**EXTRACT**:

My original data sources were derived from:

1. https://www.kaggle.com/paultimothymooney/denver-crime-data

2. https://crime.denverpost.com/neighborhood

3. https://www.criminaljustice.ny.gov/crimnet/ojsa/countycrimestats.htm

The data, originally, was formatted as a CSV file.

**TRANSFORM**:

I transformed each DataFrame by pulling specific columns out of each CSV. I created variables and filtered out any columns irrelevant to my analysis. I pulled the columns related to the neighborhood/county, and crime count based on each neighborhood/county. I also pulled the offense types for each CSV to analyze the different types of offenses – as I would want to look at the crime rate but also, eventually, analyze the rate of the various types of crimes among certain neighborhoods/counties. For cleaning purposes, I renamed the columns extracted and set the index to neighborhood/county. I noticed that the CSV had duplicate neighborhoods/counties, so I aggregated the sum of neighborhoods based on the crime count and created a new DataFrame with the neighborhoods and aggregated crime rates.

**LOAD**:

The type of final production database to load the data into is relational. The DataFrames were loaded into PostgreSQL. PostgreSQL would allow me to store and access data points related to one another and I would be able to represent the data in tables/rows. A total of 8 tables were loaded into the database. I chose to load two types of DataFrames for each CSV (the DataFrame that was transformed with the filtered columns and the DataFrame with the aggregated total for each neighborhood/county). The transformed DataFrames would allow me to join on specific occurrences based on the place of occurrence/geographical location, as well as analyze the types of crimes (i.e. violent, property). The DataFrame with the aggregated totals would allow me to look at just the neighborhood/county and analyze the crime rate among each neighborhood/county.