**Coursera Capstone Project - Report**

1. 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.

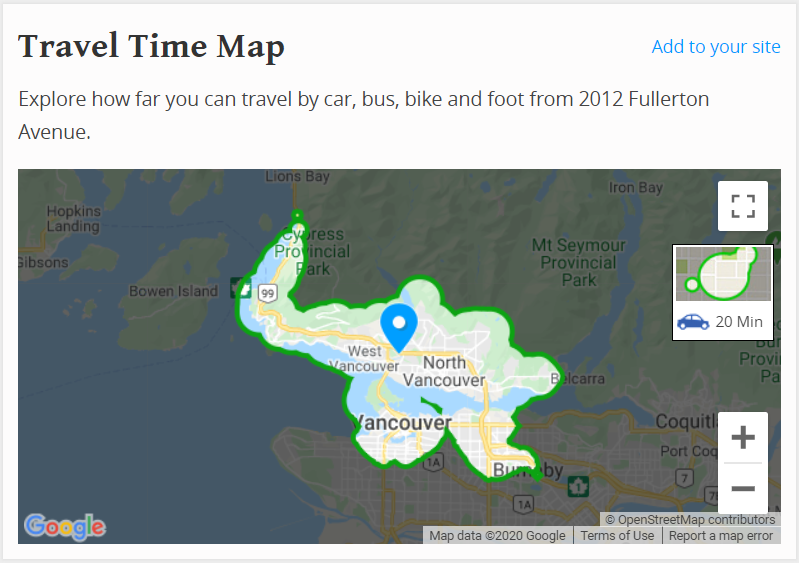
Because the shop is now the last one in Vancouver catering to the North American model railroad enthusiast (there are still a couple stores in town that cater to European style model railroading) , I believe that the store will be a ‘destination’ outlet, meaning customers will be prepared to make specific 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.

2. 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).

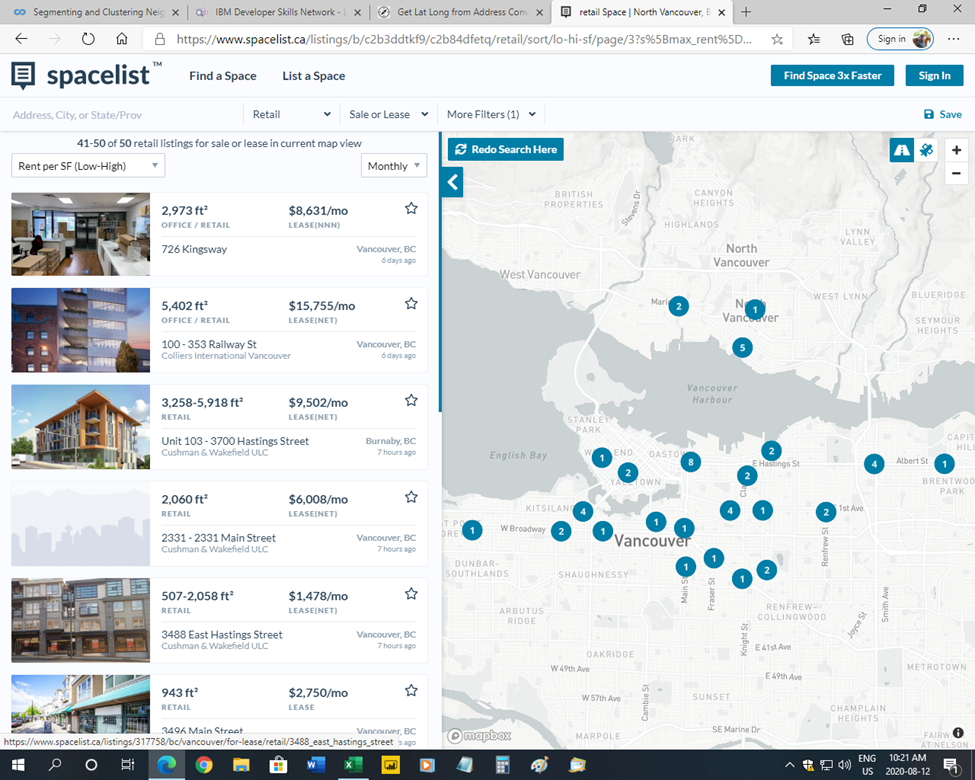
I will be using the web site Walk Score (<https://www.walkscore.com/>) to define the study region. They have a travel time map feature that identifies how far I can travel by car from my home. The results are show below:



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.

