**PUI2016 Extra Credit Project Proposal**

**Vision Zero Crash Data Analysis**

**<Alexey Kalinin, GitHub handle – ak6129, NYU ID - ak6129>**

**Problem Description:** The question for this project is to find out tendency for fatalities due to road accidents in NYC for period from 2009 to 2016, localize some high-risk locations for 5 boroughs for 3 major groups – pedestrians, bikers, and motor vehicle occupants, and explore what is the change for last years for these locations. To answer the stated question, I’m going to analyze a data provided by Department of transportation of NYC.

# **Data:** The main resource for the data for this project is DOT specifically Vision Zero Data Feeds. The project will include two sets of data: fatalities distributed by year and fatalities distributed by month. Both data sets consist of total number of fatalities for period from 2009 to 2016, distribution of fatalities among 3 major groups, year or month for registered fatalities, and location within 5 boroughs in NYC. Considering that provided data sets cover period from 2009 - 2016 and information presented is sufficient I could conclude that it is enough to answer the question.

## **Analysis:** The plan is to visualize distribution of data set, to conduct Spatial analysis for fatalities data for 3 major groups registered for several years, to spot the most dangerous locations, and determine dynamics of fatalities throughout the years. I’ll identify the most dangerous fatalities through thresholding method of time series analysis, display fatalities trend by OLS regression of rolling mean (smoothing), and map fatalities hot spots after spatial autocorrelation.

**References**: The main reference for this project is Vision Zero for NYC and initiatives of DOT.

**Deliverable:** As the outcome of the project I expect to conclude the main dynamics or changes in distribution of fatalities within boroughs and how visible that changes are in terms of locations. I plan to plot a map of NYC with determined locations for each borough and distribution of high level fatalities among pedestrians, bikers, and motor vehicle occupants.