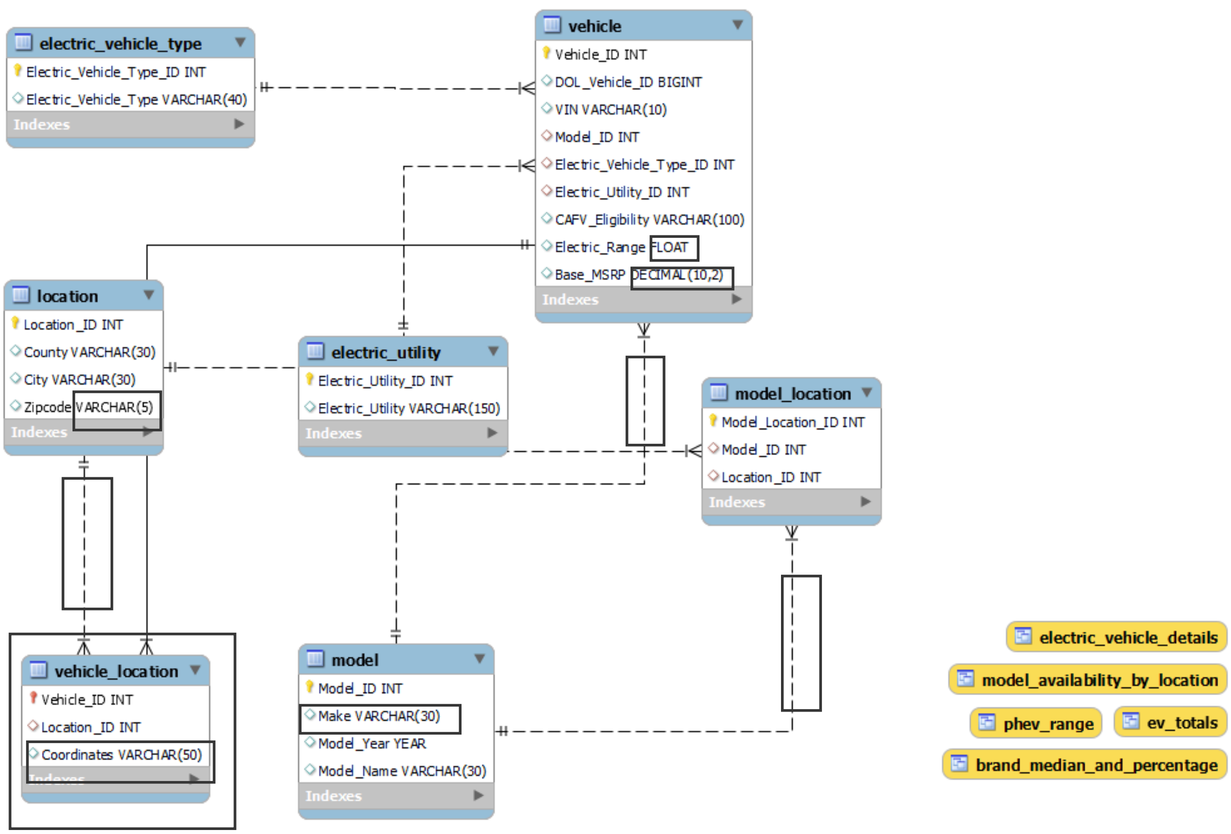
Introduction:

The group project focuses on the topic of Electric Vehicle Population in the state of Washington. This utilizes the Electric Vehicle Population Map by Postal Code database. Thus the database, sourced from the official Washington State open data portal and last updated on February 14, 2024, contains over 174 thousand rows and 17 columns. The dataset provides a comprehensive record of Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) registered in the state of Washington.

The purpose of our project is to identify popular population areas for electric vehicle sales and operation. By analyzing the database attributes, such as the vehicle mode, manufacture, type, and geographical location, we aim to help businesses and corporations strategically locate their facilities in areas with high demand. Our database aims to assist manufacturers in installing the necessary infrastructure in areas with a strong user base for their products.

Database Description

Logical Design:



We decided to make the logical design for our database show the essential information about our database in a concise way for users to understand the relationships present with our data. We decided to have seven entities that each represented a different category of information that was present in our data. The seven entities we finalized on were: vehicle, location, electric_vehicle_type, electric_utility, vehicle_location, model_location, and model. This allowed us to be able to analyze our data in a more effective way for queries. Our logical design also allowed us to better evaluate what our plans for our database would be and understand our database and sample data.

Physical Database:

The seven entities we finalized on for our database were: vehicle, location, electric_vehicle_type, electric_utility, vehicle_location, model, and model_location. The vehicle entity has attributes that would commonly be associated with the statistics of an individual EV. The location table has attributes that represent different elements of the address a EV is registered to. The electric_vehicle_type table has the information on the specific type of EV, Battery Electric Vehicles (BEVs) or Plug-in Hybrid Electric Vehicles (PHEVs), an entry could be and has the primary key for the Electric_Vehicle_Type_ID. The electric_utility table shows the different electric power retail services providing electricity for EVs in the locations they are registered at and has the primary key for Electric_Utility_ID. Vehicle_location shows the specific GPS coordinates of the registered location of an EV. The model table has attributes for the different types of models an EV in the database could be. The model_location table keeps track of the locations that certain EV models can be found in.

