ETL Project

Maria-Elena Gorini, Bhargav Andipara, Bhoomika Patel, Saniya Sule.

**Extraction**

We used 2 datasets from the public platform

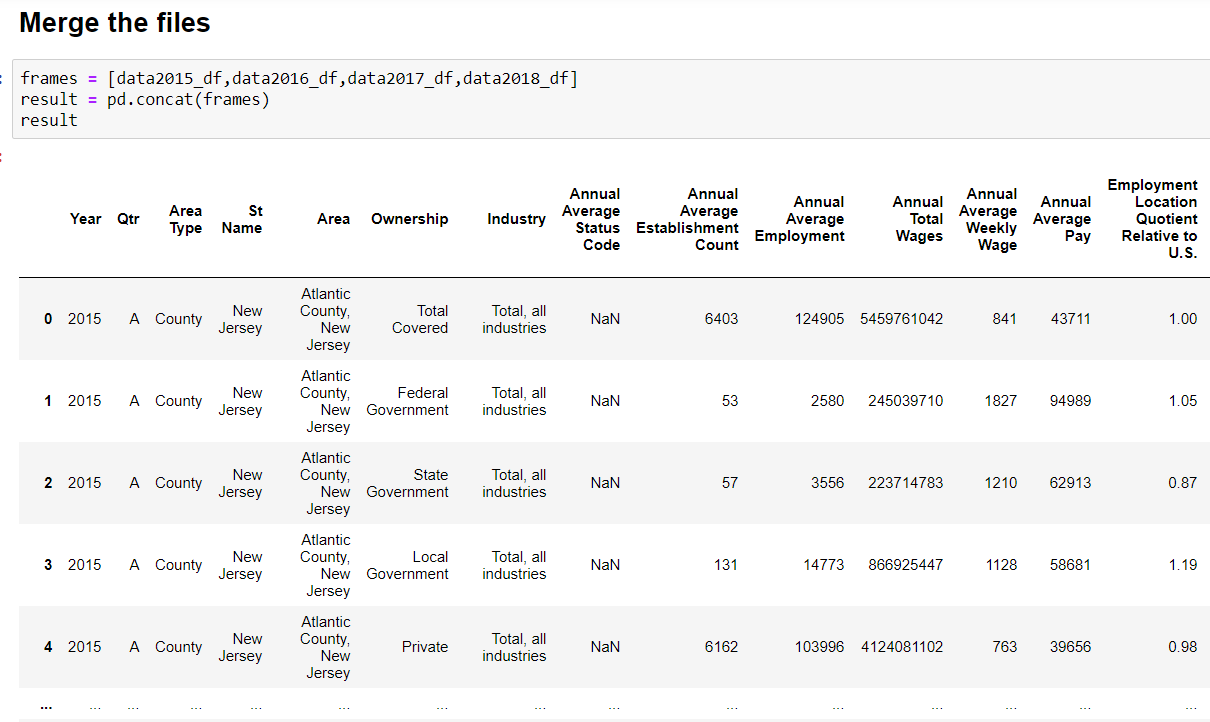
<https://www.bls.gov/> and <https://www.zillow.com/>

All of our data was based on county through all the States ranging over various years from 2015 to 2018.We extracted the data with API, csv and text.

