ETL Technical Report

Andrew Cowan, Daniel Fernandez, Kadien Peart

The proliferation of electric vehicles into the passenger car market necessitates providing documentation on the performance capabilities of these vehicles. Not all electric vehicles operate the same or have the same amount of power output or range, so as with conventional gasoline and diesel-powered engines, the abilities of each need to be compared. An individual who has a long commute and high frequency of trips necessitates a vehicle which has a longer range with a bigger battery. This project sets up data sources to perform data analysis which provides clarity for future electric vehicle purchasers.

To begin, we took two data sources from Kaggle, EV\_Performance\_data.csv and General\_EV\_data.csv, and intended to merge the two sources. We decided to perform the Extract and Transform steps in this project in a Jupyter Notebook file, since we were using two CSV files, we determined that the Jupyter notebook seemed the best development environment to perform these functions.

From there we began transforming the data. The General\_EV\_data.csv file had the manufacturer and model of the vehicles split, but the EV\_Performance\_data.csv file had the same information in one column. After reading these two files into separate DataFrames, we split the Name column in the Performance DataFrame from the DataFrame in order to break up the Make and Model, to rejoin with the Performance DataFrame. After this step was completed, we were able to merge the two DataFrames into one, dropping any rows which had any null values, returning a DataFrame with 64 rows of cleaned, non-null data.

To load the data, we connected the DataFrame to a PostgreSQL database that we created separately. From there, we created a table, filled the table with the data from the merged DataFrame, and are now able to query the database in PG Admin or from a Jupyter notebook.

The process was set out to make it easy for users to be able to compare different models of electric vehicles to suit their needs and budgets and we have accomplished this. If the user wishes to see all vehicles that have a range greater than 300 km and a top speed of over 200 km/h, that query is easy to run now. If the user wishes to find all vehicles under €50,000 or £50,000, they can find all of those models available in their currency.