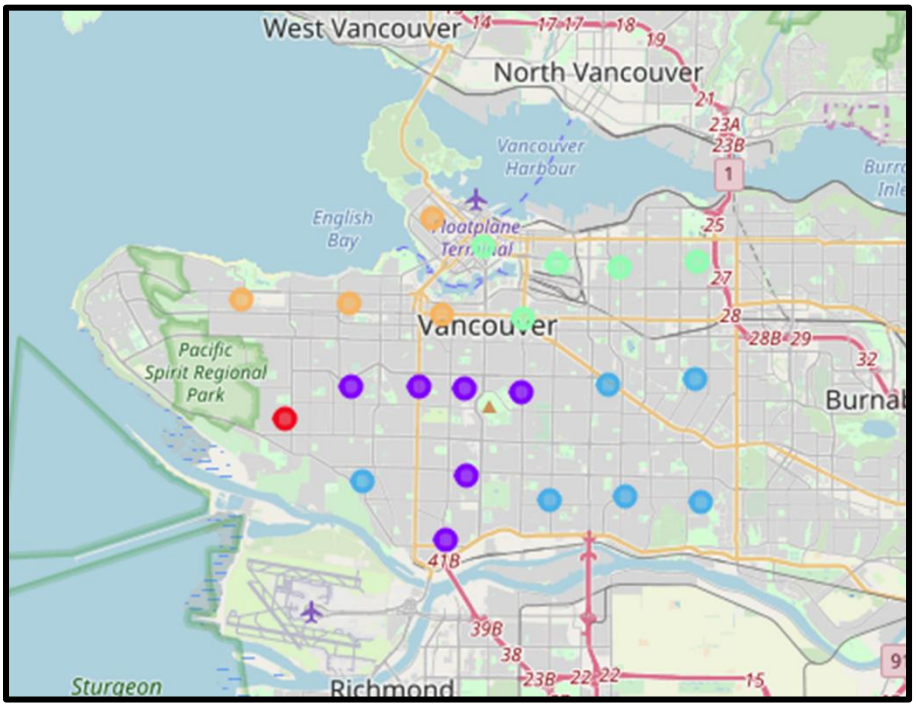

Bus Tours inc.

INTERNATIONAL BUS TOUR COMPANY EXPANSION PROPOSAL

Lucas Muller | IBM Data Science Professional Certificate – Coursera | 12 January 2019

Introduction

An international bus tour company wants a way to expand to new cities. They are offering bus tours to popular trending locations of a given city over 5 days. In order to have a starting point they want a model that will take the trending locations of each neighborhood and cluster them into similar clusters based on k-means. The bus tour consists of 5 days so they will need 5 clusters of venues, one for each day of the tour.

Data

To solve this problem, all that is needed is a list of GPS center points for each neighborhood of a given city. From the Vancouver website a KML file was downloaded, see figure 1 and 2.

