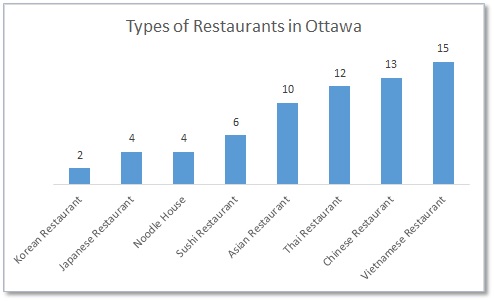
# **Coursera Capstone Assignment-Report**

## 1.1 Introduction – Background

Ottawa, Ontario has a limited variety of Asian cuisines compared to cities like Toronto and Vancouver. Currently, there are total of 66 Asian restaurants "registered" with Foursquare[[1]](#footnote-1). Ottawa is mainly dominated by Asian, Thai, Chinese and Vietnamese Restaurants.



The ratings (out of 10) for these restaurants are not very high.



The overall average rating for Asian cuisine is 6.2.

## 1.2 Introduction – Business Problem

With the growing Asian community, there is a craving for good Asian restaurants. Ann, the client, just moved to Ottawa and is looking to open an Asian restaurant. She would like to find a location with high population density, a high ratio of Asian population, and a low number of popular Asian restaurants as competition.

Where should Ann open her Asian restaurant?

## 2. Data

In order for Ann to make an informed decision as to where she should open her restaurant, we need to gather some data to determine where the current Asian restaurants are located as well as population data. Ideally, Ann would like to open her restaurant in a highly populated area with a high Asian ratio with low Asian restaurant competition.

1.Neighbourhood boundary data (<http://data.ottawa.ca>).

* This data will be used to visualize Ottawa's neighbourhoods.

2.Ottawa Population data (http://data.ottawa.ca/group/demographics).

* Population data by neighbourhood.
* This data will be used to help determine the neighbourhoods with the highest population density.

3.Asian Population data by ward (http://data.ottawa.ca/group/demographics).

* Unfortunately, there was no data by neighbourhood.
* This data will be used to determine the areas in Ottawa where there is the highest Asian population.

4.Ottawa restaurant information will be gathered from Foursquare.

* Data gathered: Venue ID, Venue Name, Address, Geo Coordinates, Venue Category
* This data will be used to illustrate where the restaurants are located in Ottawa.

5.Asian restaurant details will be gathered from Foursquare. "Asian restaurants" are defined as Noodle House, Chinese, Sushi, Vietnamese, Thai, Asian, Japanese, Korean categories.

* Subset of Ottawa restaurants.
* Specific venue data gathered: Venue ID, Ratings, Stats, Price, Likes.
* This data will be used to group (clustered) similar venues together to get a better idea of the Asian restaurants in Ottawa.

## 3. Methodology

The Methodology section will describe the main components of our analysis and prediction system. The Methodology section comprises four stages:

1. Collect and Inspect Data
2. Explore and Understand Data
3. Data preparation and preprocessing
4. Modeling

### 3.1 Collect and Inspect Data

#### 3.1.1 Data Cleaning

Data was downloaded or scraped (then saved) from multiple sources as mentioned in the Data section. The neighbourhood, ward boundary and population data were from government official websites and therefore, very little cleaning had to be performed. The population data came from Canada’s Census data, and only the Asian populations was extracted from the large dataset. Neighbourhood boundary files were merged with population data. Ward boundary files were merged with Asian population. Unfortunately, I was not able to find Asian population by neighbourhood.

There were some challenges with retrieving data from Foursquare. The rate limit restricted the number of records I could pull in a day. As a result, data needed to pulled over multiple days and combined into one dataset.

Using a reduced list of postal code geo-coordinates for the city of Ottawa enabled me to request a data pull from Foursquare. Using geo-coordinates, a radius of 10,000 m and a query on ‘food’ venues, a list of Ottawa restaurants and some details was retrieved (i.e. Venue ID, Venue Name, Address, Geo Coordinates, and Venue Category). Any duplicate records were removed from the dataset.

Foursquare Premium API was used to gather more details about specific venues. Id, name, location, location, stats, hours, contact, menu, price, rating, and likes were gathered. After some data inspection, it was decided that hours and menu would not be used due to the inconsistent data and many records having missing values. Due to the rate limit restrictions, the data pull needed to run a number of times to capture all venue details. The datasets were combined into one.

