**Predicting the Ideal Restaurant Location in Toronto**

Briana Brown

July 5th, 2020

1. **Introduction**
   1. **The Background**

Toronto, being the most populous city in Canada, is an astonishing and prime location not only for the financial sector but also for tourism. Toronto, being Canada’s leading destination for tourism and welcome to over 27 million people each year, is the prime location for restauranteurs. Being that this location boasts superb attractions like that of music, sports, and other events; this city is the perfect place to open a restaurant. The importance of restaurants to the city of Toronto is great and only continues to grow as food is and has always been the center of all great and growing cities. As such, each new restaurant brings its own unique value to the city, both for the economy and for the restaurant owners themselves. Therefore, it is advantageous for potential restaurant owners to accurately predict where exactly in the city of Toronto would be the best place to open a restaurant. For example, this information can be used to target boroughs or neighborhoods in which potential owners can place restaurants.

* 1. **Problem**

Data that might contribute to determining the best neighborhood in Toronto to open a restaurant might include which areas have the highest amount of foot traffic/ greatest accessibility and spending power, and also the lowest crime rate. The project aims to predict whether a restaurant that is placed in a certain area will thrive based on these data.

* 1. **Interest**

The likelihood that restaurateurs, future restauranteurs, and investors would be interested in the accurate prediction of where exactly in the city of Toronto would be best to open a restaurant is very high. This is because opening a restaurant in a prime location would likely guarantee more foot traffic to the restaurant and therefore more money. Others who are interested in dining at the best restaurants in the city may also find this information to be beneficial.

1. **Data Acquisition and Cleaning**

2.1 **Data Sources**

The statistics for crime can be found using the Toronto police maps crime data, the population and income data can be found using the Toronto census data, and the statistics regarding most common venue type can be found using the data from foursquare API. This data was acquiesced using the “rating, “stats,” and “price” details of the Foursquare API venue details. It must be noted however, that the Toronto police maps data set is currently lacking data for the year 2020 because the stats stop at 2019 and the year 2020 has not yet ended. The census data is also older being from 2016. However, this data is recent enough that it is still relevant to the inquiry and able to be used to answer the question of where it is best in Toronto to open a restaurant.

2.2 **Data Cleaning**

The data was downloaded from the Toronto police maps database, the Toronto census database, and also scraped from the Foursquare API venue details database and then placed into one table. The data available was from as early as 2006, however, to make sure that the analysis was as up to date as possible, only data values from after 2015 were used in this sample.

There was one main issue with all of the datasets. First, the dataset from the Toronto Police Maps database for crime rates was very large. This led to datasets being difficult to work with and sometimes rendered unloadable by Python. Therefore, the data for crime averages by type for each neighborhood was merged with the data from the Toronto city dataset to include population, area size, Latitude and Longitude of the neighborhood, and the income of each neighborhood, the table had to be scaled down in size both before and after merging. To decrease the size of the data frame before merging, the following was dropped from the crime dataset: 'reporteddate', 'premisetype', 'ucr\_code', 'ucr\_ext', 'reportedmonth', 'reportedday', 'reporteddayofyear', 'ObjectId', and 'occurrencedate.' After merging, the dataset was still very large and included millions of rows and several columns. Therefore, it was decided that it would be best to filter the data further by dropping the columns from the set and filtering the set to include only data from 2015 or after as this data is the most recent and the most accurate in terms of the crime makeup of each neighborhood. The following columns were dropped from the merged Toronto city data and Toronto crime data sets because they were proven to be irrelevant for the question at hand: 'OBJECTID', 'Population', 'Assault\_2015', 'Assault\_2016', 'Assault\_2017', 'Assault\_2018', 'Assault\_2019', 'TheftOver\_2015', 'TheftOver\_2016', 'TheftOver\_2017', 'TheftOver\_2018', 'TheftOver\_2019', 'AutoTheft\_2014', 'Assault\_2014', 'AutoTheft\_2015', 'AutoTheft\_2016', 'AutoTheft\_2017', 'AutoTheft\_2018', 'AutoTheft\_2019', 'BreakandEnter\_2014', 'BreakandEnter\_2015', 'BreakandEnter\_2016', 'BreakandEnter\_2017', 'BreakandEnter\_2018', 'BreakandEnter\_2019', 'Robbery\_2014', 'Robbery\_2015', 'Robbery\_2016', 'Robbery\_2017', 'Robbery\_2018', 'Robbery\_2019', 'Homicide\_2014', 'Homicide\_2015', 'Homicide\_2016', 'Homicide\_2017', 'Homicide\_2018', 'Homicide\_2019', 'Shape\_\_Area', 'Shape\_\_Length', 'TheftOver\_Rate\_2019', 'TheftOver\_2014', 'Robbery\_Rate\_2019', 'Homicide\_Rate\_2019', 'BreakandEnter\_Rate\_2019', 'AutoTheft\_Rate\_2019', and 'Assault\_Rate\_2019.'

