

**Estimating the Best
Location for a
Restaurant Business
in Toronto**

A Tasty Opportunity

Problem

- The Toronto downtown core has a ton of restaurants of all sizes and cuisine types because of high foot traffic and these locations also have a relatively high cost of operation which comes as no surprise. Individuals who are interested in starting a cost-efficient restaurant business with potentially low cost of operation should ideally take all Toronto neighborhoods into consideration in order to pick the optimal location.

Opportunity

- This study will explore neighborhood population, crime rate, the number of existing businesses in the area to determine restaurant market saturation as factors to inform those interested in opening a new restaurant business to choose the best location in Toronto.

Methodology

Top 3 factors that we need to take into consideration in order to have a successful restaurant in a Toronto neighbourhood: Less number of existing restaurant businesses

1. Neighborhood population
2. Low crime rate.

For criteria #1, the Foursquare API was used to find the top 10 most common nearby venues. Toronto crime data was also leveraged to further narrow down our search for the neighborhood which fits criteria #2 and #3 best.

Target Audience

- The target audience of this study is anyone who is looking to start a restaurant business in the Toronto area and would like to pick a location with a relatively low crime rate so that their business has a better chance of thriving.

Data Acquisition

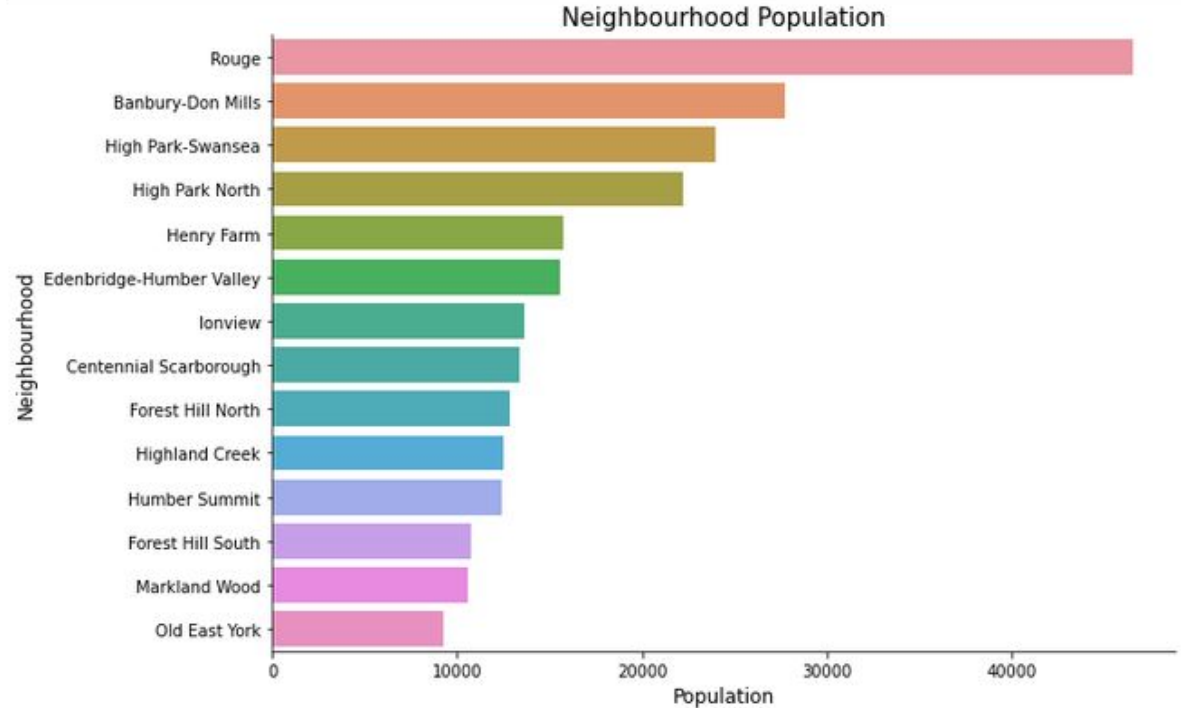
- Toronto Neighbourhoods Boundary File from Kaggle was used as the source for crime data which includes the following:
 - 2014-2019 Crime Data by Neighbourhood (specifically yearly counts for Assault, Auto Theft, Break and Enter, Robbery, Theft Over and Homicide).
 - Neighbourhood population based on the 2016 Census.
- Geopy.Geocoders library was used to get the coordinates for each of the neighbourhoods in the crime data.
- Foursquare API to enrich the dataset by finding nearby venue categories for each neighbourhood which was used to create summary of existing restaurants and businesses in each Toronto neighbourhood.
- The data was used to:
 - To normalize the data from yearly crime count to crime per 1000 people
 - To use it as an indicator for restaurant business demand.

