Team 3: Alexandra Plassaras, Brandon Wolff, Zachary Heinemann, and Stephanie Langeland

GR5069: Topics in Applied Data Science for Social Sciences

Professor Morales

February 22, 2017

Revised Project Description One Pager (Refer to the following link for the project plan/backlog: <https://github.com/amp5/QMSS_G5069_Applied_D_S/projects/1>)

**Project Description**

NYPD Stop, Question and Frisk data will be clustered by location to visualize the volume of stop and frisk events in New York City in 2015. After completing this exploratory data analysis, locations of interest (i.e., by volume of events) will be chosen to build forecasting models that predict the likelihood of being stopped and frisked in these locations. The goal of the study is to understand the impact of human bias in police officers when making the decision the stop, question, and frisk someone based on the demography of the vicinity.

**Insight**

Various relationships will be examined to understand if trends exist between the ethnicity and gender of the police officer and person being frisked, probable cause for being frisked versus popular crimes in that area, and use of force. These results will help identify trends in stop and frisk cases, which may lead to changes in police officer training aimed at reducing human bias and increasing subjectivity when making the decision to stop, question, and frisk. Depending on the results of such training based on this project, the NYPD may reconsider the stop, question, and frisk policy.

**Research Strategy**

The data will be cleaned and merged into one dataframe or tibble. Thereafter, variables/interaction terms will be added. In order to begin the analysis, the latitude and longitude coordinates for each observation will be determined from the addresses provided in the original dataset. Various data visualizations will be created by mapping all coordinates and calculating which spatial clusters, at the city block level, are statistically significant to identify regions of focus. Based on these focus locations, the final step will be to forecast the likelihood of being stopped and frisked based on location and demography.

**Data**

NYPD Stop, Question and Frisk data are collected by New York City legal guidelines. The dataset represents all interactions between police officers and civilians as well as the nature of each interaction i.e., use of force, time of day, date, and address. Since these data are collected by legal guidelines, the dataset should be complete. However, political pressure may cause police officers to omit events, threatening the completeness of the data. This potential missingness may introduce bias in the forecasting models skewing the prediction results.

**Output**

The prediction results of the forecasting models will examine whether use of force is associated with ethnicity, gender, probable cause, and/or other exogenous factors. The goal of this study is to identify whether human bias is evident in police officers’ decisions to stop and frisk individuals. The output of this data analysis will conclude with NYPD policy recommendations to reduce human bias and increase subjectivity when making the decision to stop, question, and frisk.