Blake Even

5/6/21

Crime Data In Toronto

Introduction

Description of the Problem

You and your family are getting ready to move to the city of Toronto in Canada, but you are concerned about your family's safety. Which neighborhoods in this city would be the safest for you and your family? Primarily, which of the neighborhoods have the lowest amount of crime?

As of 2017, Toronto has a population of about 2.93 million. This population is split among about 140 Neighborhoods, and each neighborhood has their own individual crime statistics. We would want to find a way to compile the individual crimes of each of the 140 neighborhoods to determine the safest of them all. This data would also be useful for the city if they were to decide to build a new school and are considering the safest possible location.

While crime data alone would be useful to influence your decision, you would also want convenience in shopping, food, and maybe even entertainment within walking or driving range. So, we would need to figure out the safest neighborhood in terms of crime that also has stores and such that you may need. To figure this out, we would need to look at venue data via Foursquare API.

Data

Data Needed

For this Project, I will need:

- A list of Neighborhoods
- The Latitude and Longitude of the neighborhoods
- Toronto's Crime Data
- A GeoJSon map of Toronto divided by neighborhoods
- And, while safety is the most important, we'll also need venue data to influence the decision as a tie-breaker

Data Acquisition

- For the list of Toronto neighborhoods, I grabbed the data from The List of postal codes in Canada that start with M which are all the postal codes in Toronto, and I'll exclude the "Not assigned" codes as they aren't assigned to any borough or neighborhood.
- For the latitude and longitude, I am using the library Geocoders on Python to find the coordinates of each and every neighborhood and store them onto new columns on a Pandas data frame with their respective neighborhoods.

- For the crime data, I will use the “Toronto Police Data Crime Rates By Neighborhood” data set found on the website Kaggle. This data set includes Toronto’s crime data occurring from 2014 to 2019, it also includes the coordinates of each crime, as well as the offence itself. And for the map divided by neighborhoods, I used this map on the Toronto city website.
- For the choropleth map, I used a GeoJson file from this GitHub repository to create a folium map that accurately divided the city of Toronto by the neighborhoods to appropriately color code the map.
- And lastly, for the venue data, Foursquare API will provide the information about venues and geolocation I will need, so we will use it to get Toronto neighborhood venue data.

Methodology

Neighborhoods

To begin, I retrieved the list of Toronto Neighborhoods from Wikipedia and added it to a data frame. I then used Geocoders to add the individual geographical coordinates of every neighborhood to the data frame. Lastly, I created a folium map marking each of the neighborhoods’ locations to be used for later.

Crime Data

To get the Crime data I need, I downloaded the Toronto police data I found on Kaggle and created another data frame. I then sorted the neighborhoods by how often they appeared on the crime data frame and sorted it in both ascending and descending order to find the neighborhoods with the highest and lowest crime and created 2 bar graphs to display the results. Finally, I displayed the crime data on a choropleth map divided by neighborhoods, where yellow is the safest neighborhoods, and red is the least safe.

Venues Data

Finally, it's time to get the venue data. First, I used the Foursquare API and retrieved all the venues within the city and created another data frame of the venues with their coordinates and what venue category they each fall into and created a folium map with marker clusters of the venues. Finally, I clustered the venue data by general categories in each neighborhood and displayed them on the crime data choropleth map we created earlier in the notebook to find the safest neighborhoods with enough essential venues nearby.