Sample Data:

The sample data that we used for our database came from a dataset titled Electric Vehicle Population Data. It was a dataset that showed information on all the Electric Vehicles (EVs) that were registered through the Washington State Department of Licensing as of 2024. The original dataset had over 177,000 different entries with 17 different attributes so we did not have to seek out another dataset for our sample data. However, 177,000 entries was an excessively large amount of entries to work with. So we decided we would randomly select 500 entries from the Washington dataset and use those entries as the sample data for our database. Our group made a program using Java that essentially runs through the CSV file of the Washington dataset, randomly selects 500 entries, and writes the data for each of the selected entries into a new CSV file called “result.csv”. That result file became the sample data we used for the remainder of the project. Below is some sample data from our vehicle table.

Vehicle_ID	DOL_Vehicle_ID	VIN	Model_ID	Electric_Vehicle_Type_...	Electric_Utility_...	CAFV_Eligibility	Electric_Range	Base_MSRP
1	244677373	5YJ3E1EA4P	138	1	21	Eligibility unknown as battery range has not bee...	0	0.00
2	193053423	WBY7Z4C55J	41	2	30	Clean Alternative Fuel Vehicle Eligible	97	0.00
3	277095129	3C3CFFGE0G	17	1	30	Clean Alternative Fuel Vehicle Eligible	84	0.00
4	217944807	3FMTK3SU1N	108	1	1	Eligibility unknown as battery range has not bee...	0	0.00
5	262813967	1C4RJYD61P	130	2	21	Not eligible due to low battery range	25	0.00
6	253553576	WA1F2AFY8L	80	2	12	Not eligible due to low battery range	20	0.00
7	213412132	5YJ3E1EBXN	104	1	21	Eligibility unknown as battery range has not bee...	0	0.00
8	125048003	5YJ3E1EA1J	43	1	30	Clean Alternative Fuel Vehicle Eligible	215	0.00
9	216575607	5UXTA6C00N	119	2	31	Clean Alternative Fuel Vehicle Eligible	30	0.00
10	204666633	7SAYGDEEXN	107	1	31	Eligibility unknown as battery range has not bee...	0	0.00
11	255002688	KM8KNDAF1P	135	1	21	Eligibility unknown as battery range has not bee...	0	0.00
12	148930099	1N4AZ1CP4K	60	1	31	Clean Alternative Fuel Vehicle Eligible	150	0.00
13	229939447	JTMABABA0P	151	1	30	Eligibility unknown as battery range has not bee...	0	0.00
14	146858190	5YJ3E1EA4M	85	1	21	Eligibility unknown as battery range has not bee...	0	0.00
15	174368358	1N4AZ0CPXF	15	1	30	Clean Alternative Fuel Vehicle Eligible	84	0.00

Views/Queries:

Overall, we ended up creating five queries (saved as views) about our database to demonstrate various operations that could be done with our database. The first query Electric_Vehicle_Details creates a view that retrieves details of electric vehicles, including their IDs, DOL vehicle IDs, VINs, make, model name, model year, electric vehicle type, electric range, and base MSRP. It filters out unspecified electric ranges and base MSRPs and sorts the results by electric range in descending order. The second query is called Model_Availability_By_Location. It calculates the availability of each model by counting the number of unique locations where each model is available. It includes the model ID, make, model year, model name, and the count of unique locations. The PHEV_Range query selects distinct electric vehicle details for plug-in hybrid electric vehicles (PHEV) with a range greater than 0. It includes the make, model, model year, electric vehicle type, and electric range. The EV_Totals query calculates the total number of vehicles for each electric vehicle type (PHEVs and BEVs) and identifies the most popular model for each type based on the count of vehicles. The last query creates a view called Brand_Median_and_Percentage which calculates the median electric range in miles and the percentage of total cars for each of the most popular brands. It excludes electric ranges of 0 and handles cases where there's no data available.

View Name	Req. A	Req. B	Req. C	Req. D	Req. E
Electric_Vehicle_Details	X	X	X		
Model_Availability_By_Location	X		X	X	
PHEV_Range	X	X			X
EV_Totals	X		X		

Brand_Median_and_Percentage	X	X	X	X	X

Changes from the Original Design:

Our ERD has undergone significant improvements since the initial proposal, reflecting refined relationships between entities and an optimized structure for efficiency. These enhancements ensure that our database accurately captures the complexities of EV usage in Washington. Additionally, we've implemented several queries within our database system to extract valuable insights from the extensive dataset. These queries are designed to calculate essential metrics and facilitate a deeper understanding of electric vehicle (EV) adoption trends in Washington. One such query calculates the percentage of total cars registered in Washington that are electric vehicles, providing a key indicator of the state's transition towards sustainable transportation. Furthermore, we've developed queries to join pertinent tables such as "Vehicle," "Model," and "EV Type," enabling us to analyze the distribution of EV models and types across different regions of Washington. This integrated approach allows us to identify popular EV models and discern any geographical disparities in their adoption rates. Moreover, sophisticated queries have been implemented to calculate the mean range of electric vehicles in our dataset, providing insights into technological advancements and practical considerations associated with EV ownership. These enhancements, along with updates to our data sample and ERD, reflect our commitment to continuously refining our project's scope and methodology to provide valuable insights into electric vehicle usage in Washington. Our progress has been a collaborative effort, with input from TAs and our team shaping the project's direction. Through iterative refinement and peer/self-reviews, we've enhanced the database's relevance, incorporating best practices and addressing challenges along the way.

