**Optimal Approach to Open a Restaurant in Toronto**

1. **Introduction/Business Problem**

A client is looking into starting a restaurant focused on Japanese cuisine. The location would be in Toronto, Canada, but the client would need to know which neighborhood in the city would offer the most benefits to the business operation, and of course, to maximize revenue.

Our objective in this project is to gather information on the different neighborhoods in Toronto, and advise our client with the most strategic location based on demographics of residents, area population, and average income. We will be using the Foursquare API for exploring nearby venues and retrieving demographic data from Toronto’s Open Data Portal.

1. **Data**

### a. Data Sources

* The neighborhoods with their respective postal codes and boroughs were scraped from Wikipedia.
* Geographical coordinates for each neighborhood were extracted from the file provided to us in last week's assignment.
* Toronto’s demographic data, such as median household income, total population, and population of different ethnicity for each neighbourhood is retrieved from Toronto’s Open Data Portal.
* We will also be using the Foursquare API to explore the list of different venues nearby in each neighbourhood.

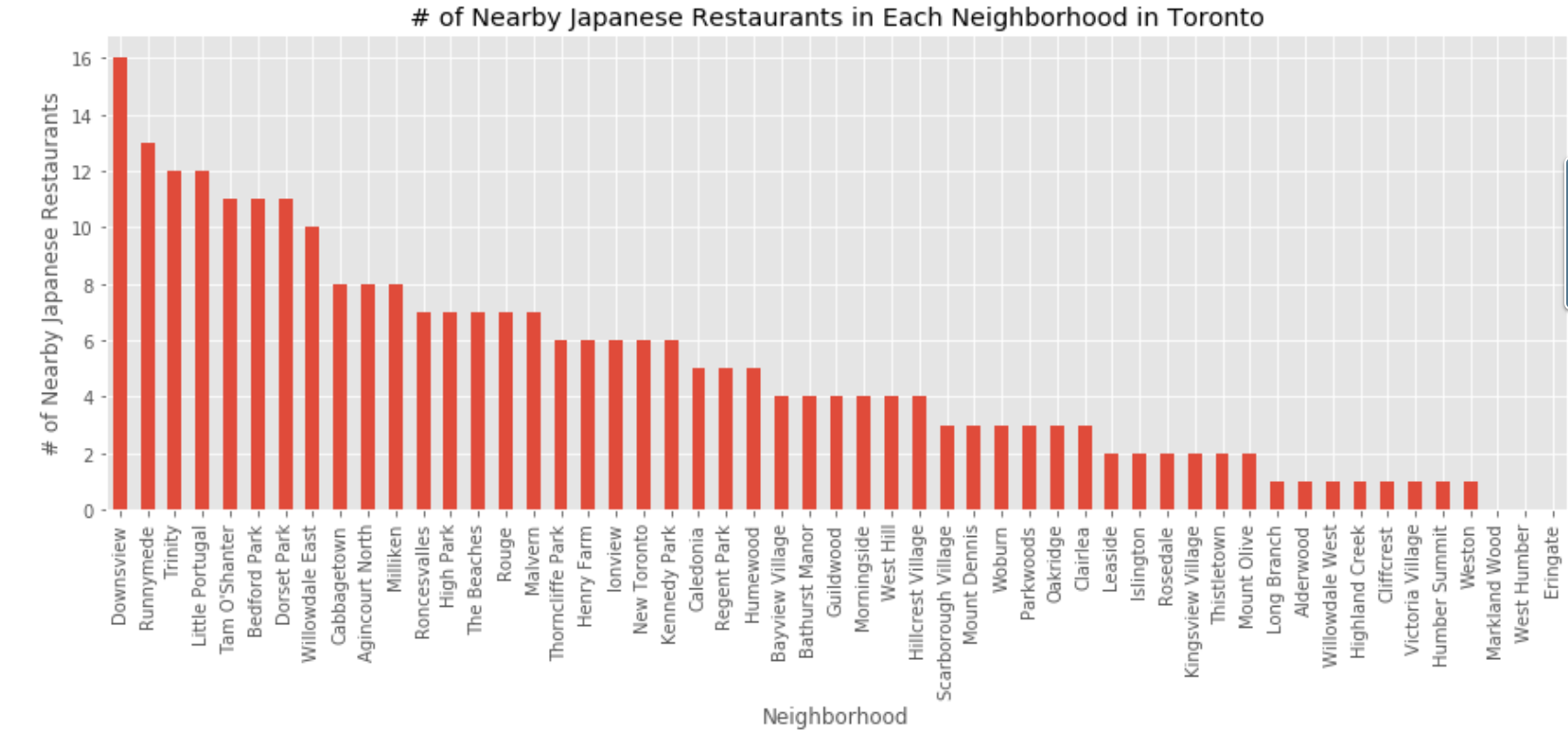
### b. Data Cleaning

* From the Wikipedia data, we will remove neighborhoods where their boroughs or postal code were labeled as not assigned
* We will remove any neighborhoods that do not have any demographics data from Toronto's Open Data Portal

