Capstone Project

The Battle of Neighborhoods

Introduction

Investors are always looking for new ventures to increase their worth and one of the most popular small business ideas is a restaurant. As we all know, one of the most important factors in a business is its location and many factors can affect this choice. A poor decision can result in a huge loss of finances. I believe entrepreneurs in the restaurant industry will be interested in this study.

I believe any potential small business investor with be interested in this study.

For this project, I will use Population and Crime data, and the FourSquare tool to acquire the locations of restaurants and a setlist of general venues that promote traffic for Chicago and Toronto. My intention is to find any connection between proximity to regular pedestrian traffic and the relative safety of different areas in the city.

Data

For this investigation, I have acquired major crime records from the official websites of the police departments in Toronto and Chicago. this was

made simple by the fact that these sites provide their databases freely in a few formats.

Population data was acquired through Census websites for Toronto and Chicago

Four-Square was used to acquire locations of a wide range of venues for the cities of Toronto and Chicago.

Methodology

Feature selection

The crime index is highly generalized and often does not indicate the level of crime in an area as perceived by the inhabitants. Because of this, I have decided to modify this approach and instead of using the overall crime statistics for a area, I have collected data on crimes for each area, acquired from yearly police records.

Quite obviously, as well, the connection between human traffic business, places of activity is obvious, and humans always need to eat. I used Four-Square to create a database of all results that come from a search using the key words: venue, restaurants, office, shop, sport, movie.

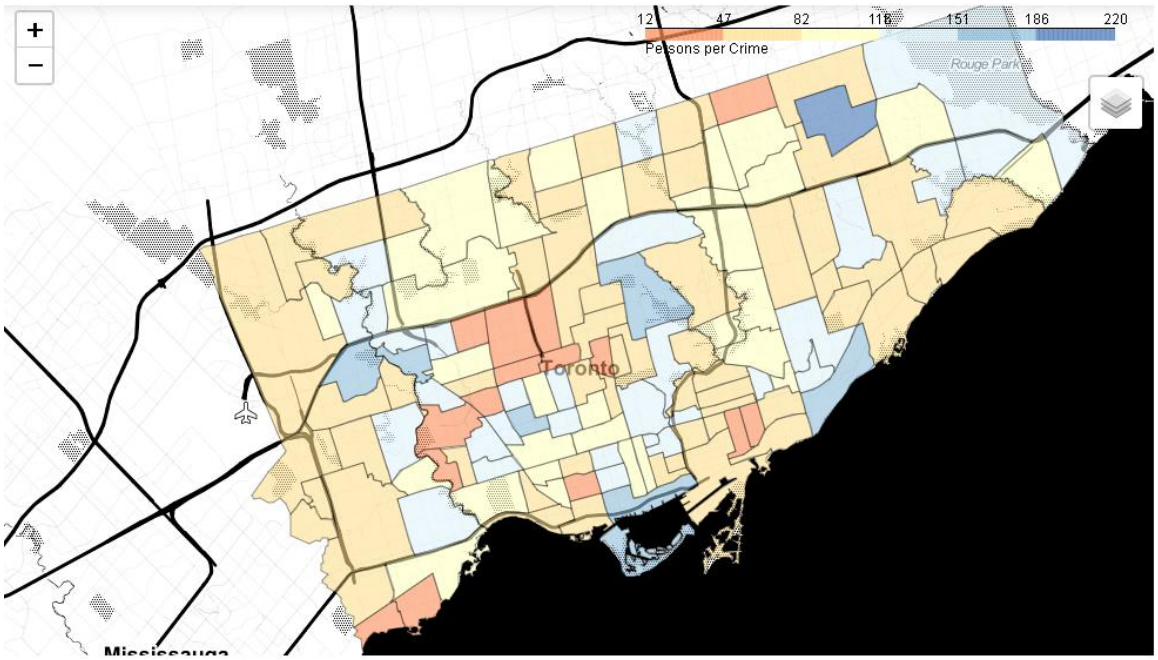
I believe this is a comprehensive enough list of activity drivers to investigate the relationship between these and the existence of restaurants and fast food outlets.

	Unnamed: 0	index	name	lat	lng	categories	address	city	state	country	id	labels	Clus_Db
0	0	8113	Fusia Asian Kitchen	43.656170	-79.383549	Food Court	NaN	Toronto	NaN	NaN	4bb0feb7f964a52021713ce3	0	green
1	1	8114	Fine Asian Bowl	43.655387	-79.380326	Vietnamese Restaurant	NaN	Toronto	NaN	NaN	4aea2b14f964a520ebb921e3	0	green
2	2	8116	SenseAsian Pan-Asian Resto	43.649610	-79.378912	Asian Restaurant	NaN	Toronto	NaN	NaN	4dd3d24345dd98b61e74f6af	0	green
3	3	8120	Ali Baba's	43.654916	-79.387172	Middle Eastern Restaurant	NaN	Toronto	NaN	NaN	4ddd83c788779c82beb061fc	0	green
4	4	8126	Lee Chen Asian Bistro	43.647010	-79.386536	Asian Restaurant	NaN	Toronto	NaN	NaN	582e3fa85e56b4259dc5f751	0	green

Hypothesis

Because individuals interact more closely within a city, the level of crime in different areas will also affect their behavior, which in turn, affects businesses. For this, I used **Folium** and **Choropleth** maps to highlight areas of low to high crime, as would be perceived by the population.

