**ETL Alternative Fuel Vehicle Data**

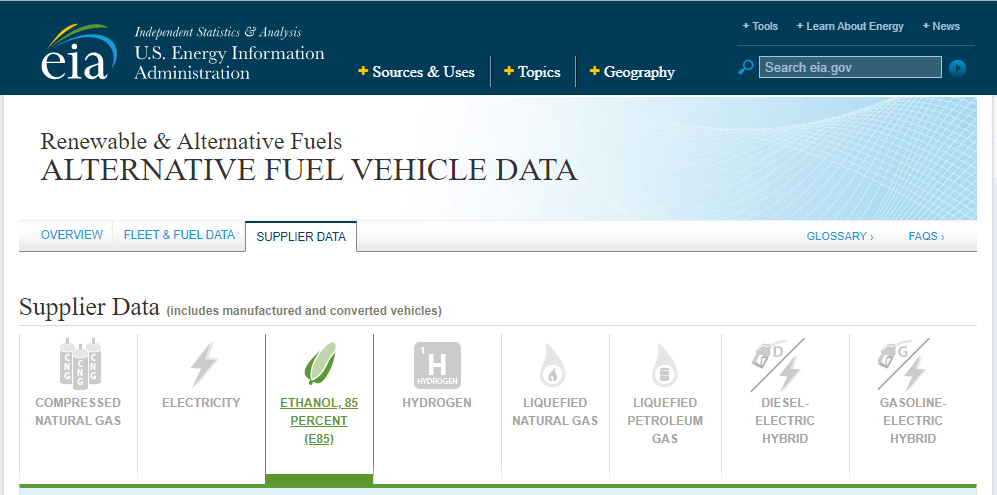
By Shahzad Nassir and Oswald Vinueza

**Data Sources and Extraction**

We used csv files obtained from the EIA website containing data for alternative fuels such as compressed natural gas, electricity, ethanol, and hydrogen. Each csv file contains the year, fuel type, number of vehicles, weight class, vehicle type and engine configuration.

The EIA website contained two types of data relating to alternative fuels, one from a supplier side and one from a user side. Suppliers include original equipment manufacturers and aftermarket vehicle converters. The supplier side data shows how many vehicles can operate on an alternate fuel whereas the user side data shows how many AFVs (Alternative Fueled Vehicles) are in use based on surveys done by the federal and state governments, alternative fuel providers, and transit companies.

For this project, we chose the supplier side and focused our data preparation and possible uses on how many AFVs are available in the auto market.



We loaded the csv files into Jupiter Notebook and started the clean up process.

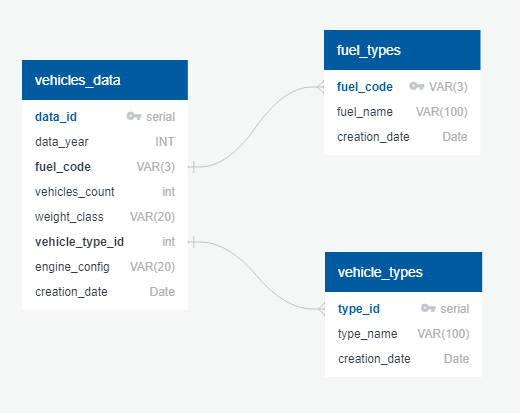
**Data Cleaning and Normalization**

We cleaned the data by equalizing differences in formatting for the various csv files that otherwise would have prevented us form properly grouping the data. After completing the cleaning process we proceeded to normalize the data.

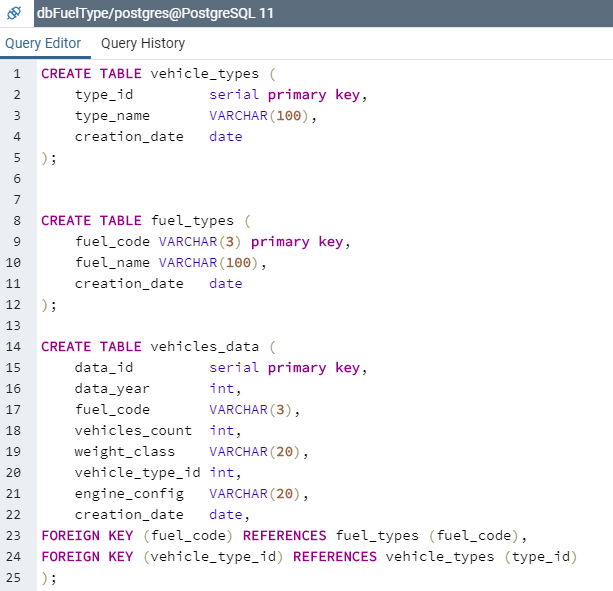
We took a closer look at the data and determined that the fuel type and the vehicle type information should be separated into 2 different tables. As new types of fuels or vehicle types are added into the EIA website in the future, it will be easier and more efficient to add these new types into the 2 separate tables.

For the fuel\_types table we created a primary key extracted from the fuel\_code information and for the vehicles\_types and the data\_id the primary keys were autogenerated by the database as serial numbers.

We used QuickDBD to create a picture of the model schemata of our database containing 3 tables – vehicles\_data, fuels\_types and vehicle\_types.



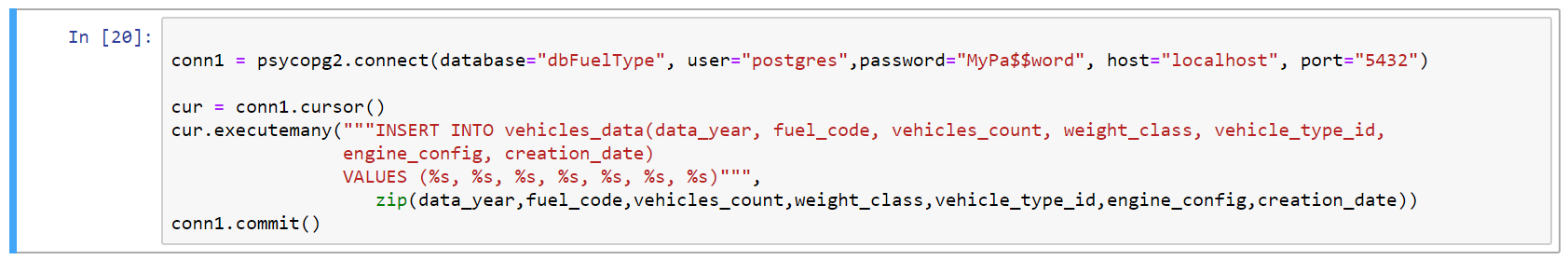
After completing the process mentioned above we created the tables in PostgreSQL.



**Data Connection and Loading**

We created an engine to connect the 3 normalized tables to a PostgreSQL database.

We loaded the normalized tables containing the alternative vehicle fuel data from Jupiter Notebook to PostgreSQL.



**Queries**

After making sure the data loaded correctly, we performed some queries to group the information into categories including fuel type, vehicle type and weight class so as to highlight the data that we think an end user would be interested to see.



By performing these queries we are able to determine the quantities supplied of alternative fuel vehicles as well as understand changes in supply year over year.

**Hypothetical Use Cases**

Which are the main fuel types (LNG, hydrogen, ethanol etc.) that auto manufacturers have focused on adding as a compatible fuel for their line of vehicles?

Are there more alternative fuel compatible sedans as opposed to large SUVS?

Do heavy duty vehicles such as transit buses tend to be more alternative fuel compatible compared to light duty vehicles, if there are regulatory policies set by the government to encourage cleaner air for mass transit?