Name: Bhargava Gaggainpali

Course: DSC 540 Data Preparation

Title: Monroe County Crash Data

Project Milestone 5 - Merging the Data and Storing in a Database/Visualizing Data

I have selected the ‘Monroe County Crash Data’ for the final project as part of the course. This dataset is the list of all the car crash or accidents happened in the county from 2003 to 2015. I have considered the data for the year 2015 for analysis as part of CSV file (one of the data sources). Retrieved the Holiday list from the give website to get the list of the holidays for the year of 2015 (source website data). Using API to the Weather details to get the weather conditions of the county for the whole year of 2015 (source API data). Using data from 3 data sources, I tried to find the trend of the car accidents across the year to see if the Holiday season or the weather conditions are relating to the increase or decrease of the car accidents.

I was little familiar of importing the data from files but reading the data from API and Websites using web scraping was a good learning. Initially I troubled in the web scraping methods to implement as ever website is structured differently. Representing the data using graphs and charts makes more sense and easy understanding of the data. Discussions and responses on other topics helped in gaining knowledge and others opinion on the topic. I believe doing the project in pieces as part of Project milestones helped in understanding and implementing all pieces of Data Wrangling. I would like to explore and practice more on the data wrangling steps, as they are very important before implementing any Machine Learning algorithms.