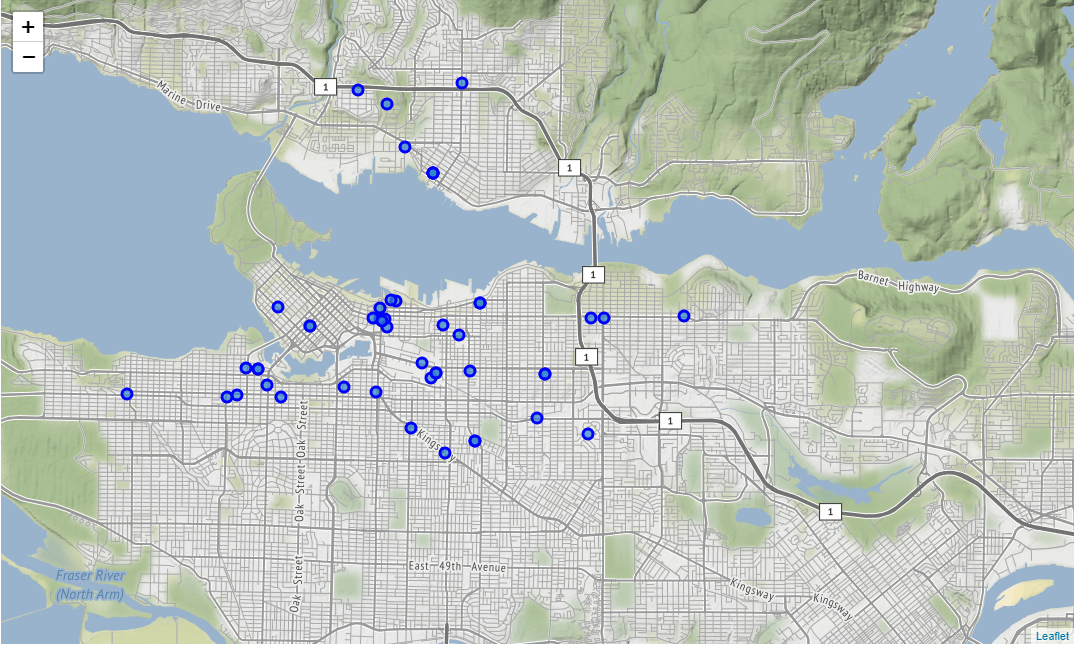
3. Methodology

I used a Jupyter Notebook in the IBM Watson Cloud, an account available as part of my studies. I will also use GitHub as a remote repository to store the related data csv files, the completed Jupyter Notebook, this report and the presentation document.

After getting the Latitude and Longitude data for the 51 retail sites from Geocodio, I then read the csv file into a dataframe. The result looks as follows (only first 5 records shown):



I then used python Folium library to visualize the location details on a map of Vancouver, see below:



Next, the Foursquare location API (<https://foursquare.com>) was used to identify venues nearby to each of the potential retail locations using the ‘explore’ feature. I set the limit to a maximum of 100 venues and a radius of 500 metres. Here is an example of the sorted and filtered result for the first record: 2825 Grandview Highway, Vancouver. Note, only the first 5 venues are shown. This address had at total of 17 venues identified.



In total, 1917 venues were returned by Foursquare, with 201 unique venue categories.

