Capstone Project 1 Proposal

Vehicle Collisions in NYC, 2015.Jan-2017.Feb

1. What is the problem you want to solve?

Living around the NYC area for a while, I’ve seen a lot of different traffic situations. Narrow streets and complicated traffic rules, as well as unreasonable traffic light durations and fearless pedestrians maybe all contributing to the vehicle collisions in NYC. Based on this dataset it’s possible to find out the most dangerous locations in 5 boroughs, the relations between number of collisions and time of a day (since there’re morning and evening peaks in a city like New York) or a year (due to influence of weather). Then it will lead us to think about what can be done to prevent more collisions.

1. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

I hope the NYPD and NYC residents care about the result since it’s closely related to their safety. Based on my analysis they’ll pay more attention to when and where it’s more likely to crash their cars or get hurt. It could also help the government to make decisions on modifying the traffic lights or street planning in future.

1. What data are you going to use for this? How will you acquire this data?

I’m going with this dataset on Kaggle: <https://www.kaggle.com/nypd/vehicle-collisions/data>

This was uploaded and then updated by NYPD, the latest data was the end of February, 2017.

1. In brief, outline your approach to solving this problem (knowing that this might change later).

After downloading data from the website I’ll first do data wrangling to figure out helpful information from the dataset. Then I’ll definitely run Exploratory Data Analysis as well as Statistical Analysis to generate many plots to help to visualize the results. Then I’ll compile the results so I can tell my story.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

There will be a report on Jupyter notebook that includes the code and plots, as well as a slide deck.