ETL-Project Report

Extract:

The crime database came in an xls file from DataWorld. We uploaded into pandas and converted it to a data frame using pd.read\_excel function.

Transform:

We renamed the columns in our data frames to be simpler. We dropped columns that were not necessary to our analysis. In order to merge the two dataframes together, we had to have our key be city, state. To do this, we split the state column to get the state abbreviations by themselves and then joined that with the city column. For cities that were on the border and had two states, we dropped the second state and kept the first one.

Load: