Introduction

An international 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 give them 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.

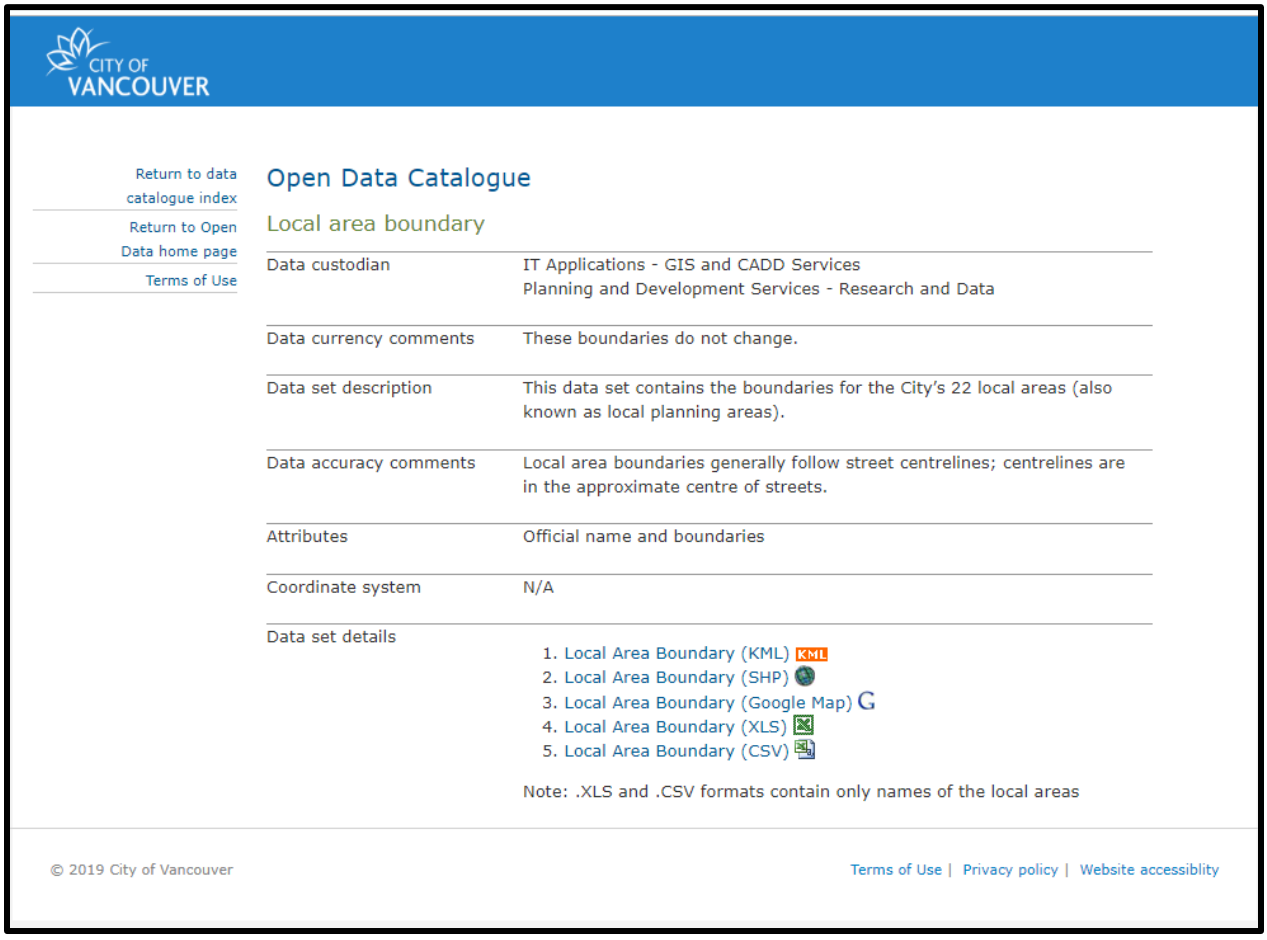

Figure 1: webpage where neighborhoods of Vancouver where acquired. The KML file was used as the CSV only contained the neighborhood names with no GPS data.

