RECOMMENDATION

For opening **new ramen restaurants in Toronto**

Ziming Li

TABLE OF CONTENTS

1. INTRODUCTION

As the largest city in Canada, Toronto combines multi cultures, traditions and foods from all over the world. One of the most popular food you can find in the city is Japanese ramen. While walking on areas like downtown or North York, you will often see find some ramen places that are full of people line up outside, despite winter or summer.

For owners who wish to open a restaurant that offers Japanese ramen, they need to know where the other ramen places located. This is important because it provides owners an overall picture of where their competitors and customers will be mainly located. They also need to know what type of the restaurant should be opened in that area.

This project analyzes all restaurants that serve ramen in Toronto's boroughs and neighborhoods, and groups similar kind of areas into a cluster, based on ramen places' rating, price tier and restaurant type, and eventually give owners a guideline of where and how the new ramen places can be located.

2. DATA

For this project, the following data will be used for analysis and modeling:

- ***Toronto postal code & neighborhood data*** *– this dataset provides postal code, boroughs and neighborhoods for the entire city*
- ***Geospatial data*** *– this dataset maps each postal code to latitude and longitude*
- ***Venue data returned from Foursquare API*** *– this dataset is an addon based on neighborhood data. It provides venues info including location, reviews, likes, etc. In our case, we will focus mainly on restaurants that serve ramens within each neighborhood. We will also get ratings and price tier info of each restaurant/venue.*

3. METHODOLOGY

3.1 Data Exploring

The original (raw) data we have is the borough & neighborhood of Toronto provided by Wikipedia. This is a good starting point as we want to analysis the location and restaurant type based on each area of Toronto city. We then use Geospatial data, which is provided by city of Toronto website, to map each neighborhoods' zip code into a pair of latitude and longitude. Once the mapping is done, our dataset looks like following as an example:

Postal Code	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Rouge, Malvern	43.806	-79.194
M1C	Scarborough	Highland Creek	43.784	-79.16
M1H	Scarborough	Cedarbrae	43.77	-79.23

Then we use Foursquare API to find actual restaurant info for each borough. Foursquare API is a map API that can provide developer venue (store, restaurant, etc.) information, including venue type, price tier, reviews, photos and much more. In our dataset, we need restaurant information including restaurant name, restaurant geo location – latitude & longitude, restaurant category, price tier and ratings. Once the restaurant detail is explored, our dataset now has Toronto neighborhoods, latitude, longitude, and all details of all restaurants that serves ramen in that neighborhood, looks like following:

Neighbor	Lat	Lng	Venue	Venue Category	Price Tier	Rating
L'Amoreaux	43.7	-79.3	Yamamoto Japanese Cuisine 山本盛世	Japanese Restaurant	2	5.5
Willowdale	43.7	-79.3	KINTON RAMEN	Ramen Restaurant	0	6.7

3.2 Data Analysis

As the result of data exploring from previous section, the data now has both Toronto city neighborhoods info, but also any price tier and rating info for all restaurants that serve Japanese Ramen in each neighborhood.

However, there are some empty or invalid data: Price Tier and rating data is missing for some restaurants. This could be due to the fact that those restaurants are relatively new and is lacking user or owner's information, or it could be that Foursquare does not have those data in server. To fix that, we set those missing values to 0. While this fix might not be very intuitive and meaningful for each restaurant, it is still valid when we group restaurants by the neighborhoods and take the average value of price tier and ratings. This is because we eventually want to do district analysis, so we mainly want to analyze the average performance of restaurants from each neighborhood.

Then we apply one hot encode for the categorical value – restaurant category. We also keep price tier and rating fields as both are numerical values.

Next, based on the frequency of restaurant type, we calculate the most common type, second common type, and third common type of ramen restaurant and attach it into each row (neighborhood) of data.

Finally, we group the data by neighborhood and take the average (mean) value of each field, including price tier and ratings, so that now each row contains the overall performance of all Japanese ramen places within the same neighborhood.

Now that the dataset is transformed into featuring dataset, with one row per neighborhood info in numerical form, we are ready for configuring model and fitting the model with the dataset.

3.3   Modeling

To find out how other ramen restaurants behaves in each neighborhood, how similar some restaurants are, we need to group them into different clusters. Therefore, we use K-Means Clustering algorithm.

In order to determine the best cluster numbers (i.e. k), I tried different k from 1 to 15, and plotted the distortion versus k which is an elbow plot.

The Elbow Method showing the optimal k
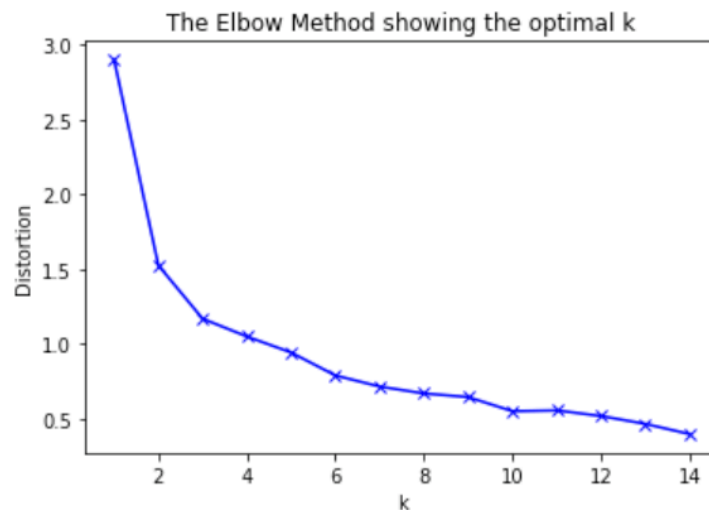
Based on the plot, the distortion decreases massively when k is from 1 to 4, and then decreases at slower pace when K is around 6. Therefore, K is chosen to be 6.