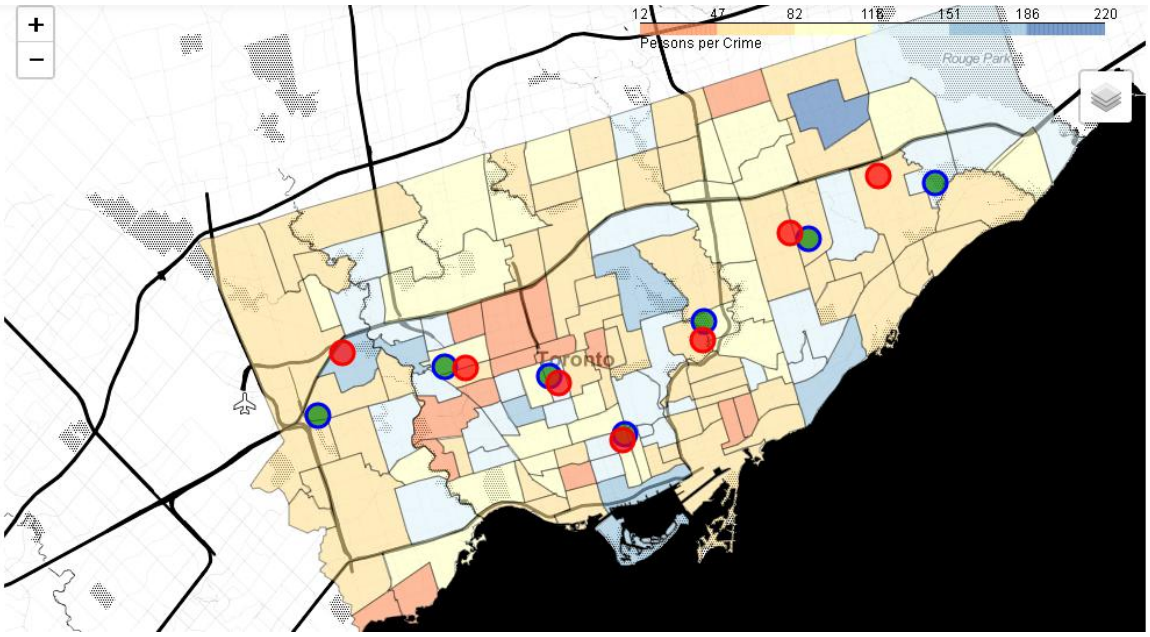
To do this, i simply find the crime rate and population of each neighborhood,and in very much the same way typical crime index is calculated, I devise the crime index for each neighborhood. Using Choropleth, its quite useful to have the relative crime index for comparisons.



Businesses and clustering

After showing the areas of low to high crime on the map, I will then plot the restaurant and general business areas, and use **K-means** to find the center of the main clusters.

While **K-means** is not the most intelligent way to cluster elements, using it to explore the data can prove insightful. As we can see below, there does seem to be a pattern to it, following the areas of lower crime-rate.



I will mainly rely on DBSCAN to perform more intelligent clustering of businesses and restaurants. The idea is to establish areas f concentrated business activity by grouping the general businesses and restaurants according to parameters that yield the least non-grouped elements.

I will then remove the non-grouped elements from both sets, combine those databases and use DBSCAN to again formulate clusters based on the least amount of non-grouped elements.

I will then plot each cluster on a bar or pie chart and see which of them has the lowest restaurant to general business ratio. I believe these areas will be good selections for opening a new business. This last choice may also be influenced by the local crime rate.

Results

The maps clearly show a correlation between general business locations and restaurants. There is also more than a vague correlation between areas of low crime and high business activity, especially in the case of Chicago.

Toronto

Image 1 shows the business cluster centers of Toronto, found using K-means. *Red circles* show the centers for restaurants, and *blue-green* ones show the centers for general businesses. The shaded areas on the map represent the level of crime for each zone, with bluer shaded meaning low crime, red shades mean more.

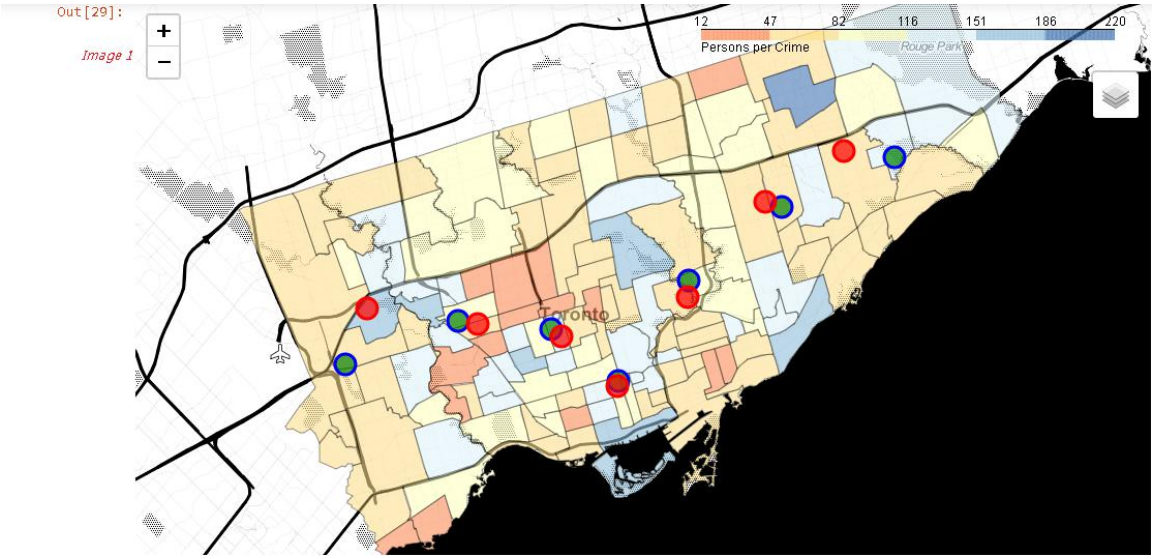


Image 2 shows the main restaurant clusters *(green)* in Toronto, acquired using DBSCAN. This gives a good idea of which areas are more densely packed with restaurants.