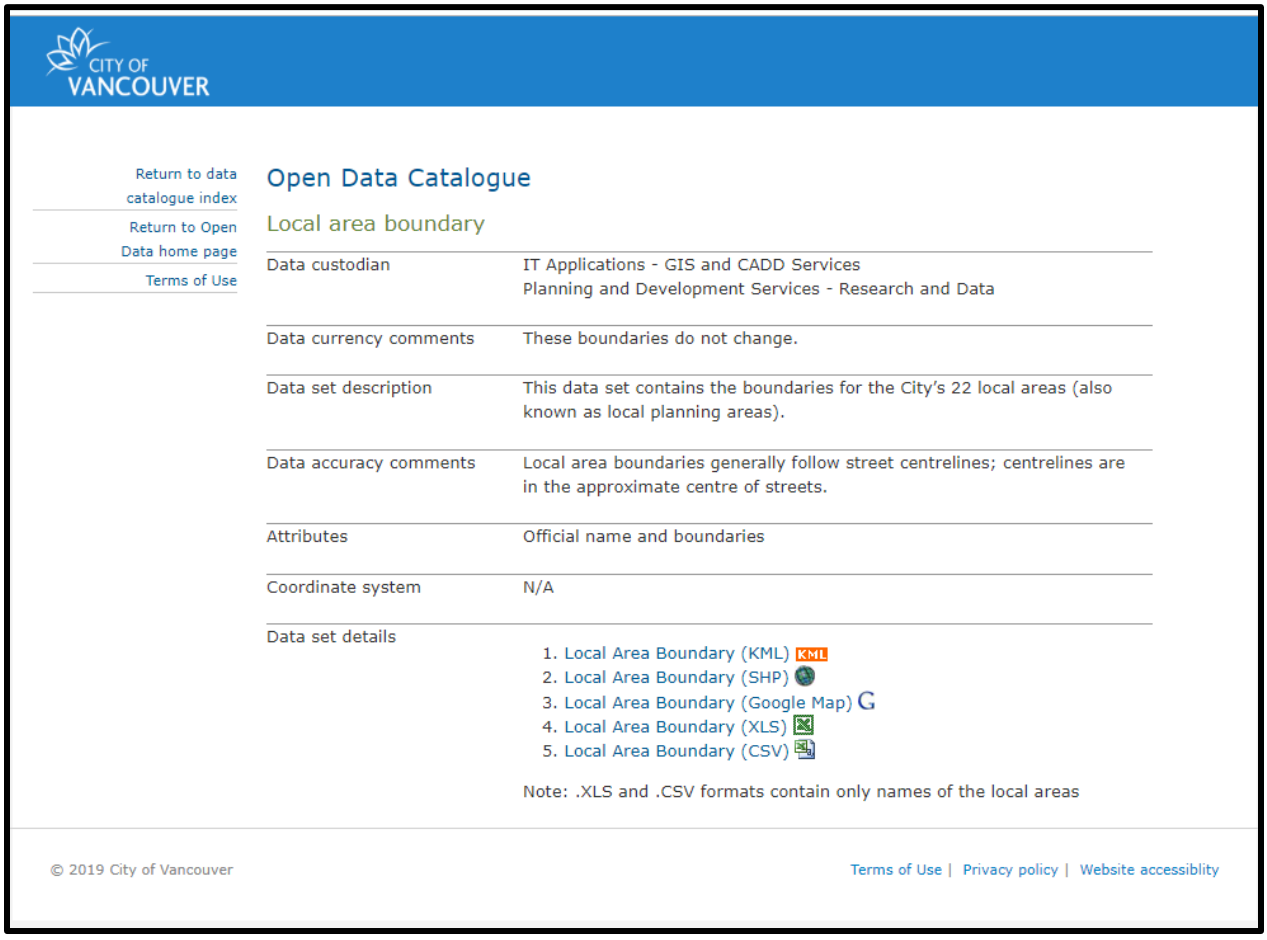

Figure 1: webpage where neighborhoods of Vancouver where acquired. The KML file was used as the CSV only contained the neighborhood names with no GPS data