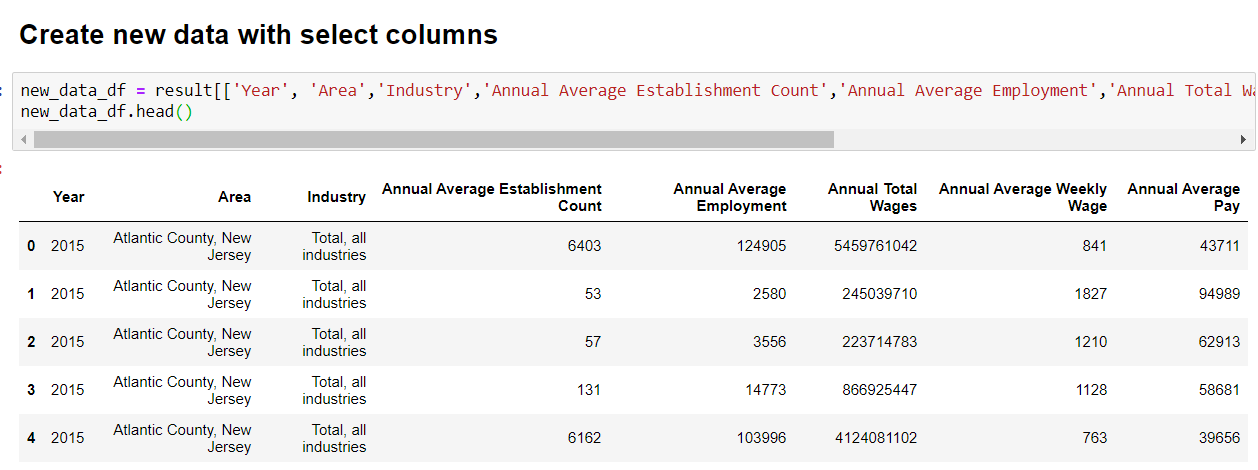
**Transformation**

Our first steps in cleaning up the datasets involved figuring out which variables were not relevant and cleaning up the data

1.Csv - we merged the data for years and concatenated it.