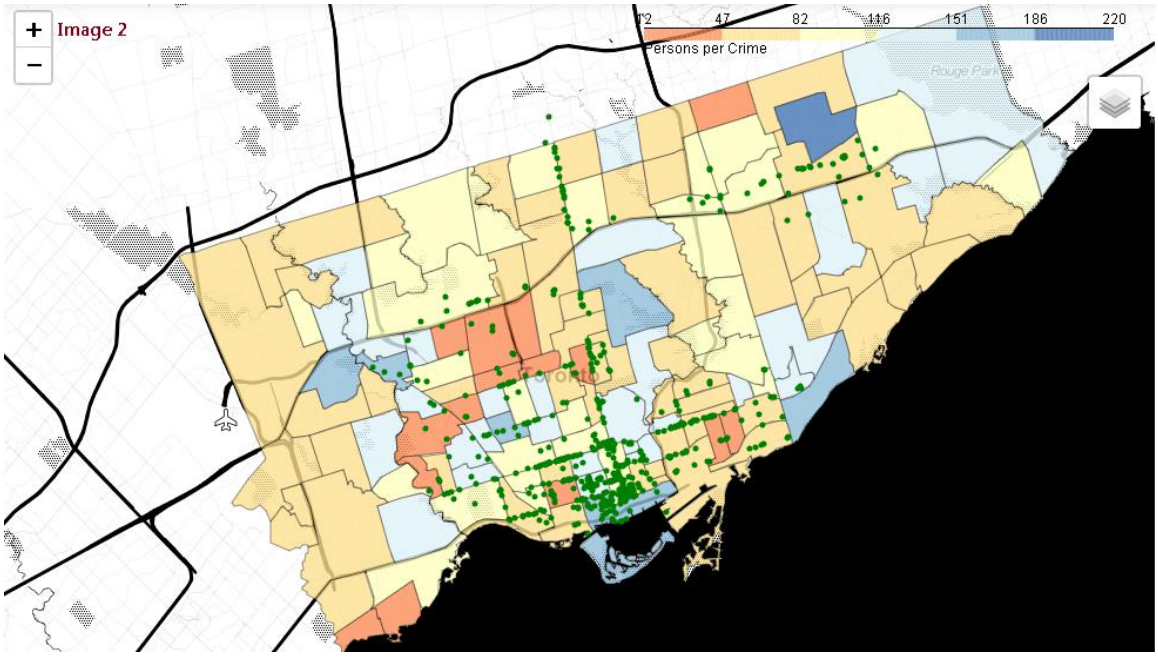
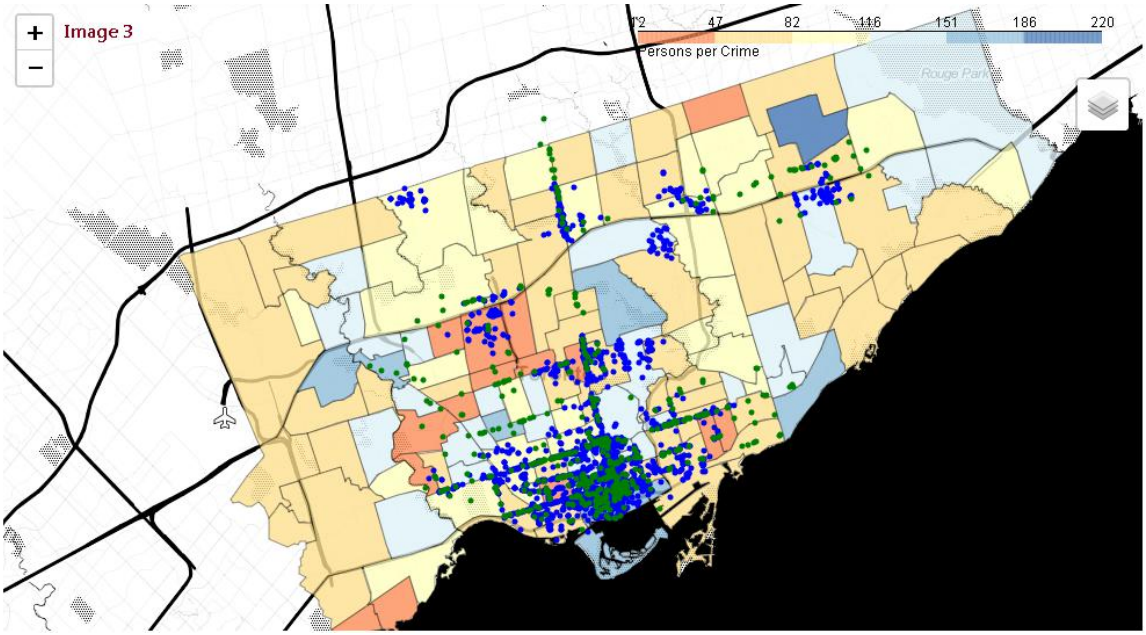
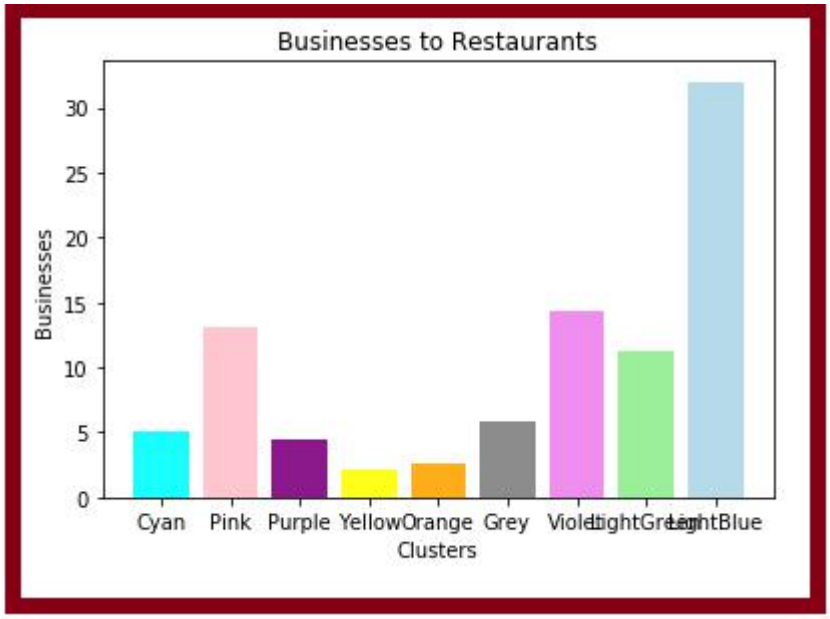


