**ETL-Project**

Extraction

We used two datasets. One from Kaggle and a one from a government website to compare Avocado prices (2015 - 2018) .

Source of data:

<https://www.kaggle.com/neuromusic/avocado-prices?select=avocado.csv>

<https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_w.htm>

Transformation:

After that we started cleaning up the datasets and removed columns/data that was either too specific or not relevant to the subject at hand. For example, we removed the information regarding bags of avocados due to not knowing how many avocados were in each bag. We also were not sure if the price per avocado in bags was the same or different compared to consumers purchasing individual avocados. We also used the grades of gas as a whole instead of drilling down into conventional and reformulated gas types because we were unsure which areas of the country utilizes conventional gas and which areas use reformulated gas.

Load:

Transferring our final output into a Database is the last step done in this project. We created a database and respective tables to match the columns from the final Pandas DataFrame using MYSQL and then we loaded the result.

Team: Robert Hartley, Jakob Brucks, Alex Cole, Abdulla Alrashdi and Jeff Bauman