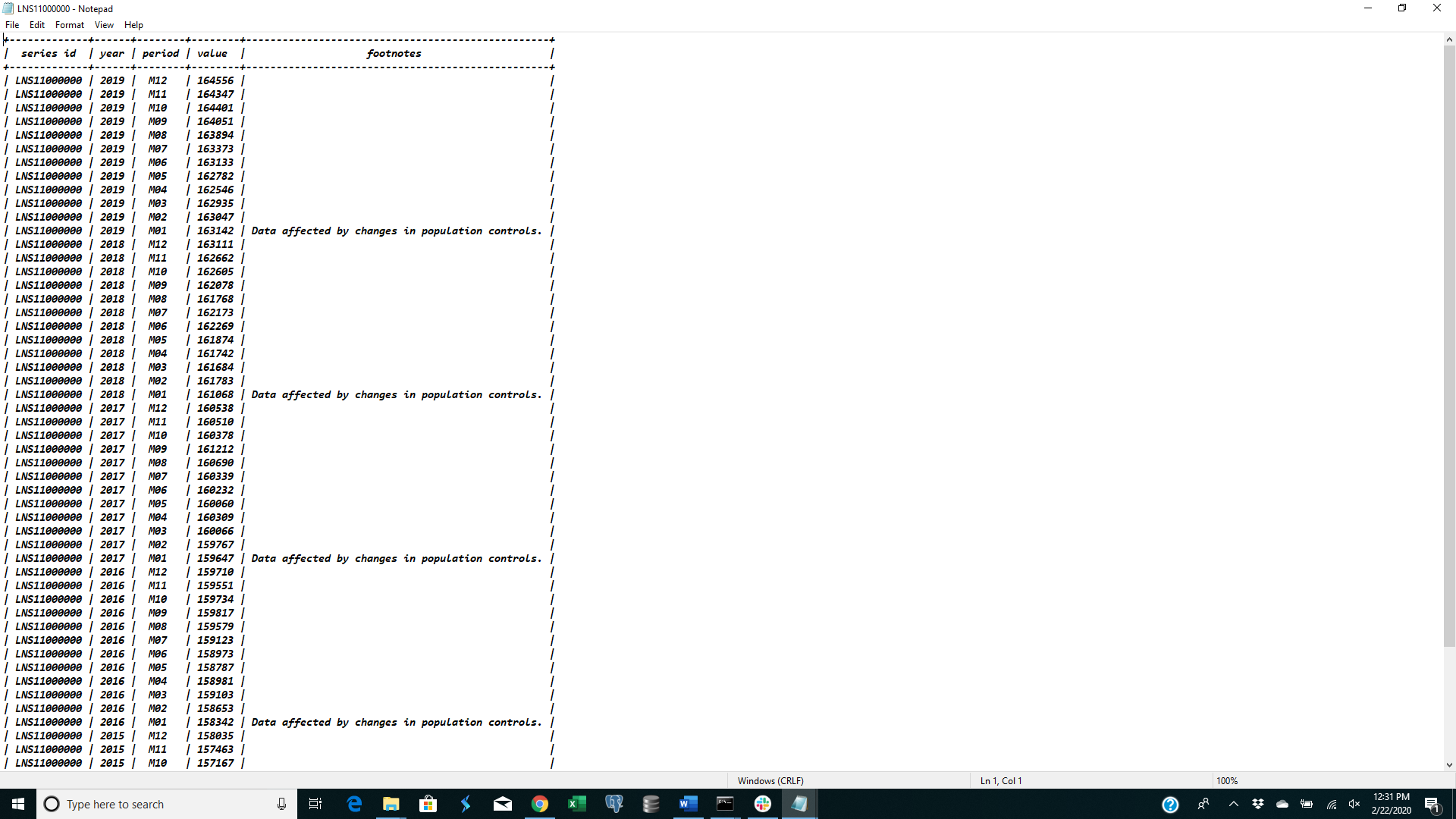


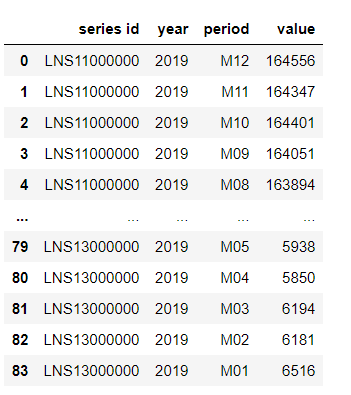
We created a new data with the required columns

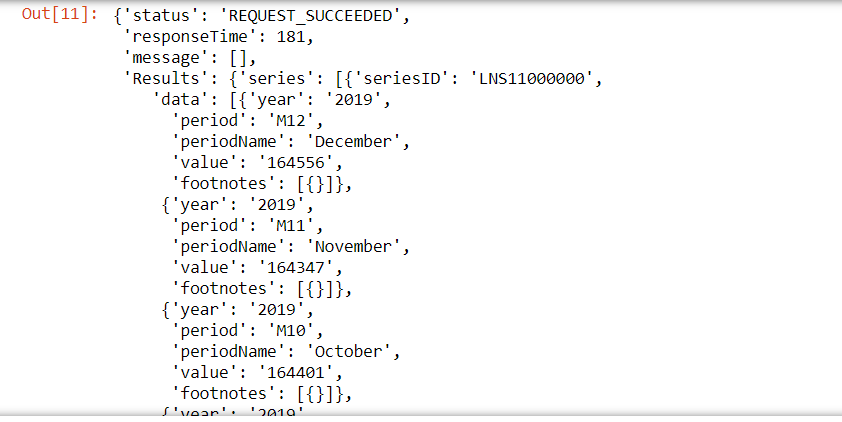


2.. Zillow API -XML Data extract from Zillow



3. BLS API – We extracted data from BLS through API and stored and created to data frame for Total, employment and unemployment and then created the data frame. 





