Investigating Motor Vehicle Crashes of New York City: Where, When, Why

Project Proposal By: Team 1
Casey Delaney
Manasa Bobba
Sudheendra Katikar

1. The title of our dataset is 'Motor Vehicle Collisions - Crashes'. It consists of information from all police reported motor vehicle collisions in NYC over the period of 2012-2021, and has over 1.8 million records. The data is very informative since it's generated from a police report template that is quite comprehensive. The dataset was downloaded from data.cityofnewyork.us and can be found here.

2. We will try to answer some questions that arise out of reviewing the data, and possibly conduct some prescriptive analysis
   - Are certain localities more prone to accidents than others?
   - Do certain vehicles tend to have a recurring cause for the accident?
   - Does a street get particularly busy during any hour of the day?

3. Potential Methods-
   Given the type of our dataset, we are going to implement unsupervised learning methods. Most of our work will involve finding correlations and association between different attributes. Some methods we will use are-
   - Chi^2 test
   - Goodness of fit
   - Test of independence

4. Measuring Success-
   - Pearson correlation coefficient