ETL Project Outline

Group #2

Katherine Morriss, Cindy Pappas, Ergin Bostanci, & Matt Mullinazi

Datasets: CSV files each for “Crime in Los Angeles” and “Crimes in Boston”

Source: Kaggle

Project: We will use Jupyter notebooks to read in the CSV files as pandas data frames. Then we will need to clean up the data frames as the columns in each source are not exact. We will rename columns and drop duplicated entries. Then we will join the columns based on crime type. We will use mySQL to create two tables and merge the CSV data into those. Then we will connect these two through our Jupyter notebook.