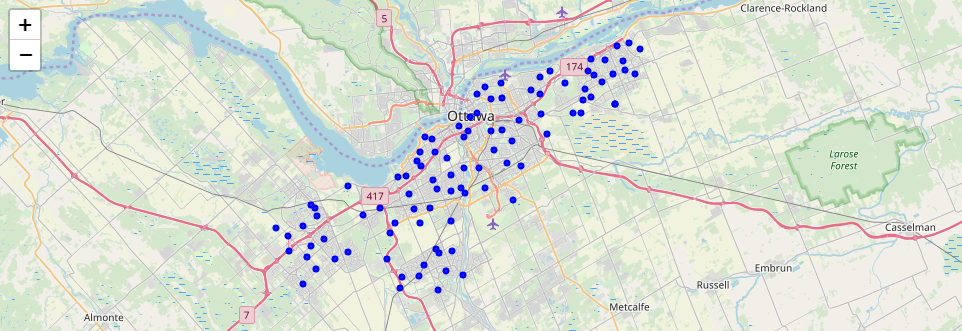
#### 3.1.1 Feature Selection

Ratings, likes, stats, price, and online presence features were used.

### 3.2 Explore and Understand Data

#### 3.2.1 Select Ottawa’s Geo coordinates for Foursquare API

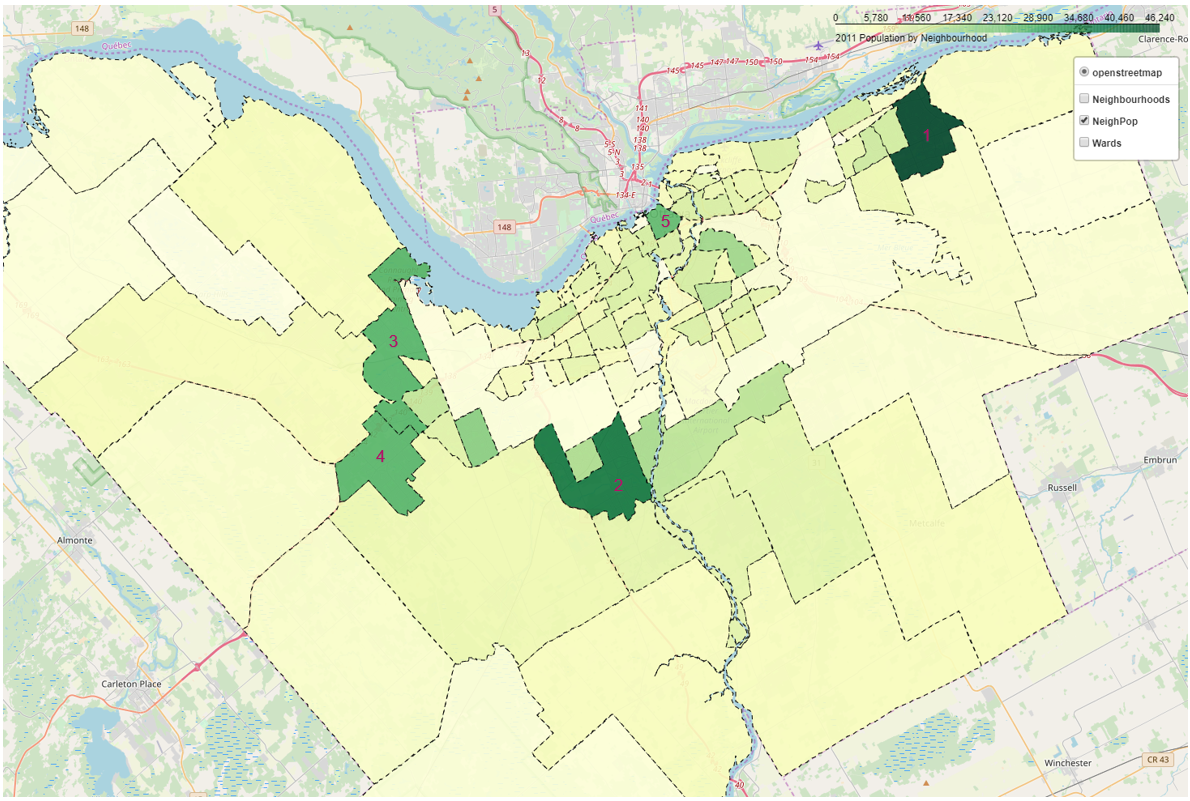
After downloading Ottawa’s postal code geo-coordinates and getting all the geo-coordinates mapped, it was determined that the markers were too dense. There were too many markers and as a result, a reduced file was created with about a quarter of the data. Below is a map with geo-coordinate points used in the Foursquare API. The points are fairly evenly spread out in areas of development (i.e. no dense clusters of blue dots).



