The Battle of Neighborhoods: Chicago v. New York City

Adnaan Azeez

Introduction

Using venue data, neighbourhood demographics and crime data, can we determine which neighbourhood in Chicago and New York City is the safest from property crimes to open a new venue?

Suppose an entrepreneur wants to find out whether the location he/she is opening a new venue carries a high risk of property crime. We attempt to answer the question using k-Means Clustering and Linear Regression to find the best location to open a new venue safe from crime.

We will first organize the data from our datasets described in the Data Section, visualize the data and begin clustering analysis to explore the characteristics of both cities. Finally, we will perform a regression on crime rate against neighbourhood characteristics to determine significant determinants of crime and another regression against venue variables to determine the risk of property crime.

Further, the results of the analysis could provide useful insights for public planning officials for neighbourhood-level development and crime prevention by focusing education and other poverty-reducing resources.

Venue data is gathered using the Foursquare API, neighbourhood demographics are gathered from various sources from the US Census Bureau, and crime data is taken from the relevant police authority databases. The data processing and sources are described in detail in the Data section, followed by the Methodology, Results & Discussion, and Conclusion sections.

Data

Neighbourhood Data

First, we need to get location data for the two cities' neighbourhoods, this is completed using the `BeautifulSoup4` HTML scraper to get the data from:

- Chicago: https://en.wikipedia.org/wiki/Community_areas_in_Chicago#List_of_community_areas

We then use `geocoder` to assign coordinates to each neighbourhood and save to a data-frame

For New York however, we use an existing dataset that already contains all the geographical data as geocoder gives inconsistent results using the Bing API. The data-set is taken from https://geo.nyu.edu/catalog/nyu_2451_34572.

	Neighborhood	Latitude	Longitude
0	Rogers Park	42.010311	-87.670135
1	West Ridge	42.003090	-87.694931
2	Uptown	41.967789	-87.652428
3	Lincoln Square	41.974934	-87.687935
4	North Center	41.946815	-87.683388

Chicago Neighbourhood Data-frame

Venue Data

We then explore each neighbourhood with the coordinates acquired using the Foursquare API and the explore endpoint, a free call request, we are only interested in the venue id, coordinates, and category response fields:

```
https://api.foursquare.com/v2/venues/explore?&client_id=CLIENT_ID&client_se
cret=CLIENT_SECRET&v=VERSION&ll=LAT,LNG&radius=1000&limit=500
```

Initially, I had tried to also use the popularity, tips, hours, and popular hours response fields from the details endpoint, a premium call request, however with only 500 call requests per day with a Sandbox account, it proved to be unfeasible with over 25,000 venues which would take ~50 days to complete.

We get a max of 100 of the most popular venues at the time the notebook is run for each neighbourhood in a radius of 1km (500 meters in the case of New York, because its neighbourhood borders are smaller). We also delete any duplicates arising from overlapping neighbourhood radii.

A sample response looks like [this](). We use the json package to read the JSON file and extract the relevant fields, namely:

- `request["response"]['groups'][0]['items']['venue']['name']`
- `request["response"]['groups'][0]['items']['venue']['location']['lat']`
- `request["response"]['groups'][0]['items']['venue']['location']['lng']`
- `request["response"]['groups'][0]['items']['venue']['categories'][0]['name']`
- `request["response"]['groups'][0]['items']['venue']['id']`

We iterate through every neighbourhood to get a maximum of 100 venues per neighbourhood and assign it to the data-frame `c_venues` for Chicago, or `ny_venues` for NYC. See examples below.

[20]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	id
0	Rogers Park	42.010311	-87.670135	Morse Fresh Market	42.008087	-87.667041	Grocery Store	4ad3bfc1f964a52017e620e3
1	Rogers Park	42.010311	-87.670135	El Famous Burrito	42.010421	-87.674204	Mexican Restaurant	4b6ed827f964a5200bcd2ce3
2	Rogers Park	42.010311	-87.670135	Lifeline Theatre	42.007372	-87.666284	Theater	4afe1044f964a520962d22e3
3	Rogers Park	42.010311	-87.670135	Rogers Park Social	42.007360	-87.666265	Bar	536300da498ee44e63dcbc1d
4	Rogers Park	42.010311	-87.670135	Glenwood Sunday Market	42.008525	-87.666251	Farmers Market	516ae9c5498e8ffe6820c740

Chicago Venues Data-frame

[23]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	id
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop	4c537892fd2ea593cb077a28
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant	4d375ce799fe8eec99fd2355
2	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant	4c9e50e38afca09379b2ff15
3	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant	4c10f6aece57c92804a682d2
4	Wakefield	40.894705	-73.847201	Jimbo's	40.891740	-73.858226	Burger Joint	4c1bed4eb306c928140763b7

NYC Venues Data-Frame

Crime Data

We will use public arrest data published by the respective police authorities of Chicago and New York. The datasets include the following fields: unique id, arrest date, primary description, secondary description, address, and geographical location.

The datasets are taken from:

- Chicago: https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6

- New York: https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u/data

We import the `.csv` files as data-frames, clean up and remove any unnecessary columns:

[44]:

	id	date	primary	secondary	Latitude	Longitude
0	JD205528	04/09/2020 02:00:00 PM	CRIMINAL DAMAGE	TO VEHICLE	41.841609	-87.658034
1	JD177980	03/08/2020 02:15:00 AM	CRIMINAL TRESPASS	TO LAND	41.777671	-87.615561
2	JD218694	04/27/2020 10:49:00 PM	CRIMINAL DAMAGE	CRIMINAL DEFACEMENT	41.969365	-87.728061
3	JD218865	04/27/2020 06:20:00 PM	CRIMINAL DAMAGE	TO VEHICLE	41.803121	-87.609460
4	JD218362	04/27/2020 02:04:00 PM	BATTERY	DOMESTIC BATTERY SIMPLE	41.733015	-87.552709

Crime Data-frame

And finally, since we are not interested in all crimes, only property crimes, we drop all rows that do not include **THEFT, CRIMINAL DAMAGE, BURGLARY, ROBBERY, CRIMINAL TRESPASS, ARSON,** and **INTIMIDATION** for the Chicago dataset. And all rows without **PETIT LARCENY, CRIMINAL MISCHIEF AND RELATED OF, GRAND LARCENY, ROBBERY, FORGERY, BURGLARY, OTHER OFFENSES RELATED TO THEF, CRIMINAL TRESPASS, POSSESSION OF STOLEN PROPERTY, OFFENSES INVOLVING FRAUD, FRAUDS, BURGLAR'S TOOLS, THEFT OF SERVICES, THEFT-FRAUD,** and **ARSON** for the NYC dataset

This reduces the dataset to between 60,000–100,000 entries, we reduce it further and take a `random.sample` of about 20,000 entries for each dataset.

Finally, to compare neighbourhoods, we have to assign each crime and its location to a neighbourhood in the city, this is completed using `geopy.distance.geodesic`. We loop through each crime and measure its geodesic distance (see geodesic distance) to every neighbourhood and assume that whichever neighbourhood is the least distance away is the crime's corresponding neighbourhood. The final dataset should look like this:

[68]:

	id	date	primary	secondary	Latitude	Longitude	Neighborhood
8	JD218434	04/27/2020 03:50:00 PM	THEFT	OVER $500	41.751434	-87.716963	Ashburn
9	JD218189	04/27/2020 10:00:00 AM	CRIMINAL TRESPASS	TO RESIDENCE	41.924870	-87.690677	Logan Square
10	JD218465	04/27/2020 02:00:00 PM	THEFT	$500 AND UNDER	41.774994	-87.602763	Woodlawn
21	JD218727	04/27/2020 05:25:00 AM	CRIMINAL TRESPASS	TO RESIDENCE	41.723610	-87.537090	East Side
24	JD219697	04/27/2020 07:00:00 PM	THEFT	$500 AND UNDER	41.985809	-87.822379	Norwood Park

Final Crime Data-frame

Demographic Data

Demographic data for Chicago and NYC were collected from the US Census Bureau and compiled from various sources.

- Chicago: [Population and Ethnicity](), [Unemployment and Poverty Rate](), and [PCI and Population aged 25 without High School Diploma]()
- New York: [NYU Furman Center CoreData.nyc]()

The data has been compiled into CSV files: `chicago_demographics.csv` and `nyc_demographics.csv` which are included in the [GitHub repository]() which include the 5-year averages (2013-2018) for each variable. We then import them as data-frames:

[75]:

	Neighborhood	unemployment	poverty rate	population	pop aged 25 without hsd	asian_pop	black_pop	hispanic_pop	white_pop	other_pop	pci
0	Rogers Park	0.075	0.227	53346.0	0.182	0.043227	0.238012	0.240355	0.448543	0.029862	23939
1	West Ridge	0.079	0.151	77074.0	0.208	0.214755	0.128682	0.202117	0.408802	0.045644	23040
2	Uptown	0.077	0.227	55042.0	0.118	0.106010	0.178355	0.158315	0.526471	0.030849	35787
3	Lincoln Square	0.068	0.095	47927.0	0.134	0.103094	0.068124	0.176915	0.610992	0.040875	37524
4	North Center	0.045	0.071	36157.0	0.045	0.044805	0.086622	0.099649	0.730232	0.038692	57123

Chicago Demographics Data-frame

[76]:

	Neighborhood	unemployment	poverty rate	population	pop aged 25 without hsd	asian_pop	black_pop	hispanic_pop	white_pop	other_pop	median income
0	Astoria	0.059342	0.149257	171100	0.140745	0.154099	0.067739	0.267112	0.480503	0.030547	64569.48051
1	Bay Ridge	0.060594	0.163629	123691	0.182881	0.250729	0.020173	0.171766	0.535353	0.021979	68877.27333
2	Bayside / Little Neck	0.045744	0.080287	117271	0.117773	0.438910	0.019587	0.116965	0.401430	0.023108	83002.87339
3	Bedford Stuyvesant	0.087078	0.284157	145895	0.179234	0.032716	0.494498	0.187020	0.262244	0.023522	48657.07822
4	Bensonhurst	0.067998	0.184279	191525	0.256549	0.403927	0.010737	0.158654	0.404201	0.022481	53834.77753

NYC Demographics Data-frame

Note that, for NYC, we are looking at sub-borough areas now instead of neighbourhoods (the only data that was available), for clustering, we will continue using the neighbourhoods, but for the regression we will use sub-borough areas instead. And we have median income instead of PCI, for individual city clustering, this will not be an issue, however when we perform the regression analysis, we will have to drop the income columns

Methodology

k-Means Clustering

Next, we prepare the venue dataset for the k-Means Clustering analysis, we're only interested in the venue categories in each neighbourhood, so we use one-hot encoding, where we make each venue be the function of every category in the dataset, a category takes the value of 0 if the venue does not belong in the category and 1 if it does. We then group the venues by neighbourhood and take the mean value of each category and sort the categories in descending order.

