ETL Project Technical Report

 **Data Sources**

For our project we extracted data from the city of Chicago related to traffic crashes and red light camera violations.

Traffic Crashes: <https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if>

Red Light Camera Violations: <https://data.cityofchicago.org/Transportation/Red-Light-Camera-Violations/spqx-js37>

 **Transformations**

We read in two files, one for traffic crashes and a second file for red light violations. To prepare to join these two datasets we needed decompose the address column by splitting out the address by its elements. Using the resulting split elements, we grouped by street name and direction to get the total number of crashes and violations on each street. Following this step, we then joined these two output datasets on street name and direction to give us our final dataset.

 **Database Type**

Our database is relational, we used SQLAlchemy to store our database as a “.SQLite” file.

 **Database Schema**

There are two tables in our database. The first table includes crash data and the second table includes red light camera violations data.