Link: https://data.vancouver.ca/datacatalogue/localareaboundary.htm

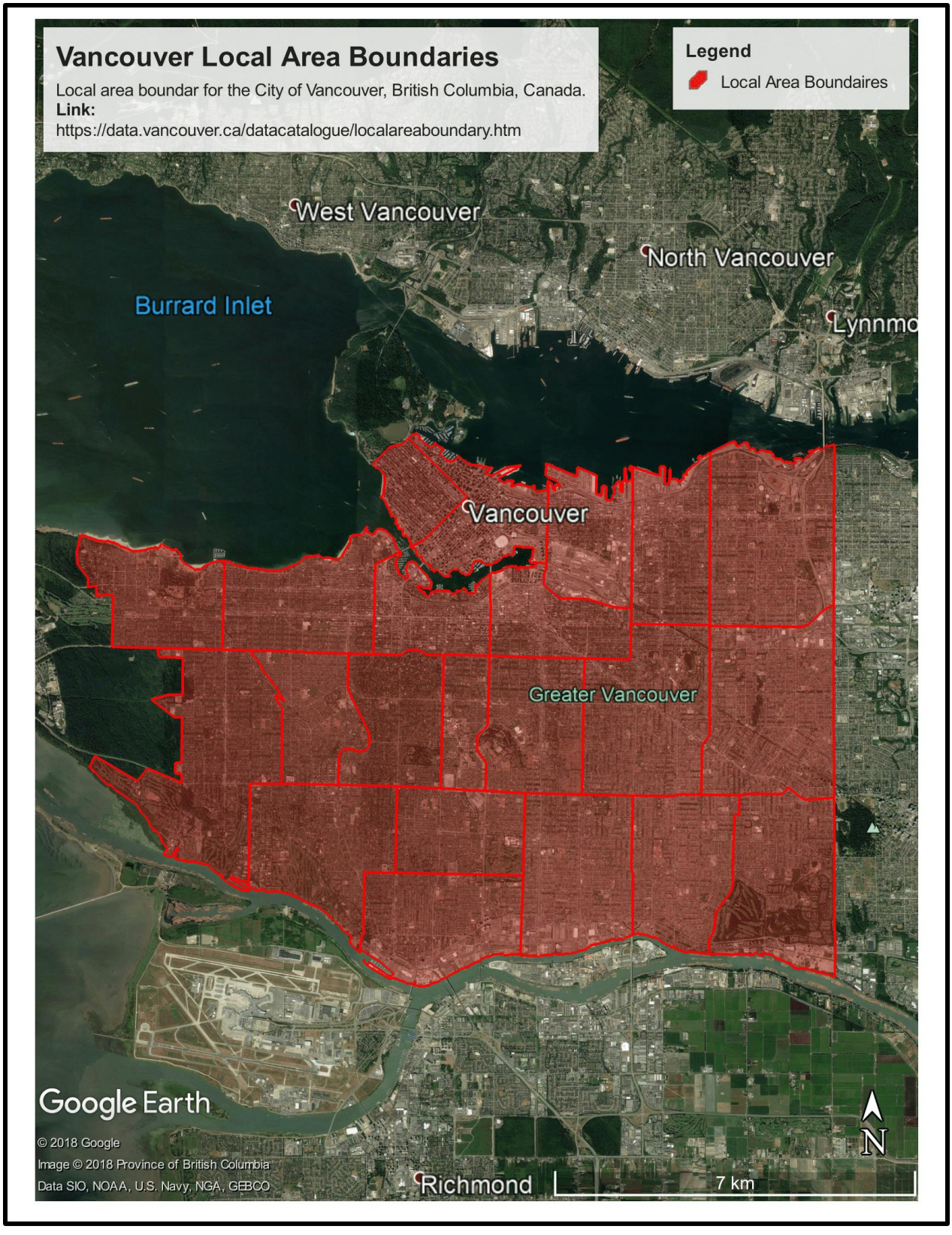

Figure 2: KML file of Vancouver neighborhoods.