We do the same for the crime dataset except on the basis of crime categories.

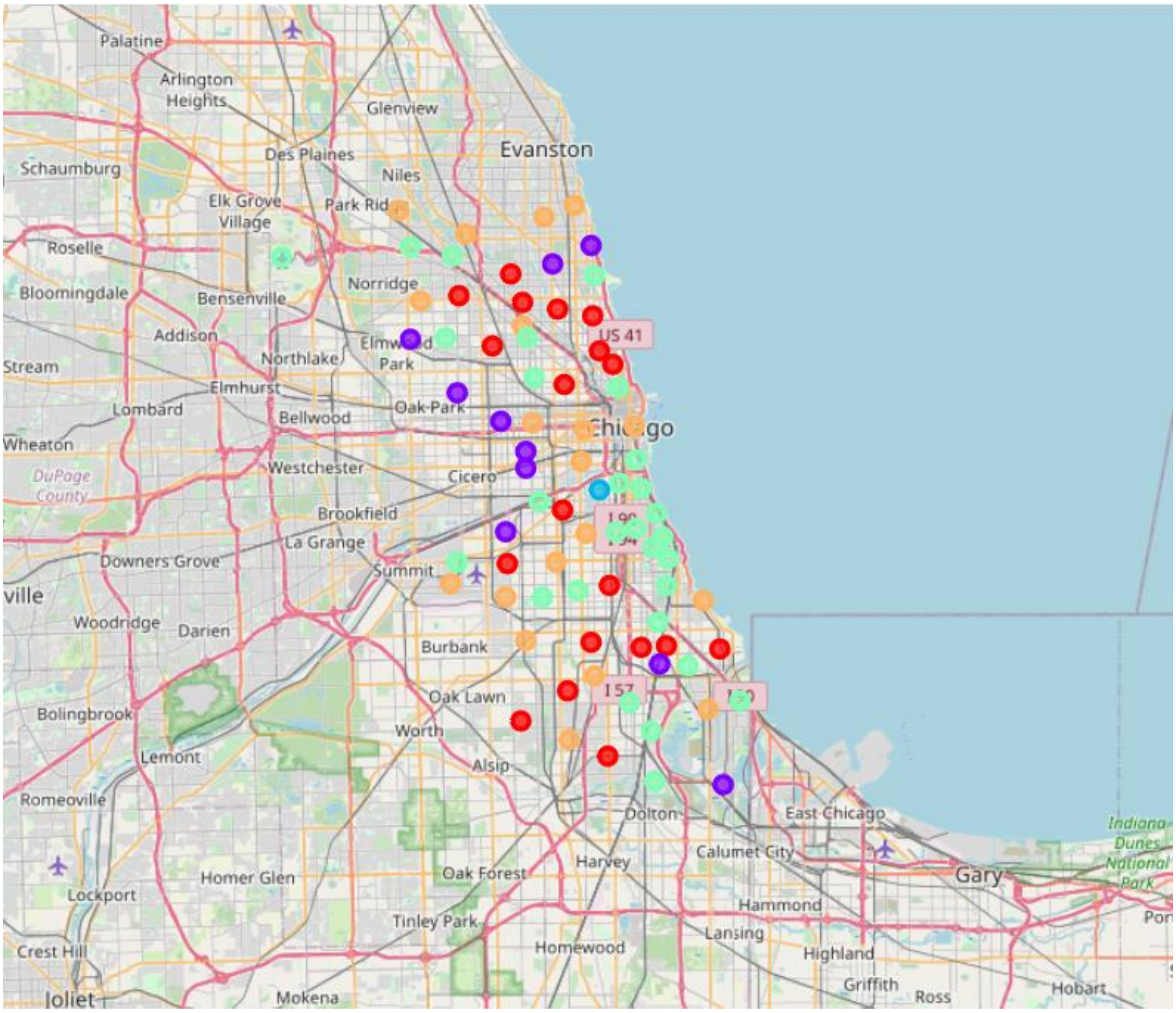
	Neighborhood	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	American Restaurant	Amphitheater	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Astoria	0.0	0.000000	0.0	0.0	0.00	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.010000
1	Bay Ridge	0.0	0.000000	0.0	0.0	0.03	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.010000
2	Bayside / Little Neck	0.0	0.000000	0.0	0.0	0.03	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.020000
3	Bedford Stuyvesant	0.0	0.012048	0.0	0.0	0.00	0.0	0.0	0.0	0.012048	0.0	0.012048	0.0	0.0	0.012048
4	Bellerose / Rosedale	0.0	0.000000	0.0	0.0	0.00	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000