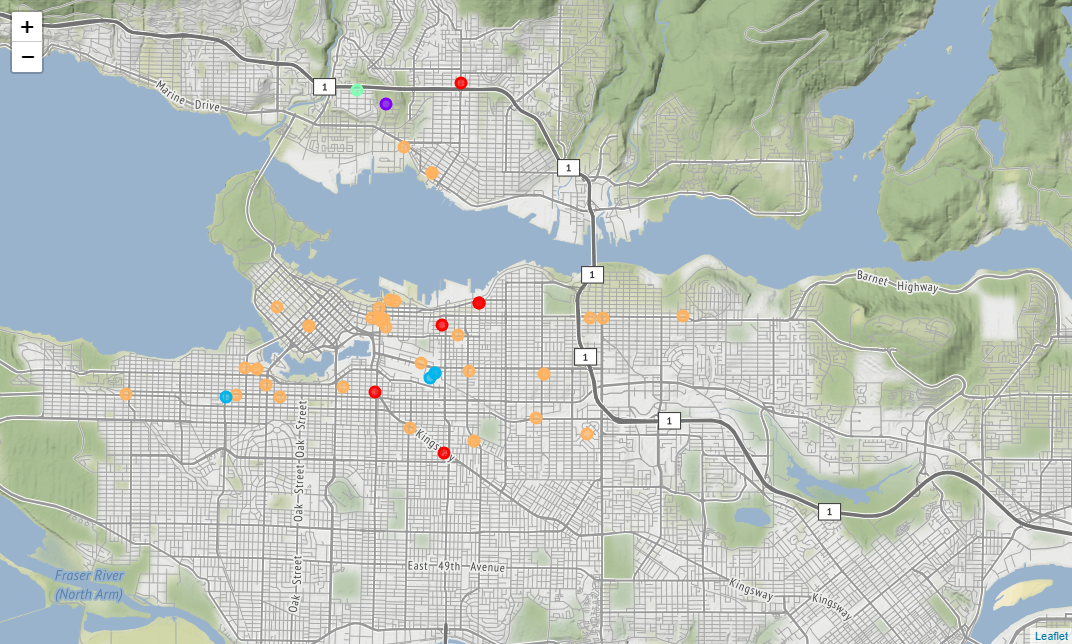
The next step was to do one-hot encoding of the venues, group by retail location, and calculate the mean of frequency of occurrence of each venue category. A new dataframe was then created that displayed the top 10 venues for each location, see below (only first 5 records shown):



Then a k-means cluster analysis was done on the location data to group the 51 locations into 5 similar clusters. After the cluster data was merged with the original retail site and venue data, the resulting dataframe looked as follows (only first 10 records shown):



Then, the 51 retail locations were visualized on a map of Vancouver, colouring the location markers by the cluster. This was with Python in a Jupyter Notebook, using Folium Map, see below:



4. Results

The k-means cluster analysis grouped the 51 potential retail sites into 5 similar groups. Note, because the Python k-means cluster analysis package indexes the 5 clusters with 0-4, I will refer to them as they are assigned in the computer.

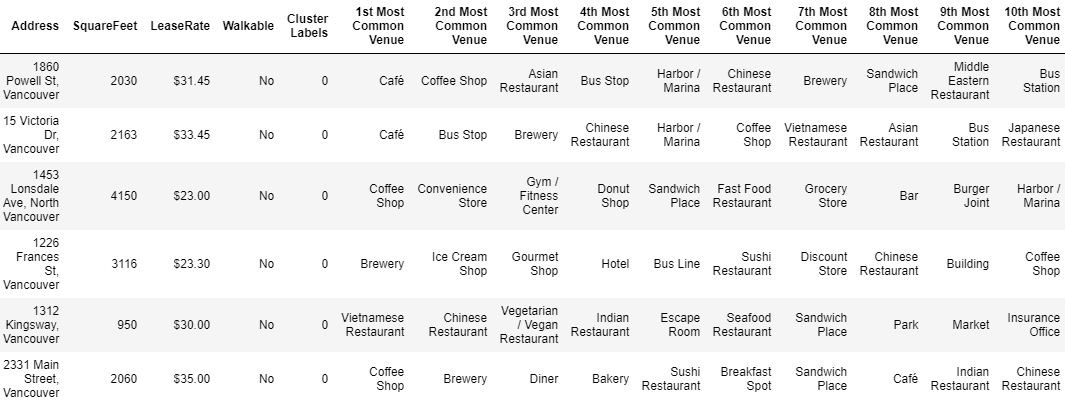
Summary:

* Cluster #0 – red markers, 6 retail sites
* Cluster #1 – purple marker, 1 retail site
* Cluster #2 – blue markers, 4 retail sites
* Cluster #3 – light blue marker, 1 retail sites
* Cluster #4 – orange marker, 39 retail sites

A more detailed description of each cluster follows.