1. **Methodology**

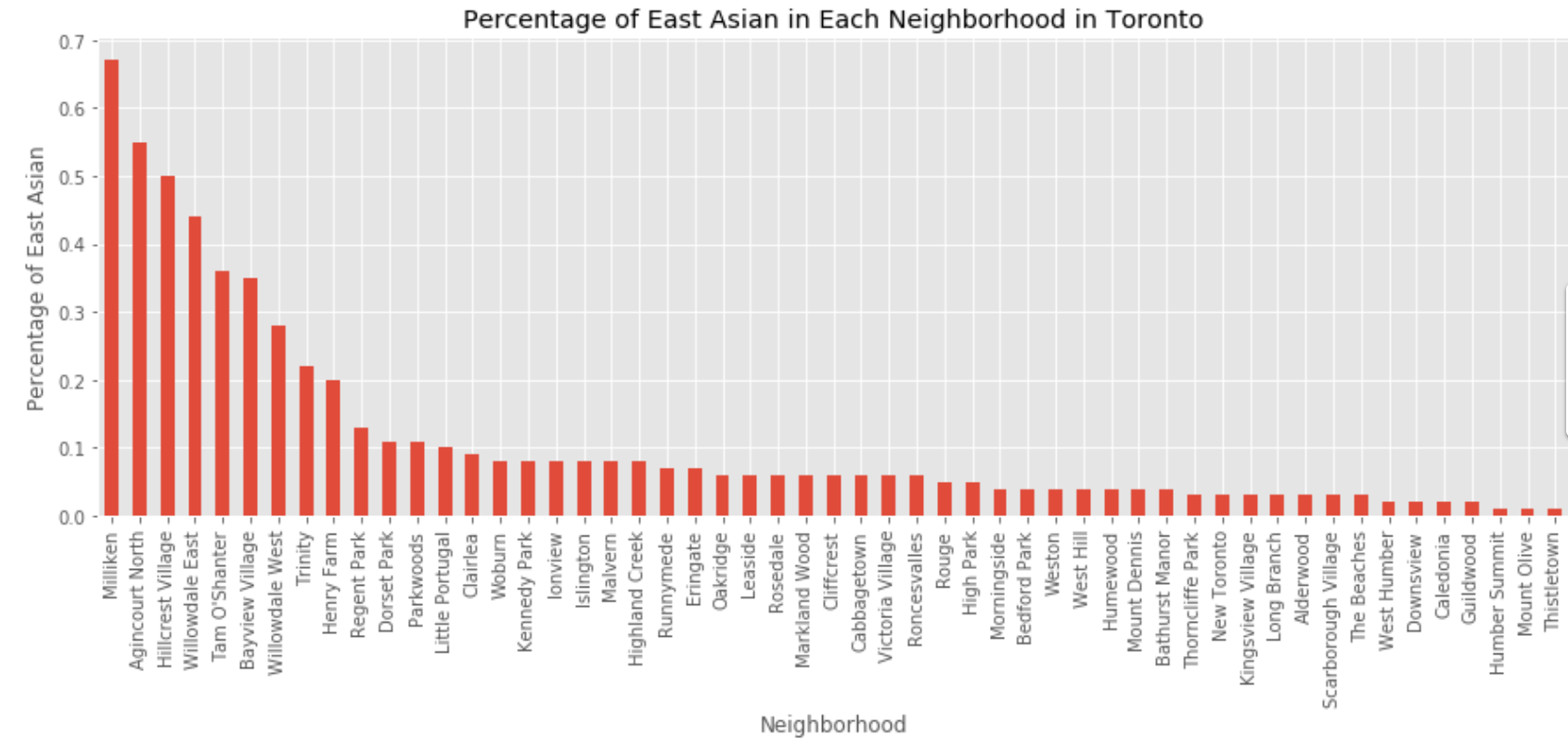
### a. Why analyze number of Japanese restaurants in each neighborhood

We used the Foursquare API, mainly the Venue Explore endpoint, to retrieve nearby Japanese restaurants within 1000 meters of each neighborhood. The assumption is that if there is a high number of similar restaurants nearby, this can generate a lot of competition for our client's new business, which can go in a negative way if other restaurants already have many loyal customers.



