Jake Thaker

**EXTRACT**

We found crime data sourced from data.world, Kaggle, and FBI.gov. All of our files were .xls documents, and we had to convert them into .csv files for easier data extraction within the jupyter notebook.[[1]](#endnote-1)

**TRANSFORM**

We had to clean all three data sources, as they had extraneous rows and columns for our final database. As an example, we filtered the Kaggle source to fetch only data from 2011, to match our other two sources which were pulled from the same year. The data.world file had additional summary rows at the top of the database, which were unnecessary for our analysis. Finally, we renamed all of the columns for ease of use within our final steps in SQL.

**LOAD**

The final database consisted of data from data.world and the FBI website. We joined the tables on the type of crimes committed, and our columns summarized the relationships to the victim as well as the weapons used in the crime. This data was chosen, as it required the least cleaning, and the data was relevant for comparison.

1. <https://data.world/ucr/expanded-homicide-data-tables>

   <https://www.kaggle.com/murderaccountability/homicide-reports>

   <https://ucr.fbi.gov/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/offenses-known-to-law-enforcement/expanded/expanded-homicide-data> [↑](#endnote-ref-1)