ETL Project

Hannah G. Lee

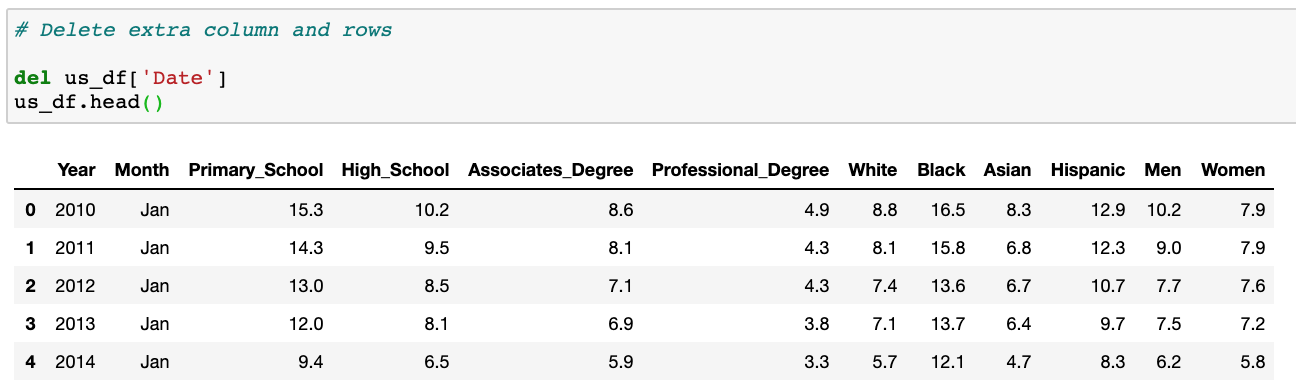
**Extract**

The focus on this project was to obtain data concerning unemployment rates over the past 10 years in the United States and Ohio unemployment rates, including data from 2020 (during the effects of Covid-19). The following two sources were utilized to obtain data:

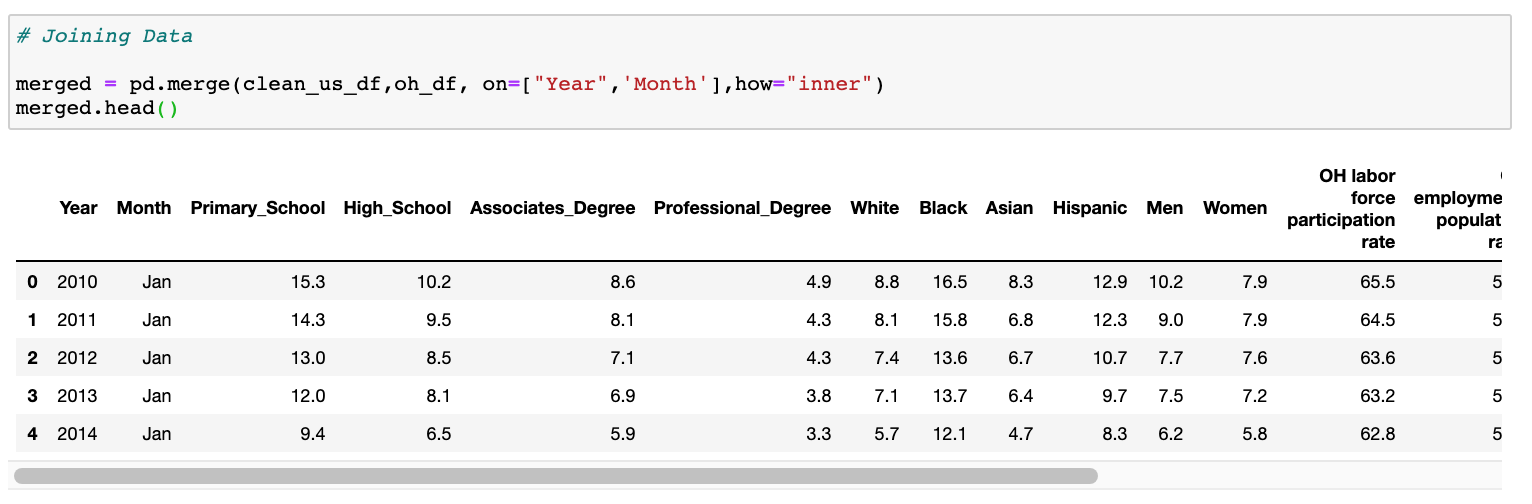
* Kaggle (unemployment\_data\_us.csv)
  + <https://www.kaggle.com/aniruddhasshirahatti/us-unemployment-dataset-2010-2020?select=unemployment_data_us.csv>
* U.S. Bureau of Labor Statistics (OhioUnemployment.xlsx)
  + <https://www.bls.gov/web/laus/laumstrk.htm>

**Transform**

Cleaning the datasets involved reviewing the data and eliminating unnecessary information, such as dropping NaN values and duplicate information (dates were listed in multiple columns).