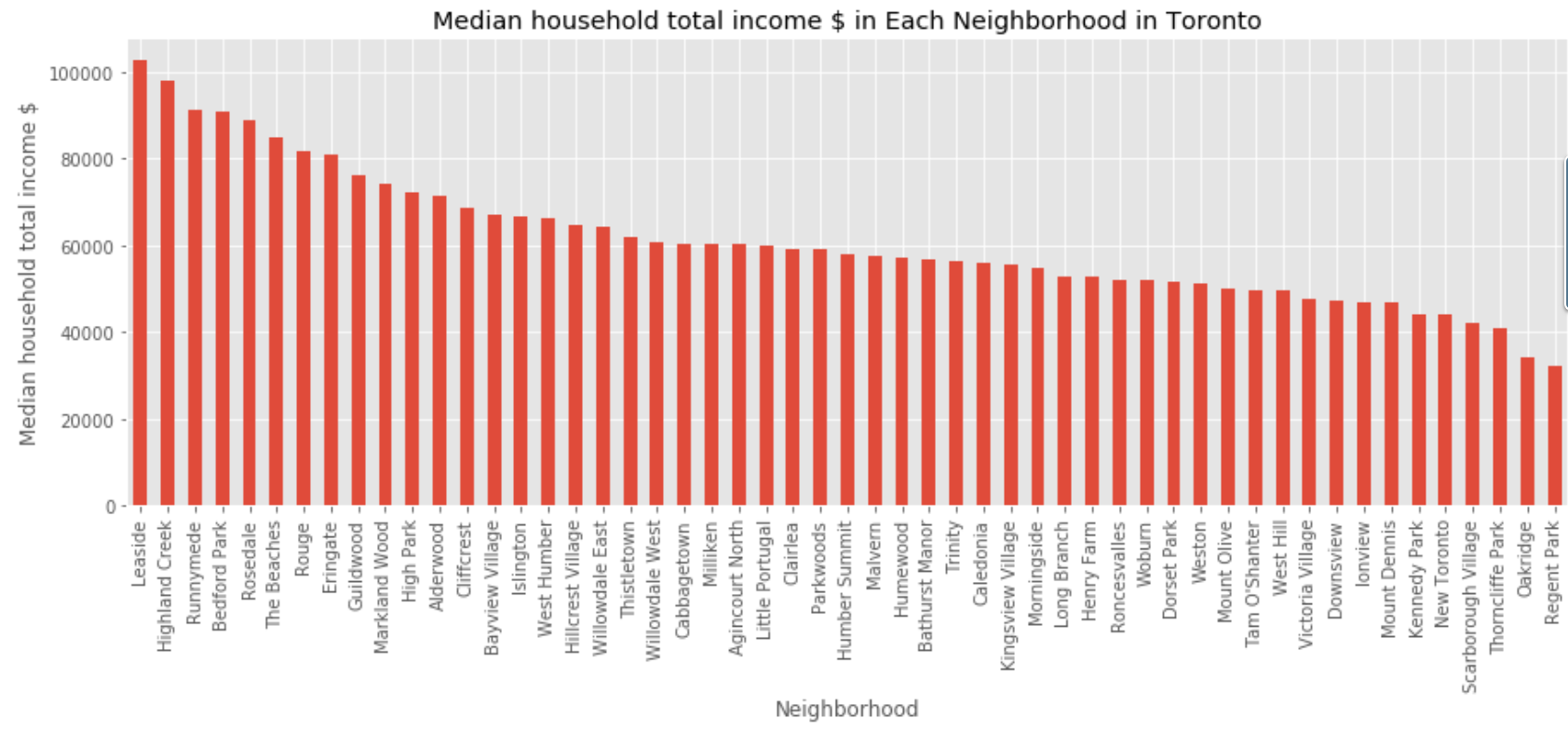
### b. Why analyze the ratio of East Asian population to total population in each neighborhood

We assume there is a positive correlation between the percentage of East Asian population and the demand for Japanese cuisine; therefore, it would benefit our client's business growth if the restaurant were to start up at neighborhood with a relatively densely populated location among East Asians.



### c. Why analyze the median household total income in each neighborhood

Japanese cuisine prices usually tend to be more suitable for middle class income earners and up; therefore, analyzing the median household total income can let us know that if these relative prices could potentially be scaled too expensive for some neighborhoods.