Data Cleaning

- The crime data from Kaggle contained 140 Toronto neighbourhoods and 60 features.
- Neighbourhoods with unobtainable coordinates were dropped.
- In the final dataset, there were 109 neighbourhoods and their crime records along with their coordinates.
- Our dataset had 2498 rows with 268 features and contained Neighbourhood names along with their various venue categories.
- The dataset was grouped by Neighbourhood and used the mean value for the venue category.

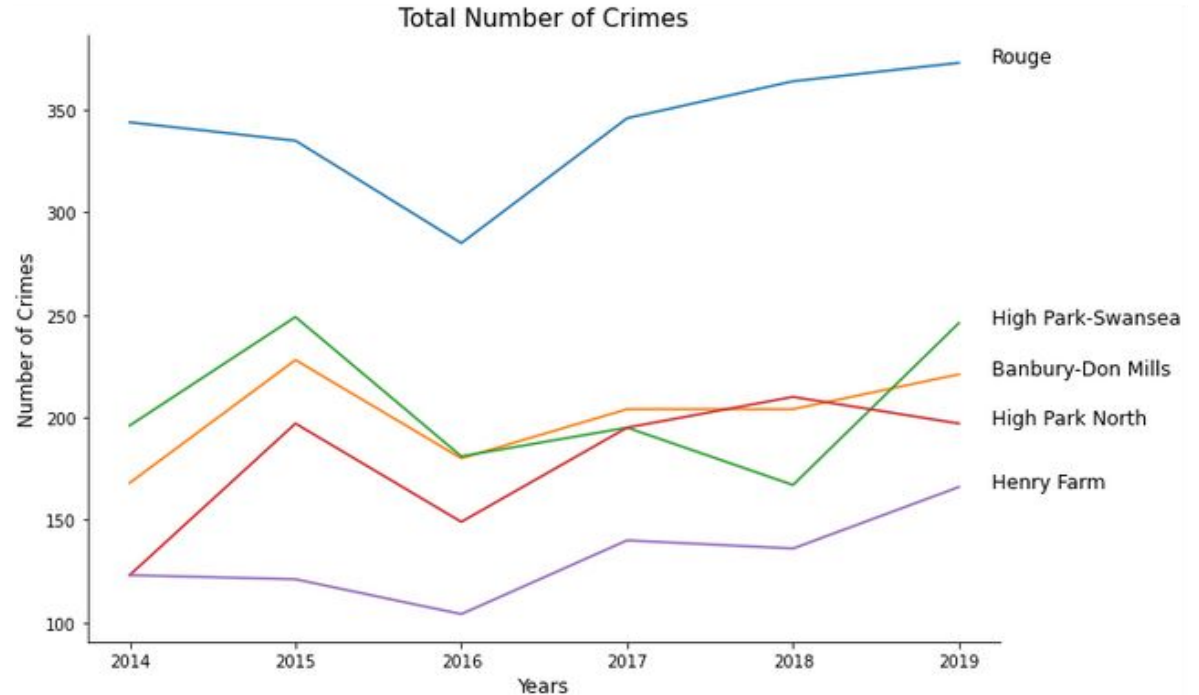
More People, Bigger Market

- The most 5 populated neighbourhoods within cluster 1 were identified.
- Rouge, Banbury- Don Mills, High Park Swansea, High Park North, and Henry Farm were identified as the top 5 most populated neighbourhoods within cluster 1.