There was much more data that was found to be irrelevant after the merging of the new dataset with the dataset of the Toronto census. This data included information from the following columns: 'PostalCode', 'Population\_Density', '< 5k','5k - 10k','10k - 15k', '15k - 20k','20k - 25k','25k - 30k','30k - 35k','35k - 40k','40k - 45k','45k - 50k', '50k - 60k','60k - 70k','70k - 80k','80k - 90k','90k - 100k','100k - 125k','125k - 150k', '150k - 200k','> 200k', 'TheftOver\_AVG', 'Robbery\_AVG', 'BreakandEnter\_AVG', 'AutoTheft\_AVG', 'Assault\_AVG', 'offence', and duplicate columns such as 'reportedyear', 'Hood\_ID\_x', 'Hood\_ID\_y', 'Longitude\_y', 'Latitude\_y', 'MCI', 'Neighbourhood', 'TheftOver\_CHG', 'Robbery\_CHG', 'Homicide\_CHG', 'BreakandEnter\_CHG', 'AutoTheft\_CHG', and 'Assault\_CHG.'

Upon viewing the new merged dataset, I decided to keep only the homicide average by borough and neighborhood because the it is important to factor in the most dangerous of crime when attempting to look for the best place to open up a restaurant since the ideal location will be a safe one. Therefore, to make the dataset even smaller and much easier to work with, the columns: 'TheftOver\_AVG', 'Robbery\_AVG', 'BreakandEnter\_AVG', 'AutoTheft\_AVG', 'Assault\_AVG', 'offence', and 'reportedyear' were all dropped from the data set.

The average spending power of each neighborhood was also calculated

After the issue of an over ten million row dataset was fixed, I checked for outliers in the data. There appeared to be no extreme outliers in the data, nor were there any missing values in the dataset. After cleaning, the dataset still included over seven million rows of data.

2.3 **Feature Selection**

Because there are some common venue categories in boroughs, I used an unsupervised K-means algorithm to cluster the boroughs together. The number of top venues then had to be grouped and then sorted to include only specific restaurant types such as ‘American,’ ‘Asian,’ or ‘Columbian.’ The total number of each venue type was then returned to show the total number of restaurants by restaurant type. After this process was completed, the remaining table included a few redundant features. The dataset included over seven million rows of data, much of which included repeats of the amount of spending power possessed by each borough as well as the average number of homicides that occurred in each borough. Therefore, to make the dataset smaller again and easier to work with, the data was filtered to only include boroughs with a spending power equal to or above 3.0, a homicide average equal to or less than 0.5, and boroughs where the total number of restaurants by neighborhood was equal to or less than 15.

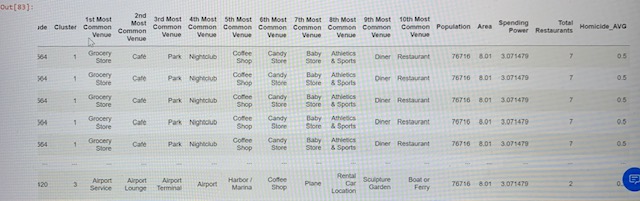
This was done so that a specific neighborhood in which to open a restaurant could be identified in one ideal borough. Therefore, the dataset was narrowed down to the borough of Downtown Toronto after filtering and this dataset included less than 8,000 rows of data. It was then decided to narrow down which neighborhood in the borough of Downtown Toronto would be best to open a restaurant as it gives a more specific location rather than just a general area.