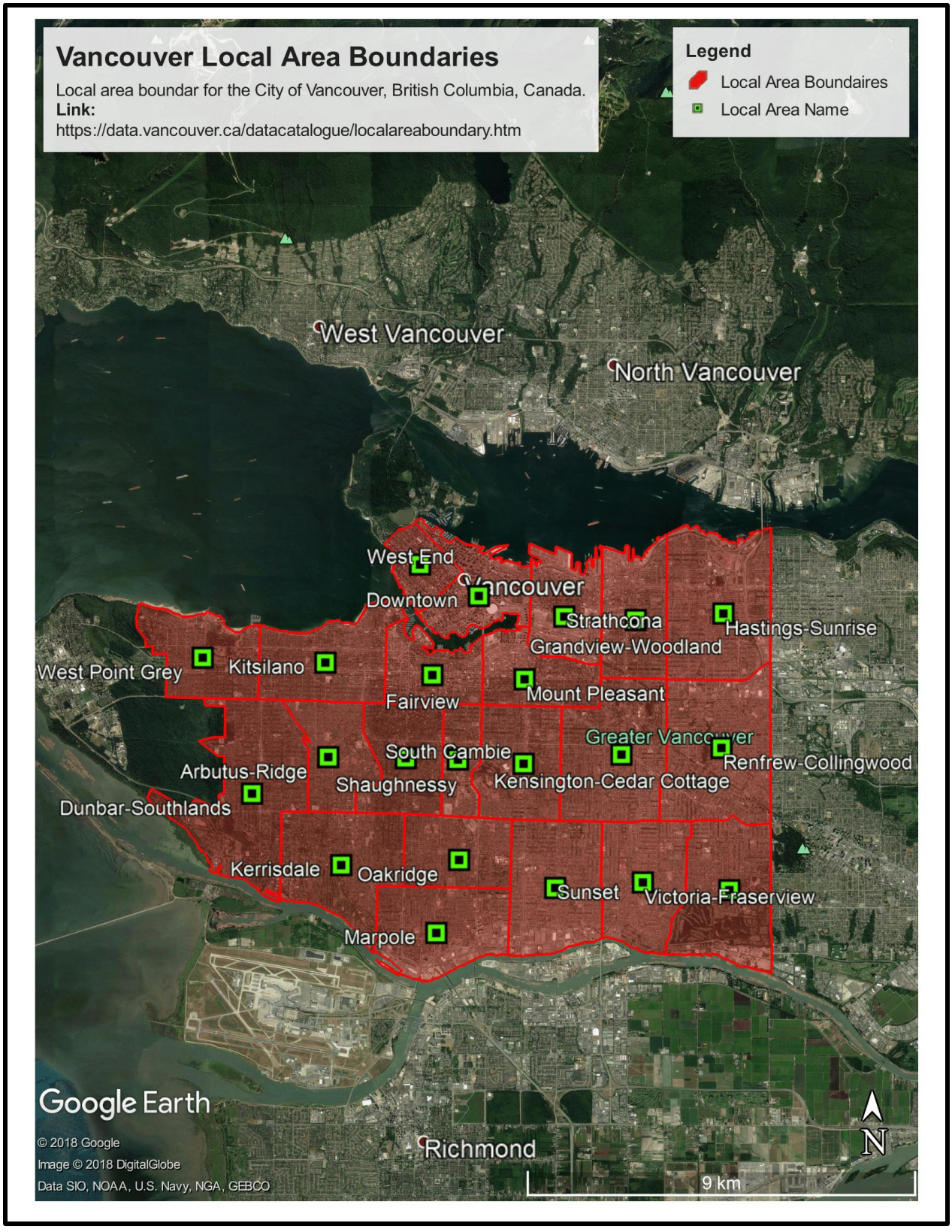

Figure 3: Neighborhood polygons for the city of Vancouver and center points.

From figure 3 a KML of file of the center points was saved and exported to http://www.gpsvisualizer.com/. Then the convert KML to a CSV function was used. Please see final product on git hub https://github.com/lmuller92/Van_hoods. Which was then imported into a juniper notebook and the following table was displayed (see table 1).

Table 1: Neighborhoods and corresponding GPS coordinates. Table head

	latitude	longitude	name
0	49.246316	-123.163438	Arbutus-Ridge
1	49.279594	-123.115711	Downtown
2	49.238770	-123.187580	Dunbar-Southlands
3	49.263254	-123.130439	Fairview
4	49.274615	-123.065973	Grandview-Woodland

The coordinates for the neighborhoods (Table: 1) can then be run through the four-square API call to acquire nearby trending venues and group them in to 5 clusters using the K-means Clustering model. This will give the tour company a starting point to set up a bus tour in any city.

Methodology

Once the neighborhood GPS data has been acquired for any given city the foursquare API call can be used to acquire the 10 most common "Trending" venues around each neighborhood GPS point. The **explore section = trending** end point was used for this call. The Radius was set to 2000m with a limit of 100 venues to be returned. A loop was created to do the same for each neighborhood.

The returned venues are then grouped using a hot encoding method to display the top 5 venues for each neighborhood and put into a pandas data frame, see table 2.

Table 2: Data frame for the categories of top 5 trending venues for each neighborhood. Table head

name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Arbutus-Ridge	Coffee Shop	Sandwich Place	Grocery Store	Sushi Restaurant	Chinese Restaurant
Downtown	Coffee Shop	Japanese Restaurant	Restaurant	Seafood Restaurant	Sandwich Place
Dunbar-Southlands	Golf Course	Sushi Restaurant	Coffee Shop	Bakery	Park
Fairview	Japanese Restaurant	Restaurant	Coffee Shop	Bakery	Café
Grandview-Woodland	Coffee Shop	Brewery	Pizza Place	Sushi Restaurant	Café