### d. Predictive Modeling : KMeans Clustering

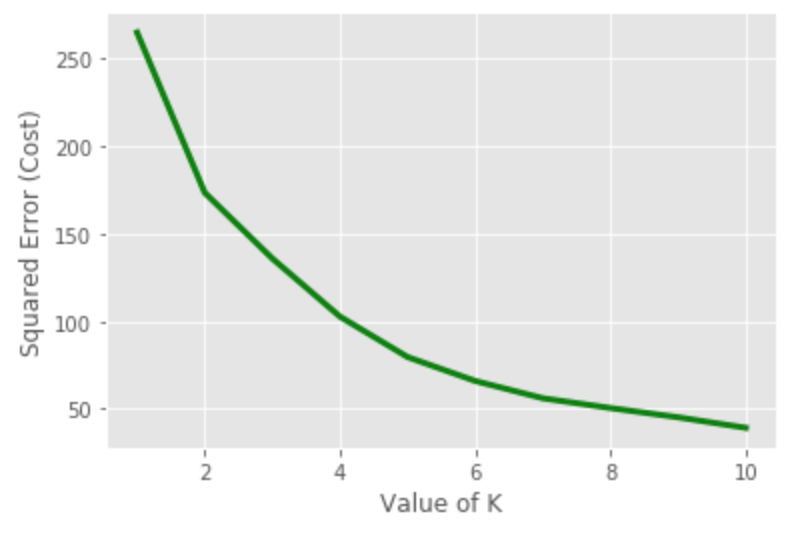
#### Data Preprocessing

In order to segment the Toronto neighborhoods into groups based on their numerical distribution and magnitudes, a good approach for this unsupervised learning task is using KMeans Clustering.

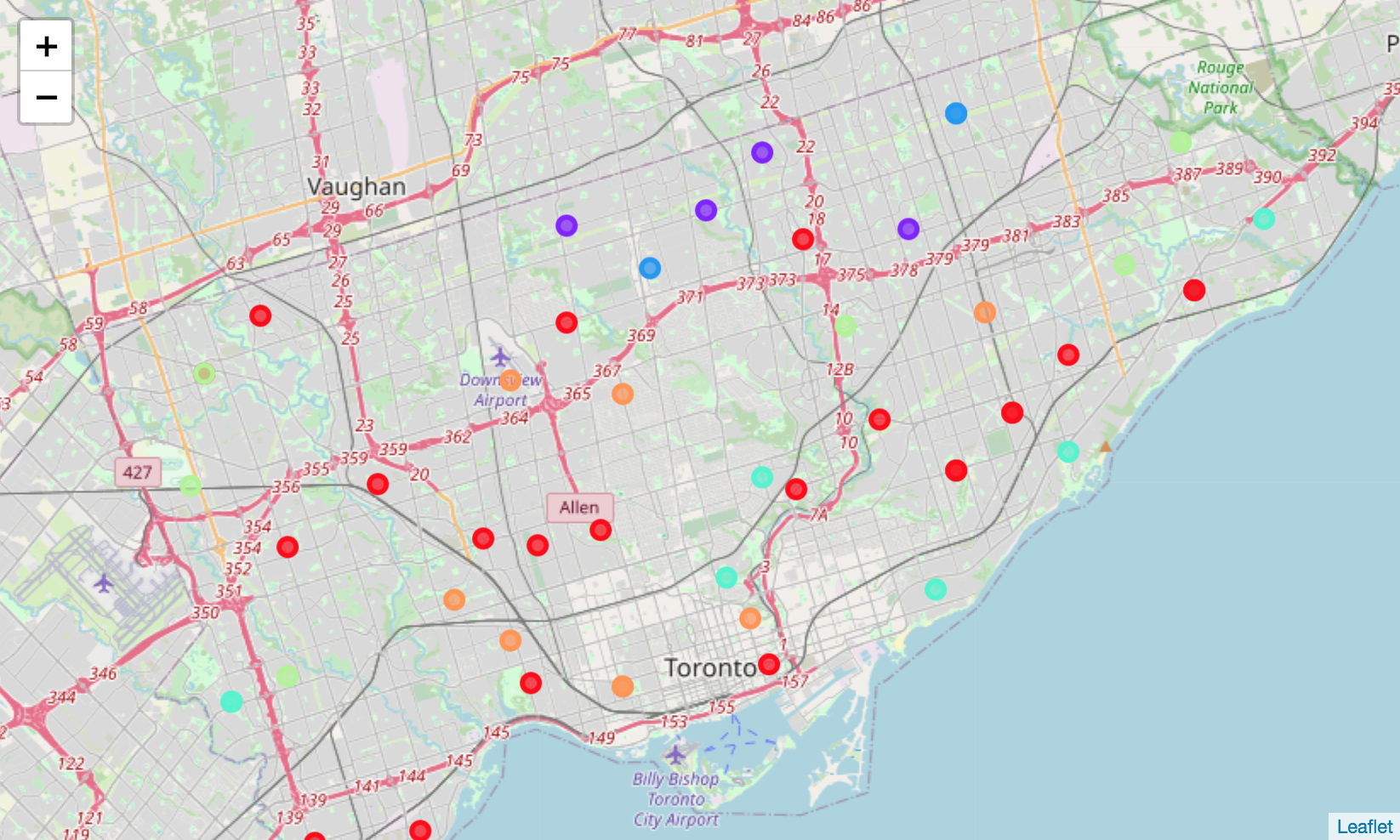
We must first normalize our quantitative features to a common scale, so that our model can interpret the values equally. For this occasion, we use the StandardScaler to normalize our data.

