Coursera Applied Data Science Capstone

Data to help drive home purchasing decisions in Vancouver Canada

By: Taij Taggar

Introduction: Business Problem

The Vancouver Real Estate market is among the hottest globally. Just last year, Vancouver was ranked the 4th most expensive place to purchase a home in the world. Real Estate agents and companies must be providing their clients with the best recommendations to ensure company growth and client satisfaction.

If Real Estate Agencies were able to utilize data, they could provide their clients with tailored neighbourhoods and areas to choose from. These recommendations could be based on amenities in the neighbourhood in addition to safety metrics and quality from life. Home buyers will be able to use this information in addition to their budget to target certain neighbourhoods to purchase a home. With home buying being such a large investment in Vancouver, it is important that all stakeholders are utilizing all information available to them.

Data:

Based on the problem above, the following data would be ideal to use in such a project

- Neighbourhood names and locations in Vancouver
- Crime data for these neighbourhoods in order to advise on safety
- Census information for the neighbourhoods to further understand demographics, population and general makeup of the neighbourhoods
- Information on amenities in these neighbourhoods to tailor to clients persona

Utilizing all this information is beyond the scope of this project. For this project, the following sources of data will be used.

- Vancouver Police Crime Data: Downloaded as a csv and added to project workbook
- 2016 census data: Downloaded as a csv and added to project workbook. *Note that in this project only population data is used. Utilization of other census data is beyond the scope of this project
- Foursquare API: Used as a source to find information on amenities in surrounding neighbourhoods

Methodology:

The data will be gathered and used in the following way

Part 1: Crime Data Wrangling

- VPD Data will be uploaded to the workbook. Neighbourhoods will be accessed and crimes will be summed to find the total number of crimes per neighbourhood. Crime data from 2018 will and onwards will only be used.

Part 2: Merge with Census Data

- Population census data will be added to the workbook.The neighbourhoods will be accessed and reconciled with the VPD data to ensure neighbourhoods match. If there are discrepancies between the two sets of neighbourhoods, neighbourhoods will be merged based on geographic location in order to match.
- Once neighbourhoods match in VPD and Census data, these two datasets will be merged so the "Crime per population" can be calculated. This makes it possible to compare neighbourhood crime rates even if populations are different.

Part 3: Defining geographic locations for neighbourhoods

- Latitude and Longitude will be generated for each neighbourhood by using geocoded to determine coordinates.

Part 4: Foursquare API integration to find venues

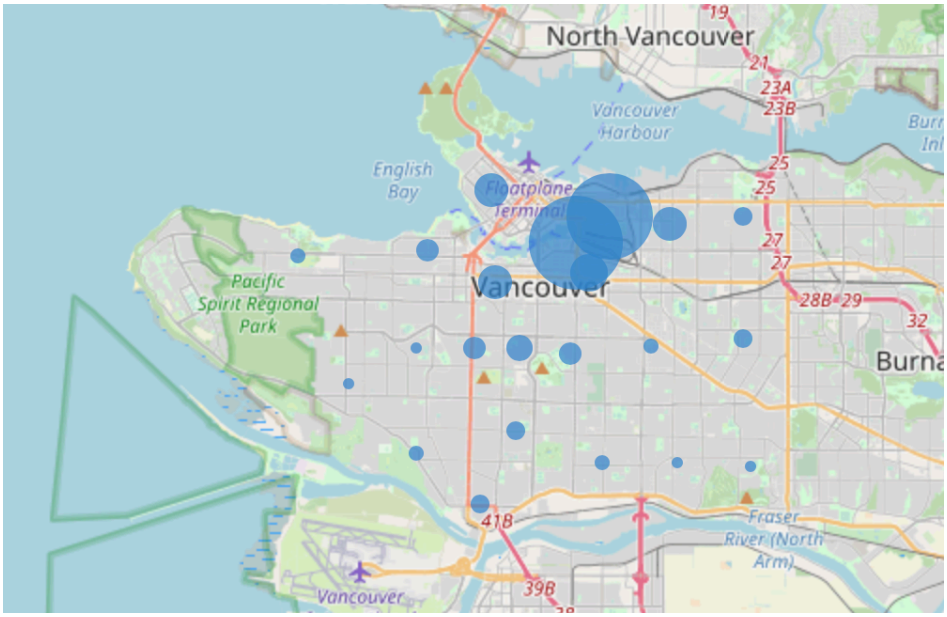
- The foursquare API will be used to find venue locations in neighbourhoods in Vancouver