After discarding all of the redundant features and filtering the data set down to less than 8,000 rows of data, the data was clustered again by neighborhood using an unsupervised method. Being that the data was already clustered by top venue and sorted to include only restaurant types, this unsupervised k-means cluster algorithm was used in which to cluster neighborhoods together to show the total number of restaurants, spending power, homicide average, and population number of each neighborhood.

This makes sense because the ideal place to open up a restaurant in Downtown Toronto will be a neighborhood with the highest spending power/ foot traffic, the lowest rate of dangerous crime (homicide average), and a lower number of restaurants in the area which means less competition for the owner of the restaurant. In the end, this is why the following features were selected.

Table 1. Simple feature selection during data cleaning – Merged data table with cluster labels





1. Exploratory Data Analysis

3.1 **Relationship between number of restaurants and Neighborhood**

Because the ideal restaurant location will be one in which there is little to no competition, the relationship between the number of total restaurants and each neighborhood of the borough ‘Downtown Toronto’ was analyzed. According to the data, the neighborhood of Rosedale has no restaurants in the area compared to the other neighborhoods in Downtown Toronto. This can be seen in the figure below.

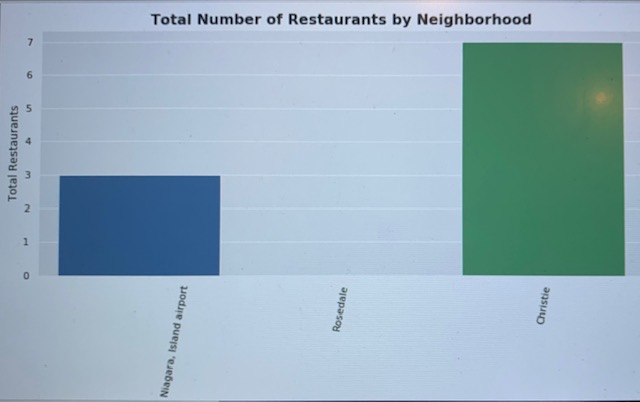


Figure 1. Bar graph of the total number of restaurants in each Downtown Toronto Neighborhood

3.2 **Homicide Average by neighborhood**

It is also important that the homicide average of the neighborhood in which the restaurant is opened is safe as well being that it means there will be less likelihood of thefts, murders, and also higher instances of rates of property value as well. Therefore, it was decided thatthe relationship between each neighborhood of Downtown Toronto and their homicide averages should be analyzed. According to bar graph below the neighborhood with the lowest homicide average is Christie with an average of 0.375 compared to 0.400 in the neighborhoods of Niagara and Rosedale. This can be seen in the figure below.



Figure 2. Bar graph of the average homicide rate in each neighborhood of Downtown Toronto

3.3 **Relationship between total number of restaurants and average spending power**

One other relationship that is important to analyze when attempting to find out where it would be best to open a restaurant in Downtown Toronto is to look at the relationship between the total number of restaurants in each neighborhood and the average spending power of each area. This was done because a high amount of spending power means more foot traffic and more people who spend their money on goods and services. Meanwhile, a high number of restaurants means more competition for the owner. Therefore, it would be best to see which neighborhoods have a high amount of spending power but a low number of restaurants in the area so that there is less competition for business. According to the bar graph below, the neighborhood with the highest amount of spending power paired with the lowest number of restaurants in the area is the neighborhood of Rosedale.



Figure 3. Bar graph depicting the average amount of spending power vs the total number of restaurants by neighborhood in Downtown Toronto

1. Predictive Modeling

4.1 **Solution to the problem**

According to the information given by the data table as well as the relationships shown betwixt the total number of restaurants, homicide average, and spending power of each neighborhood in Downtown Toronto, the best neighborhood in which to place a restaurant is the neighborhood of Rosedale. Rosedale has a fairly low average deadly crime rate of less than 0.5% and also has a high amount of foot traffic/ spending power for the area while also having no restaurants meaning no restaurant competition. Therefore, the ideal restaurant neighborhood location is the neighborhood located in cluster two, Rosedale.

