# **The Battle of Neighborhoods Final Report**

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

Toronto is the most well-known city in Canada and also one of the most densely populated. This makes it a very promising location to start a new restaurant, as the returns on investment are lucrative. However, given the multitude of options and competitors, it is necessary to undertake adequate amounts of research to ensure that the investment is fruitful.

* 1. **The Business Problem**

The key business problem that persists while opening up a new restaurant is choosing the right neighborhood for the business venture. There are numerous benefits associated with selecting a good location, including greater customer footfall, more security, lower marketing costs, as well as a greater level of overall profitability. On the other hand, choosing a less favorable location may result in greater costs and lower profitability, even if the quality of the food and ambiance within the restaurant are excellent.

Researching neighborhood data in Toronto to understand the types of restaurants that exist will enable a budding restaurant entrepreneur to make the right decision in opening up her restaurant. This project will examine Foursquare venue and location data to find the entrepreneur’s ideal location for the new restaurant. This project takes a look at the frequency of finding Japanese restaurants in Toronto. The criteria for setting up a new restaurant is that there needs to be a hotel and a park in the same neighborhood.

* 1. **Audience and Interest**

This research will be particularly valuable to new entrepreneurs who are seeking to open restaurants in Toronto. The objective is to pinpoint and recommend which neighborhood will be the best choice in Toronto and to rationalize why it is being recommended. Although this research project looks at Japanese restaurants, the findings will be useful to those setting up other restaurants as well. This project will also be of interest to others in the restaurant industry, such as those who are seeking to expand their business or those who aspire to relocate due to challenges in their existing location.

1. **Data Sources**
   1. **Data Acquisition and Cleaning**

The data requirements for this project are taken from:

* Neighborhood information regarding the boroughs, neighborhoods, and postcodes will be scraped from Wikipedia and a dataframe will be created: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* The geographical coordinates will be taken from a CSV file shared on this course h[ttp://cocl.us/Geospatial\_data](http://cocl.us/Geospatial_data)\_
* The Foursquare API will be used to find information regarding various locations and venues.

The data required extensive cleaning and modification to curate the findings. Firstly, the Wikipedia data was scraped to find the boroughs in Toronto, and the BeautifulSoup package was then applied to transform it into a pandas dataframe for further analysis. This is a python package which is used for parsing XML and HTML documents, which creates a parse tree for pages which can then be used to extract data, and is hence very useful for scraping websites.

For the purpose of this research, it was unnecessary to use all the information available in the dataset, hence the dataset was narrowed down to include only the relevant information. The top ten rows of this dataframe, which has thirty-nine neighborhoods across seven boroughs is depicted below. The postal code, which is not highly relevant, was omitted, and Figure 2 shows the results of the top five rows of this further refined dataframe.

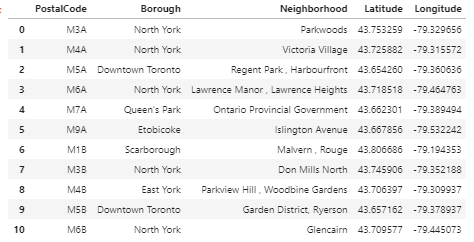


Fig 1: The first ten rows of the dataframe

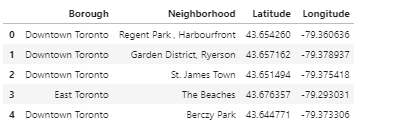
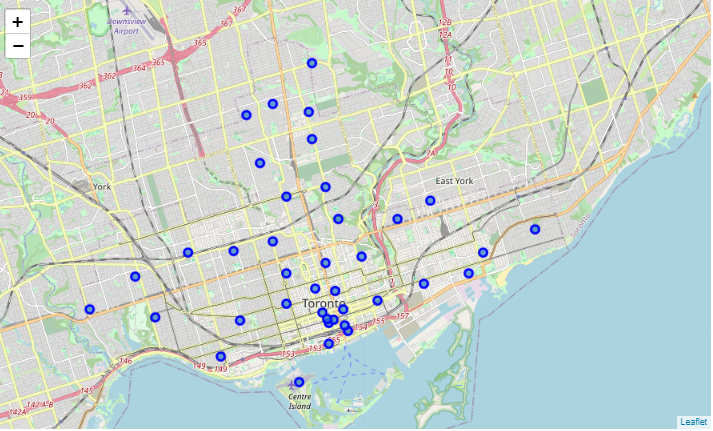


Fig 2: Top five rows of cleaned dataframe

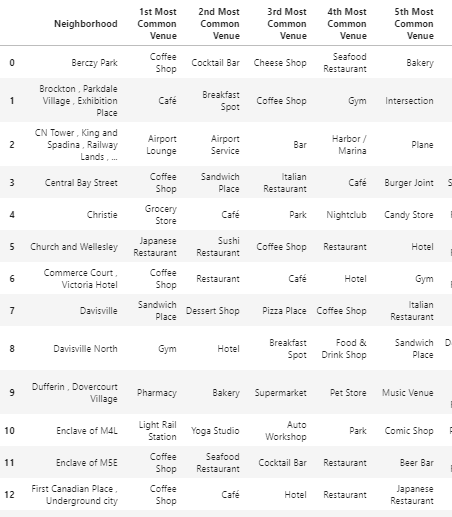
1. **Methodology**
   1. **Exploratory Data Analysis**