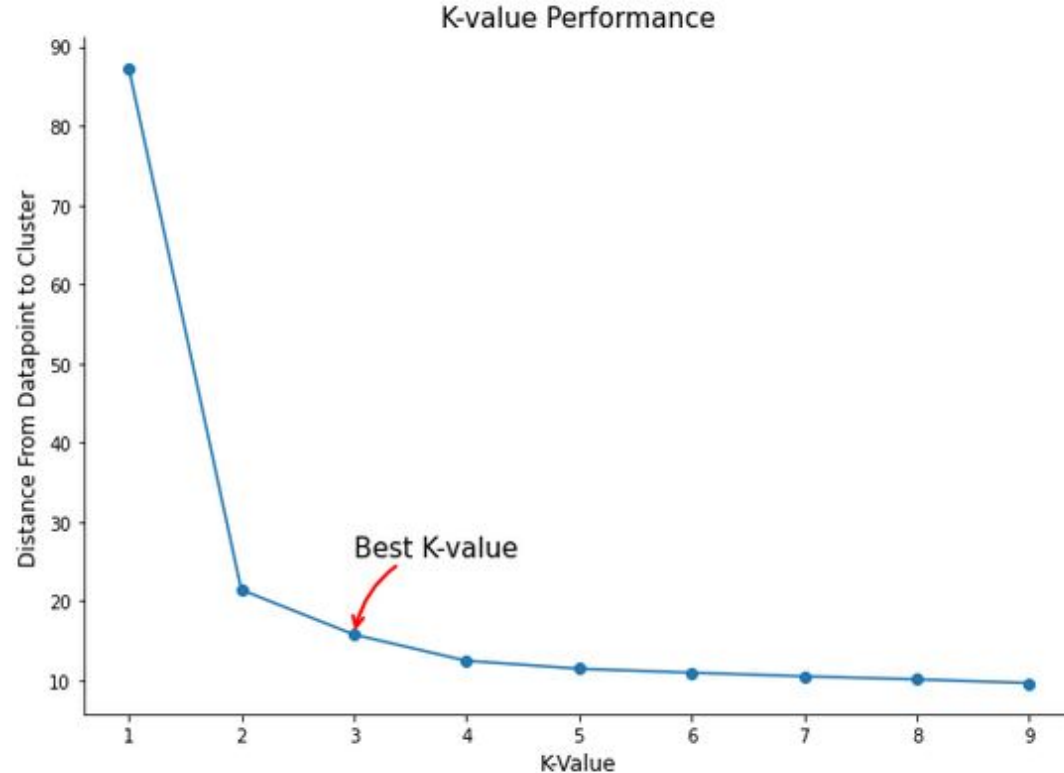
The Dish on Crime

- These 5 neighbourhoods were then analyzed to visualize the crime trend.



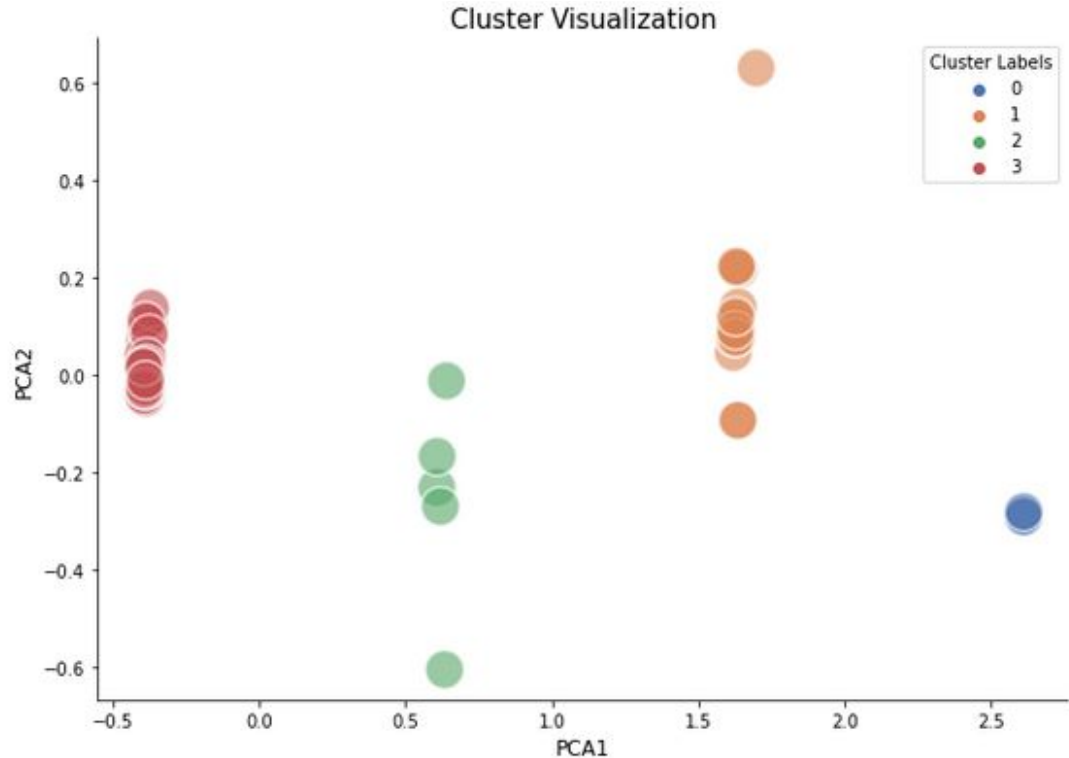
Finding the Best K-Value for K-Means Clustering

