ETL-Project Report

EXTRACT- We extracted the original set of data from Kaggle.

<https://www.kaggle.com/usdot/nhtsa-traffic-fatalities>

This was imported as a CSV. The second and third set of data came from an API. The APIs are CitiPy API wrapper and Open Weather API wrapper.

TRANSFORM- The initial CSV was cleaned by dropping redundant columns. We dropped all columns except the State ID number, latitude and longitude. We zipped the latitude and longitude columns to get a new column with a latlng(zipped latitude and longitude) We ran the latitude and longitude through the CitiPy API to retrieve the city name. Then we ran the city name through the Open Weather API wrapper to retrieve the temperature. We created a data frame that included the State ID, City Name and Temperature.

WE PICKED ONLY THE FIRST 100 APPEARANCES OF ACCIDENTS IN THE CSV FILE

The data frame was then converted into a list of dictionaries.

LOAD- We loaded the final data frame into a Mongo Data Base. We chose a non-relational database for our final production so that each instance of an accident has the ability to be retrieved individually by state code or city name. This will allow a user to view the temperature in that city for all instances of accidents.

Further ETL transformation could be done by adding a Gmaps Api to give some information about each city and maybe pictures