ETL Project Report

We used the Town of Cary’s CSV on traffic accidents and the CSV we previously used from Kaggle for our Project 1. We used Jupyter notebooks to clean up the data and make the same columns across both data frames for ease of comparison. The Town of Cary CSV pulls official accident data from the police department, whereas the CSV from Kaggle is filled with data reported from apps such as Mapquest. We thought it would be interesting to compare these two data sets to see if the Kaggle dataset matches the official dataset that’s submitted. We loaded our cleaned up data into PGAdmin for further analysis. We chose SQL because it’s a straight-forward language so we thought it would be more efficient to use in a comparison analysis.