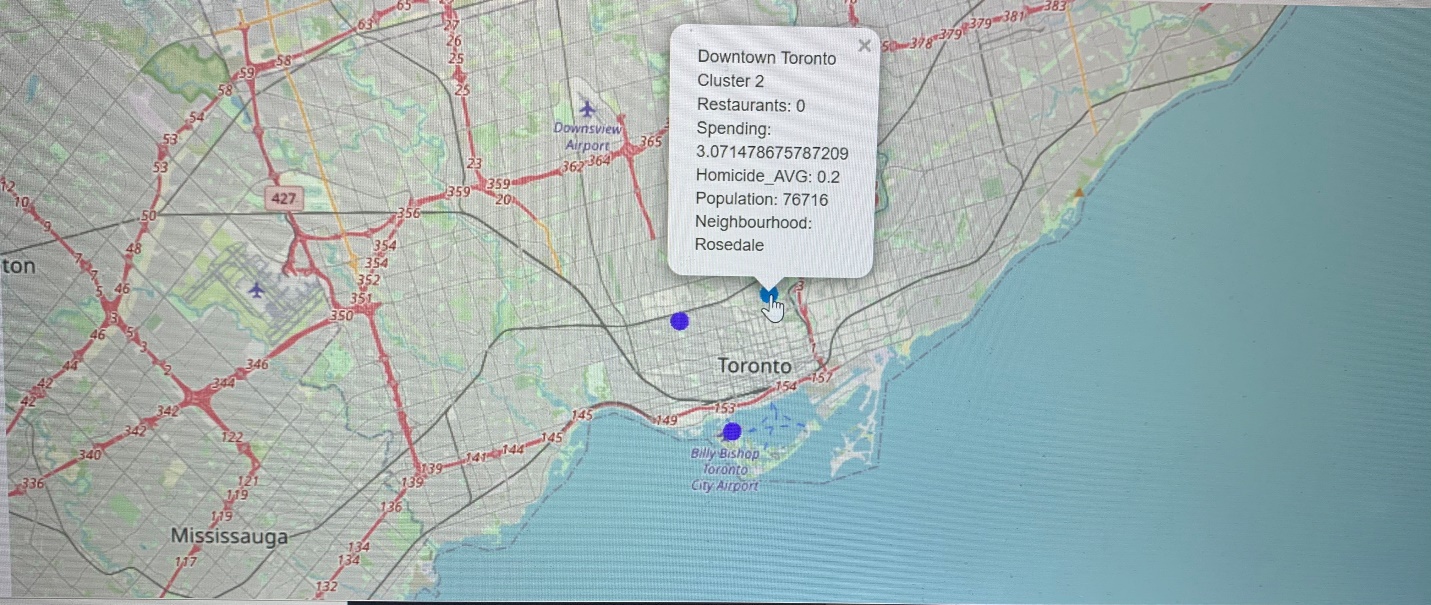


Figure 4. Cluster map of Rosedale including homicide average, spending power, and total number of restaurants

1. Conclusions

In conclusion, I analyzed the relationships between each neighborhood’s spending power, homicide average, and total number of restaurants in the area. I clustered the data of top venues by borough into the total number of sorted venues separated by restaurant types such as ‘Asian,’ ‘Columbian,’ or ‘American.’ I identified that the best borough in which to open up a restaurant after data cleaning was the borough of Downtown Toronto. I then narrowed this down further by analyzing the relationships between each neighborhood’s total number of restaurants, homicide average, as well as the relationship each neighborhood had between the total number of restaurants and spending power. These maps and graphs can be very useful in identifying the best area in which to open a restaurant because the city of Toronto has an ever expanding tourism industry.

Therefore, investors and restauranteurs interested in opening a restaurant in a prime location would be interested in knowing which area of the city would be best in which to open a restaurant as it can maximize profit. Such business people would therefore, be interested in learning which areas have the lowest crime rates, the highest amount of foot traffic and spending power, and the least amount of competition as was analyzed and discussed in this report. As a result, it was found upon analyzation of the data that the absolute best neighborhood in which to open a restaurant in the city of Toronto is the Rosedale neighborhood.

1. Future Directions

While I was able to determine that the neighborhood of Rosedale is the prime location in which to open a restaurant in the borough of Downtown Toronto, I think the models could benefit from using more updated census data as the city of Toronto is continuously developing due to tourism as well as other business ventures and changes in demographics. Also, it would be interesting to see how property values affect spending power and foot traffic of these areas and further how they interact with the total number of restaurants that are found in each area. These interactions could allow potential investors and business owners to better assess the potential profit values and margins when looking for neighborhoods in which to open up a restaurant.