Neighbourhood Focused Grocery Store Merchandising Mix for Toronto, Ontario.

1.0　Problem Statement

In this project, we analyze these neighbourhoods and try to discover how we can use the concentration of restaurants and demographic information from the 2016 census to build a better focused merchandising mix for different neighbourhoods.

While the census data shows ethnicity, restaurant information gives a more dynamic view of the influences and tastes that may shape their actual grocery shopping behaviour

- Grocery retailers face increasing competition from other brick-and-mortar stores and also from on-line retailers.

- Retailers in highly cosmopolitanism urban centres like Toronto, Ontario face additional challenges trying to maximize revenue from an increasingly diverse demographic.

- Unlike on-line retailers, grocery stores are limited in the variety of merchandise they can display. Additionally, high real-estate prices limit the size of new stores.

- According to the 2016 census bout half of Toronto residents were born outside of Canada. With people from similar ethnicities living together in similar neighbourhoods.

1.1 Target Audience

Who would benefit from this project?

- Existing national grocery chains that want to change their merchandising mix to better target customers in different neighbourhoods.

- New grocery chains who are looking to design a merchandising strategy that reflects the differing neighbourhood demographics in Toronto, Ontario.

- Entrepreneurs looking at establishing ethnic focused grocery stores

- Marketers who work for the grocery industry and will like to better target their direct marketing and other location based advertising to better reflect local requirements.

2.0 Data

- Toronto neighborhoods by webscraping data from Wikipedia here.

- Geospatial data obtained from a dataset (Geospatial_data.csv).

- Demographics of Toronto Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.. This shows the ethnic origin groups with

- Dataset from Geocodio.io (geo_data_geocodio_fcb7866a8350d4ec45a3dd27f1805bdbcebff253.csv) which contains Postcode and geospatial data for each riding.

- Using the the food category in the **Foursquare API** to obtain food establishments in each neighbourhood.

- We then merge our Restaurant Dataframe with our Demographics dataframe to obtain a single Dataframe.

- To get a better idea of the nature of each neighbourhood, we also prepared another DataFrame showing the top five most common Ethnicity as well as the top most common restaurant type in each neighborhood to give a more intuitive understanding of the characteristics of each neighbourhood.
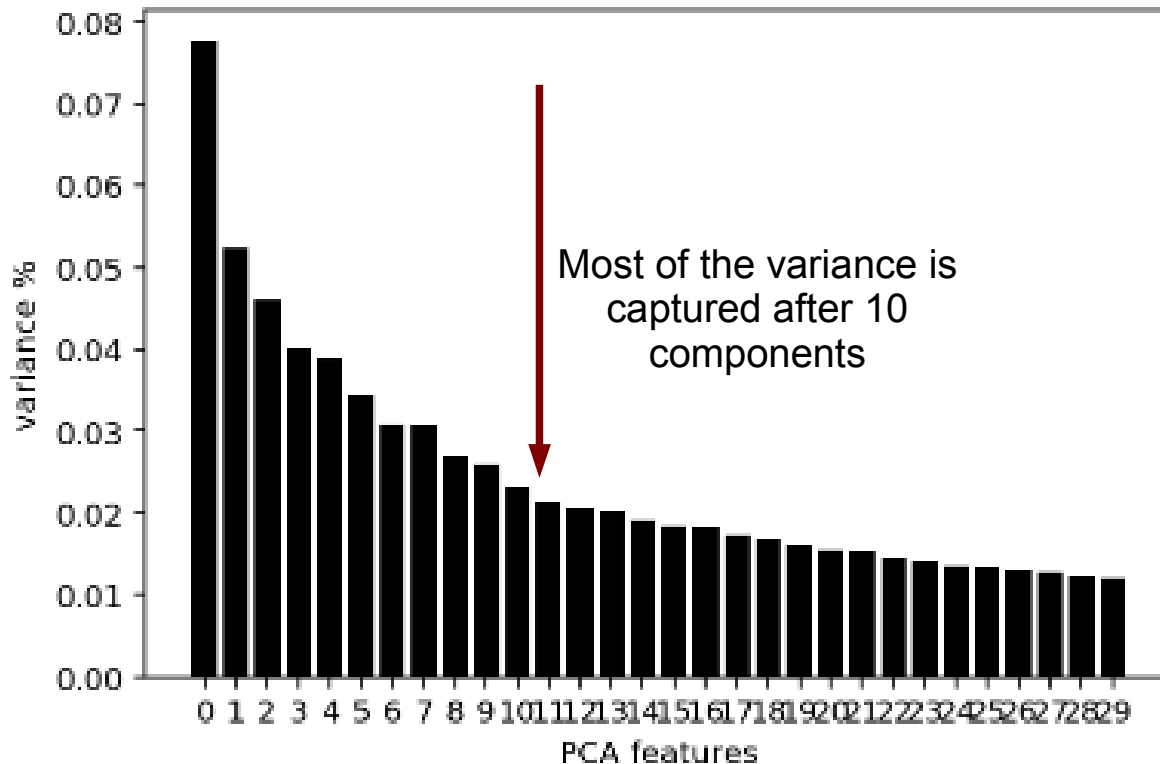
3.1 Methodology - correlation