#### 3.2.2 Neighbourhood population

After graphing Ottawa’s neighbourhood population, the top 5 most populated neighbourhoods were identified[[2]](#footnote-2).

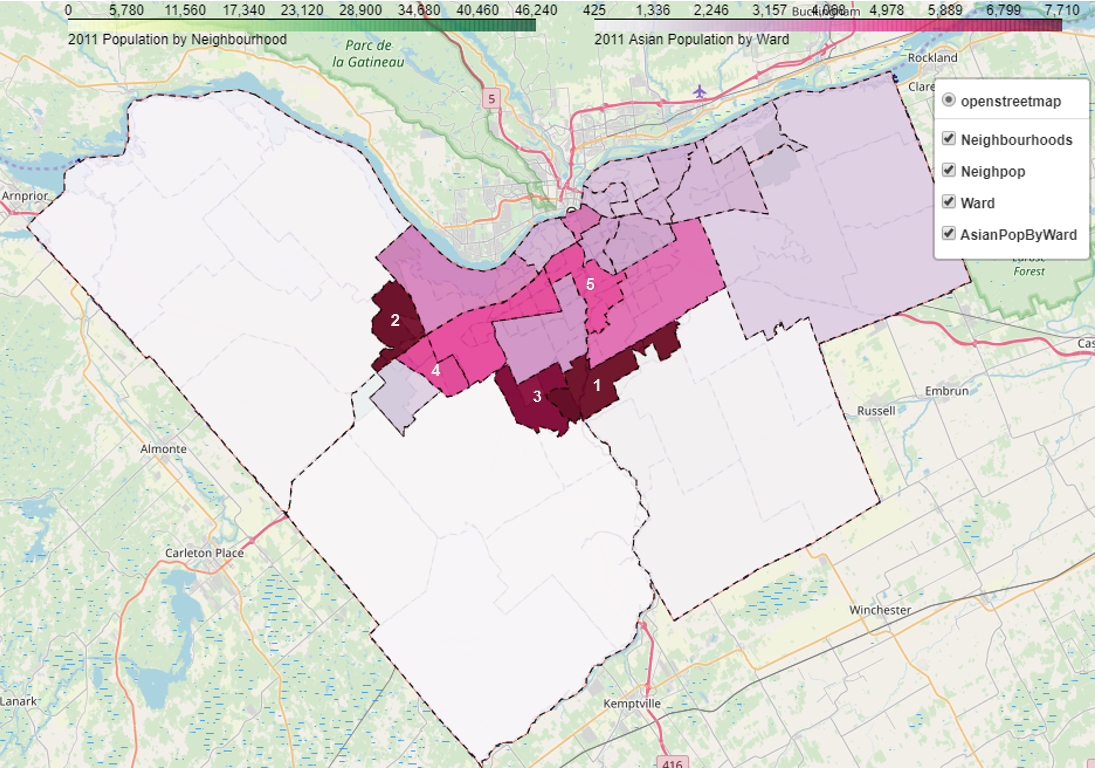
1. Orléans Avalon - Notting Gate - Fallingbrook - Gardenway South
2. Stonebridge - Halfmoon Bay - Heart's Desire
3. Kanata Lakes - Marchwood Lakeside - Morgan's Grant- Kanata North Business Park
4. Stittsville
5. Centretown