#### Building a KMeans Clustering Model

One of the important aspects of instantiating a KMeans Clustering model is to specify the most optimal number of clusters for the algorithm to label. Using the Elbow method, we calculated the squared error for a range of 1-10 for our K number of clusters. It was determined that K=6 would deliver the best results for our predictive algorithm.



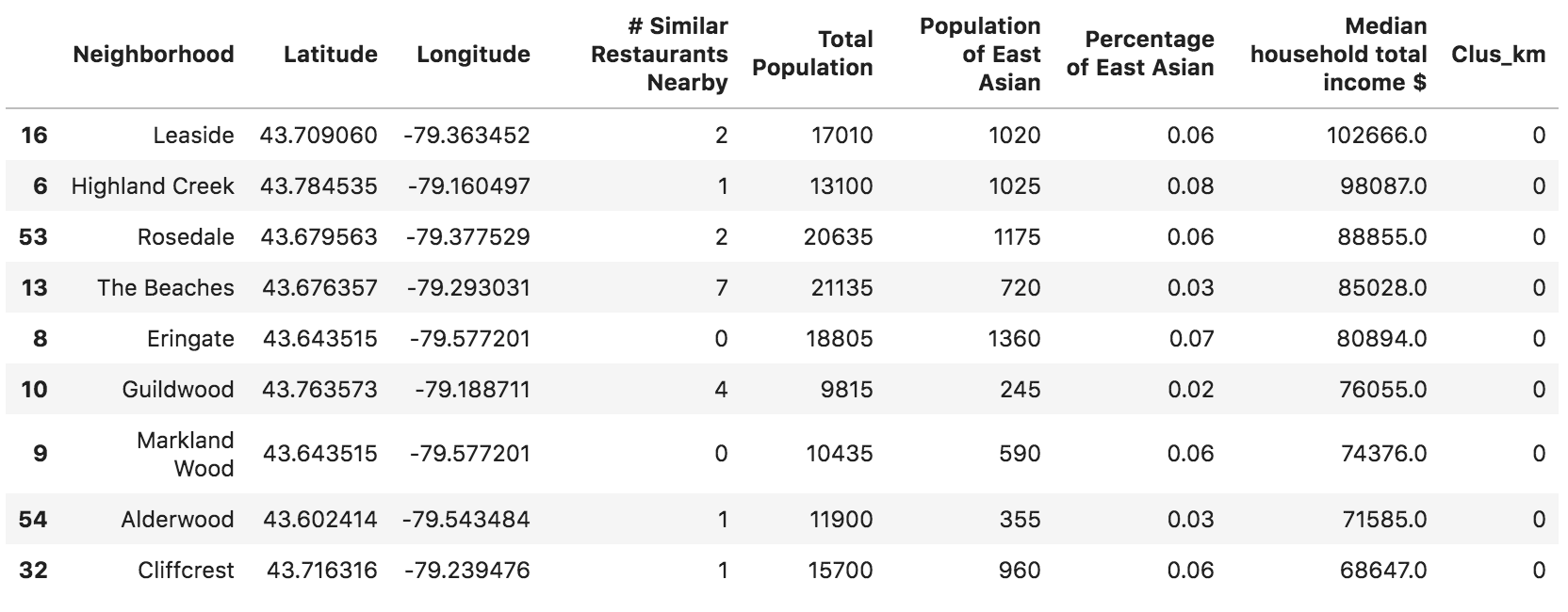
Next we go ahead and fit our standardized features into the KMeans Cluster model to segment each neighborhood into a cluster.