Results

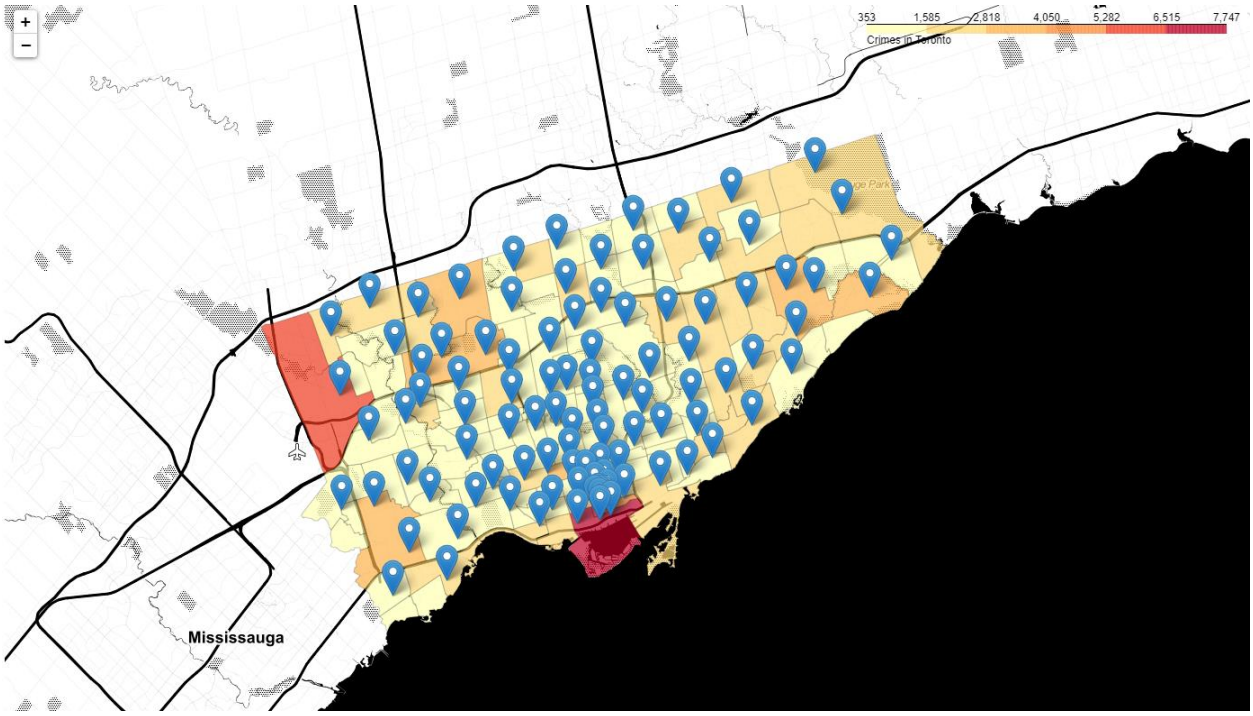
Crime Results

After careful examination of the data, 103 neighborhoods were identified in the City of Toronto, Ontario in Canada. After examining additional data sets, more neighborhoods were identified. Data was extracted from multiple repositories in order to get the most accurate results. Latitudes and longitudes were also used to appropriately map the neighborhoods.

The intention was to uncover the neighborhoods with the lowest crime in order to determine the lowest crime areas, data was taken from a Toronto Police data set from Kaggle. It was determined that the neighborhoods with the highest crime are Bay Street Corridor and Waterfront Communities-The Island, and the lowest are Woodbine-Lumsden and Lambton Baby Point, Lambton Baby Point therefore has the lowest amount of crime in Toronto.

Figure 1

Crime Choropleth Map



Venue Results

By using Foursquare API, the venues in Toronto and their coordinates were identified. The general categories for the venues were also defined for easier clustering. The results of analyzing the venue data was the production of an extremely detailed marker cluster map. I was also determined that the Toronto area is rich in shop and service with a number of about 2500. The Toronto area has a

very low number of arts and entertainment, college and university, and nightlife spot venues all under 500.

