Exploring the Correlation Between the Crime Rate and Avenues in NY City

Aoune Barhoumi

June 16, 2020

1. Introduction/Business Problem

The question I want to answer in this project is the following: is there a correlation between the crime rate and the avenues in New York boroughs. It is know that the crime, in general, is highly correlated to many factors such as poverty, unemployment, low educational attainment. However, there might be a correlation between crime rate and the type of avenues in certain areas. For example, areas with high crime rate might have more clubs, bars whereas area with low crime rate might have more museums and parks. I am not saying that this is the truth but the project will allow me to check this hypothesis and determine this correlation.

Once we determine the avenues that are more prevalent in safe areas, we could probably generalize the correlation to areas where crime data is not available and predict the safety level of the area based on the type of avenues. I believe this analysis will be fun and interesting and I am excited to see the results.

The audience could be NY city authorities, social scientists, and criminologists. This information will help them better understand factors that govern the crime rate in the city and how we can control the crime rate by controlling the type of avenues.

2. Data Acquisitions

Data Sources:

Crime map based on NYPD complaint data current from the NYC open data (https://data.cityofnewyork.us/Public-Safety/Crime-Map-/5jvd-shfj). This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) for all complete quarters from 1974 till 2020. It has 108058 crimes with information about type of the crime, borough, date and time, and location (lat, lon). That data can be easily exported as csv and used for analysis.

NY city boroughs: Brooklyn, Manhattan, Bronx, Queens, Staten Island, is ranked based on the total number of crimes committed in that specific borough. The dataset categorizes crimes in 3 levels: Violation, Misdemeanor, and Felony. Violation, also known as infraction, is the most minor offense. Speeding ticket and public intoxication are examples of offenses that may fall under violation. Felony, on the other hand, are the most serious of offenses and require a more

thorough classification. Misdemeanor is more serious than a violation but less severe than a felony.

NY city boroughs can also be ranked based on the level of the committed crime. A coefficient will be associated for each crime level (1 for violation, 2 for misdemeanor, and 3 for felony) and a total point will be assigned to each borough.

I will use the Foursquare to determine the venues in different NY city 5 boroughs. The venues will be categorized based on type for each borough and a correlation between the crime rate and venue type will be extracted.

Data cleaning:

The data did not require a lot of cleaning. Minor columns dropping and renaming was required to create easy to understand dataframes for analysis

3. Results and Discussion

First, I ranked NY city 5 boroughs based on total number of crimes committed.

	Total_Number_of_Crimes
BROOKLYN	31028
MANHATTAN	26911
BRONX	23409
QUEENS	22097
STATEN ISLAND	4535

Figure 1: NY city 5 boroughs ranked based on the total number of committed crimes between the years 1974 and 2020

The total number of crimes can be also grouped by crime level. This grouping provides better understanding of the severity of the crimes committed in each borough

Boroughs	Level_of_offense	Total_Number_of_Crime
BRONX	FELONY	6699
	MISDEMEANOR	12570
	VIOLATION	4140
BROOKLYN	FELONY	10339
	MISDEMEANOR	15691
	VIOLATION	4998
MANHATTAN	FELONY	8492
	MISDEMEANOR	14782
	VIOLATION	3637
QUEENS	FELONY	7242
	MISDEMEANOR	11170
	VIOLATION	3685
STATEN ISLAND	FELONY	1090
	MISDEMEANOR	2544
	VIOLATION	901

Figure 2: Total number of crimes committed in NY city boroughs grouped by crime level.

The generate a most comprehensive ranking, I assigned a coefficient for each crime level based on its severity (1 for violation, 2 for misdemeanor, and 3 for felony). I ranked the boroughs based on the total number of points.

	Boroughs	Level_of_offense	Offense_desc	Total
3	BROOKLYN	FELONY	10339	67397
4	BROOKLYN	MISDEMEANOR	15691	67397
5	BROOKLYN	VIOLATION	4998	67397
6	MANHATTAN	FELONY	8492	58677
7	MANHATTAN	MISDEMEANOR	14782	58677
8	MANHATTAN	VIOLATION	3637	58677
0	BRONX	FELONY	6699	49377
1	BRONX	MISDEMEANOR	12570	49377
2	BRONX	VIOLATION	4140	49377
9	QUEENS	FELONY	7242	47751
10	QUEENS	MISDEMEANOR	11170	47751
11	QUEENS	VIOLATION	3685	47751
12	STATEN ISLAND	FELONY	1090	9259
13	STATEN ISLAND	MISDEMEANOR	2544	9259
14	STATEN ISLAND	VIOLATION	901	9259

Figure 3: NY city boroughs ranked based on total points

Both ranking (total number of crimes and total number of points) agrees and the final ranking would be:

1. Brooklyn
2. Manhattan
3. Bronx
4. Queens
5. Staten Island

Now, let us explore the venues in the 5 boroughs in NY city. First, I explored the venues based on boroughs directly. That using the coordinate of the boroughs in the search. The number of venues with this search was not sufficient.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
BRONX	25	25	25	25	25	25
BROOKLYN	23	23	23	23	23	23
MANHATTAN	30	30	30	30	30	30
QUEENS	5	5	5	5	5	5
STATEN ISLAND	5	5	5	5	5	5

Figure 4: Total number of venues in the 5 boroughs when searched with the boroughs coordinates.

To solve this problem, I searched the avenues using the neighborhoods in each borough. This search yielded a sufficient number of venues that could be grouped based on type for all boroughs. I generated a dataframe composed of the number of different venues type in each borough. 1

	Brooklyn	Manhattan	Bronx	Queens	Staten Island
Pizza Place	130.0	85.0	98.0	89.0	57.0
Coffee Shop	98.0	143.0	14.0	36.0	19.0
Bar	84.0	59.0	16.0	48.0	16.0
Deli / Bodega	68.0	33.0	47.0	71.0	40.0
Italian Restaurant	64.0	112.0	39.0	38.0	42.0
Bakery	63.0	70.0	23.0	56.0	8.0
Grocery Store	57.0	33.0	35.0	38.0	19.0
Chinese Restaurant	53.0	41.0	37.0	57.0	18.0
Mexican Restaurant	49.0	57.0	23.0	32.0	13.0
Ice Cream Shop	48.0	39.0	14.0	24.0	14.0
Café	48.0	86.0	6.0	25.0	6.0
American Restaurant	45.0	71.0	9.0	21.0	13.0
Park	43.0	71.0	23.0	25.0	15.0
Sandwich Place	41.0	53.0	37.0	43.0	20.0
Bagel Shop	40.0	25.0	3.0	23.0	22.0
Pharmacy	40.0	18.0	47.0	49.0	21.0
Donut Shop	38.0	12.0	45.0	52.0	17.0
Bank	37.0	10.0	33.0	46.0	17.0
Caribbean Restaurant	35.0	11.0	13.0	19.0	1.0
Sushi Restaurant	29.0	45.0	2.0	23.0	11.0

Figure 5: Venues in the 5 boroughs of NY city

Conclusion:

The data did not show a clear correlation between venues and crime rates in NY city boroughs. From this data analysis, the crime rate is not related to the type of venues in the city. It is possible that the correlation is not trivial and requires more intensive analysis where we analyze every neighborhood in NY city and try to explore possible trends.