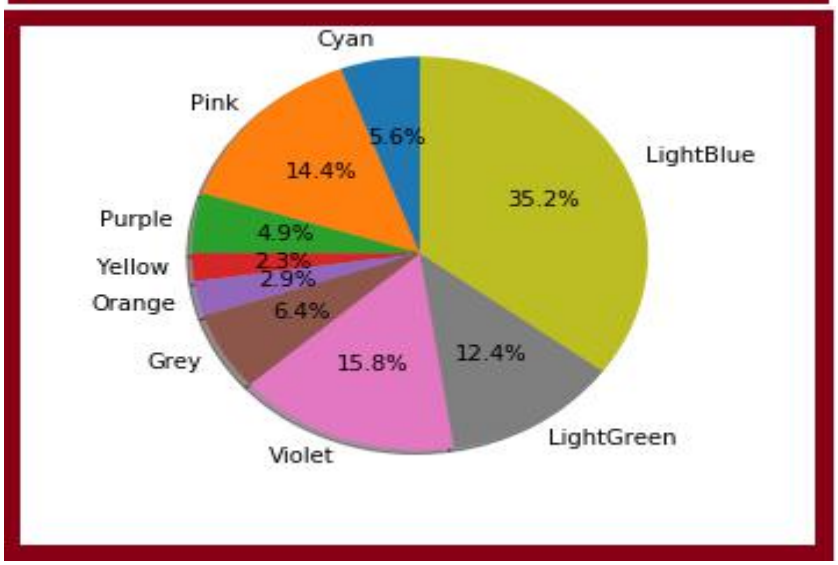
Image 3 contains the major clusters for both restaurants *(green)* and general businesses *(blue)*. It can be seen that both clusters coincide, and also its more clear that businesses generally follow areas of low crime.



Further Clustering

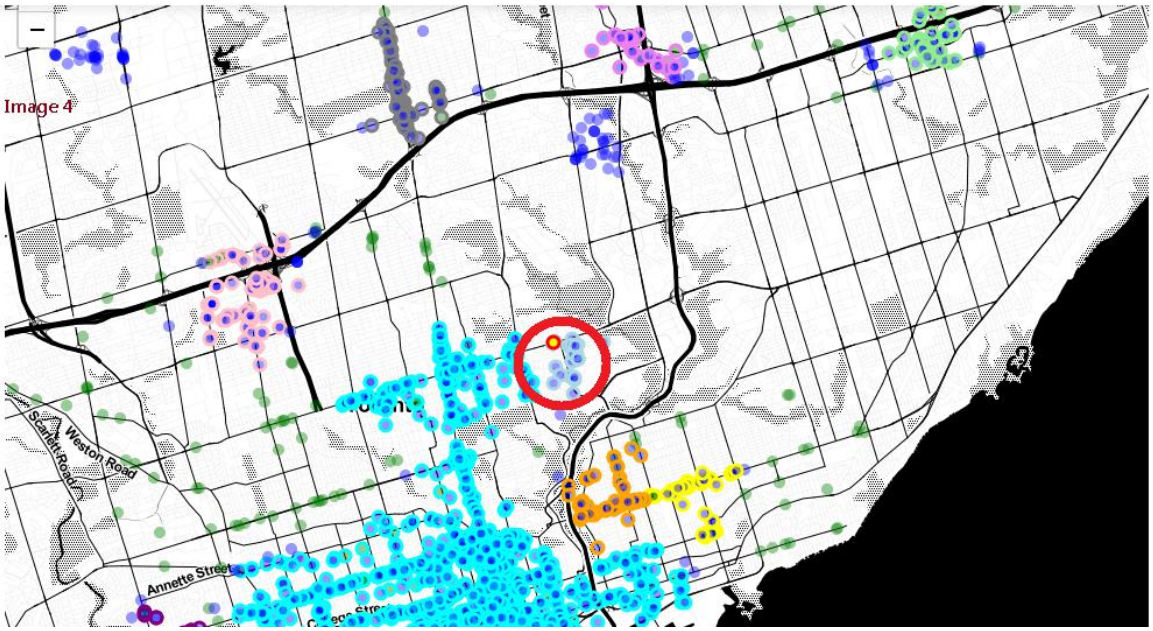
At this point, I have the major clusters of businesses and restaurants, and will combine then into a single database. With this, I will then plot each cluster on a bar chart and pie chart, according to the ratio of restaurants to general businesses. This will show me which areas have the lowest amount of restaurants per business and so may be able to support more.