Using the information from table 2 a k means clustering model was used to group the neighborhoods based on trending venues. five clusters where set with a random state = 0, see table 3. The link to the code is uploaded to git hub https://github.com/lmuller92/IBM-Final-Capstone-Project/blob/master/Trending.ipynb

Table 3: Neighborhoods and corresponding GPS coordinates with assigned clusters. Table head

latitude	longitude	name	Cluster Labels
49.246316	-123.163438	Arbutus-Ridge	1
49.279594	-123.115711	Downtown	3
49.238770	-123.187580	Dunbar-Southlands	0
49.263254	-123.130439	Fairview	4
49.274615	-123.065973	Grandview-Woodland	3

Results

From table 3 a map can be generated showing the clusters, see figure 4. From this map we can see the 5 distinct clusters for trending venues in the City of Vancouver.

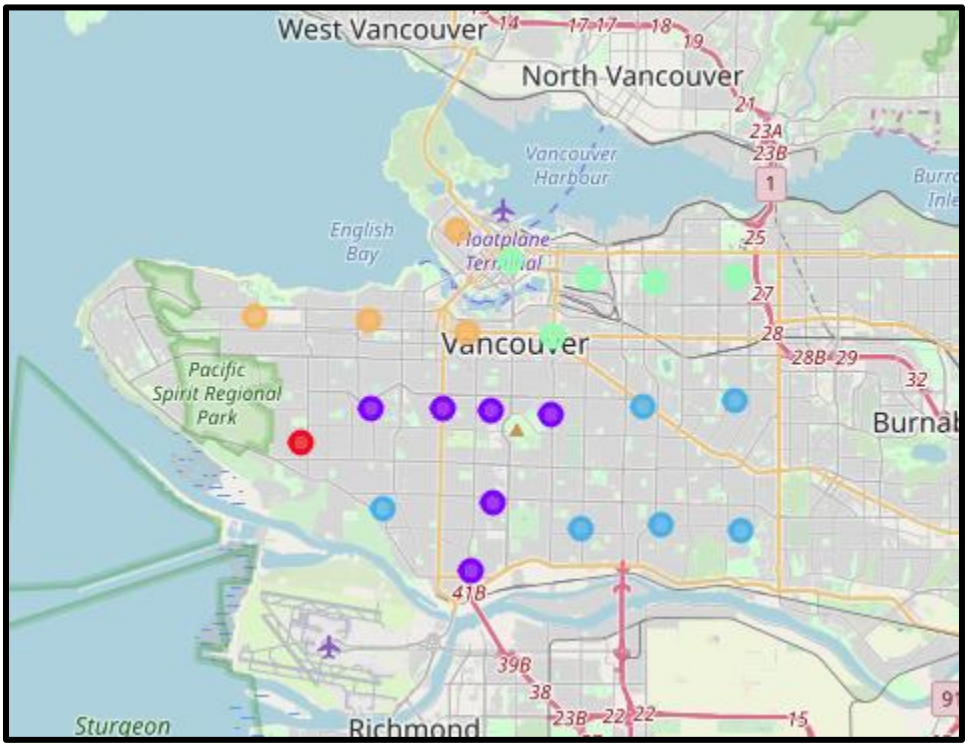

Figure 4: map of clusters from the k-means clustering model based on the top 5 venues for each neighborhood.

Discussion

From the results, a bus tour company can apply this model to any city and produce a starting point for a bus tour, with no prior knowledge of the city. Each day of the tour can be spent in a different cluster. Future refinement of the model could use the foursquare API to acquire the '**top picks'** for each cluster, allowing for better decision making when planning out where the tour should spend its time. One problem with this model could be that that the foursquare **explore section = trending** end point category may not be a good representation of what tourists want to see. Using other endpoints such as **explore section = outdoors** or **explore section = sites** may be a better solution.

Conclusion

This model could be applied to any city where the GPS locations of a neighborhood are known. As it stands the model breaks the neighborhoods into 5 clusters of similar trending venues. The bus tour will be 5 days long and each day will be spent in a cluster.

Further refinement of the model could help choose what venues to visit in each cluster, by using the top picks end point. This model will cut down on research time and allow a company to expand faster than the competitors. This could also be applied to individuals who would like to go traveling, but are unfamiliar with a given city.