**4. Results**

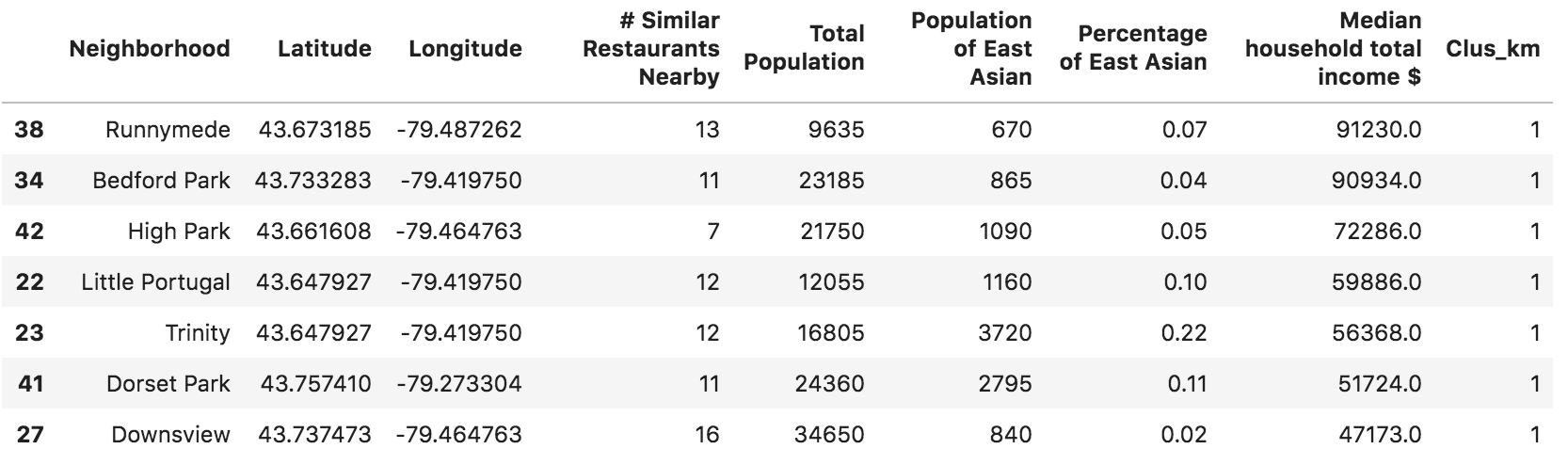
**Cluster 0** segments the neighborhoods with the highest level of spending power among the clusters and low competition in Japanese Cuisine. However, it also shows that there is a low number in East Asian population, leaving us with an assumption that Japanese cuisine is most likely not in high demand, which could explain why competition was low in the area.

* Median Household Income: High
* Percentage of East Asian: Low
* Similar Restaurants Nearby: Low



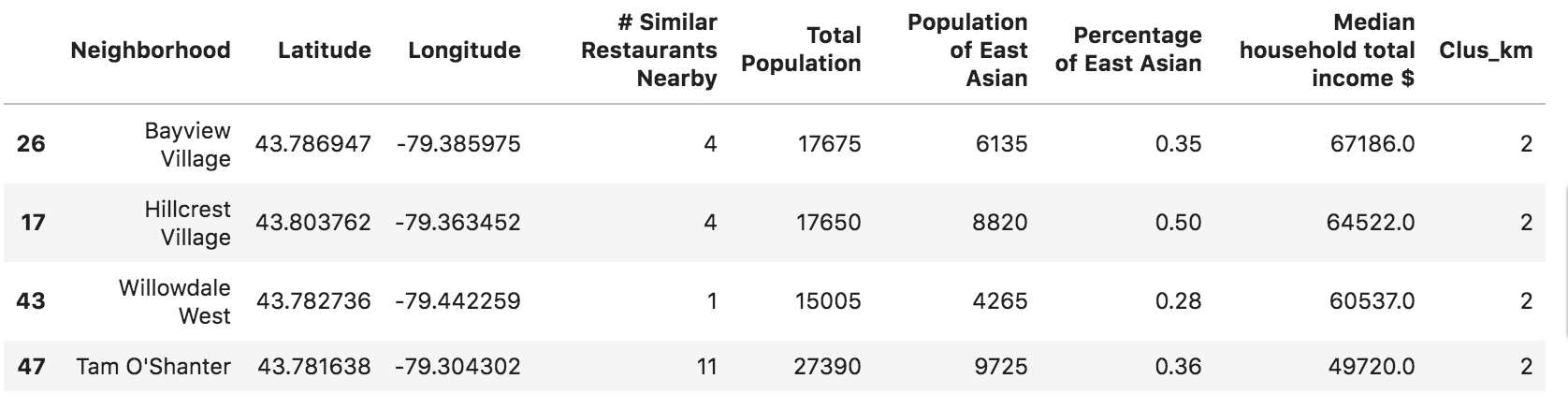
**Custer 1** shows the group with a medium level spending power. Although, this would be adequate for our business, but both the high number of competitions nearby and a low population of our target customer brings negative value to the growth of our business.

* Median Household Income: Medium
* Percentage of East Asian: Low
* Similar Restaurants Nearby: High



