**Project Report**

**E**xtract:

The original data source(s) were formatted as CSV files located at the links below.

* + <https://catalog.data.gov/dataset/motor-vehicle-collisions-person>
  + <https://catalog.data.gov/dataset/motor-vehicle-collisions-vehicles>

**T**ransform:

The data was originally compiled into two large tables with 20+ columns each. In order to get the data into a normalized form the data cleaning included creating multiple tables as well as dropping duplicate accident and person entries.

**L**oad:

The final database, accidents\_db, consists of 5 tables:

* + - Person – information on each person involved in any collision in the data
    - Crash – information on the collisions listed in the data
    - Car – information on the vehicles involved in the collisions listed in the data
    - Damage – information on damage(s) to vehicles or property as a result of the collisions in the data.

The final production database is a relational database in postgres. The decision was made to create a relational database due to the relationship of the source data tables to one another. This allowed the parsing of data to each individual table to include id’s that relate across the database, as such the relational database was the clear choice.