**Extraction**

3 datasets from Kaggle and Data World. Based on county through all the States to 2015.

• Diversity Index from Kaggle.

• Unemployment from Kaggle.

• Median Income by county from Data World.

**Transformation**

Cleaning up the irrelevant variables.

**Load**

Transfer output into a Database. Create a database and tables to match the columns. Using MYSQL and then connect to the database using SQLAlchemy.

First, I use Pandas. I imported the diversity index worksheet from Kaggle to Python and got the data frame.

Then I created a new data frame by restating county and state. The reason for doing this is because the original location data has both county and state. So, I would like to split them.

The next step is to group up the counties’ date to each state.

The second step is to import unemployment rate dataset from Kaggle.

Because there is extra information for year and month, I cleaned the data frame.

In order to make two previous data frames merge, I need to standardize the state information by changing the unemployment data’s state to abbreviation format. So, I created a dictionary and used “loc” function to substitute my state information.

Like what I did for the first data frame, the next step is to group up the counties’ date to each state.

The third step is to import median income data by county from Data World. Like the previous ones, I also cleaned the data frame to leave the useful information only.

I rename “state code” to “state”

The next step is to group up the counties’ date to each state.

Once you have all three data frames all grouped up by state/county order, you can merge the data frame easily. Finally, I re-ordered the data frames’ items.

The Next Step is to load. The initial plan is to use PostgreSQL. But after multiple tries, I later figured my PostgreSQL setup was not correct and could not link both using the given instructions.

I later found out that this could be resolved by using MySQL and installing pymysql. Instead of using given instructions, I used “engine = create\_engine("mysql://root:@localhost/Project2")” to create the engine.

But I do not have MySQL in my computer. But eventually, with MySQL, I should be able to link both even if the error is there.