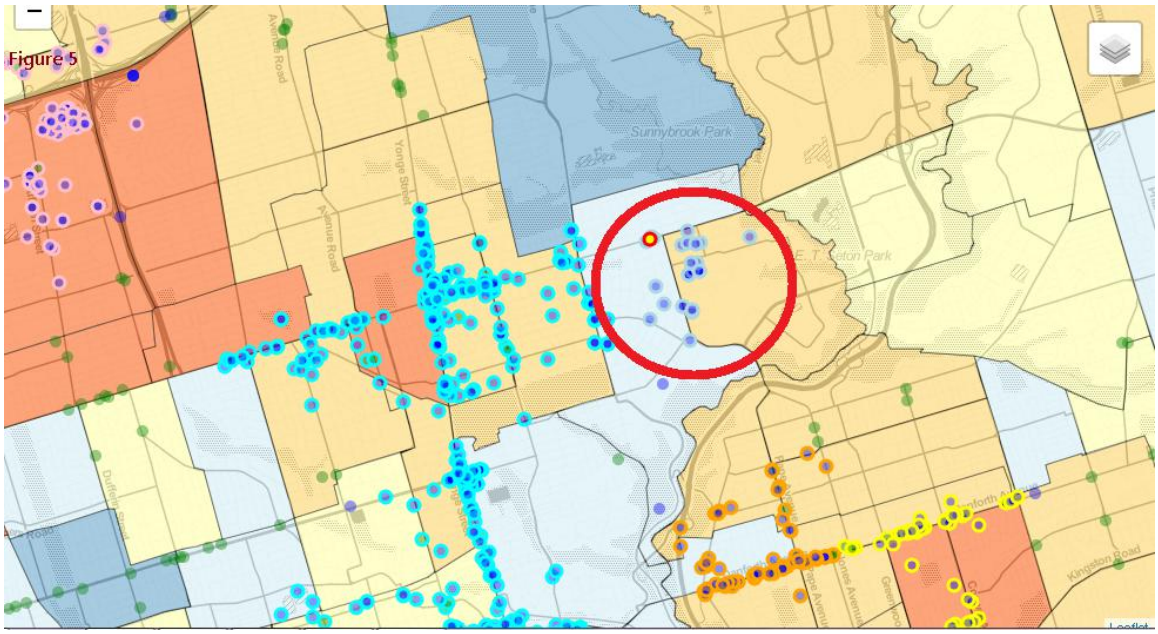


The pie chart above and the bar graph before clearly show the cluster marked Light blue as having the lowest ratio of restaurants to General Businesses. Now I locate that cluster on the map. This is made easy by my having already switched all the cluster numbers in the database with unique colors.

Image 4 and Image 5 show the spot I have selected for starting a business (red-yellow circle), based on the cluster identified as having the most favorable restaurant to General Businesses ratio. Visual inspection of the map and the group clusters also gives one a very good gauge of where this could be as well.
As we can see, It is located in an area containing no restaurant cluster, so competition will not be fierce. *(Circles shaded blue are general businesses, restaurants are shaded green)* It is also close to many active businesses, meaning it can benefit from traffic generated as a result.



We can also appreciate the lower rate of crime in the immediate area, as seen by the blue and light brown shaded areas on the map(Figure 5) below.



Chicago

The Above steps were repeated for Chicago, and since the theme of this project was “ Battle of the Neighborhoods”, I found it interesting to test my theory of crime and its effect on businesses by comparing a city of high crime to a city of low crime, Chicago versus Toronto.

Figure 1 shows the business cluster centers of Chicago, found using K-means. *Red circles* show the centers for restaurants, and *blue-green* ones show the centers for general businesses. The shaded areas on the map represent the level of crime for each zone, where bluer shades mean less crime, red shades mean more.

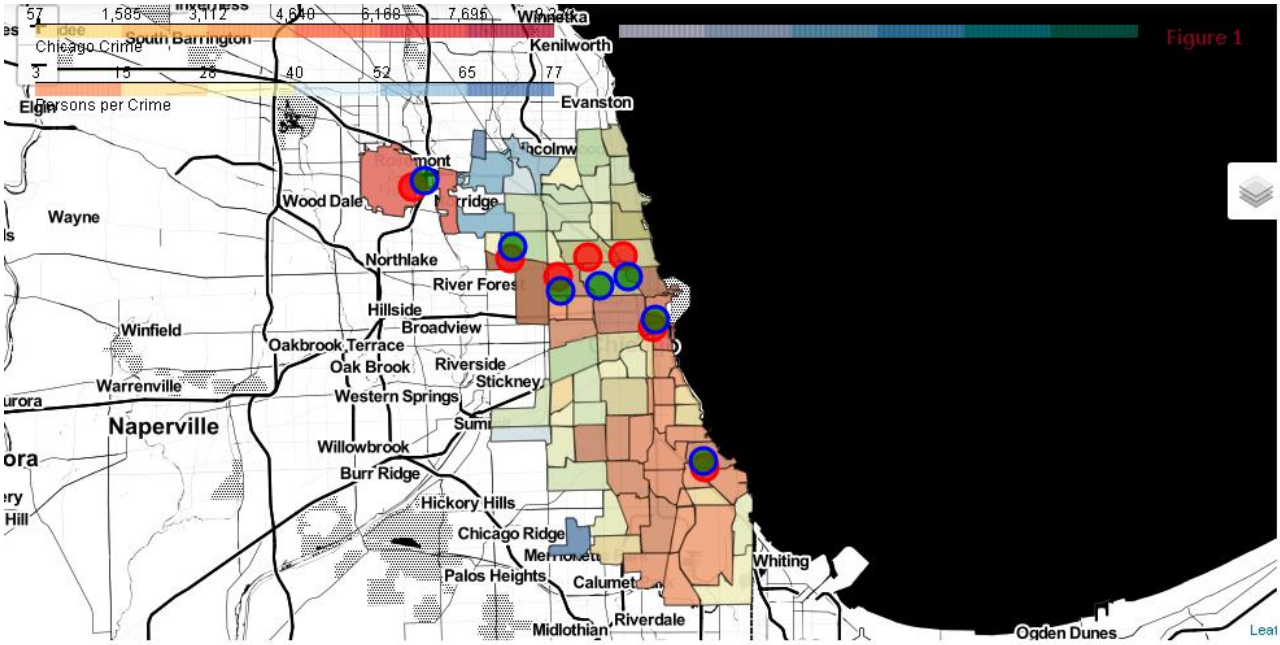


Figure 2 shows the main restaurant clusters in Chicago *(green)*, acquired using DBSCAN. This gives a good idea of which areas are more densely packed with restaurants.

