ENGDATA101 Project Proposal

Mehrad Haghshenas

November 26, 2022

This work is the proposal for the final project of the "**Introduction to Data Science"** course.

1. **Research Question**

   To what extent does the presence of police stations contribute to deterring crime in New York City (USA)?

2. **Background**

   It is found that "Regional access to local police infrastructure matters for crime outcomes." Accordingly, it is shown that closing local police stations can increase crime, particularly car theft and residential burglary. *(Blesse & Diegmann, 2022)* In recent years, there has been an ongoing trend of closing down local police stations in several countries. Such decisions are rooted in two primary reasons: 1- centralizing police organizations and law enforcement. 2- Due to the recession and Covid-19 pandemic, governments are looking for approaches to improve their economic climate. Cutting down on public expenses is known to be less costly than raising additional revenues. *(Alesina et al., 2019)* Therefore, policymakers tend to centralize police agencies mainly for budgetary reasons. *(Blesse & Diegmann, 2022)*

3. **Goal**

   The project aims to explore the effects of the presence and proximity of police stations on crime in New York City. This will indeed give an overview of whether regional access to local police stations can decrease the rate of crime or not. For this aim, two datasets are used: The primary dataset used for this goal is the NYPD crime dataset. The dataset can be found on *(NYPD Complaint Data Current (Year to Date)*, n.d.). The second dataset is the list of all NYC police stations, along with their locations. In brief, the task is creating a visualization to indicate whether being approximate to police stations or not has any impact on the number of crimes. In other words, as people might assume, *does the number of crimes decrease in the neighborhoods where a presence of a patrol is noted*? The intention is to use a variety of plots including geospatial plots to create a compelling and informative graph. The datasets will be wrangled and tidied as needed, and any plots for further clarification will be created.

4. **Dataset**

   As mentioned, there are two different datasets.

   1- The NYPD crime dataset is a breakdown of every felony, misdemeanor, and violation crime denoted by the New York Police Department during 2022. Precisely, the data consists of all the crimes in NYU between 1 January 2022 and 30 September 2022.

The Office of Management Analysis and Planning reviews and updates this data every quarter. This dataset has 396979 observations and 36 predictors, namely:

CMPLNT_NUM, ADDR_PCT_CD, BORO_NM, CMPLNT_FR_DT, CMPLNT_FR_TM, CMPLNT_TO_DT, CMPLNT_TO_TM, CRM_ATPT_CPTD_CD, HADEVELOPT, HOUSING_PSA, JURISDICTION_CODE, JURIS_DESC, KY_CD, LAW_CAT_CD, LOC_OF_OCCUR_DESC, OFNS_DES, PARKS_NM, PATROL_BORO ,PD_CD, PD_DESC, PREM_TYP_DESC, RPT_DT, STATION_NAME, SUSP_AGE_GROUP, SUSP_RACE, SUSP_SEX, TRANSIT_DISTRICT, VIC_AGE_GROUP, VIC_RACE, VIC_SEX, X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lat_Lon, New Georeferenced Column.

2- The second dataset consists of the police stations in NYC, with their longitude and latitude. Such a dataset was not found on the internet. Therefore, a list of all police stations in New York was obtained. (*Precincts - NYPD*, n.d.) Moreover, a dataset will be created using the address of the stations and Google Maps. It will hold 77 observations (the number of precincts) and three predictors: station name, latitude, and longitude.

5. Motivation

On a general level, the results of this project can be interesting to see whether the presence of a local police station in a location/neighborhood will reduce the number of crimes. On a larger scale, the results can be used for rethinking or consolidating the notion of centralizing police forces.

On a more personal level, this dataset was not the first choice. Initially, the choice was the "US accidents dataset," which consisted of all the car accidents from 2016-2019. After plotting the data, the far eastern and western cities were observed to have higher accident rates than mid-located cities. Specifically, California and Miami were hotspots of car accidents. Further research was intended to determine whether the poor quality of road infrastructure leads to this result—however, no datasets were found on the quality of roads in United States cities. I, therefore, concluded on this dataset, which also meets the purpose of working with geospatial data. Moreover, it is controversial whether having more police stations will lead to the safety of residents or not. The intention is only to examine the correlation between the presence of police stations and the crime rate. Note that the education levels of residents in specific neighborhoods, policies & laws, and family background can influence the crime rate. *All these aspects are excluded from the project, and only the effect of police stations on the crime rate will be examined.*

6. Reference

Blesse, S., & Diegmann, A. (2022). The place-based effects of police stations on crime: Evidence from

station closures. Journal of Public Economics, 207, 104605.

https://doi.org/10.1016/j.jpubeco.2022.104605

Alesina, A., Favero, C., & Giavazzi, F. (2019). *Austerity: When It Works and When It Doesn't (Illustrated)*. Princeton University Press.

NYPD Complaint Data Current (Year To Date). (n.d.). NYC Open Data. https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243/data

Precincts - NYPD. (n.d.). https://www.nyc.gov/site/nypd/bureaus/patrol/precincts-landing.page