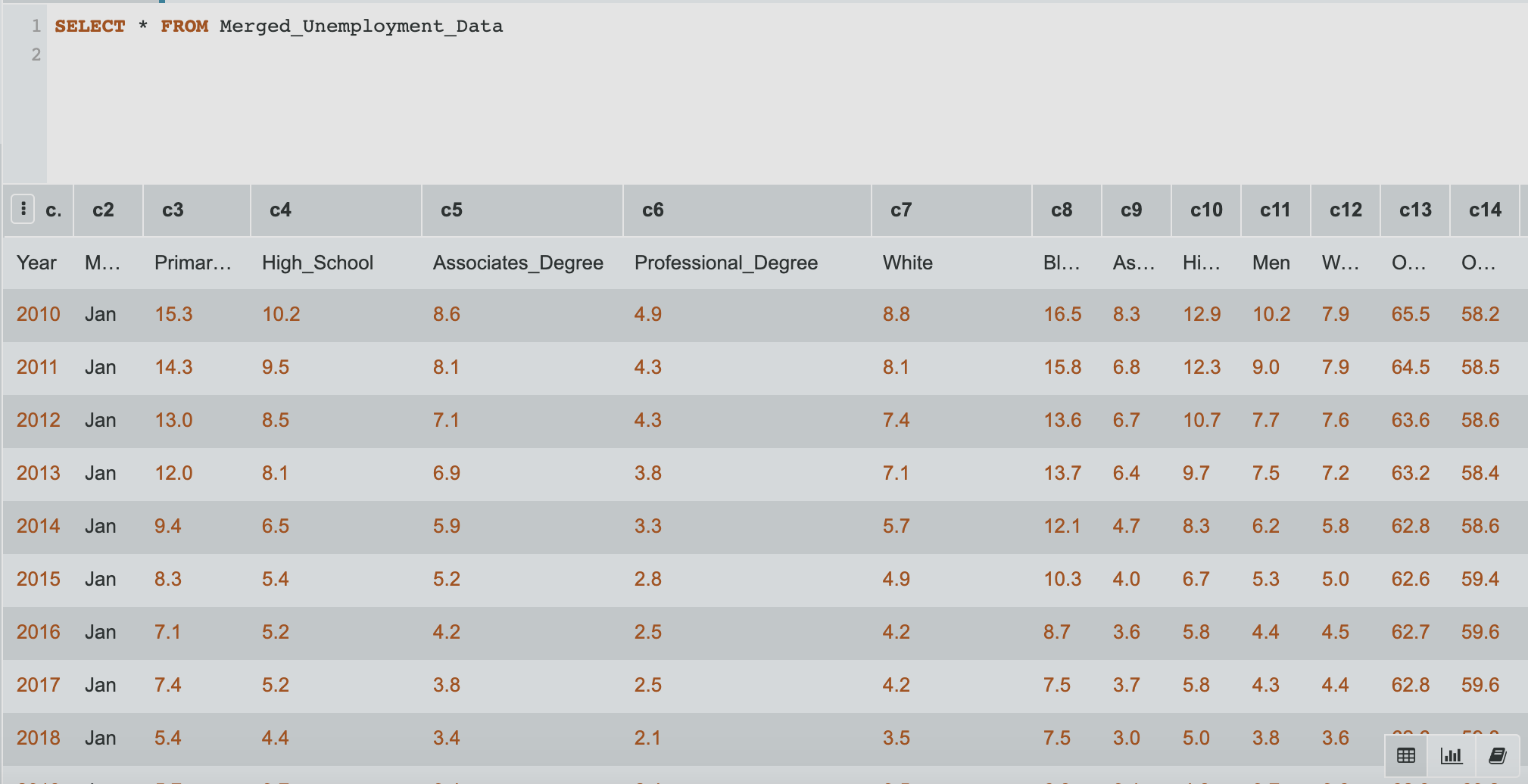


After cleaning the datasets, the data frames were merged on “Year” and “Month” using an inner join.



**Load**

Finally, the data was loaded into a database and SQLite was used to create various queries.



A picture containing table

Description automatically generated