Figure 2

Marker Cluster Map

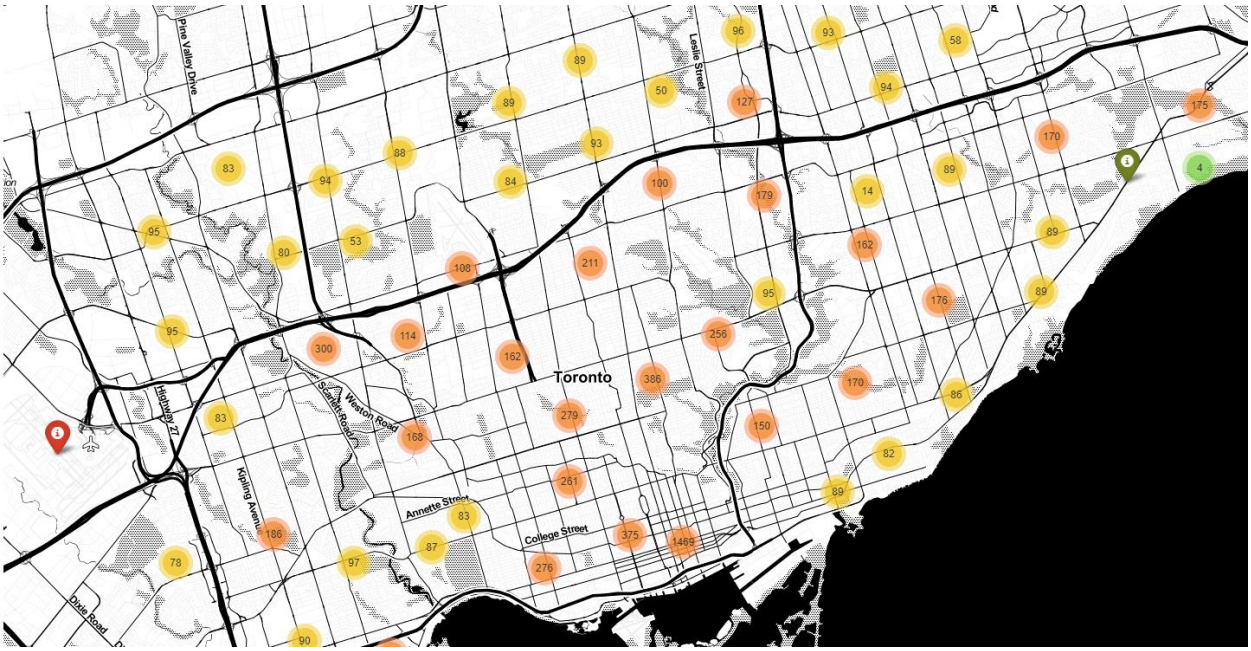
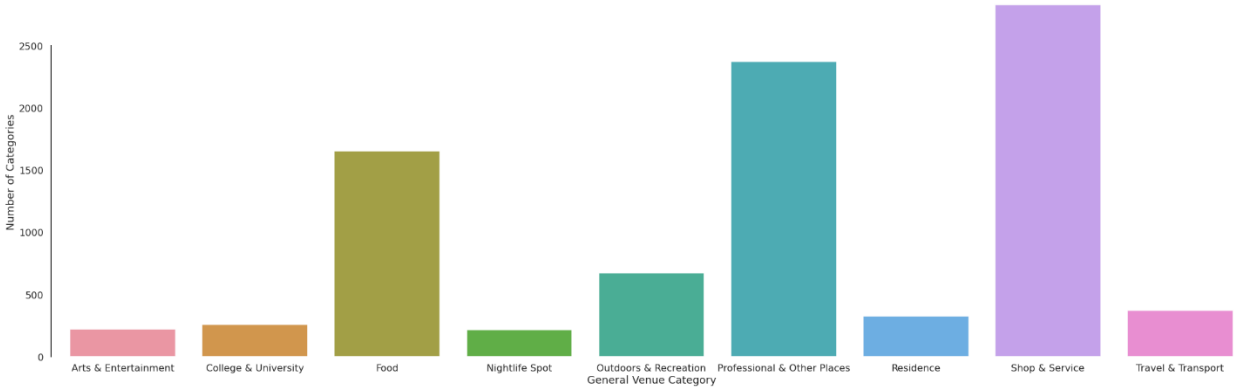


Figure 3

Venues Bar Graph



Discussion

Crime Discussion

From the crime data I have collected, it should be simple to determine and recommend the safest neighborhoods. From the bar graphs alone (see Figures 4 and 5), we can already recommend moving to neighborhoods like Lambton Baby Point, Woodbine-Lumsden, Maple leaf, Guildwood, or Yonge-St. Clair, while discouraging moving to neighborhoods such as Waterfront Communities-The Island, Bay Street Corridor, Church-Yonge Corridor, West Humber-Clairville, and Moss Park. Or, if you don't care about living in the absolute safest neighborhoods, you can also determine where you will move based off the color coding of the choropleth map (see Figure 1).

