ETL-Project Report

EXTRACT- We extracted the original set of data from Kaggle. This was imported as a CSV. The second and third set of data came from an API. The APIs are CitiPy and Open Weather API wrapper.

TRANSFORM- The initial CSV was cleaned by dropping non important columns. We dropped all columns except the State ID number, latitude and longitude.. We zipped the latitude and longitude together. We ran the latitude and longitude through the CitiPy API to retrieve the city name. Then we used the city name to retrieve the temperature from the Open Weather API Wrapper. We created a data frame that included the State ID, City Name and Temperature. The data frame was 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. This will allow a user to view the temperature in that state for all instances of accidents.