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

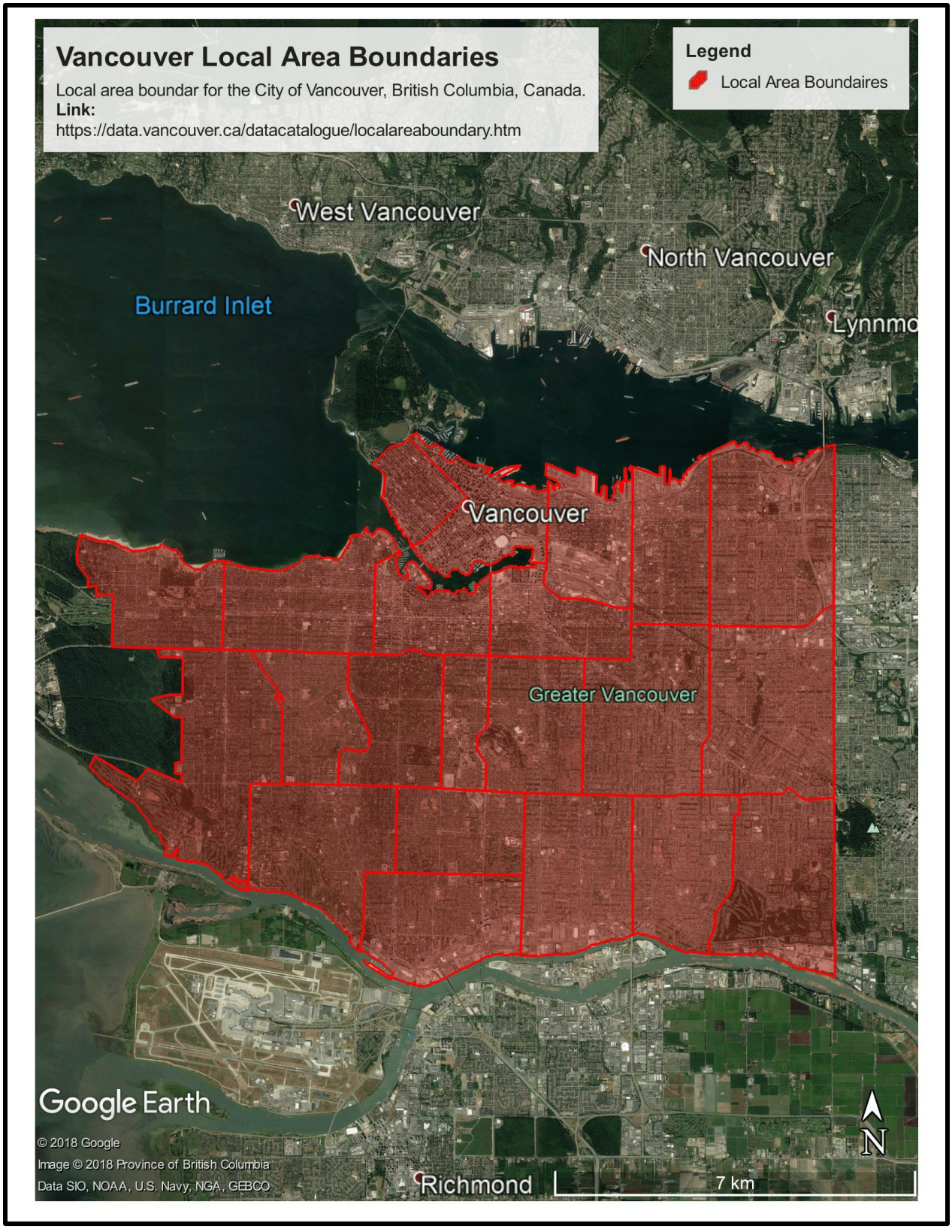

Figure 2: KML file of Vancouver neighborhoods.