Figure 4

Lowest Crime Neighborhoods

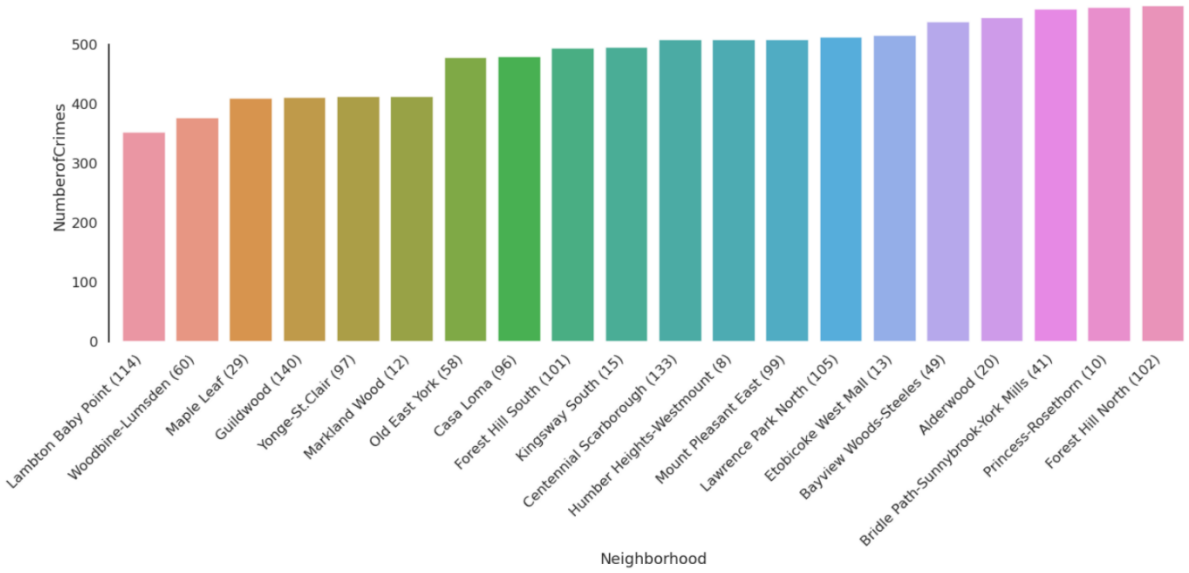
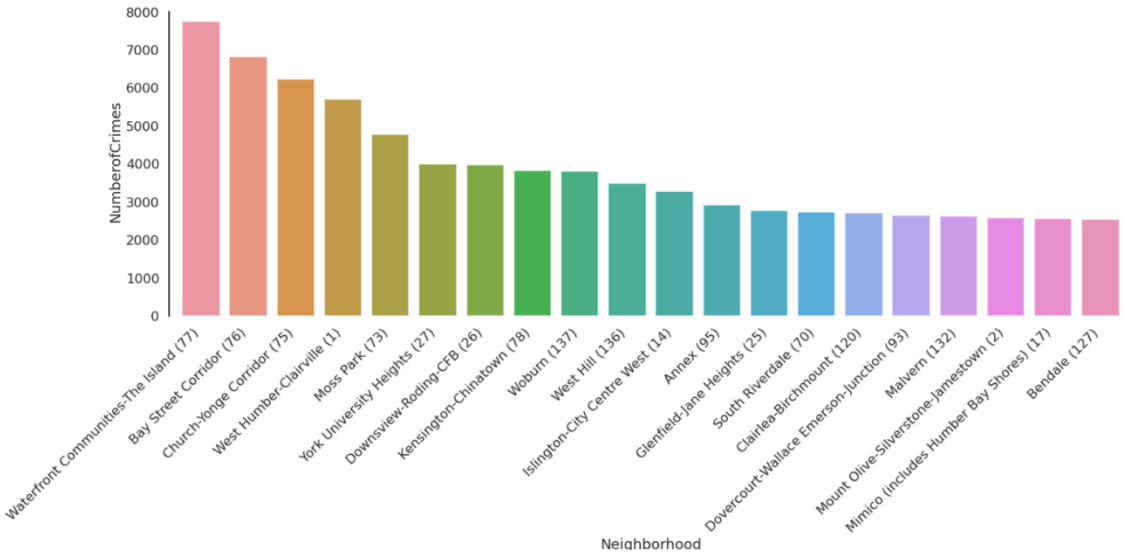