Venues One-hot Encoded

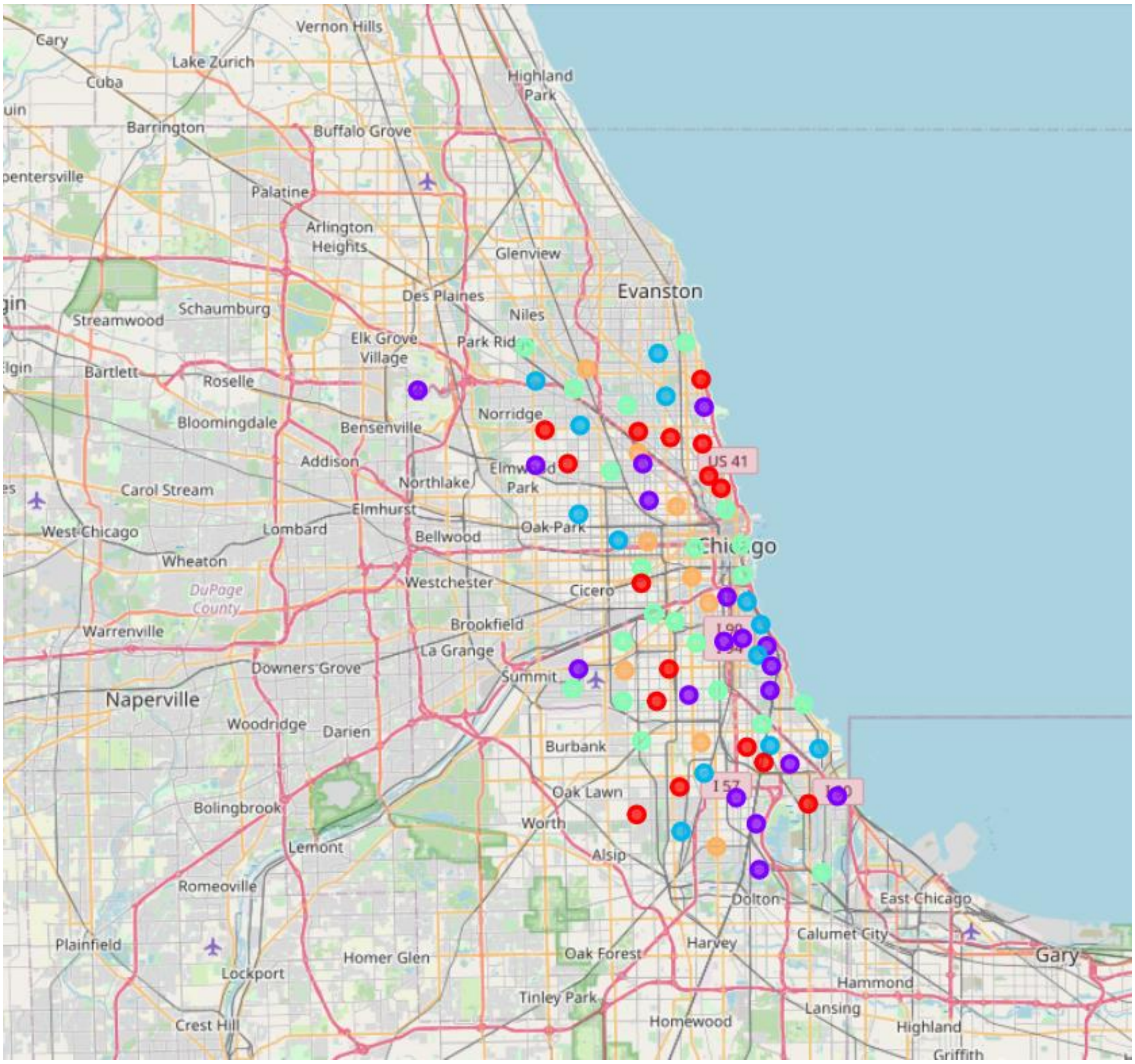
	Neighborhood	ARSON	BURGLAR'S TOOLS	BURGLARY	CRIMINAL TRESPASS	FORGERY	FRAUDS	GRAND LARCENY	OFFENSES INVOLVING FRAUD	OTHER OFFENSES RELATED TO THEF	PETIT LARCENY	POSSESSION OF STOLEN PROPERTY	ROBBERY	THEFT OF SERVICES	TI FI
0	Astoria	0.000000	0.010753	0.107527	0.053763	0.043011	0.026882	0.193548	0.016129	0.026882	0.333333	0.037634	0.150538	0.000000	0.0
1	Bay Ridge	0.000000	0.026846	0.026846	0.067114	0.100671	0.020134	0.080537	0.020134	0.026846	0.402685	0.073826	0.154362	0.000000	0.0
2	Bayside / Little Neck	0.006803	0.013605	0.122449	0.020408	0.040816	0.034014	0.244898	0.013605	0.013605	0.231293	0.040816	0.204082	0.000000	0.0
3	Bedford Stuyvesant	0.000000	0.019417	0.119741	0.038835	0.122977	0.022654	0.090615	0.006472	0.064725	0.294498	0.025890	0.181230	0.012945	0.0
4	Bellerose / Rosedale	0.000000	0.007576	0.075758	0.106061	0.219697	0.015152	0.068182	0.030303	0.015152	0.287879	0.098485	0.068182	0.000000	0.0

