**Eating Around Toronto**

**- Exploration of Restaurants Clustering in the City of Toronto**

**Yu Chen**

**December 18, 2020**

1. **Introduction**

**1.1Background:**

One thing people care about international diversified cities is the

gourmet. No matter you are looking for a restaurant to feed your hunger or you are looking for the most trendy spot to run a restaurant, it is necessary to understand the city’s Gourmet Map. The category of restaurant is important, for example, Asian restaurants may cluster in certain places where Asian people congregate and Asian flavor is welcomed. Different districts, or neighborhoods represent different demographic features, therefore may have their distinctive appetite for restaurants. However, some neighborhoods maybe a mosaic of various ethics thus various types of restaurants prospers there. What‘s more, according to the functions of different neighborhoods, wine bars, fast food restaurants, dinner restaurant will find their own suitable locations.

This project helps to give you an overall view of Toronto Restaurants and tries to identify the possible clustering in its cosmopolitan areas. It will also give guidance about the location selection regarding opening a restaurant.

**1.2 Problem:**

By utilizing the Foursquare neighborhood API, we are able to discover the venues in neighborhoods. But we need to get the coordinates of every neighborhood we want to explore. We also need to choose the right endpoints to conduct proper queries. In order to further explore the gourmet map, we need to get the restaurants among all venues and build a clustering model to distinguish them.

**1.3 Interest**

Obviously, people who plan to open a restaurant in urban Toronto will be interested in exploring the gourmet map. Tourists, local residents will all would like to know where to find their favorite restaurant.

**2. Data acquisition and cleaning**

**2.1 Data sources**

Coordinates data of Toronto neighborhood was obtained from Wikipedia.com. Venue information comes from Foursquare.com developer version.

**2.2 Data cleaning**

Download the postal code of every borough and neighborhood from url='https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M'. Read in as a table, check the ‘not assigned’ values, get rid of these values. Import the Geocoder library to extract the coordinates information, match it to the neighborhoods. Since the area we try to discover is Toronto urban, we filtered and obtained the boroughs which contain “Toronto”.

**2.3 Data preparation**

Next, we want to know what the json file from Foursquares looks like. We prepared the API. Developer version needs a user credential and user ID. Other fields in the API URL includes version, latitude and longitude. Try with a specific pair of latitude and longitude. Examine the structure of the json file, obtain the layer that contains the venues information.

Using the query “restaurant” to further filter the venues whose type is “restaurant”. Get all the restaurants neighborhood by neighborhood using a loop, reading neighborhood’s coordinates from the table we get from data cleaning step.

Therefore, we have obtained all the restaurants in Toronto urban with their coordinates, categories as well as the neighborhoods in which they are located.

**3. Exploratory Data Analysis**

Playing with this Toronto restaurants list, we found there are 55 different categories of restaurants in total. There are several neighborhoods, such as Garden District, Central Bay Street and others have more than 40 restaurants there, which shows many people choose to eat there. We want to further explore whether there are some similarities between some neighborhoods regarding the restaurants.

To count the number of each category in every neighborhood, we use get\_dummies() method from panda library. This function helps calculate the statistics of restaurants, so that we can easily get the average number of each category in each neighborhood.

We sort within each category to get the top 10 categories in each neighborhood.

We can see in some neighborhoods, Chinese restaurants rank 1 while in some neighborhoods , Sushi restaurants rank 1; In some neighborhoods, the bars, nightclubs are the most frequent restaurants while in some other neighborhoods, dinner restaurants are the majority.

**4. Clustering Modeling:**

Here we want to use unsupervised method to classify those neighborhoods. K-means is a good automatic classifier which calculate the Euclidean Distance of different features. The method randomly selects a pre-defined K centroids, calculate the distance which results in K initial clusters. Then move the K centroids to the mean of each cluster, recalculate the distance we have new clusters. Repeat this process until the total distance to each centroid reaches the minimum. The python library provides the package kmeans from sklearn clusters to perform the classification.

We apply the restaurant data to the classifier, the target are four restaurants clusters, the features are 55 restaurants categories. We use the K-means classifier the 32 neighborhoods into 4 clusters.

**5. Conclusion**

K-Means Classifier put the neighborhoods into 4 clusters as follow, we retrieved the top 10 restaurants categories of each cluster to better understand how the groups are divided.

Cluster 1 consists of 3 neighborhoods, the similarity of these 3 neighborhoods is that there are a lot of Indian, Italian, Greek, French restaurants, and ice cream shops and Gastropubs in these neighborhoods.

Cluster 2 consists of 28 neighborhoods, where the Chinese, Vietnam, Japanese, and Korean restaurants rank top among various restaurants.

Cluster 3 consists of only one neighborhood, where the Asian restaurant rank top 1 among all others.

Cluster 4 consists of only one neighborhood, where Portuguese restaurant rank first among all others.

This classification gives instruction on the diversification and distribution of various types of restaurants in Toronto urban areas. As a international cosmopolitan, Toronto has all kinds of restaurants from all over the world. With an understanding of the restaurants clusters, tourists or citizens can be easier to make a choice when deciding where to eat, and it will also help entrepreneurs to make decisions regarding where to start an restaurant business.