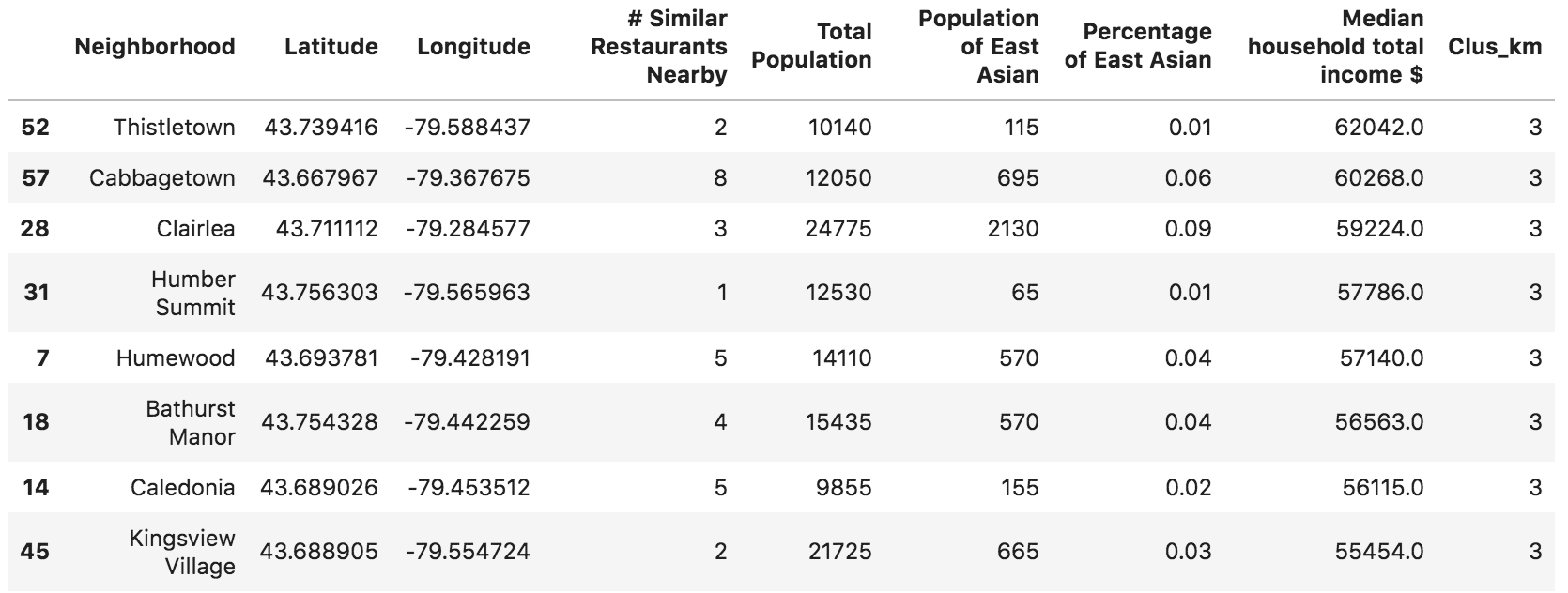
**Cluster 2** shows a medium level of spending power and low competition. The locations are fairly populated with East Asians, which does show potential value; however this segment is expressing a greater population of our target customers can further increase the demand for Japanese Cuisine in the area.

* Median Household Income: Medium
* Percentage of East Asian: Medium
* Similar Restaurants Nearby: Low



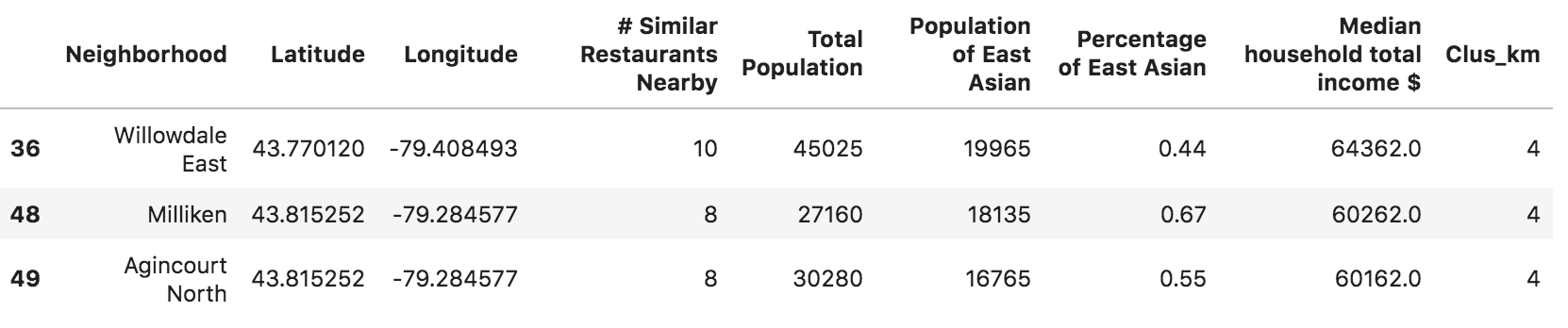
**Cluster 3** describes the group of neighborhood to have a relatively low spending power in each household compared to neighborhoods in other clusters. Although, there is a low number of competition in the area, there also is a low population of the target customers for our client's business.

