



## Introduction

The recent unprecedented increase in incidents and crimes in mass transit stations has heightened the anxiety and worries among daily commuters who depend on the public transit system. Identifying hot spots of criminal activity or predicting the potential next attacks is invaluable in the fight against crime, not only to respond timely to incidents but to prevent any if possible, by mobilizing the police resources beforehand.

Most recently, under Local Law 11 of 2012, the Police Department of the City of New York (NYPD) released an incident level data set detailing all reports of major felonies reported in New York City, updated quarterly. Included in this data set are the date and time of the incident, the crime committed, and the GPS coordinates. In this project, we examine the prevalence of felonies committed in Manhattan, and their proximity to the New York City mass transit system (including subway and bus stations.) The analysis of these datasets will enable end users (including the NYPD, community boards, and citizens alike) to make better, and more well informed decisions in allocating police and community resources and planning better routes for citizens while commuting.

#### Methods

To begin the analysis, data was acquired from the NYC Open Data Portal, along with the NYS MTA and analyzed in SAS for statistical analysis, and with QGIS for spatial analysis.



# **Identifying and Predicting Crime Hotspots in Mass Transit Stations using Big Data Analytics**

Austin Krauza and Dr. Soon Ae Chun School of Business, College of Staten Island, City University of New York



Based on the map below, it is evident that there are great hotspots of criminal activity around main transportation hubs- i.e. Penn Station, Times Square. However, upon running statistical tests, including linear regression, the number of fare swipes (weighted by station count) and number of subway entrances only accounts for 34.12% of the variability in the data. Therefore, it can be seen that there are some other factors that are influencing the hotspots of criminal activity that are arising in these census tracts. These factors can include population, demographics (gender/race,) and income levels.



### **Odds Ratios For "Subway Crimes"**

Crime	Total Crimes	Within .25 miles of Subway	More than .25 miles of Subway	Odds Ratio
GRAND LARCENY	15524	13382	2142	1.272
FELONY ASSAULT	3526	2806	720	0.6457
BURGLARY	2735	2342	393	1.0628
GRAND LARCENY OF MOTOR VEHICLE	799	609	190	0.5558
MURDER & NON-NEGL. MANSLAUGHTER	39	29	10	0.5133
RAPE	247	225	22	1.8209
ROBBERY	3129	2692	437	1.1045
Total	25999	22085	3914	-

The above chart shows the odds ratios, or increased probability of a certain felony offense occurring within a quarter mile of a subway station. (For reference, 5 blocks= 1 quarter mile) So, for example, it is 1.82 times more likely for a rape offense to occur within .25 miles of a subway station. However, it is no more likely for a felony assault to occur within a quarter mile. There are several hypothesized reasons for the overall increased odds ratios, including easier means of escape and an increased number of potential targets.

Based on our the results of the research that we have conducted, a greater number of patrols needs to be assigned to the areas of mass transportation hubs. 84.945% of the crimes studied in this project were within great proximity to a subway. While it appears that most locations in Manhattan are within a quarter mile of subway stations, patrol units can be dispatched to monitor subway entrances to look for suspicious persons, and potentially stop their escape.

**References** [1] NYPD 7 Major Felony Incidents- New York City Socrata Open Data Portal-[2] MTA Weekly Fare Swipe Data- New York State MTA Data Mine



## Conclusions