Crimes One-hot Encoded

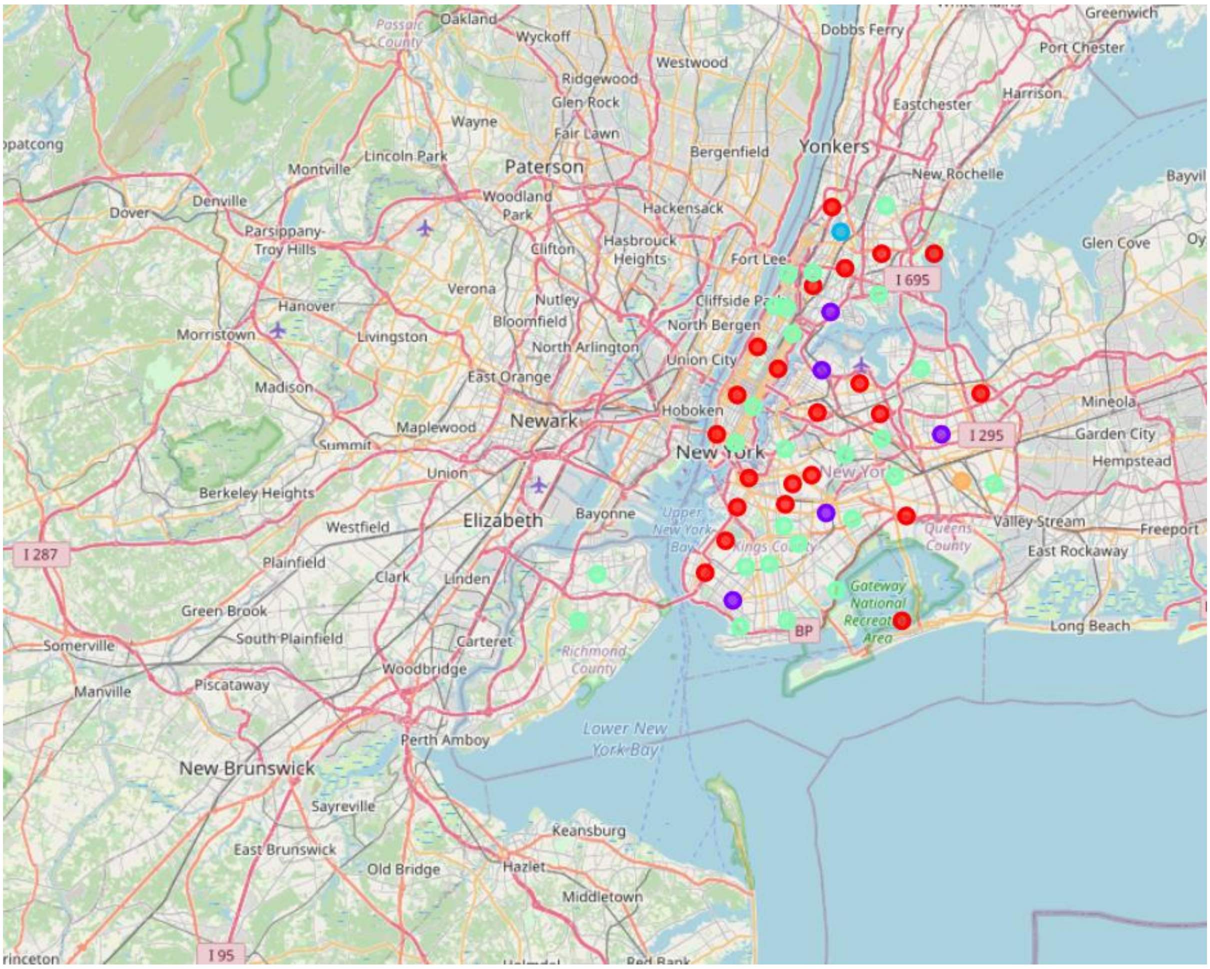
We can now begin the clustering process; we choose a k of 5 which is close to the optimum k for all datasets and we require a consistent number of clusters for the regression analysis. After acquiring the cluster labels, we then use folium to map the neighbourhoods with different colour depending on their cluster labels:



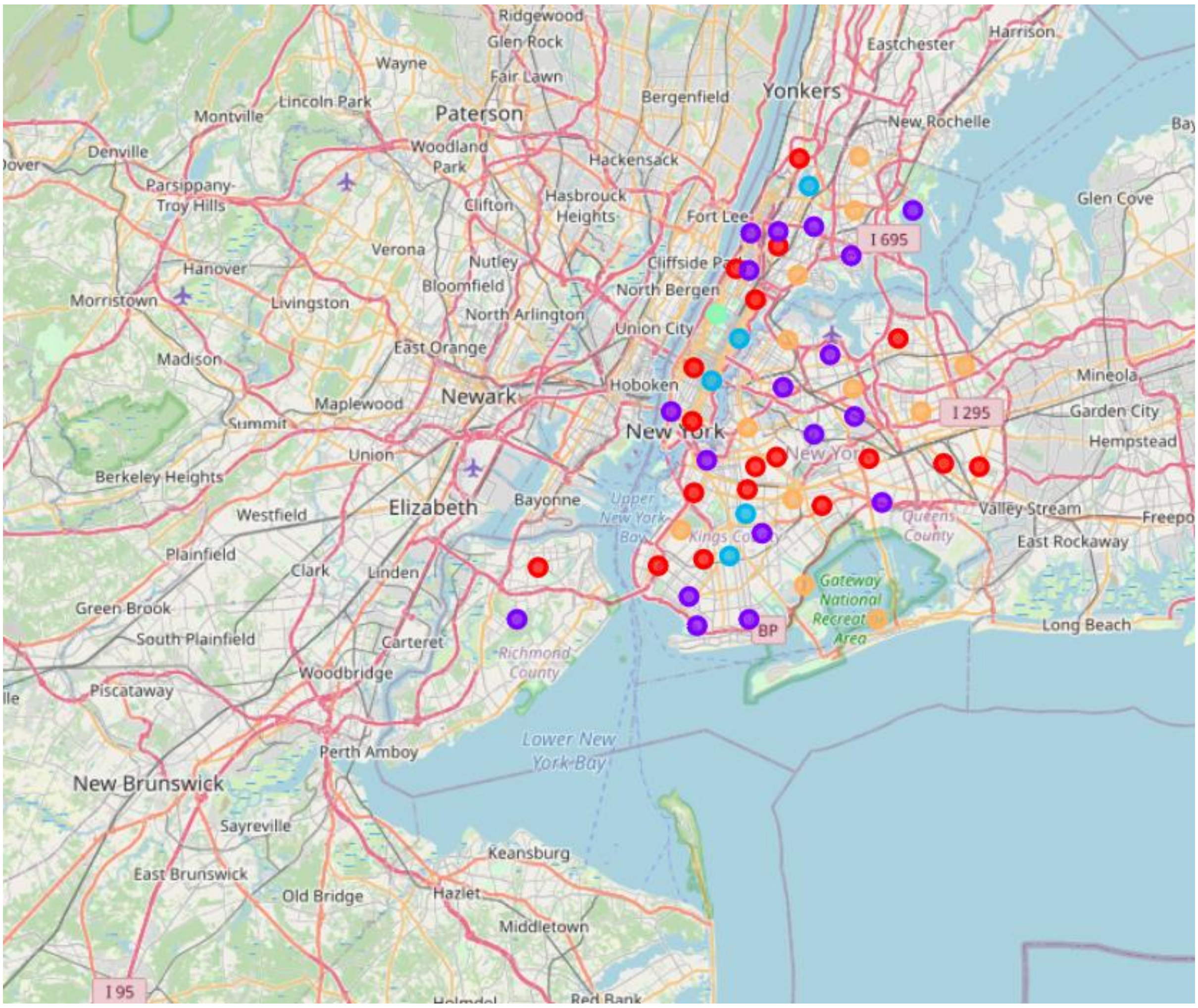
Chicago Venues Clustered



Chicago Crimes Clustered



NYC Venues Clustered



NYC Crimes Clustered

A basic description of each cluster is provided in the notebook inside the GitHub repository. Overall, NYC has a more diverse array of venue categories in its clusters than Chicago does.

Linear Regression

For the regression analysis, we first need to prepare a data-frame with all the variable we want to explore. In this model, we will be using the data-frame created from the demographic dataset along with the crime ratio of each neighbourhood/sub-borough which we get simply by dividing the number of crimes in each neighbourhood by the total number of crimes in the city. We also add columns for the cluster labels from the venue clustering to include the effects of the venue categories. Our dependent variable (y) is the crime ratio, while our independent variables (X's) are unemployment rate; poverty rate; population size; population aged 25 without a high-school diploma; Asian population ratio, Black population ratio, Hispanic, population ratio, White population ratio, and Others population ratio,4; Per Capita Income (median income for NYC); and the dummy variable for venue cluster labels 1–4 (venue cluster labelled 0 is not included as it is used as the benchmark dummy, avoiding the Dummy Variable Trap and perfect collinearity).

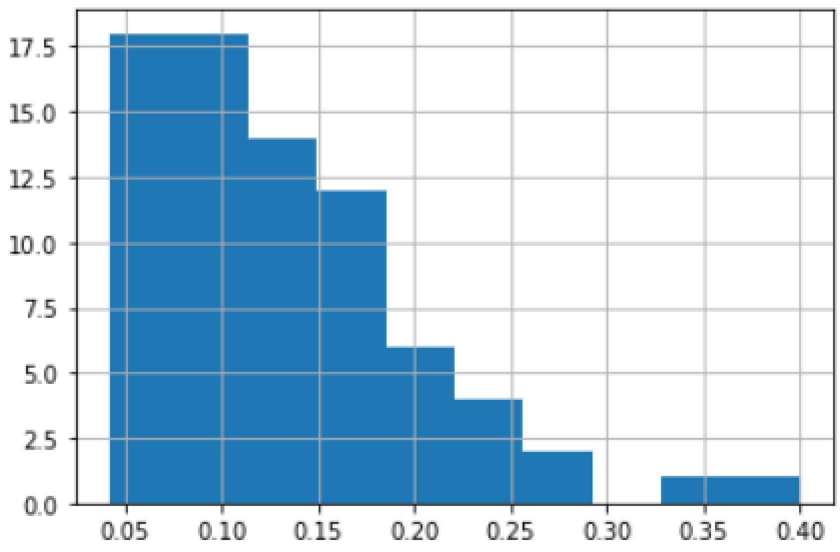
	unemployment	poverty_rate	population	without_hsd	asian_pop	black_pop	hispanic_pop	white_pop	other_pop	pci	cluster_1	cluster_2	cluster_3	cluster_4	crime_ratio
0	0.075	0.227	53346.0	0.182	0.043227	0.238012	0.240355	0.448543	0.029862	23939.0	0	0	0	1	0.019490
1	0.079	0.151	77074.0	0.208	0.214755	0.128682	0.202117	0.408802	0.045644	23040.0	0	0	0	1	0.013914
2	0.077	0.227	55042.0	0.118	0.106010	0.178355	0.158315	0.526471	0.030849	35787.0	0	0	1	0	0.012024
3	0.068	0.095	47927.0	0.134	0.103094	0.068124	0.176915	0.610992	0.040875	37524.0	1	0	0	0	0.010715
4	0.045	0.071	36157.0	0.045	0.044805	0.086622	0.099649	0.730232	0.038692	57123.0	0	0	0	0	0.011490