#### 3.2.3 Asian population

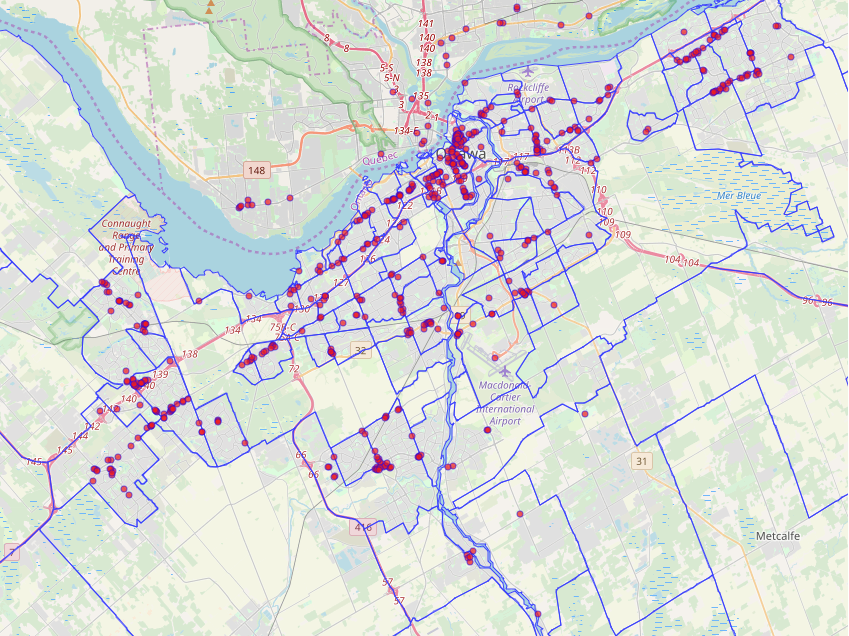
After graphing Ottawa’s Asian population, the top 5 most populated wards were identified[[3]](#footnote-3).

1. Gloucester-South Nepean
2. Kanata North
3. Barrhaven
4. Kanata South
5. College



#### 3.2.4 Ottawa Restaurants

After mapping the restaurant geo-coordinates gathered from Foursquare, the locations can be easily seen on this map. There are a total of 561 restaurants registered with Foursquare.

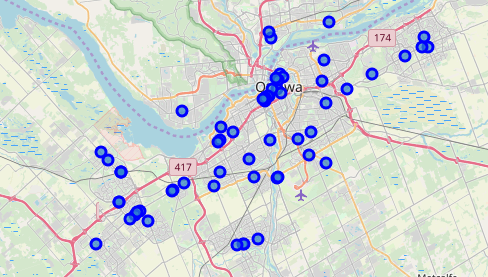


As expected, the downtown Ottawa location has a dense arrangement of restaurants.

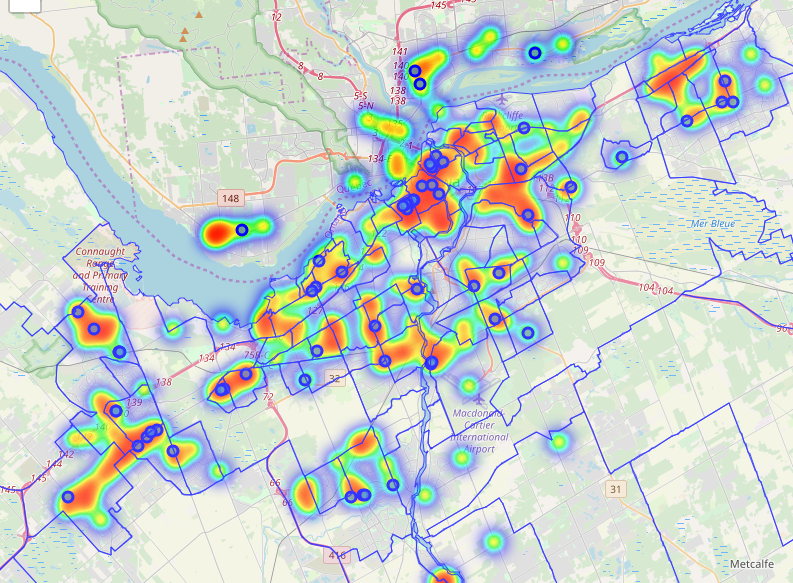
Refer to Appendix 3 for restaurant categories. Ottawa is mainly dominated by fast food restaurants, sandwich and pizza places.

