Clustering the Neighborhoods of Toronto

# 1. Introduction

1.1 Background

Toronto is the capital of Ontario and is the most populous city in Canada with a population of 2,731,571 in 2016. It is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. The cuisine of Toronto reflects the multicultural diversity of the city by being the host to restaurants belonging to various cuisines from all over the world like Chinese, Vietnamese, Korean, Indian, Italian etc. There are neighborhoods where the dominant cuisine is based on the demographics of the people living there. Certain neighborhoods are famous for certain types of cuisines. For example, Chinatown is famous for Chinese cuisine, Koreatown is famous for Korean cuisine and so on.

## 1.2 Problem

According to Yelp, a recommendation app, there are around 15,822 restaurants and 103 neighborhoods in Toronto. Therefore, tourists who are new to Toronto need help in deciding what neighborhood to visit in order to try the cuisine it is famous for. Tourist guides find it helpful if they know what neighborhoods are popular for specific cuisines so that they can recommend them to the tourists.

## 1.3 Interested audience

Tourists to Toronto would be interested in finding the neighborhoods in which their favorite cuisine is famous. Tourist guides find it helpful in recommending those neighborhoods based on the interests of their clients. Entrepreneurs can find what the dominant cuisine of a neighborhood is in order to decide if they want to establish restaurant of a different cuisine to avoid competition or choose a different neighborhood.

# 2. Data acquisition and cleaning

## 2.1 Data Sources

Data about the boroughs, neighborhoods and postal codes in Toronto taken from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>.

The coordinates of each neighborhood according to its postal code is given in <https://github.com/kamis1r/Coursera_Capstone/blob/master/Geospatial_Coordinates.csv> .

There are 11 boroughs and 103 neighborhoods in total.

Details of restaurants can be obtained by refining location data given by Foursquare API. To do this, firstly, a search request is made to tabulate all venues in a particular neighborhood. The venue ID from the table is then used to make a request to obtain details of the venue like the menu, price, rating, category. Since we are clustering neighborhoods based on the most liked category of restaurants, we are only interested in the “category” field of the data.

## 2.2 Data Cleaning

Rows where boroughs are not assigned are dropped. If a row has a borough and a postal code, but has no neighborhood, the borough is assigned as the neighborhood. The data in the geospatial coordinates table did not need any cleaning. Venue details obtained from Foursquare are filtered to get only the venue ID, venue neighborhood, category of the venue has.

We only use neighborhoods that belong to boroughs that contain the word “Toronto” in their names.

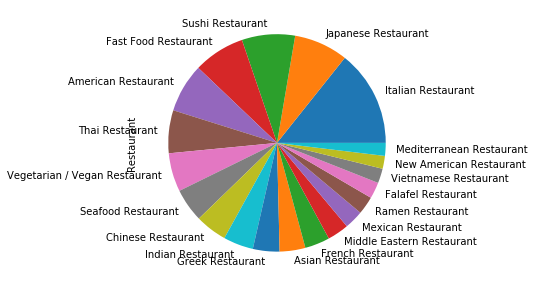
A screenshot of a computer

Description automatically generated

# 3. Methodology

## 3.1 Exploratory Data Analysis

There are 64 unique categories of restaurant ranging from Western cuisines like American and Italian to Eastern ones like Thai, Japanese, and Indian in Toronto. Italian is the most popular followed by Japanese. The least popular cuisine among all cuisines is Afghan. Down below is a pie chart of representing the share in popularity of the top 20 popular cuisines of Toronto.



Among the Neighborhoods in Toronto, the neighborhoods with the highest number the restaurants are Chinatown, Grange Park and Kensington Market with 61 restaurants. The neighborhoods with the least number of restaurants are Roselawn, Lawrence Park with only 1 restaurant in each. The descriptive statistics of neighborhood and restaurant data is given below.

A screenshot of a cell phone

Description automatically generated

There are 38 neighborhoods that contain the word “Toronto” in the names of their Boroughs They are visualized on a map below created using folium library.

A close up of a map

Description automatically generated

## 3.2 Clustering

We use an iterative clustering algorithm called the K-means clustering algorithm by the sklearn library in Python to cluster the neighborhoods based on the frequency of the top 7 cuisines in the neighborhoods. The number of clusters is selected beforehand to be 7.

K-means is appropriate for this problem and is used because it is a type of unsupervised clustering where we do not know the distinguishing factors of the resultant clusters or groups.

The figure below is obtained after k-means clustering.

A screenshot of a cell phone

Description automatically generated

# 4. Results:

After running the K-means clustering algorithm, 7 clusters are produced that have similar cuisines within a cluster. The clusters of neighborhoods are visualized on a map as shown below with each color indicating a specific cluster using the folium package.

A picture containing text, map

Description automatically generated

# 5. Discussion:

A screenshot of a social media post

Description automatically generated

In Cluster 6, we can observe how the K-means clustering algorithm successfully grouped the neighborhoods of Christie and Roselawn since they have similar cuisines like Italian, Hawaiian, and Greek.

There are also clusters of a single neighborhood and clusters where there are more than 10 neighborhoods in them.

# Conclusion:

In this study, I analyzed the popular cuisines of different neighborhoods in Toronto using the foursquare API and clustered the neighborhoods using a machine learning algorithm called k-means clustering. I used the coordinates of the neighborhoods from a different table to visualize the clusters on a folium map. These mapped clusters can be very useful for tourist guides to recommend their clients what neighborhoods in Toronto to visit to get a taste of a cuisine of their choice. Restaurant entrepreneurs can examine the clusters closely to look at the dominant cuisines of a neighborhood and make financial decisions.