Chicago Regression Data-frame

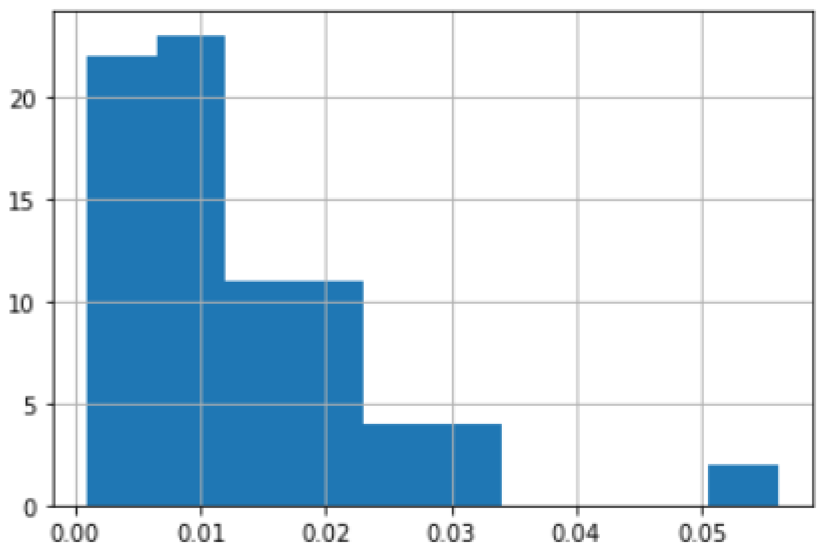
	unemployment	poverty_rate	population	without_hsd	asian_pop	black_pop	hispanic_pop	white_pop	other_pop	median income	cluster_1	cluster_2	cluster_3	cluster_4	crime_ratio
0	0.052985	0.200258	174560	0.159116	0.091365	0.212375	0.301798	0.367099	0.027363	65995.32978	0	0	1	0	0.021143
1	0.092744	0.337528	125108	0.262010	0.074297	0.292495	0.476014	0.137318	0.019876	34626.18585	0	0	1	0	0.031312
2	0.049128	0.108996	137478	0.114426	0.136783	0.042821	0.138292	0.661047	0.021056	83661.65114	0	0	1	0	0.012298
3	0.092942	0.210649	224898	0.280442	0.029888	0.080457	0.683003	0.189328	0.017324	52977.54067	0	0	1	0	0.017122
4	0.059342	0.149257	171100	0.140745	0.154099	0.067739	0.267112	0.480503	0.030547	64569.48051	1	0	0	0	0.008798

NYC Regression Data-frame

When exploring the shape of our variables, I noticed that all the variables including the dependent variable are highly skewed to the right, to solve for skewness, I standardize the data before using a logarithmic transformation which makes the data more resemble a normal distribution which is necessary for OLS regression



Unemployment Rate Before Logarithmic Transformation



Crime Ratio Before Logarithmic Transformation

Chicago

After performing the regression, for the Chicago dataset, we get an R^2-score of 0.6878 and an adjusted R^2-score of 0.6162, or that 61.62% of the variation in crime ratio is explained by this model. The coefficients of the variables are presented below:

0	unemployment	-0.016104
1	poverty_rate	0.522128
2	population	0.433557
3	without_hsd	-0.189545
4	asian_pop	0.027615
5	black_pop	-0.064823
6	hispanic_pop	0.060663
7	white_pop	-0.333351
8	other_pop	-0.060093
9	pci	0.264181
10	cluster_1	0.072007
11	cluster_2	-0.729898
12	cluster_3	-0.453846
13	cluster_4	-0.345748

Chicago Regression Coefficients

Some observations:

- Oddly, a unit increase in unemployment, decreases the property crime ratio of a neighbourhood by 0.01%.
- Poverty and population increase the crime ratio by 0.52%.
- Increases in both Black, Other, or White population decrease the crime rate
- Asian and Hispanic population increase the crime rate.
- A unit increase in PCI increases the property crime ratio by 0.26%.
- All neighbourhoods belonging to Venue Cluster 3 are the mostly likely to experience lower crime ratio in comparison to our benchmark Cluster 1.
- Cluster 2 is the most likely to experience a higher crime ratio than Cluster 1.

New York City

After performing the regression, for the NYC dataset, we get an R^2-score of 0.6996 and an adjusted R^2-score of 0.5918, or that 59.18% of the variation in crime ratio is explained by this model. The coefficients of the variables are presented below:

0	unemployment	-0.013468
1	poverty_rate	1.076453
2	population	0.125312
3	without_hsd	-0.758288
4	asian_pop	0.362318
5	black_pop	0.083436
6	hispanic_pop	0.146352
7	white_pop	-0.401183
8	other_pop	-0.117368
9	median income	0.509628
10	cluster_1	-0.142882
11	cluster_2	0.282403
12	cluster_3	-0.040216
13	cluster_4	0.424195

NYC Regression Coefficients

Some observations:

- Again, a unit increase in unemployment decreases the crime ratio by 0.01%
- Here, unit increase in poverty rate increases crime ratio by 1.07%.
- Population also has a positive correlation with crime ratio here (unit increase increases crime by 0.12%).