We have 112 features. We do a correlation heatmap to examine any correlations in the features.
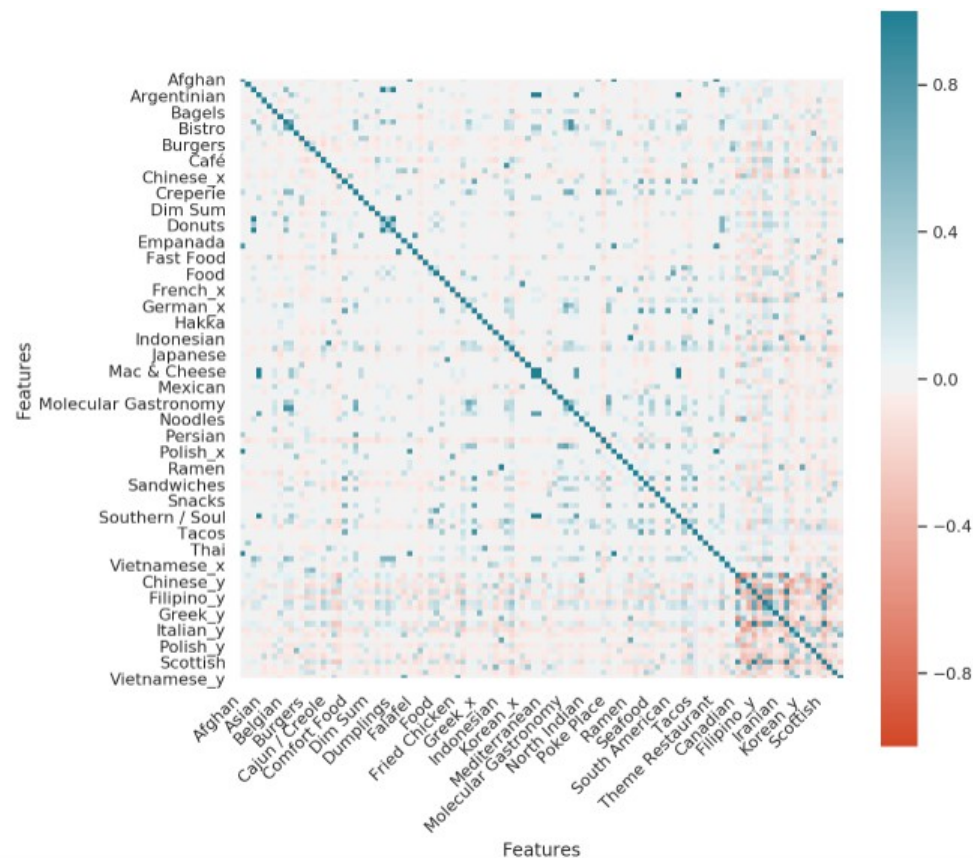
3.2 Methodology - PCA

Given the large number of features, we use Principal Component Analysis (PCA) to limit the number of features.