Data Ethics Considerations:

Our dataset is derived from the official State Open Data Portal on Electric Vehicle Population. It places importance on protecting the privacy of electric car drivers. We ensure that the data we utilize contains no identifiable information, thus safeguarding individual privacy. The Washington licensing department's methodology for data collection, encompassing customer renewing tabs, ID card updates, driver license actions, reported vehicle sales, and professional license applications, serves as the foundation for our anonymization process ("Information We Collect"). By exclusively utilizing non-public identifiable information (PII), we meticulously anonymize the data, eliminating any personally identifying details such as names and contact information. This rigorous approach ensures individual privacy is preserved while still yielding valuable insights into electric vehicle adoption patterns. Our commitment to privacy protection extends to minimizing any potential privacy or copyright concerns associated with the database.

In addition to privacy preservation, our database adheres to broader ethical considerations. The risk of leaking personal data is a genuine concern for EV owners, and in order to maintain trust with users, we have to implement advanced security measures (Kale). These include encryption and access controls, which strengthen our dataset against unauthorized access or breaches.We prioritize data minimization, retaining only essential information to fulfill our research objectives. Transparent communication and explicit consent practices underscore our commitment to respecting user privacy. Moreover, we actively address biases to ensure the accuracy and fairness of our insights. Accountability mechanisms, governance structures, and ongoing training initiatives further reinforce ethical data practices within our organization. When sharing data, we assess recipients to ensure alignment with ethical standards and secure handling practices. Collaborations are conducted with transparency and integrity, promoting responsible data use across the ecosystem.

Overall, our dataset exemplifies ethical data practices, encompassing privacy protection, transparency, security, bias mitigation, and responsible data sharing. By upholding these principles, we not only mitigate privacy and copyright risks but also contribute positively to the ethical use of data within the broader community.

Lessons Learned:

Group projects are a common and valuable experience for many people. They allow people to have an opportunity to apply learned knowledge to real-world scenarios, develop teamwork and communication skills, and learn from challenges and success. One of the biggest lessons learned throughout this project was that communication is key. As we worked together, clear and consistent communication among team members was essential to ensure that everyone was on the same page regarding the database's design, structure, and implementation. We realized that regular meetings that occur mostly on Zoom, helped us stay updated on each other's progress, address any issues or concerns promptly, and maintain cohesion within the team.

We also learned a great deal about database design and how to improve it over time. At first, our design could have been better, but through feedback from TAs we improved normalization, and our tables to ensure that our database would be efficient and work properly when used. We also learned how to correctly place constraints to enforce data integration such as a primary key constraint or foreign constraint to make sure data is unique or matches other values in another table. This project has helped us tremendously in learning about SQL and we have all made great progress in our journey.

Potential Future Work:

Like many things, there is still room for improvement in our database. One of the first areas in which we could improve is our optimization. This would involve refining our schema, revisiting indexing strategies, and optimizing SQL queries for better performance. Performance tuning would significantly improve the efficiency of the database especially as the dataset grows.

Another area in which we could improve in the future is data quality. Ensuring that the data quality is maintained over time is important. This will be done through data validation checks, a data cleansing process, and regular data quality audits to ensure that the database remains accurate and reliable.

Compliance with relevant regulations and standards would also be beneficial. Future work would be implementing compliance measures and staying updated with evolving regulations to ensure that the data remains aligned with regulations such as GDPR, HIPAA etc.

Overall, there are many areas for improvement in our dataset and the group project as of right now provides a solid foundation for future work and enhancements to the database system. By focusing on areas such as optimization, data quality, and compliance with regulations, the database can be further developed to meet the evolving requirements and deliver greater value to its users.

Bibliography

Electric vehicle population data. (2019, April 16). State of Washington. https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2/data_preview

"Information We Collect." *Information We Collect | Washington State Department of Licensing*, Washington State Department of Licensing , dol.wa.gov/about/privacy-center/information-we-collect#personallyidentifiableinformation. Accessed 18 Apr. 2024.

Kale, A. (2024, January). *E-vehicles and data privacy: Concerns and considerations*. CyberPeace. https://www.cyberpeace.org/resources/blogs/e-vehicles-and-data-privacy-concerns-and-considerations#:~:text=The%20constant%20recording%20of%20a,genuine%20concern%20for%20EV%20owners. Accessed 10 May 2024.