ETL Project

Group 2:

* Beau Nguyen
* Edel Farah
* Estuardo Balderas
* Nader Elkhatib

Hypothesis:

* Display relationship between US States’ working force population and export revenue data.

Data Resources:

* US Census Bureau
  + API: Population Data
  + CSV: State Export Revenue

ETL Steps

**Extraction**

*Census Data (via Jupyter Notebook)*

* Run API to extract 3 sets of data: 1 set of census for all states, 2 sets of census on a gender level (for the working force only)

*State Export Revenue Data (via Jupyter Notebook)*

* CSV file downloaded from US Census Bureau website
* Import CSV’s data into jupyter notebook.

**Transform**

*Census Data (via Jupyter Notebook)*

* Create 3 data-frames: names = census\_pd, male\_pd, female\_pd.
* Consolidate into 1 data-frame: name = census\_merge

*State Export Revenue Data (via Jupyter Notebook)*

* Export into Pandas and create data-frame: name = state\_export\_df
* Line graph created to display each state’s export revenue.

*Census & Export Revenue Data (via Jupyter Notebook)*

* Create 1 data-frame to reflect each state’s total working population and export revenue.

**Load**

* Export CSV file containing State Export Revenue Data.
* Export 2 data-frames from Jupyter Notebook to PostgreSQL
  + state\_export\_df
  + census\_merge

Other Sources:

* GitHub Repo: <https://github.com/nelkha2/ETL-Project->
* API Source: <https://api.census.gov/data/2017/acs/acs5/variables.html>