To obtain an optimal number of PCA component, we plot variance against PCA features

We can see that we are able to capture most of the variance in the data after 10 components. So we will use a PCA with 10 components

3.3 Methodology - K-Means

Using the PCA with 10 components, we try to find the optimal number of clusters for our k-means clustering. We do this by plotting the number of clusters against the inertia.
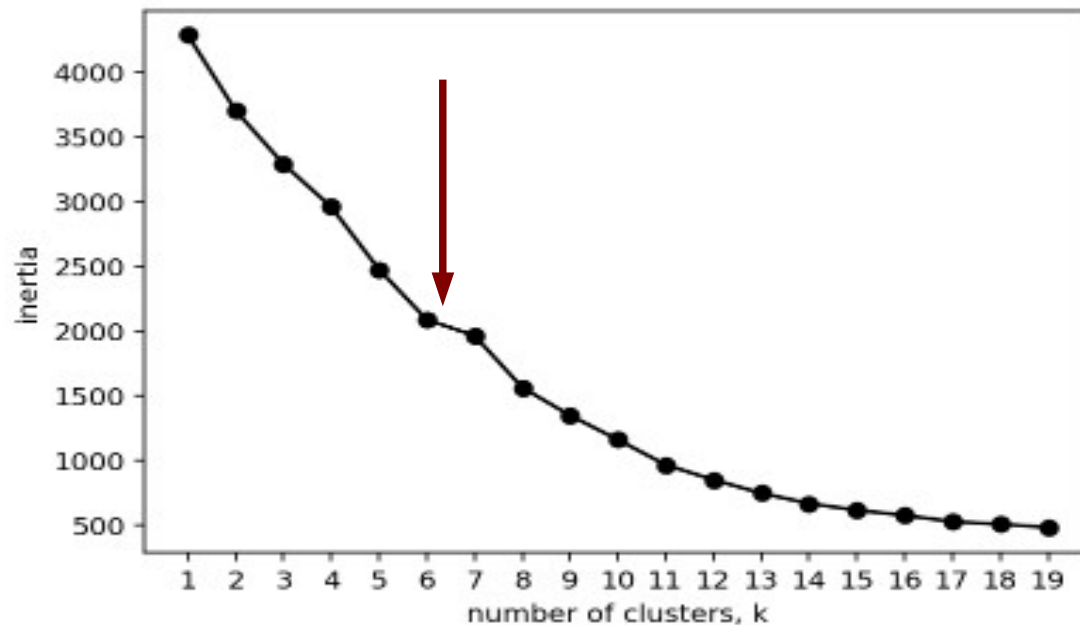
We can see that after 6 clusters the inertia more or less drops off. We therefore run k-means to cluster the neighborhood into 6 clusters.