#### 3.2.5 Asian Restaurants

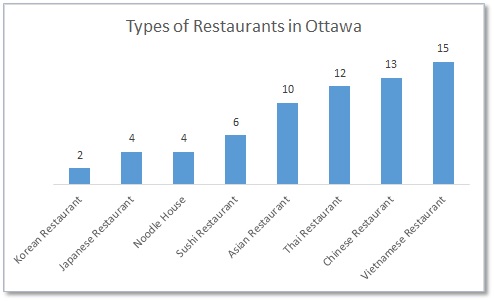
Locations of the Asian restaurants are identified in the map below. There are a total of 66 Asian restaurants registered with Foursquare.



The heat map identifies Ottawa’s densest restaurants (indicated in red). Asian restaurants tend to be located near other restaurants. Only a few restaurants reside in areas with low restaurant density.



As mentioned previous, Vietnamese, Chinese and Thai restaurants are popular in Ottawa.



### 3.3 Data preparation and preprocessing

After inspecting data gathered from Foursquare, a new category (online presence) was created from whether or not the venue had a URL, Twitter, Instagram, or Facebook name listed. Any listed value will be changed to the value of 1 and 0 represents no listing. The computed ‘online presence’ will range from 0 to 4. The rating for each venue is between 0 and 10. A number of venues had no rating, the value ‘-1’ was assigned to represent ‘no rating’.

### 3.4 Modeling

To gain insight on current competitors, K-means was used to segment the Asian restaurant data.

Experimentation was done with clusters set to 3, 4 and 5. It was determined that 5 clusters produced a model with enough distinct data.

## 4. Results

#### 4.1 Desirable neighbourhoods

Two desirable neighbourhoods stood out: Stonebridge - Halfmoon Bay - Heart's Desire (A)[[4]](#footnote-4) and Kanata Lakes - Marchwood Lakeside - Morgan's Grant- Kanata North Business Park (B).[[5]](#footnote-5) Both locations are highly populated area with a high Asian ratio.

#### 4.2 Competitor Segmentation Insights



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **Colour** | **Rating** | **Price** | **Likes** | **Online Presences** |
| Cluster 1 | Purple | 8+ | 1.5 | High | High |
| Cluster 2 | Blue | 8+ | 2.5 | Moderate | High |
| Cluster 3 | Mint | 7+ | 1.8 | Low/Moderate | Low |
| Cluster 4 | Red | 7+ | 2 | Low | Low |
| Cluster 5 | Orange | None | 1.5 | Low | None |

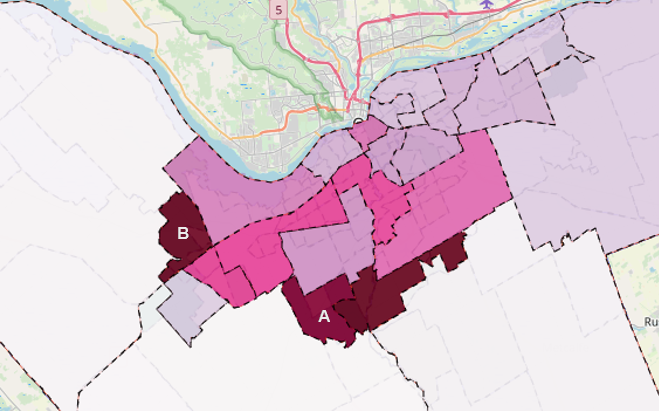
There are many Asian restaurants with low online presence with an average rating of 7+ (cluster 3-red and cluster 4-mint).

Only 2 restaurants possess high ratings, are well liked and have affordable menu items (cluster 1- purple).

Menu prices don’t seem to be correlated with ratings.

## 5. Discussion

#### 5.1 Neighbourhood of choice



Stonebridge(A)[[6]](#footnote-6) was determined to be the best location for Ann to open her restaurant and with low Asian restaurant competition. The second neighbourhood to consider is Kanata Lakes, however it has fierce competition in the nearby neighbourhoods.