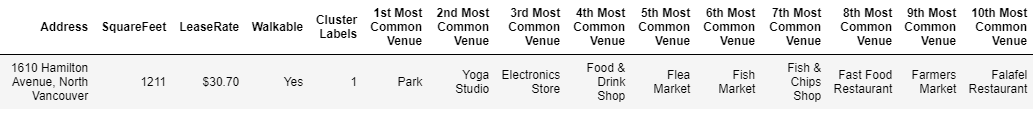
Cluster #0 – “The Café and Restaurant Crowd”

This cluster appearing in red on the above map, includes 6 retail sites, and most frequently is located next to cafés, coffee shops and restaurants. A list of the retail sites is shown below.



Cluster #1 – “Park and Mixed Stores”

The next cluster appears in purple on the above map, and includes only 1 retail site. It is located across from a park, through there are Yoga Studios and stores nearby. It also represents 1 of only 2 potential retail sites that are walkable from my home. Something to weigh strongly! The retail site detail is shown below.



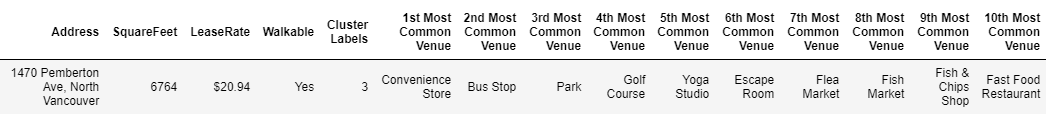
Cluster #2 – “Coffee, Furniture and Hardware”

This cluster appears in blue on the above map, and includes 4 retail sites. Like the earliest cluster, it too has many coffee shops located nearby, but they also have a mix of furniture and hardware stores as well, see below:



Cluster #3 – “Convenience Store, Park and Golf”

This cluster appears in light blue on the above map, and is the only other cluster that includes just 1 retail site. It too is located near a park, though there is a convenience store, golf course and bus stop nearby. This is also the only other site that is walkable from my home. See details below:



Cluster #4 – “All the Rest”

This last cluster appears in orange on the above map, and lumps together a mix of 39 retail sites. It is a more diverse grouping with more car oriented and big box outlets nearby. This cluster includes 2825 Grandview Highway, the current site for the Hobby Shop. Details for the first 4 sites is shown below:



5. Discussion

In general there were a fair number of available retail sites for lease. There were 51 sites included in the analysis here, though if I had chosen to accept a longer commute by car (more than 20 minutes) or consider annual lease rates of more than $35.00 per square foot, many more potential retail sites could have been included in the analysis.

By limiting the retail lease rate to roughly the lower half of what was available, I found many of the sites were located a block or two away from a main retail shopping area. There would likely be less foot traffic in these sites, but then the rent is significantly lower too. Because this is the last model railroad Hobby Shop of its kind left in the city, I think most business would be repeat customers who would make a special trip to visit the store. Therefore, I don’t think the store would benefit much from walk-in, spur of the moment, foot traffic. However, having good car, parking or transit access could be valuable.

I would also want to weigh heavily the 2 retail sites that are located within a half-hour walking distance of my home. This would be a great advantage to me, though they may be a less central location for customers. I may have to weigh my own preferences versus the business.

Lastly, I should consider the effect of the coronavirus. At the time of doing this report we have come out of a community-wide lockdown but new cases are rising again and it may be several more months or even a couple years before the effects of the pandemic are fully behind us. It is expected many businesses, particularly small retail businesses will be very hard hit. In time, this could dramatically alter the retail landscape, affecting availability and lease rates.

6. Conclusion

Although the hobby shop business has thus far not been too negatively impacted by the coronavirus, the retail landscape is changing fast. The analysis here suggests there are many suitable retail locations currently available for lease. Because the current hobby shop lease is not up till next year, there is time to assess trends in the retail market and perhaps an opportunity to catch lower lease rates in the near future.

References:

1. Walk Score - <https://www.walkscore.com/>
2. Spacelist - <https://www.spacelist.ca/>
3. Geocodio - <https://dash.geocod.io/>
4. Foursquare - <https://foursquare.com/>