* The sources of data that you will extract from.

**Extract**

* <https://www.kaggle.com/mikejohnsonjr/us-counties-diversity-index>
* <https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations>
* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).

**Transform**

1. Clean the Median Income dataset to separate the county and state as PK
2. Clean the Diversity Index dataset to separate the county and state as PK
3. Renaming the columns for both Datasets to make it more consistent
4. Develop an ERD (Schema)

**Load**

1. Develop our dataset including two Tables (diversity & median Income) in pgAdmin
2. Inner Join Data Sets through the Primary Keys by running query in pgAdmin
3. Connect the pgAdmin local host to python through Jupyter Notebook
4. Analyze the median income based on the diversity in different counties of US

* The type of final production database to load the data into (relational or non-relational).

Our final data will be loaded to a relational database (pgAdmin)

* The final tables or collections that will be used in the production database.

1. Diversity
2. Median Income

* You will be required to submit a final technical report with the above information and steps required to reproduce your ETL process.

**To Be Developed**