**Eva and Suleman**

**ETL Project**

**Final Report**

Extraction

We extracted a dataset on Houston Population from Census.gov. <https://api.census.gov/data/2017/acs/acs1/groups.html>

The dataset was retrieved by population data based on the zip codes in Harris County and loaded into dataFrame.

The dataset for the “public schools in Houston” was downloaded from <https://data.world>

Transform

The population data was cleaned by resetting the index, dropping the extra index columns, converted and saved the file in csv format.

The necessary columns from the school dataset was selected. The columns were renamed and sorted by zip codes.

The data in both population and public school in Houston data were retrieved by common zip codes. And the irrelevant columns were dropped. And both datasets were merged.

Load

With our data, a postgressql database was created, and connection was established to the database via Jupyter Noebook.

We stored the following tables:

* census\_sorted;
* cleaner;
* merged.