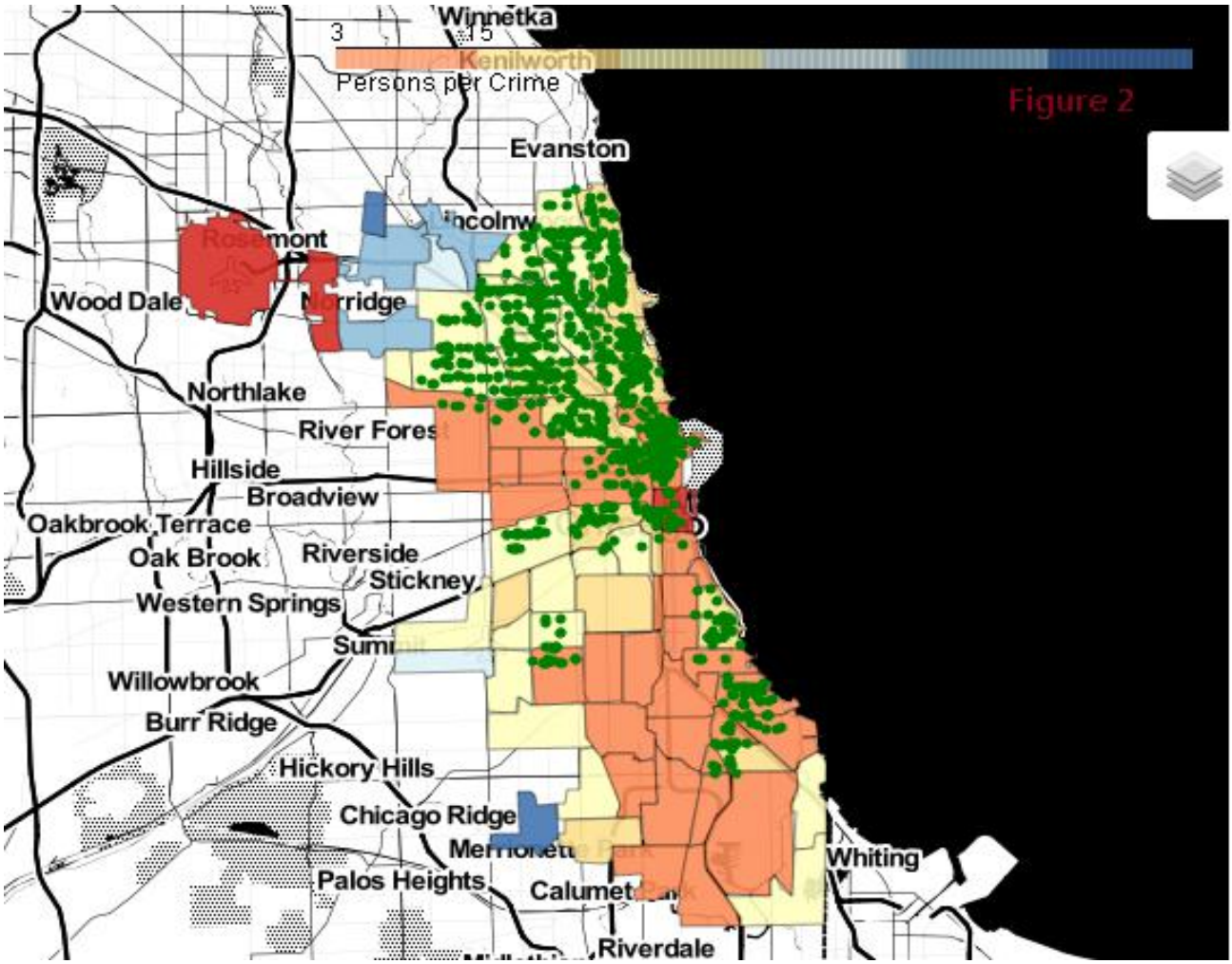
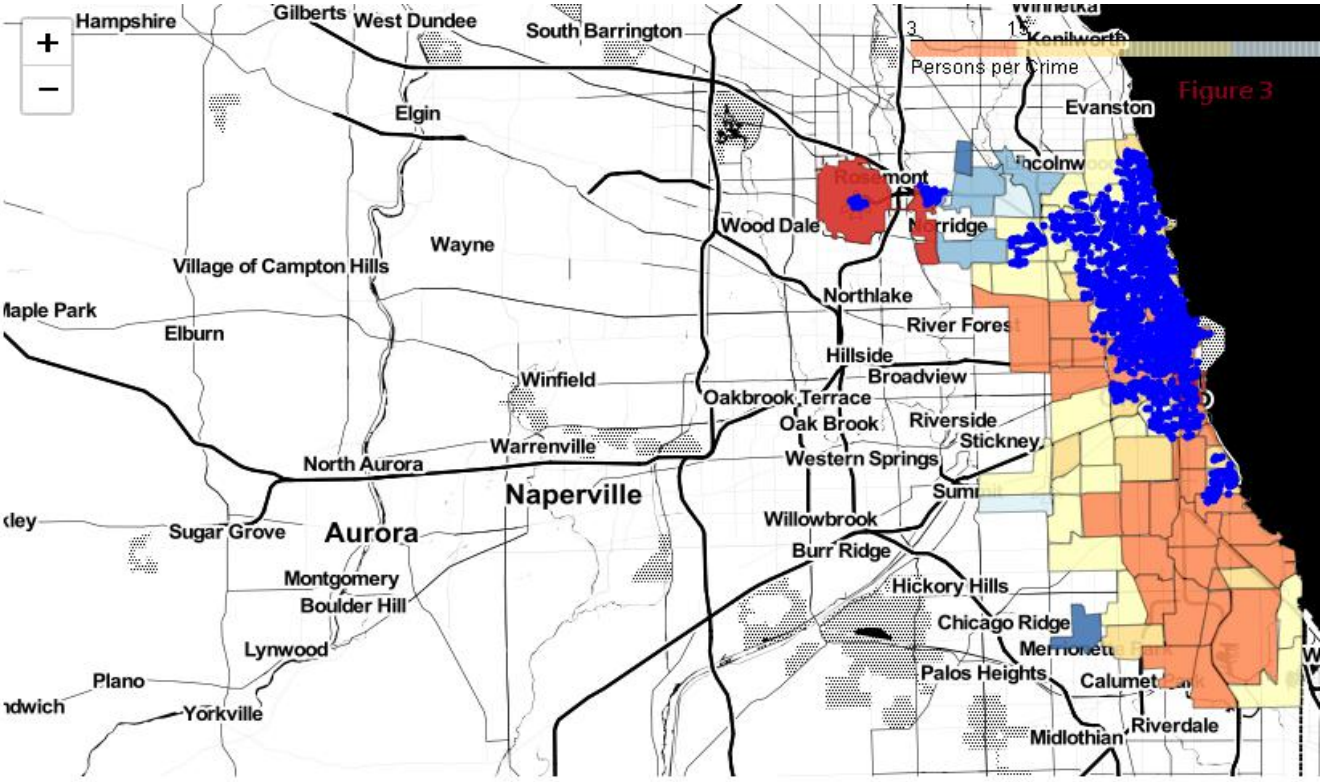