- K-means clustering was used to segregate each row based on their data points in order to visualize the market gap in terms of what businesses are currently available in the neighbourhood.
- The elbow graph method indicates that the K value of 3 is the ideal choice.
- Using inertia value from the sklearn library, we found the sum of distances of samples to their closest cluster center.



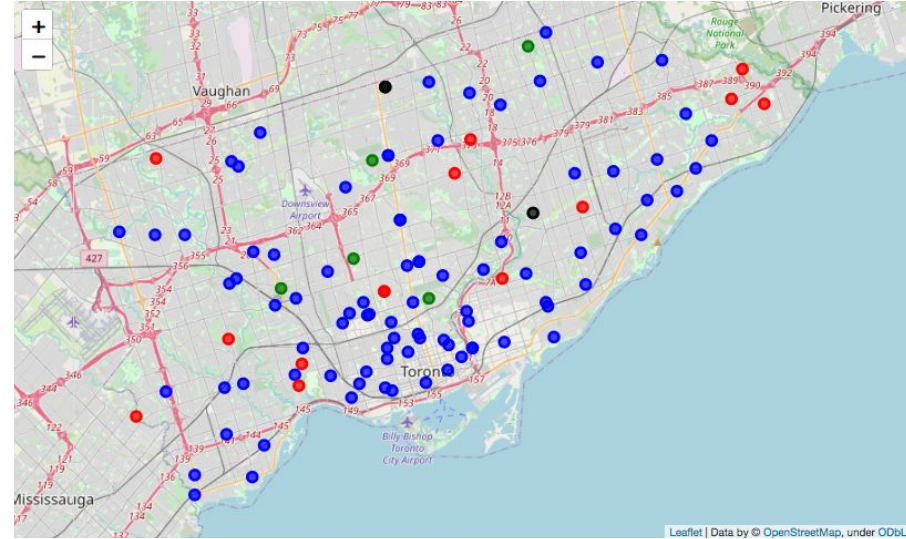
Cluster Visualization

- The PCA module was used to estimate the principal component from our features and reduce our feature down to 2 and visualize how the clusters are distributed.
- We used the combined dataframe for mapping our cluster to visualize the geographical location and their cluster category into 4 distinct clusters despite a small number of outliers.



Cluster Examination

The map plots the clusters on a geographical map to show each location in the data set. The colour indicates which cluster the neighbourhood belongs to. The red markers represent cluster 0, greens show cluster 1, blue indicate cluster 2, and black represents cluster 3.