Then we fit the model with the prepared data from previous section. Once the algorithm fitting process is completed, we use the model to assign group number for each neighborhood.

Final dataset with the clustering label looks like following:

Borough	Neigh	Lat	Lng	Cluster	1st Common	2nd Common	3rd Common
Scarborough	L'Amoreaux	43.8	-79.3	5	Japanese Restaurant	Thai Restaurant	Bakery
North York	Don Mills	43.7	-79.3	2	Ramen Restaurant	Japanese Restaurant	Giftshop

Where the **cluster** column is filled by the K-Means Clustering model and is from 0 to 5. The next section will provide a visual result of each neighborhood's clustering based on ramen places' feature.

4. RESULTS

In order to show each area's clustering and some feature info, we use map feature from folium package to draw a map of Toronto city, as well as add markers for each resulting data from previous section. The map looks like the following:
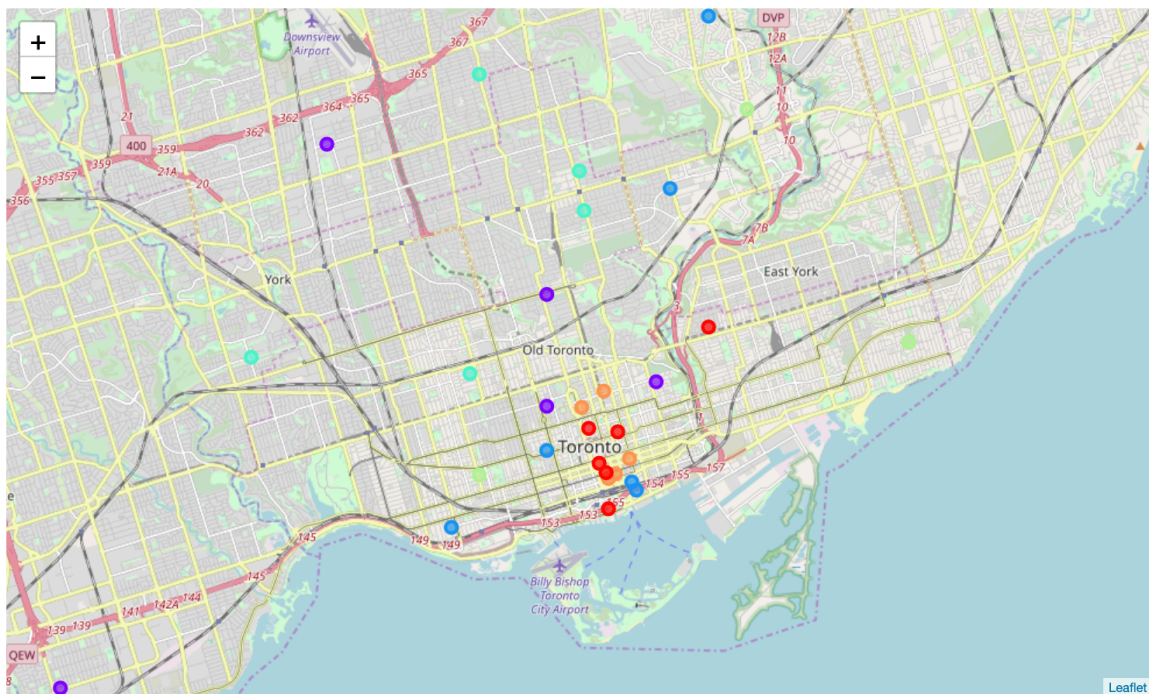
Where each color stands for a group (or cluster), meaning those markers with same color has similar performance in terms of:

1. *Similar ratings*
2. *Similar price range*
3. *Restaurant type is similar. e.g. most of them are Japanese Restaurant, second most common type is food court, etc.*

Based on the map, restaurant owners who wish to open a new ramen place can visually get the info that:

1. *For each neighborhood, what kind of restaurant offer Japanese ramen that is most common.*
2. *Once the owner has decided the price tier of menu as well as restaurant type (e.g. sushi restaurant or pure ramen place), he will know which group of restaurants his new one might belong to.*

5. DISCUSSION

Based on the distribution of all restaurants that offer ramen in Toronto, we can see clearly that:

1. *Downtown area has all kinds of clusters, meaning there are all kind of ramen restaurant type, from bakery to Japanese restaurants, and from pure ramen place to food court.*
2. *In some neighborhoods, Japanese ramen has been served at other Asian restaurants including Thai and Chinese restaurants. This could either be an innovative idea from those venues, or it could be the incorrect info that user or Foursquare API provided.*

In order to make the model more intuitive and accurate in the future, we can try other mapping services like Google Map, so that we can do a comparison of how each model clusters the ramen places with more details.

6. CONCLUSION

For this project, we aimed at store managers and owners who wish to open a new restaurant that offers one of Toronto's most popular Asian food – Japanese Ramen. We want to provide them a guide for what their restaurant type could be if they want to open the place in a neighborhood. We also provide them what are other competitors that is similar to the store they planned to open, so owners can then study those places online and gather experiences.

We started with raw geo code and neighborhood data from wiki, processed them with City of Toronto public data which added latitude and longitude. We then enhanced the data by adding restaurant information (geo code, price tier, restaurant type and rating) using Foursquare API.

Finally, we use K-Means clustering algorithm to group those similar restaurants based on their type, price range and rating, thus assigned complete data with cluster labels.

The algorithm can provide user reference info that:

1. *What group of restaurants could owner's new venue belong to?*
2. *What are his potential competitor's ramen place's type and features?*