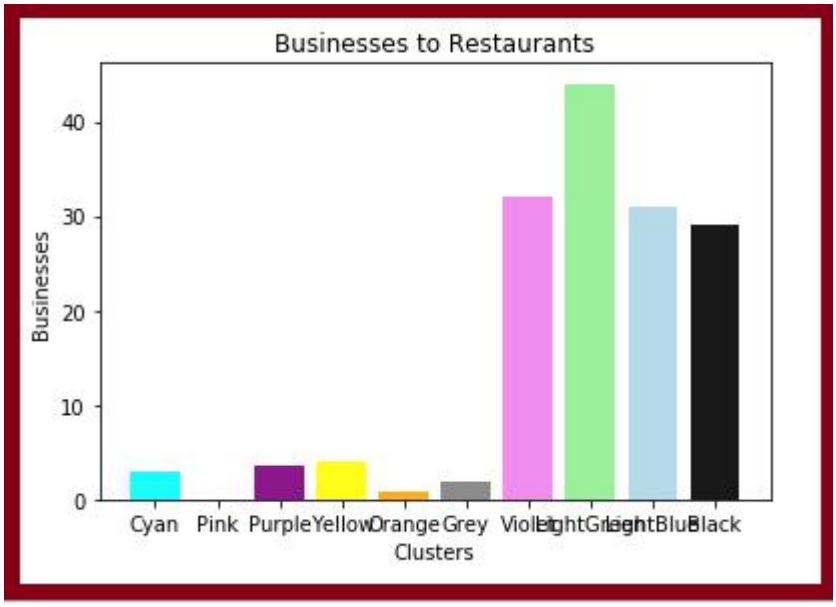
Figure 3 contains the major clusters for general businesses *(blue)*. Take note of how strongly they align with the lower crime areas of Chicago *(areas not shaded red)*.

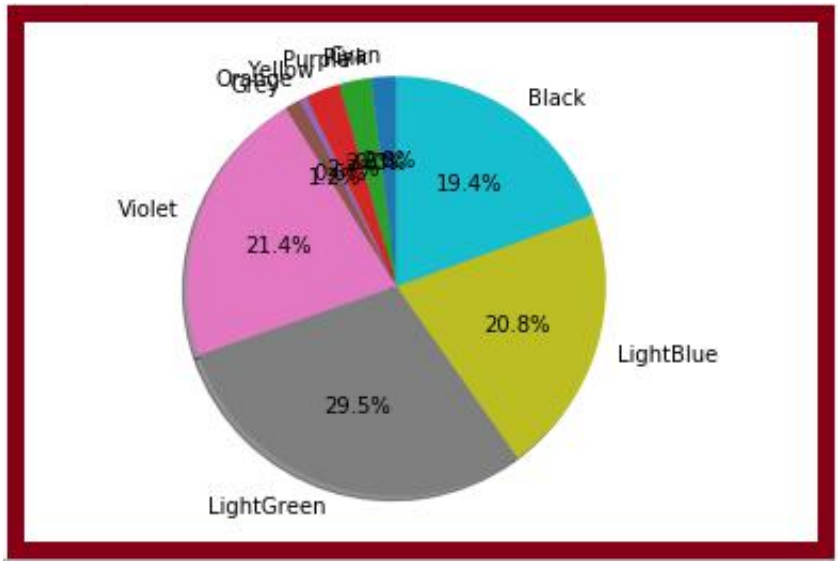


Further Clustering

At this point, just as with Toronto, I proceed to combine the two databases containing the man clusters fr Restaurants and General Business. After this, I use DBSCAN to cluster them into areas of highest activity, while keeping each element clearly identified as a restaurant or General Business.

I then use these clusters to find the Restaurant to General Businesses in each cluster, then plot them to on charts for visual inspection.





The above bar chart and pie graphs clearly show that the cluster identified by light-green has the most favorable restaurant to general business ratio. We will now take a look at the clusters on the map.