Figure 5

Highest Crime Neighborhoods

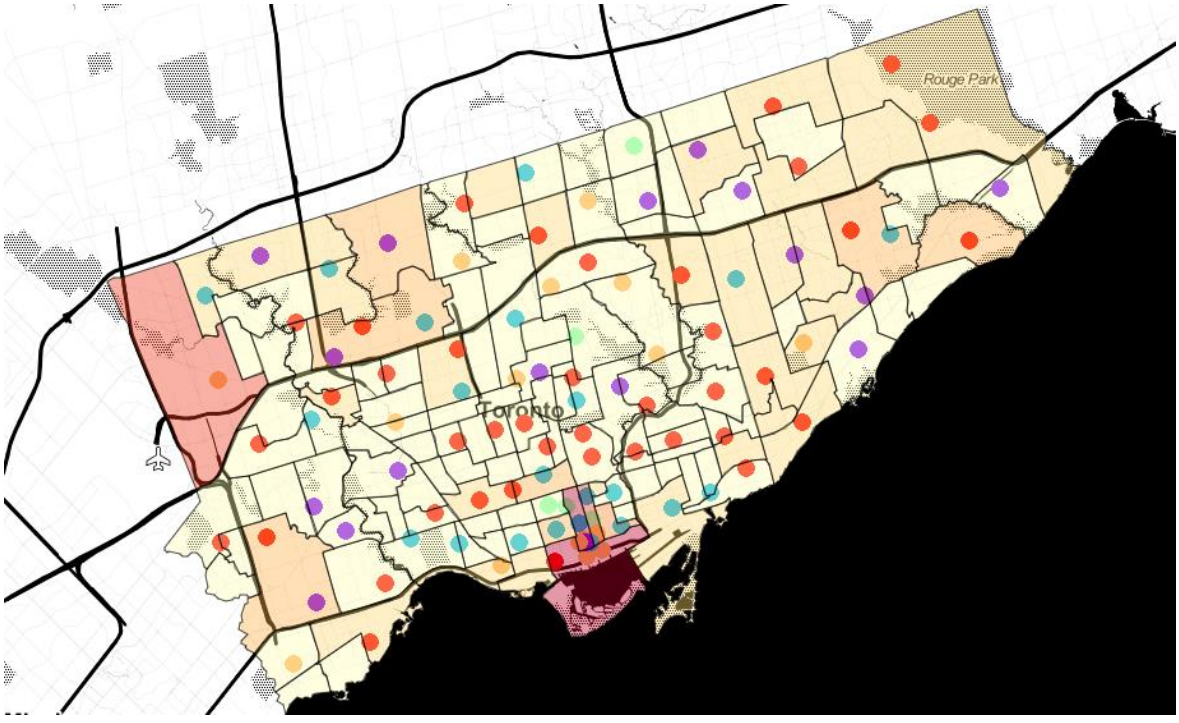


Venue Discussion

While crime data alone can be useful, we may also need venue data to help influence the final decision of which to move. Using the Marker Cluster map, you can find an area with a large number of the types of venues you need nearby to help determine your decision further. While this alone is useful for choosing where to move, you should still keep the crime data in mind. This updated choropleth map with venue clusters should help simplify the decision by merging the crime data with the venue data (see Figure 6).

Figure 6

Crime and Venue Data Map



Conclusion

After analyzing the neighborhoods, the crime data, and the venue data, I as the researcher for this project can recommend moving to neighborhoods Lambton Baby Drive or Forest Hill South. These 2 neighborhoods are within the top 10 lowest crime neighborhoods and the areas are rich in venues. I feel that these neighborhoods would be the best choices to move to because they will likely be the safest for you and your family, your house will have a high resale value for if you move again, and they would be great locations to raise a family.