**Extract:**

<https://www.kaggle.com/thaddeussegura/us-monthly-transportation-statistics> This source was downloaded as a csv.

<https://www.kaggle.com/mruanova/us-gasoline-and-diesel-retail-prices-19952021> This source was in the original xls format.

**Transform:**

Gasoline and diesel prices- File was converted from an excel worksheet to a CSV file. The dates were formatted to DATE to match the format in the Transportation Statistics file. The header row drop due and row (index) 1 was set at the new header. A source key row was also dropped as it did not have any useful information. Any data before November 28, 1994 was dropped since most of the fields were null. Columns for a specialized diesel was also dropped since there are very few data values and would just add noise to the data when compared to vehicle mileage. The column headings were also changed to be shorter and have no spaces

Transportation Statistics - The data was examined to determine how big it was and whether or not everything would be needed. Specific columns were extracted in the transportation statistics file to reduce the amount of data and use only road specific information. Items like border crossings and aviation were not needed. The date format was changed to date. The column headings were changed to be shorter, easier to read and have no spaces. Only years 1994 or sooner were extracted in order to match the other data series.

**Load:**

A sql database and a RDS was created. SQLalchemy was used to add the data files to sql and to create a table view

[myetlproject.chxgbxyu9xtv.us-east-2.rds.amazonaws.com](http://myetlproject.chxgbxyu9xtv.us-east-2.rds.amazonaws.com/)

username: FabulousErin

password: myETLproject2021

We wanted to see if the cost of gasoline and diesel is impacted by or has an impact on government spending, personal spending, and mileage. For government spending we would start at looking at bridge and pavement spending and for personal spending we would look at transportation cost index and auto sales.