CSE332 - Mini Project 1

Osmonaliev Zhanarbek

NYPD Crime Data in Manhattan Borough from 2019 to 2021

Motivation

I have been extremely interested in geo-locational data visualization for the past year. My last two projects involved data visualization on a 3D globe and pandemic simulation in a synthetic city. As for my next project, my ultimate goal is to use Machine Learning algorithms to predict the safety index based on one's current location. However, I believe that my first steps should involve less complexity and smaller scale. The dataset initially included 206,000 rows and 36 columns, which was reduced to 11,080 rows and 25 columns to only include Manhattan borough. The dataset was cleaned from empty or erroneous cells.

With this dataset, multiple visualizations can be rendered. The most straightforward visualization would involve crime mapping onto a map, where the user would be able to click on any crime incident and get additional information. As an additional feature, the heatmap of crimes can be derived from the dataset, allowing us to better understand neighborhoods in New York City.

Another approach would be to plot crime start date against the total timeline. Fitting the data into a Cumulative Distribution function would allow us to see which times of the year tend to be more dangerous than others.

Additional micro-observation can be observed if the crime start time would be plotted against a 24-hour timeframe. This perspective would allow us to derive interesting information on crime peak levels during the day.

My last but not the least approach would involve merging police station locations into the dataset to study how the proximity to the closest police station affects the crime rate around the area. The patrol office dataset will be merged into the main dataset according to the precinct assigned to a crime incident.

Dataset Background

The main dataset was sourced from NYC Open Data. The dataset includes various violation, misdemeanor and felony crimes reported to the NYPD across all boroughs in New York City. The dataset was reduced to 11,080 rows and 25 columns.

Link → NYPD Complaint Data Current (Year To Date)

Additional police station location in each precinct in Manhattan was taken from NYPD's website. It is an HTML table that was scraped and converted into .xlsx and .csv format.

Dataset Information

The total number of fields is 23, where 16 of them are quantitative and 7 are categorical. It's important to note that due to the nature of my intended project, some fields exist

merely to provide quantitative data; they serve as a source of additional information that will be displayed on the web application.

Field Name	Type	Description
CMPLNT_NUM	Quantitative	Randomly generated
ADDR_PCT_CD	Quantitative	The precinct in which the incident occurred
BORO_NM	Categorical	The name of the borough in which the incident occurred
CMPLNT_FR_DT	Quantitative	Exact date of occurrence for the reported event (or starting date of occurrence, if CMPLNT_TO_DT exists)
CMPLNT_FR_TM	Quantitative	Exact time of occurrence for the reported event (or starting time of occurrence, if CMPLNT_TO_TM exists)
CMPLNT_TO_DT	Quantitative	Ending date of occurrence for the reported event, if exact time of occurrence is unknown
CMPLNT_TO_TM	Quantitative	Ending time of occurrence for the reported event, if exact time of occurrence is unknown
JURISDICTION_CODE	Quantitative	Jurisdiction responsible for incident. Either internal, like Police(0), Transit(1), and Housing(2); or external(3), like Correction, Port Authority, etc.
LAW_CAT_CD	Categorical	Level of offense: felony, misdemeanor, violation
OFNS_DESC	Categorical	Description of offense corresponding with key code
PD_DESC	Categorical	Description of internal classification corresponding with PD code (more granular than Offense Description)
PREM_TYP_DESC	Categorical	Specific description of premises; grocery store, residence, street, etc.
RPT_DT	Quantitative	Date event was reported to police

SUSP_AGE_GROUP	Quantitative	Suspect’s Age Group
SUSP_RACE	Quantitative	Suspect’s Race Description
SUSP_SEX	Quantitative	Suspect’s Sex Description
VIC_AGE_GROUP	Quantitative	Victim’s Age Group
VIC_RACE	Categorical	Victim’s Race Description
VIC_SEX	Categorical	Victim’s Sex Description
X_COORD_CD	Quantitative	X-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Y_COORD_CD	Quantitative	Y-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Latitude	Quantitative	Midblock Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
Longitude	Quantitative	Midblock Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)