**Coursera Capstone Project**

Introduction/Business Problem

I am evaluating purchasing a model railroad hobby shop. The current owner has said he is likely going to shut it down next year when the lease is up. I have an idea to downscale the physical size of the store, to move it to a lower rent area of town, and to greatly enhance the online presence of the business.

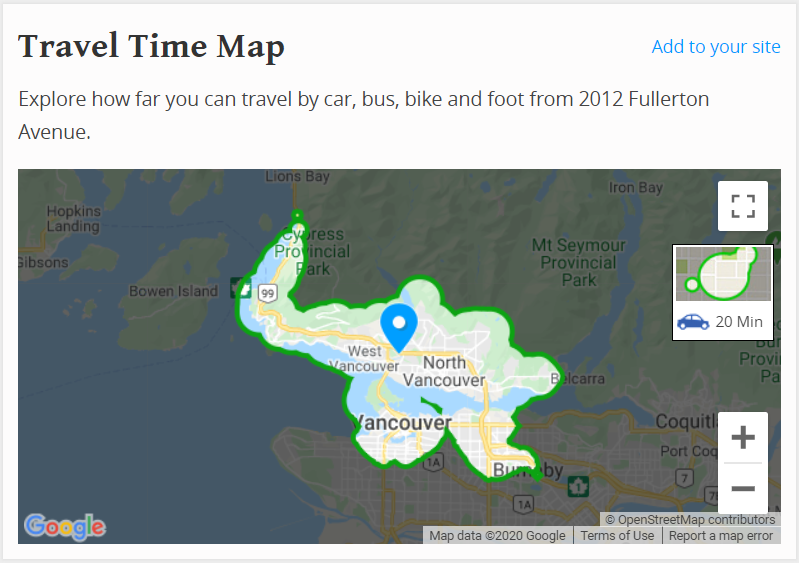
Because the shop is now the last one in Vancouver catering to the North American model railroad enthusiast (there are still a couple stores in town that cater to European style model railroading) , I believe that the store will be a ‘destination’ outlet, meaning customers will be prepared to make specific trips out to the store and it no longer requires the high visibility, and high rent, location it is at today.

This project is therefore to evaluate various retail locations around Vancouver to identify the best ones for potentially relocating this hobby store business.

Data

The first step is to identify the region of study. Because commuting time is a significant value for me, I have decided I will only consider locations that are within 20 minutes drive of my home (no traffic).

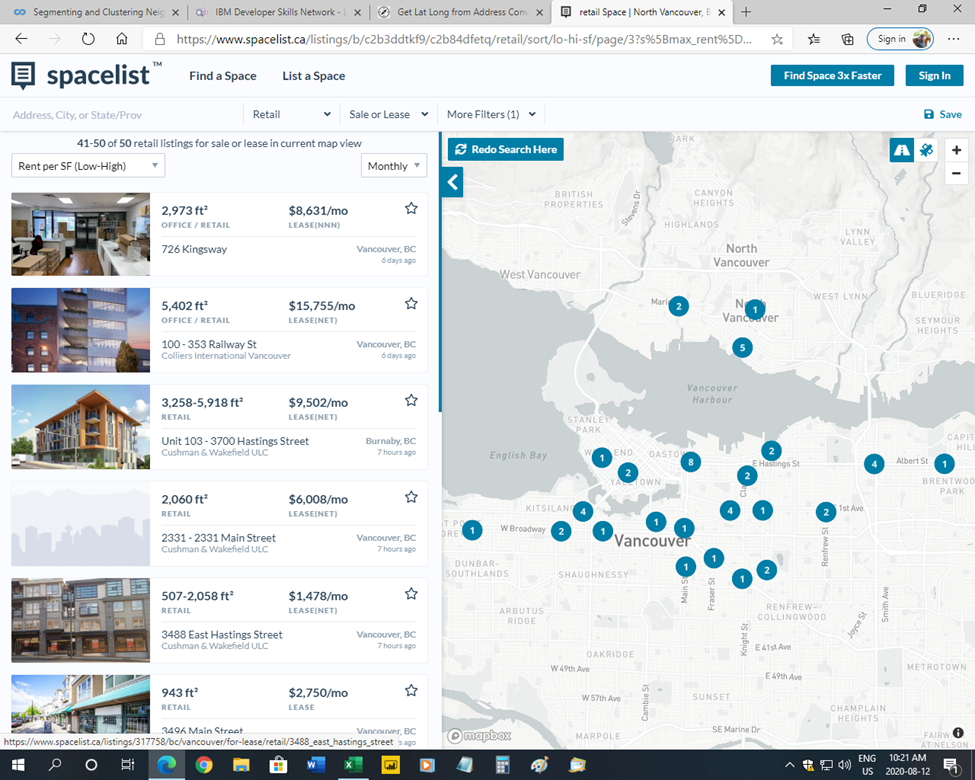
I will be using the web site Walk Score (<https://www.walkscore.com/>) to define the study region. They have a travel time map feature that identifies how far I can travel by car from my home. The results are show below:



In the second step, I will use the retail space listing site Spacelist (<https://www.spacelist.ca/>) to identify available retail sites for lease. I will use their map feature to filter for sites that correspond to the area previously identified by Walk Score.

I will further limit my criteria to locations where the annual lease rate is less than $35.00 per square foot. The search results in 50 locations, see below:

**Retail Space for Lease**



I will scrape the Spacelist web site to extract the relevant data fields for the 50 locations, namely, address, square footage, and lease rate ($/ft2/yr).

Note, I will also add in the existing address and presumed square footage and lease rate for the hobby shop, so that I can include it in the Foursquare analysis.

In the third step, I will use the web site Geocodio (<https://dash.geocod.io/>) to obtain the latitude and longitude coordinates of the 51 retail lease locations. This service is free as long as the daily usage is small. I will upload into the site the list of addresses in a csv file and the results are returned in a downloadable csv file with the latitude and longitude coordinates added as extra columns, see below:

**Retail Lease Data with Latitude and Longitude (5 of 51 shown)**



In the fourth step, I will use the Walk Score web site (<https://www.walkscore.com/>) to identify those retail locations that are within a 30 minute walk of my home. I will add this as a column to the above location data. This will be an added criterion to consider when evaluating and clustering the various retail sites.

In the fifth step, I will use Foursquare location data (<https://foursquare.com>) to identify venues nearby to each of the 51 potential retail locations. Using Python in a Jupyter Notebook, I will create a function using ‘explore’ in the Foursquare API to get the top 100 venues within a 500-meter radius of each retail location. I will then do one hot encoding of the venues, group by retail location, and calculate the mean of frequency of occurrence of each venue category. Then, I will create a new dataframe that displays the top 10 venues for each location.

In the sixth step, I will run a k-means cluster analysis on the location data to group the 51 locations into 5 similar clusters. This will be done using Python in a Jupyter Notebook.

Lastly, I will visualise the 51 retail locations on a map of Vancouver, colouring the location markers by the cluster. This will be done with Python in a Jupyter Notebook, using Folium Map.