# Kelly curtiss

# ETL Project Report

## Extract

The original data used for this project was gathered using two different sources. The first source was an API from [www.back4app.com](http://www.back4app.com) . A list of 100 car makers and models were gathered from this API from the year 2020. The second source of data was a web scrape from [www.goodcarbadcar.net](http://www.goodcarbadcar.net). This site provided several data tables on US vehicle sales by brand/make. The data table scraped provides sales figures from 2019 and 2020 and the change between the years. The html code from the scraped data table was then saved as an html file and outsourced as a csv file.

## Transform

The data from the API was saved into lists that were then transformed into columns of a pandas data frame. The csv file was imported into a pandas data frame as well. The csv file needed some cleaning. The headers needed to be renamed. The tables were then merged. The merged table also needed some cleaning. The index was reset and the NaN values were dropped.

## Load

The merged table was exported to a PgAdmin4-PostgreSQL data table. In order to confirm that the data table was in PgAdmin, a sql query was run in pandas. Finally, the merged table was exported to an Excel file for future data analysis.