4.0 Results

We are able to plot our clusters in a map as shown below.
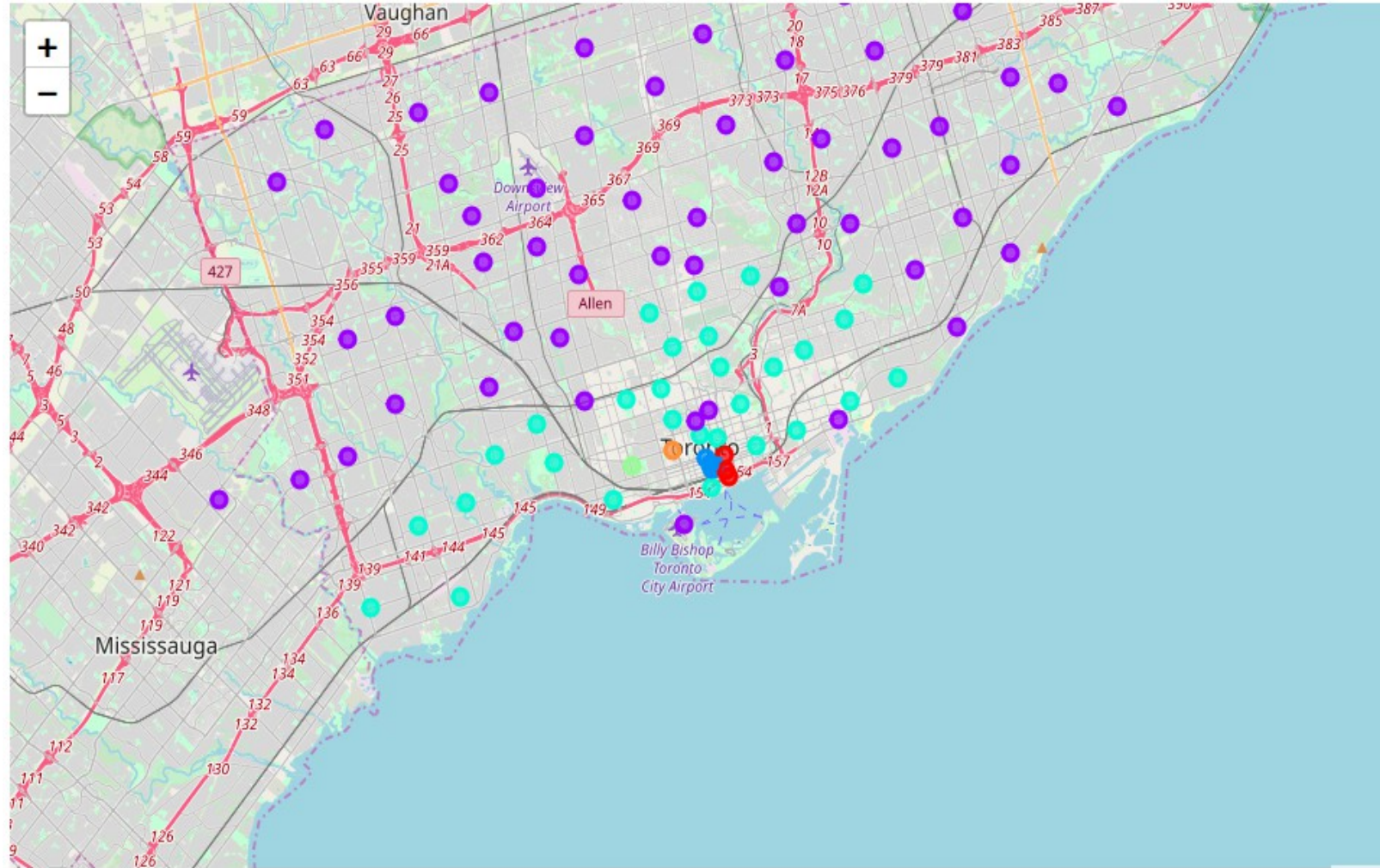
From the map, we see that there are 2 large clusters and 4 smaller ones, and the diversity of clusters seems to be greater within the downtown core which also corresponds to where real estate values are highest and where the utility of the clustering is highest.

5.0 Discussion

Looking at the clusters, we find that of the 2 largest clusters, one is predominantly English and Canadian while the other one is more diverse. This means that in planning a merchandising strategy, the English/Canadian neighbourhood cluster will have a less ethnically diverse merchandising than the other ones.

While the store segmentation can be further customized by utilizing in-store experimentation and data collection to find the particular mixes suitable for each neighbourhoods, this approach is a useful initial approach to segmenting the different neighbourhoods.

6.0 Conclusion

Cities around the world are getting more diverse and retailers face significant obstacles to serving this increasingly diverse population in an optimal manner.

This exercise attempts to illustrate how retailers may attempt to segment possible store locations based on demographics and the type of restaurants (as an indicator of local color).

The results are instructive and represents a good initial approach to customizing merchandising to fit different locations. This method may be supplemented by other techniques including in-store data gathering to further customize retail offerings.