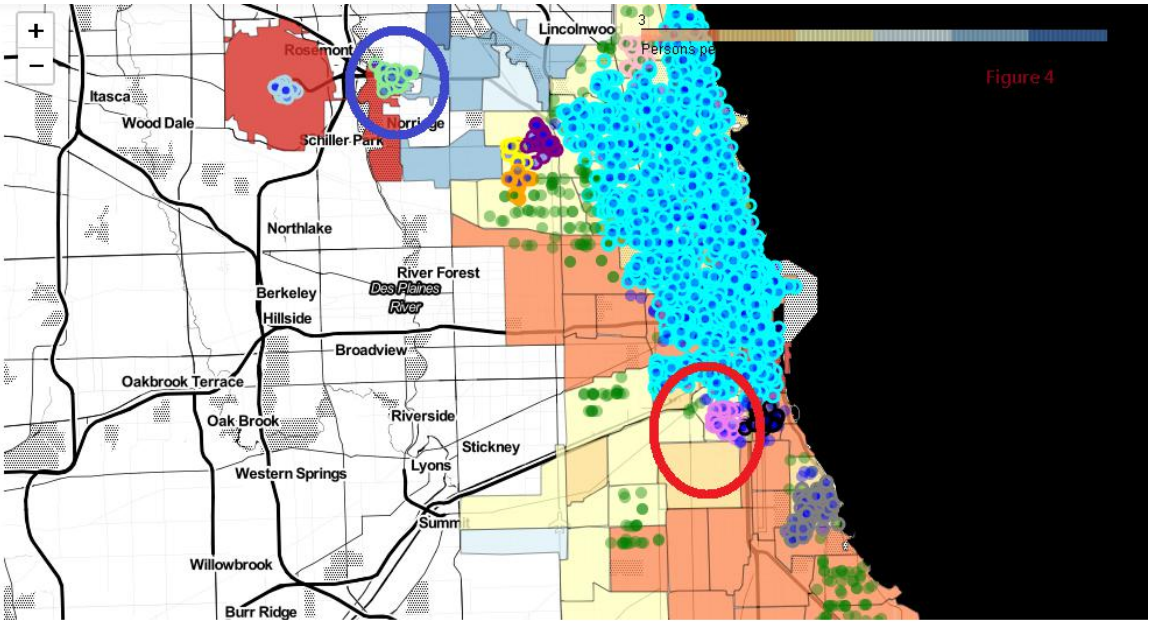


Figure 4 above shows the Light_green cluster, circled in blue, lies in an area quite remote from the main business areas of Toronto. It also lies in an area of relatively high crime. The cluster with the second-best ratio, represented by Violet, circled in red, resides in an area of low crime and is also quite close to the main business center.

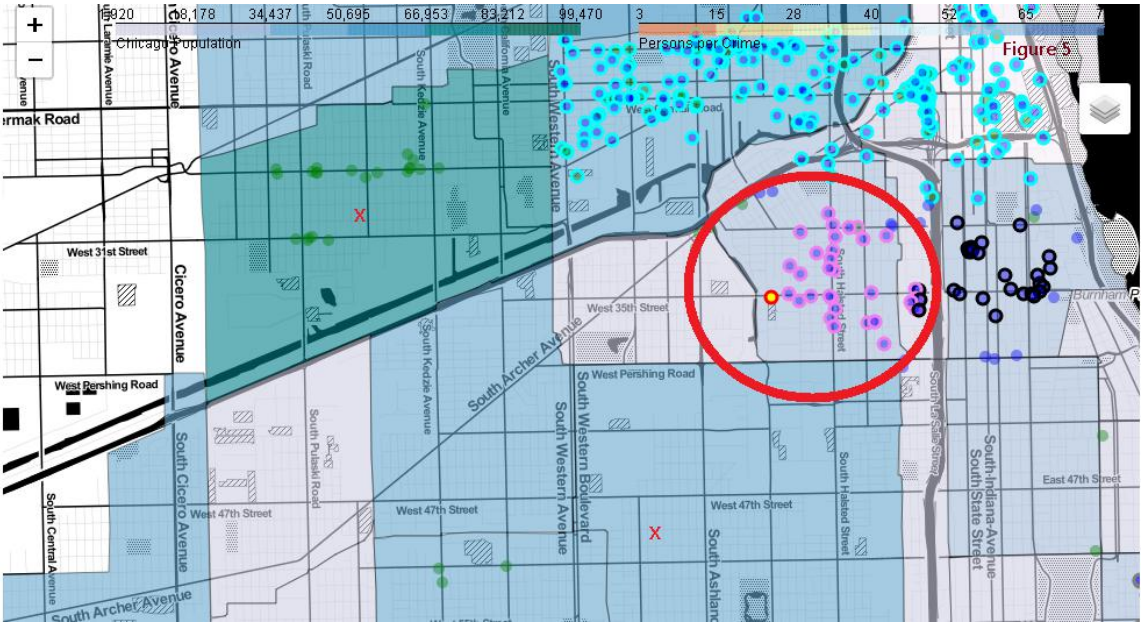


Figure 5 above shows that the immediate areas surrounding the purple cluster also happens to be highly populated. This adds to the likelihood that this area can be suitable for the opening of a Restaurant or fast food outlet.

It can also be seen that the Black Cluster is also close to the purple cluster and the black cluster's restaurant to business ratio is quite similar of violet's.

Discussion

It can be inferred from the results that that areas of ***low crime*** *and* ***high business activity*** are associated with the success of restaurants. Using both forms of machine learning revealed strong relationships between restaurant locations and the two features.

It should be possible to include regressions into this study, but I feel like the data needs to be more carefully gathered. Through visualization alone, however, clear patterns can be seen.

Crime rates played a more significant role than I expected and I intend to follow this up. It is safe to assume that if persons are able to move about more safely and freely, they are more likely to patronize businesses in that location.

Traffic generated from persons simply having to be in an area due to work, leisure activities and shopping is also quite significant, as seen from the maps. It was quite clear that clusters of high general business activity are usually accompanied by a cluster of restaurants.

Using this observation, I selected areas close to clusters of general business activity, that lacked an accompanying restaurant cluster, while remaining In a zone of low-crime rate.

Recommendations

While I believe the findings of my project to be sound, I recommend further investigation into the rent of the areas suggested. This data was not considered but it may have the power to change a 4th or 5th choice into a 1st or 2nd.

A more targeted investigation, aimed at providing recommendations for specific restaurant type should be done as a follow up.

Conclusion

Based on my findings:
1. In the city of Toronto, anywhere along the corners of Eglinton Avenue East and Bayview Avenue , is my suggestions for opening a fast food outlet or Restaurant.

Coordinates : (Lat : 43.7111887, Long : -79.3766768)

In the city of Chicago, West 31st and 35th Streets, between South Ashland Avenue and South Halsted Street, are my suggestions for opening a restaurant or fast food outlet.

Coordinates: (Lat. = 41.830389,Long. = -87.6550883),