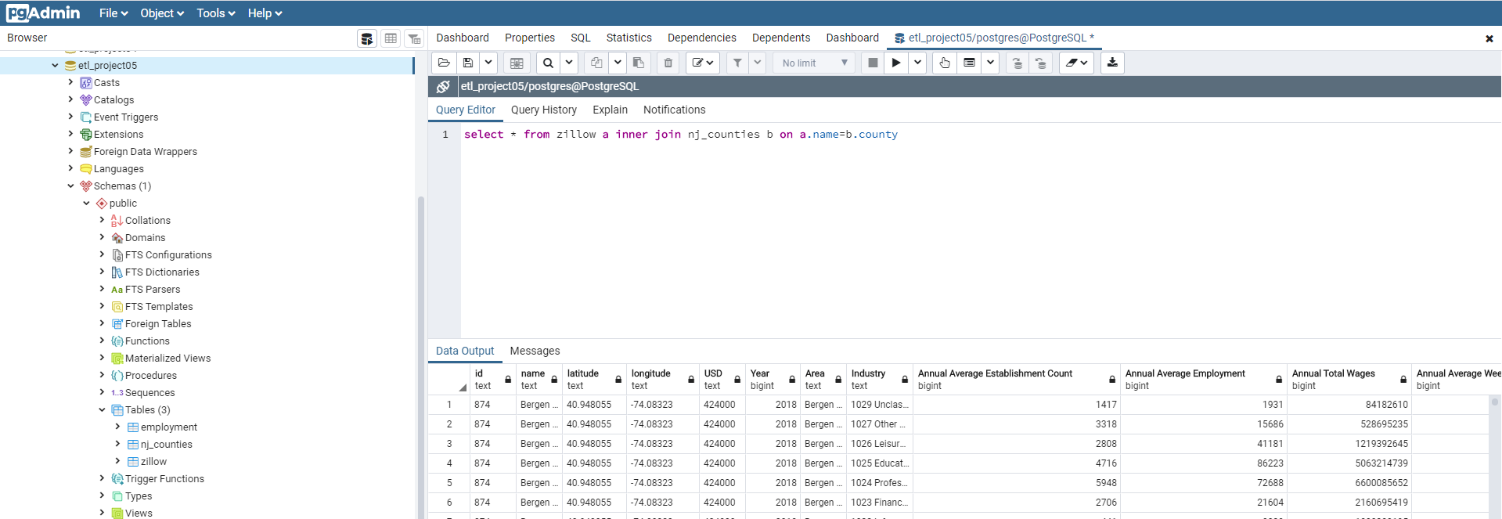
Json Data Zillow

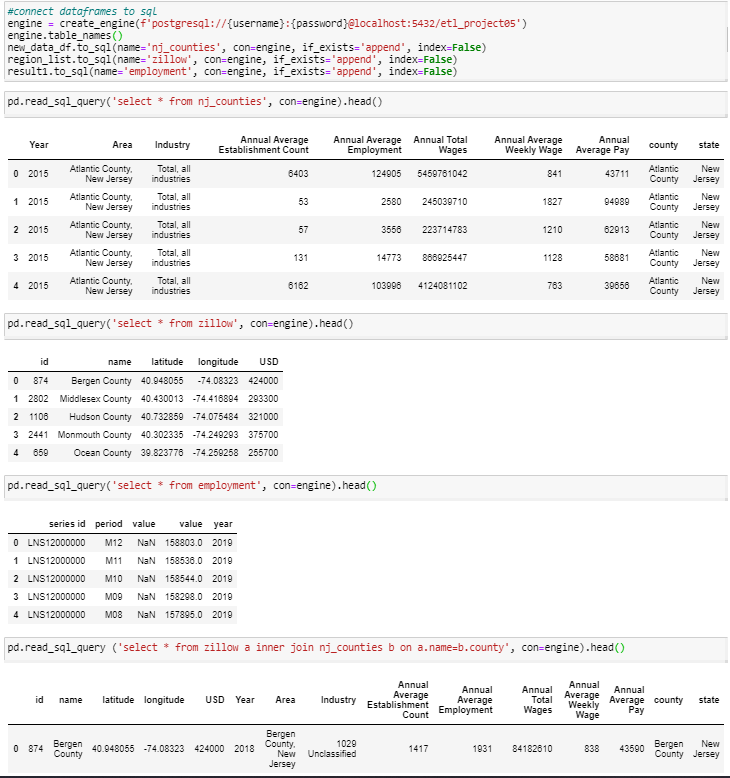


**Load**

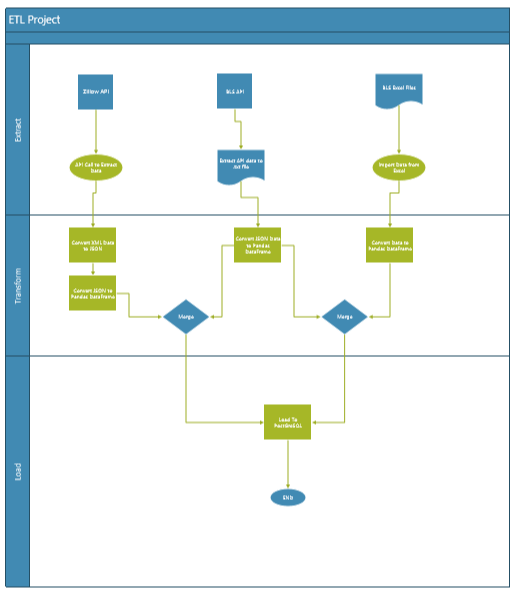
The last step was to transfer our final output into a Database. We created a database and respective tables

We queried the data using MySQL that returns desired outputs





**Summary**



We used these datasets so we could identify the unemployment rates per county for the state of new jersey including wages and median house prices. The final output has the tables for

* Civilian labor Population
* Unemployment and Employment
* Median Household Income
* Median house prices