In the neighbourhood of choice (Stonebridge), there is very low competition. The list of competitors are as follows:

|  |  |  |
| --- | --- | --- |
| **Competitors** | **Category** | **Rate** |
| China Star | Chinese Restaurant | na |
| 1000 Sushi Islands | Sushi Restaurant | 5.7 |
| Barrhaven Vietnamese Restaurant | Vietnamese Restaurant | 7.6 |

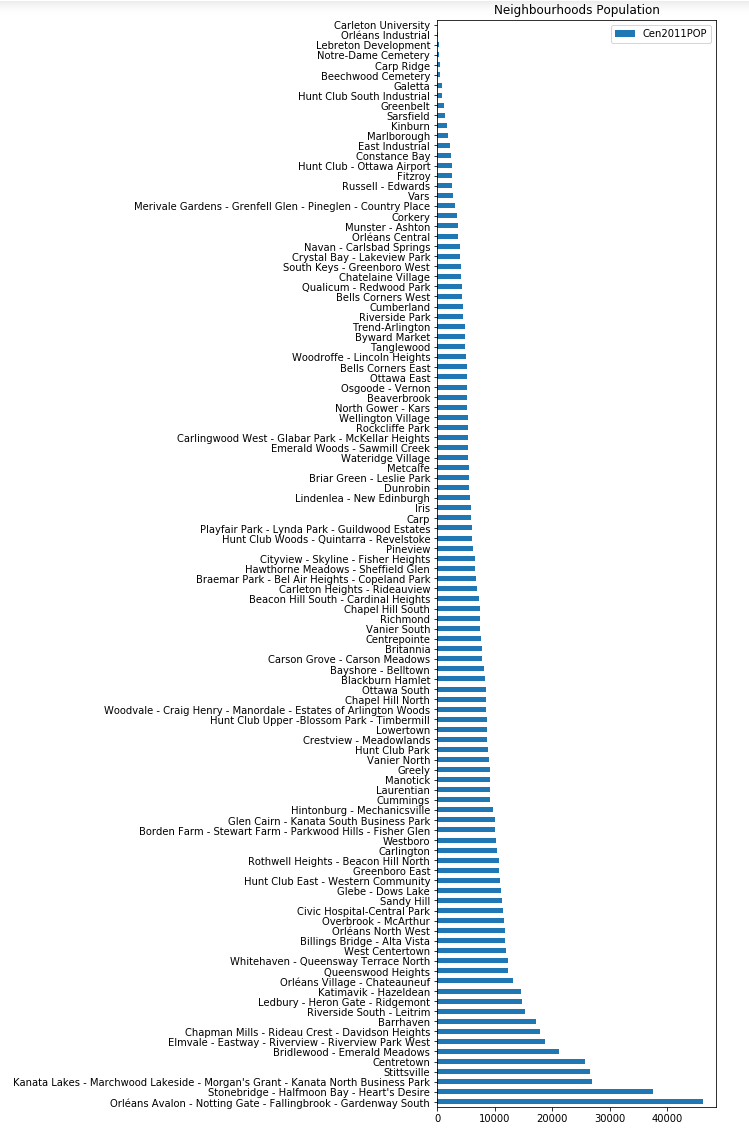
Ann should create a strong online presence by hosting a restaurant website and establishing a social media following.

Suggested restaurant categories: Korean or Japanese due to low competition and having generally higher ratings than others types of restaurants.