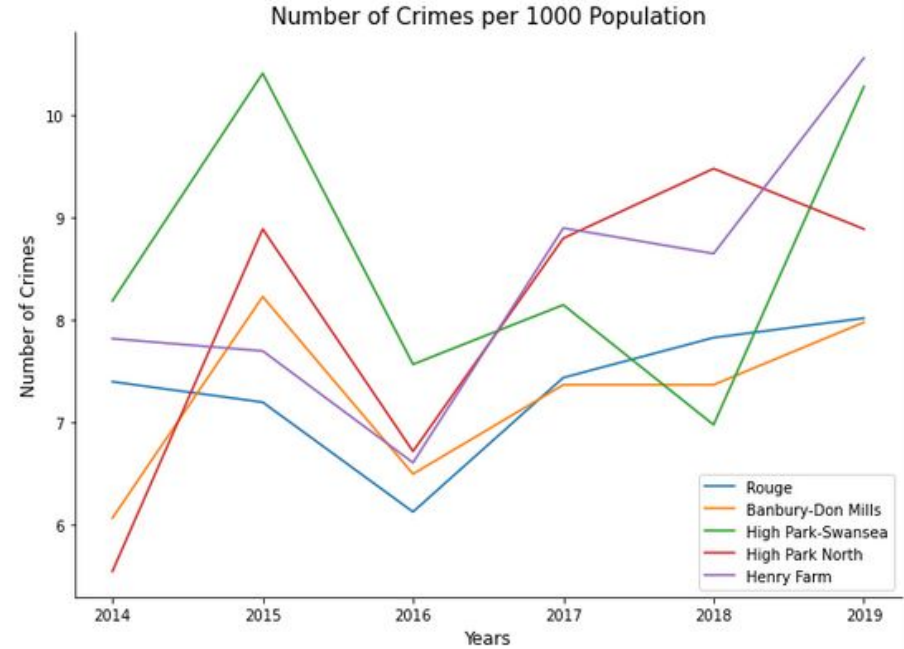


Diving Deeper on Clusters

- Cluster 0, indicated as red markers in the map below, includes a variety of restaurants, particularly eastern and middle eastern cuisine, as the top 10 most common venues. Since these neighbourhoods are oversaturated with restaurant businesses, opening a new restaurant venue in these neighbourhoods would be risky, particularly for those interested in starting an eastern or middle eastern restaurant.
- Cluster 1, highlighted as green markers in the geographical map below, have parks and recreational activity businesses as the dominant venues in that neighbourhood. Restaurant businesses rank from 6-10th most common venues in these areas therefore, cluster 1 meets requirement 1 (fewer existing restaurant businesses). A market gap exists providing the best opportunity to establish a new restaurant venue in these neighbourhoods.
- Cluster 2, shown as blue markers, includes mostly fast food restaurants and coffee shops. These businesses rank from the 2nd-5th most common venues in these neighbourhoods. Opening a restaurant in these areas would be risky.
- Cluster 3, shown as black markers in the map below, has a mix of different business types with pizza places and restaurants being the most common venues in these areas. Opening a new restaurant in these areas may be risky.

Crime as an Appetite Suppressant

- We calculated the crime per 1000 population.
- This graph indicates that Henry Farm neighbourhood crime trend is on the rise while Rouge and Banbury has a fairly stable crime trend,



Green is Go!

The top 3 requirements for the best neighbourhood to open a restaurant business includes:

1. Less number of existing restaurant businesses
2. Neighborhood population
3. Low crime rate

From this observation, neighbourhoods in Cluster 1 (green markers) provides strong support for our requirements.

The ideal neighbourhood to start a restaurant business where there is a good balance of neighbourhood population and crime rate could be either Rouge or Banbury since they both have the highest population in the cluster and also lowest crime rate. In addition, these 2 neighbourhoods also don't have many existing restaurant businesses which eliminates a lot of competition.

Further Implications

- It would be interesting to dive further into cluster 2. There are a large number of elementary schools in the areas saturated with fast food businesses and coffee shops. It would be interesting to see whether or not dine in family owned restaurants would succeed in these areas.
- In addition, it could be worth looking into other businesses that would succeed alongside middle eastern and eastern restaurants in cluster 0. For instance, perhaps opening an ethnic supplies store in this area would be beneficial, or perhaps it would be a great neighbourhood to host a cultural street festival.
- It might also be interesting to see affluence through house-hold income as a factor whether there's a reason cluster 3 has a large number of fast-food restaurants in those areas.