**Name: Bhargava Gaggainpali  
Course: DSC 540 Data Preparation**

**Project – Milestone1**

**Title: Monroe County Crash Data**

Identify Datasets

1. I choose to select 3 different data sets with a CSV file, API and a website for the project.

CSV data file from ‘catalog.data.gov’ data set ‘City of Bloomington/Traffic Data/Monroe County Crash Data\_2003 to 2015’.

1. Website data from ‘www.opm.gov’ to get the list of Holidays for a particular year.
2. API data from ‘http://api.openweathermap.org’ to get the Weather data for a city on a day.

CSV data file:

Government Data set ‘Monroe County Crash Data’ consists of the different car accidents which happened in a county for 12 years period 2003 & 2015.

Source: ‘City of Bloomington/Traffic Data/Monroe County Crash Data\_2003 to 2015’

<https://catalog.data.gov/dataset/traffic-data/resource/e46a5cc5-ed4d-4b8d-b750-18e6c9ec570e>

Number of Observations: 53943

Number of Columns: 12

Column details:

Master Record Number - Primary Key / Sequence Number

Year - Year number

Month - Month number

Day - Day number

Weekend? - Weekday / Weekend

Hour - Time of the day

Collision Type - Type of car collision

Injury Type - Type of Injury

Primary Factor - Reason for crash

Reported\_Location - Crash reported location

Latitude - Latitude coordinates

Longitude - Longitude coordinates

Website data:

Holiday list Data for a particular year. Here I am planning to pull the data for 2018 and 2019 to relate to the CSV data.

Source: [https://www.opm.gov/policy-data-oversight/pay-leave/federal-holidays/#url=2019](about:blank#url=2019)

API data:

Retrieve the Weather data from the ‘api.openweathermap.org’ for the city of ‘Louisville Kentucky’ to relate to the CSV data.

Source: [http://api.openweathermap.org/data/2.5/weather](about:blank)?

As part of the analysis, I am planning to relate the 3 different sources based on the Holiday dates and the ‘Monroe County’ weather conditions on the holiday days and count the number of car crashes. Analyze to see how the holidays, Weekends and Weekdays effect the car crash accidents of the county. Clean the data based on the several years and to certain weather conditions to see the maximum occurrence and type of incidents. Also as part of analysis, would like to see if the weather conditions and holidays are causing for more car accidents.

Reference:

<https://catalog.data.gov/dataset/traffic-data/resource/e46a5cc5-ed4d-4b8d-b750-18e6c9ec570e>

[https://www.opm.gov/policy-data-oversight/pay-leave/federal-holidays/#url=2019](about:blank#url=2019)

[http://api.openweathermap.org/data/2.5/weather](about:blank)?