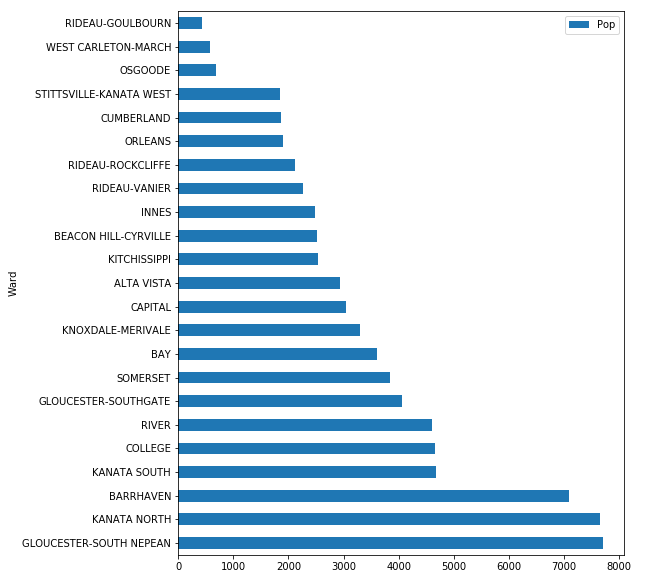
## 6. Conclusion

In this study, general population data and Asian population in Ottawa were analyzed. Two neighbourhoods were identified to have a high population density and a high Asian ratio. Utilizing Foursquare API, food venue data was gather. Venues details were pulled for restaurants that fell under the noodle House, Chinese, Sushi, Vietnamese, Thai, Asian, Japanese, and Korean restaurant category. Segmenting these restaurants using K-means model gave insight on competitors. Ann, the client could use the competitor insights to help her business succeed.

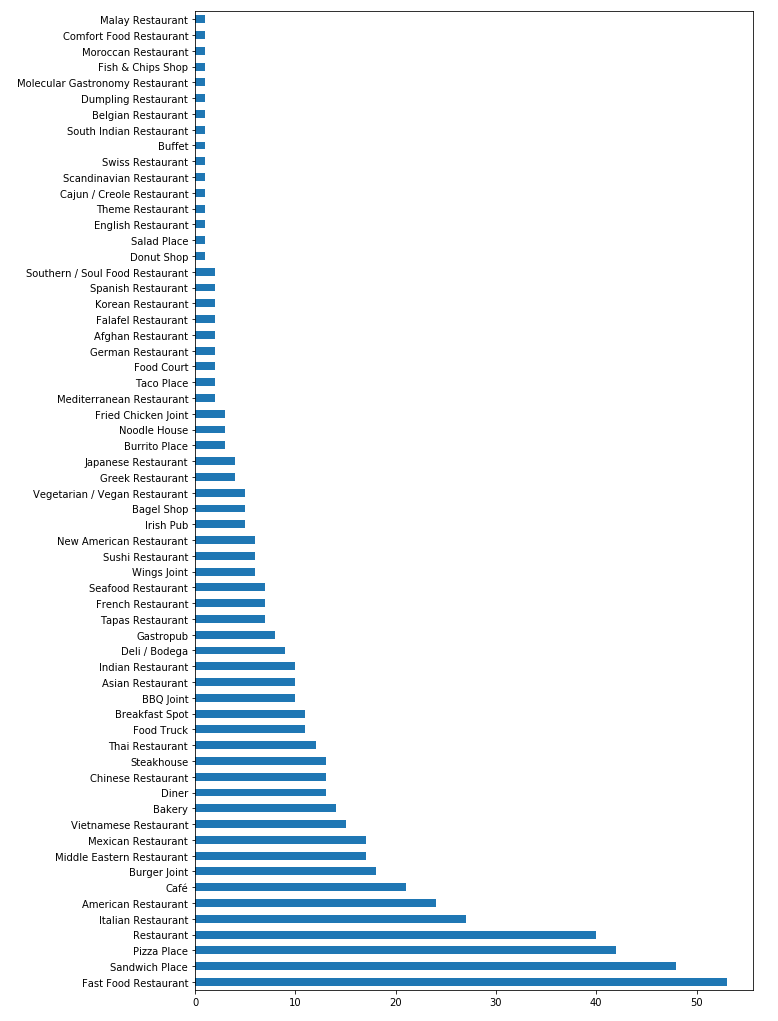
# Appendix 1 – Neighbourhood Population



# Appendix 2 – Asian population by Ward



# Appendix 3 – Restaurant Categories



1. For the scope of the capstone course, only restaurants found in the Foursquare DB will be used. [↑](#footnote-ref-1)
2. Refer to Appendix 1 for Neighbourhood population graph. [↑](#footnote-ref-2)
3. Refer to Appendix 2 for Asian population by Ward graph. [↑](#footnote-ref-3)
4. Abbreviations: Stonebridge = Stonebridge - Halfmoon Bay - Heart's Desire [↑](#footnote-ref-4)
5. Abbreviations: Kanata Lakes = Kanata Lakes - Marchwood Lakeside - Morgan's Grant- Kanata North Business Park [↑](#footnote-ref-5)
6. Abbreviations: Stonebridge = Stonebridge - Halfmoon Bay - Heart's Desire [↑](#footnote-ref-6)