The sources of data that you will extract from.

The data for restaurant inspections for Chicago and San Francisco is gathered from Kaggle.com. These are in CSV file format, the column definitions are relatively same but the data values , eg. “Low (Risk1)” in Chicago inspection is defined as “Low Risk” in San Francisco extract. The data is cleaned to address these issues and bring to common similarities using Pandas in Jupyter notebook.

Created two tables in PostgreSQL and loaded the data frames into these tables using sqlalchemy.

The data needed lot of cleaning before the table load and we limited the data values only to the Risk analysis criteria based on the hotel name and violations.

The final data is related based on the hotel name in each city.

Based on the load into the DB tables, we are able to identity the relation between violations in different cities for same name chains.