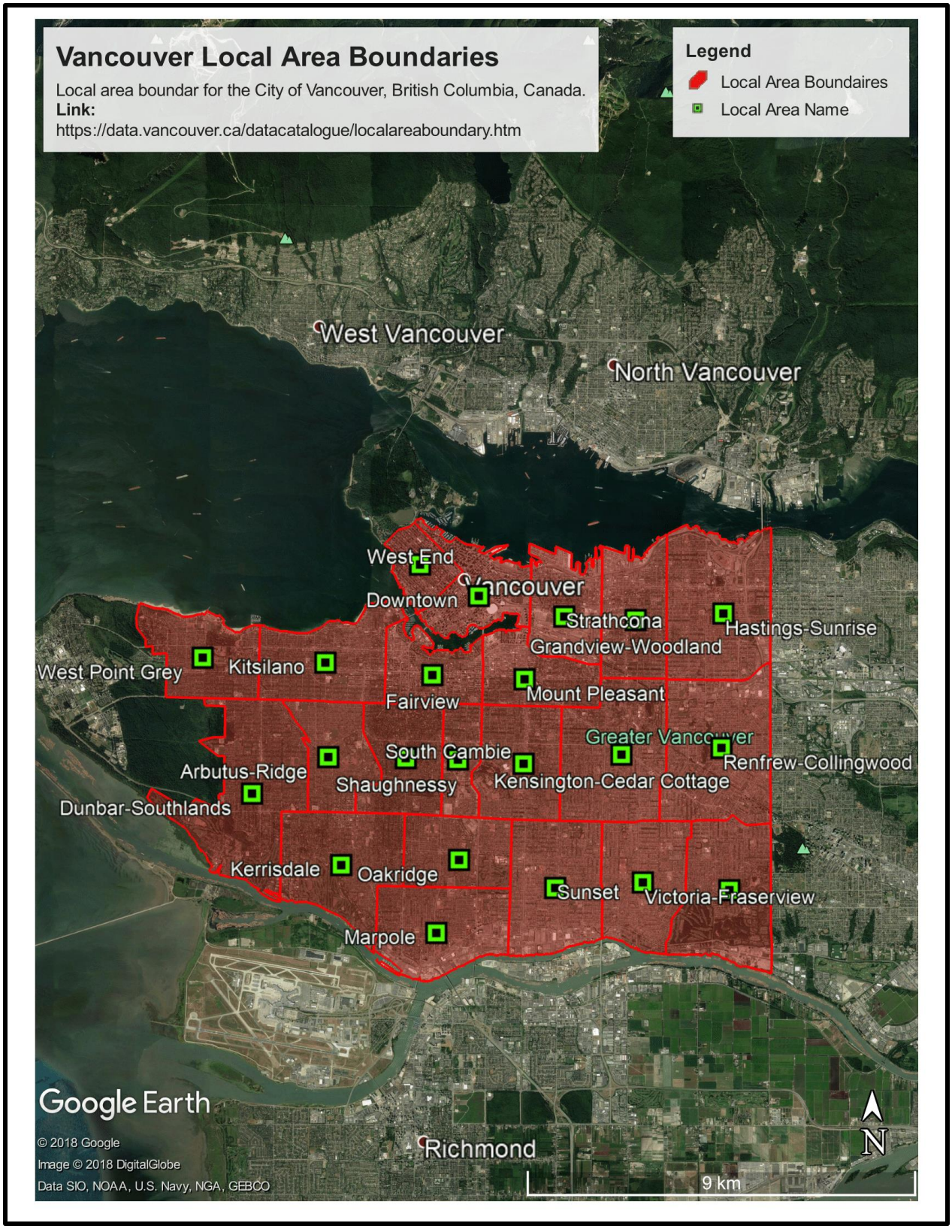

Figure 3: Neighborhood polygons for the city of Vancouver and center points.

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

Table 1:

	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 coordinated for the neighborhood names can then be run through the four-square API to acquire nearby trending venues and group them in to 5 clusters using K-means Clustering. This will give the tour company a starting point to set up a bus tour in a new city.