* Median Household Income: Low
* Percentage of East Asian: Low
* Similar Restaurants Nearby: Low



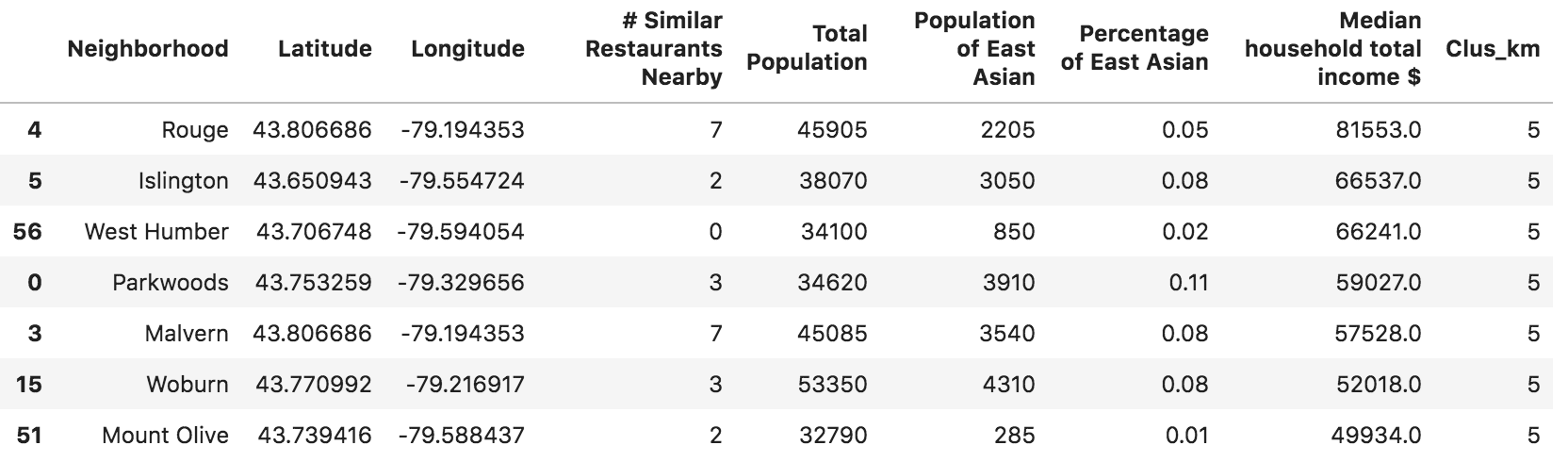
**Cluster 4** shows the neighborhoods with a great number of East Asian residents among total residents, giving our client's business a good chance to attract the local target consumers. The average household income among the group also shows that is adequate for the typical pricing for Japanese cuisine. The fair number of competitions allows our new business to stick out among the others, while also supporting the hypothesis that East Asian cuisine is somewhat popular to locals.

* Median Household Income: Medium
* Percentage of East Asian: High
* Similar Restaurants Nearby: Medium



While **Cluster 5** shows that the medium spending power is suitable for typical Japanese cuisine prices and fair number of competitions nearby, it does not seem there will be sufficient amount of the target customers living nearby.

* Median Household Income: Medium
* Percentage of East Asian: Low
* Similar Restaurants Nearby: Medium



# 5. Discussion

Milliken (in Cluster 4) appears to be the most optimal neighborhood to start a Japanese restaurant in Toronto.

In Cluster 2 and Cluster 4, where both have a medium - high ratio of East Asian population, Japanese restaurants have a good chance to stand out to the public due to competition never reaching too high. This detail provides great support in terms of future expansion and franchising the restaurant into different parts of Toronto. The flexibility allows the business to cope with changes in consumer demands in terms of relocation.

# 6. Conclusion

In this study, I have clustered the neighborhoods in Toronto based on their different characteristics, including average annual income per household, ratio of target customers, and the level of competitions nearby. Based on our algorithm, it appears that Cluster 4, especially in Milliken, would bring the most optimal setting to allow our client's business to operate.

The spending power of the residents in this location shows that the typical pricing of Japanese cuisine should not financially impact the local demand.

The high ratio of target customers nearby is a perfect indicator that our Japanese cuisine could already be in high demand and our business would be an attractive addition to the local residents.

The fair number of competitions allows our new business to stick out among the others, while also supporting the hypothesis that East Asian cuisine is somewhat popular to locals.

# 7. Reference[¶](https://jupyterlab-41.labs.cognitiveclass.ai/user/jx22553/lab?" \l "7.-Reference" \t "_self)

1. My Github for this project:

[https://github.com/jx22553/Coursera\_Capstone](https://github.com/jx22553/Coursera_Capstone" \t "_blank)

1. Wikipedia page on Toronto neighbourhoods for scraping:

[https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M" \t "_blank)

1. Toronto Open Data Portal for demographics data:

[https://open.toronto.ca/dataset/wellbeing-toronto-demographics/](https://open.toronto.ca/dataset/wellbeing-toronto-demographics/" \t "_blank)

1. Geographical coordinates of Toronto neighborhoods:

[https://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data" \t "_blank)

1. Foursquare Developer API:

[https://developer.foursquare.com](https://developer.foursquare.com/" \t "_blank)