- Meanwhile, Population aged 25 without a High-School Diploma has a negative effect on crime ratio
- In NYC, increases in White or Other Population have a negative effect on crime ratio, while all other populations increase the crime ratio in the sub-boroughs of NYC.
- A unit increase in Median Income increases the crime ratio by 0.5%.
- In comparison to Cluster 1 sub-boroughs, Cluster 2 and 4 sub-boroughs should experience lower crime ratios with Cluster 2 being the best
- While, Cluster 3 and 5 sub-boroughs experience higher crime ratios with Cluster 5 being the worst.

Results & Discussion

From each regression analyses we have seen that Unemployment, Population Aged 25 without High-School Diploma, and Income have a negative correlation with crime ratio. While, these may be erroneous results, it might be explained by the fact that higher-income neighbourhoods/sub-boroughs are more likely to be targeted in the case of property crime, this would also explain the negative correlation between Unemployment and Population Aged 25 without High-School Diploma as high-income neighbourhoods would also have lower rates of unemployment and higher rates of population that have at least achieved a high-school level of education.

In terms of this analysis' recommendations for entrepreneurs trying to choose a neighbourhood in one of these cities to start a new venue where it is safest in terms of property crime, we make the following recommendations:

Chicago

Firstly, we should choose a neighbourhood belonging to Venue Cluster 3 (`cluster_2`), as it has the least likelihood of crime and since it only has one neighbourhood in it, we recommend Bridgeport as the ideal neighbourhood to start a new venue safe from property crime.

	Neighborhood	Clusters	unemployment	poverty rate	population	pop aged 25 without hsd	asian_pop	black_pop	hispanic_pop	white_pop	other_pop	pci	ratio
59	Bridgeport	2	0.112	0.173	32543.0	0.222	0.334911	0.058538	0.301201	0.297483	0.007867	22694.0	0.006012

Chicago Neighbourhood Recommendation

New York City

Firstly, we should choose a neighbourhood belonging to Venue Cluster 2 (`cluster_1`), as it has the least likelihood of crime.

	Neighborhood	Clusters	unemployment	poverty rate	population	pop aged 25 without hsd	asian_pop	black_pop	hispanic_pop	white_pop	other_pop	median income	ratio
5	Astoria	1	0.059342	0.149257	171100	0.140745	0.154099	0.067739	0.267112	0.480503	0.030547	64569.48051	0.008798
14	Hillcrest / Fresh Meadows	1	0.080896	0.128895	163193	0.136681	0.336426	0.118590	0.198557	0.303631	0.042795	66776.88834	0.010879
27	Mott Haven / Hunts Point	1	0.120996	0.408318	163773	0.389136	0.006969	0.289725	0.670032	0.022448	0.010825	25069.69555	0.025258
44	Brownsville / Ocean Hill	1	0.131329	0.348951	118907	0.234917	0.013644	0.717856	0.224885	0.027804	0.015811	28351.88456	0.023035
53	Bensonhurst	1	0.067998	0.184279	191525	0.256549	0.403927	0.010737	0.158654	0.404201	0.022481	53834.77753	0.007568

NYC Neighbourhood Recommendations

Finally, we recommend the sub-borough with the lowest crime ratio, Bensonhurst.

	Neighborhood	Clusters	unemployment	poverty rate	population	pop aged 25 without hsd	asian_pop	black_pop	hispanic_pop	white_pop	other_pop	median income	ratio
53	Bensonhurst	1	0.067998	0.184279	191525	0.256549	0.403927	0.0107367	0.158654	0.404201	0.0224811	53834.8	0.00756787

NYC Neighbourhood Final Recommendation

Caveats and Suggestions for Improvement

In this analysis, our regression could only explain 59–61% of the variation in crime ratio as given by the Adjusted R-squared, making this model not very accurate. A possible extension of this analysis could be improving its accuracy by including more cities and years to extend the dataset and allow more observations to increase the accuracy.

The model could be further improved by including other variables such as average wage rates, venue prices in the clustering, level of police enforcement and others

Also, no hypothesis testing was performed in the preliminary model to test the significance of the variables if any, other than the basic R-squared score

It should also be noted that the venue clustering will have different results every time the notebook is run and thus change the results of the regression every time which may or may not be the same depending on Four square's venue popularity, day of the week, and time of the day. The study could be extended by running the clustering multiple times over the course of a period of time to get a more holistic view of the true dataset.

Conclusion

In this neighbourhood analysis, we have explored two cities' neighbourhoods using k-Means Clustering and Regression to examine the relationship between crime rates, demographics, and venue data. Using this examination, we have provided recommendations for starting a venue in Chicago and NYC to entrepreneurs looking to minimize the risk of property crime — Bridgeport for Chicago and Bensonhurst for NYC. We also explore the effects of the different demographic variables on a neighbourhood's crime rate which may be of interest to public planning and development and assist in lowering crime rates.