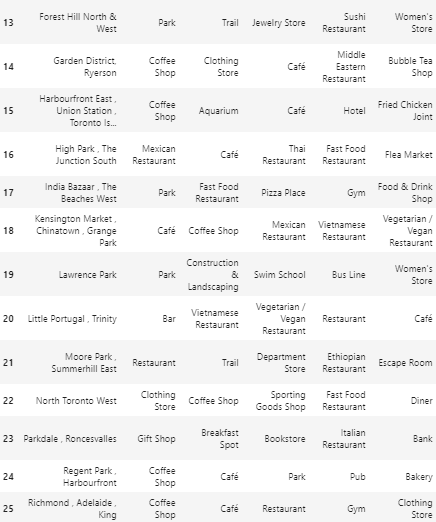
Python’s geopy and folium libraries were used to create a visual representation of the neighborhoods in Toronto. This is shown in Figure 3. The Foursquare API has been leveraged to provide information regarding the venues which are prevalent in each neighborhood. Toronto has 232 unique venue categories according to the Foursquare data, which were then inserted into a pandas dataframe.

Fig 3: Map of Toronto showing the various neighborhoods.

1. **Results**

The most popular venues in each neighborhood were then determined, using another Python staple function, numpy. This is shown in Figure 5.





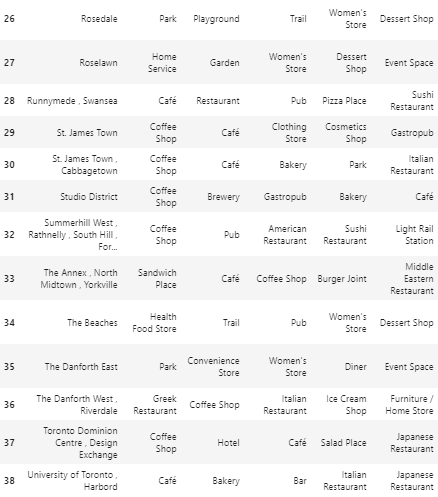


Figure 4: Most popular venues in each neighborhood in Toronto

To cluster the neighborhoods, the k-means clustering algorithm was utilized. This is an iterative form of clustering strives to divide *n* observations into *k* clusters, whereby each observation belongs to the cluster with the nearest mean. The k-means clustering of Toronto’s neighborhoods are depicted in Figure 5.

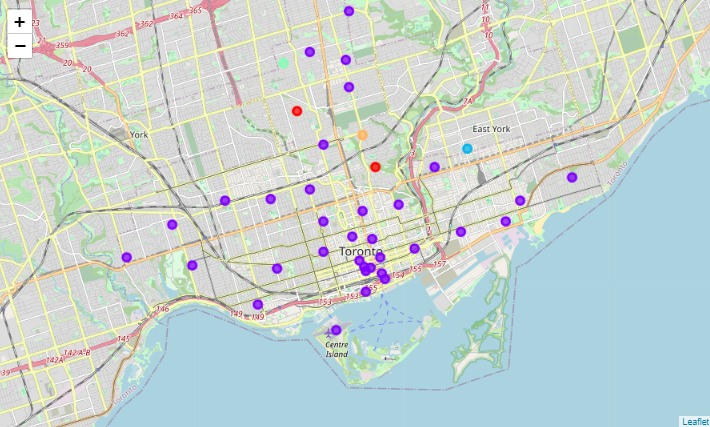


Fig 5: K-means cluster of Toronto neighborhoods

In order to decide where to open a new Japanese restaurant, it is imperative that we find the neighborhoods which already have one. For the purpose of this study, which was carried out with the impression that the restaurant to be opened is slightly upscale and trendy, it was ascertained that the neighborhood must also have a hotel and a park. The results from this are shown in Figure 6. After carrying out this analysis, it was arbitrarily decided that it would be nice to have a yoga studio in the neighborhood as well, and this is shown in Figure 7.

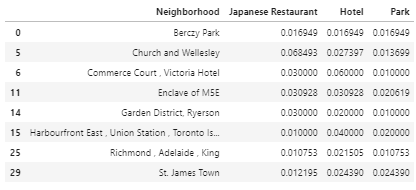


Figure 6: Neighborhoods with a Japanese restaurant, a hotel, and a park

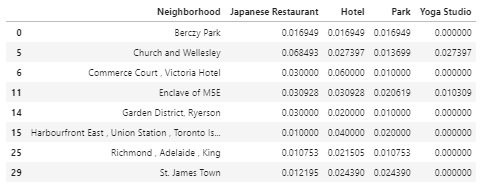


Figure 7: Neighborhoods with a Japanese restaurant, a hotel, a park, and a yoga studio

1. **Discussion**

The above results showcase that the neighborhood of Church and Wellesley may be a safe choice to open a new Japanese restaurant as it already has a many Japanese restaurants in the area, and also has a hotel, park, and yoga studio nearby. However, there are seven other neighborhoods with hotels, parks, and yoga studios, which may be better choices, as there is a possibility that the neighborhood of Church and Wellesley is already too saturated. The existing restaurants may have an established clientele and a loyal consumer base, which may make it challenging for a new restaurant to gain popularity in that neighborhood. Given that there is a scope of starting the new restaurant in seven other neighborhoods, it may be fruitful to explore these other locations first.

1. **Conclusion**

This project examined neighborhood data for the city of Toronto to narrow down choices to open a new Japanese restaurant. The Foursquare API was utilized to obtain venue information, and various Python tools such as Numpy, Geopy, and Folium were used to visualize and clean the date to gather the information that is required. Further research can be undertaken, especially that with more specific objectives, in order to make a well-informed business decision.