Part 5: K-means clustering and Analysis

- K-means clustering will be used to divide the neighbourhoods into similar groups. From this each group will be analyzed and assigned a "Persona". This persona is what the Real Estate Agencies could use to tailor based on their clientele. Each group will have crime rate available to see so the user is able to see what the safest neighbourhood in the target group they are looking at

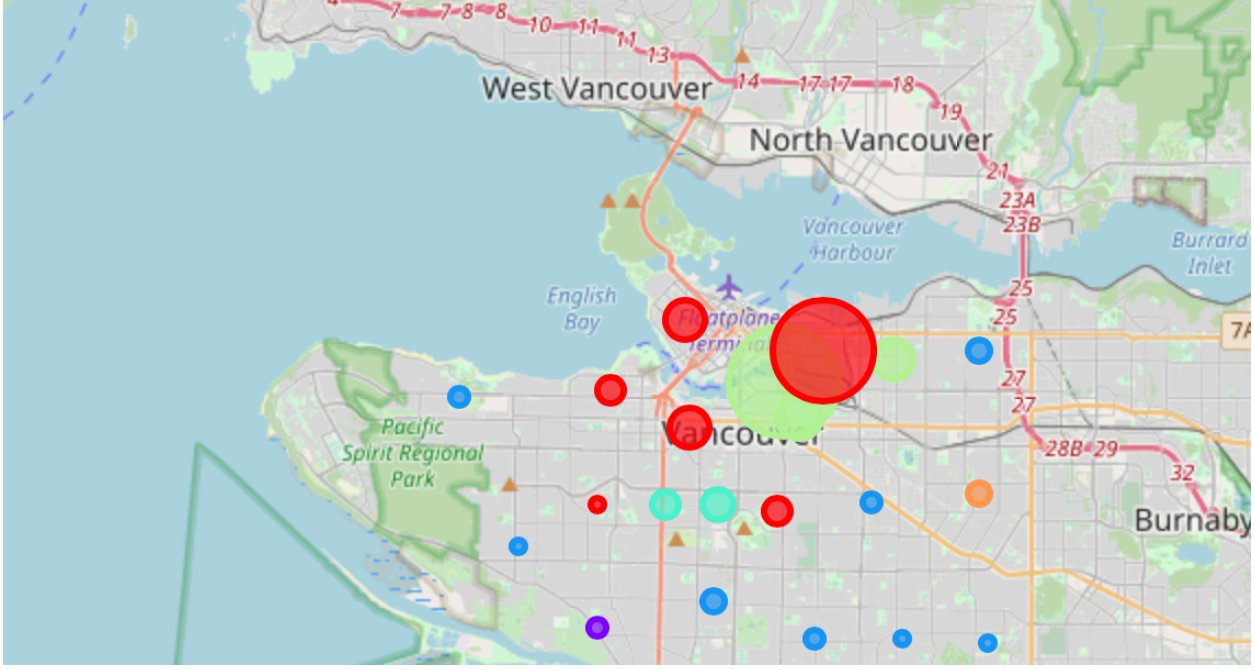
Results & Discussion:

The relative crime data is first visualized using folium in Figure 1.



	Neighbourhood	Crime per Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arbutus Ridge	0.061000	Food Truck	Sushi Restaurant	Convenience Store	Hobby Shop	Coffee Shop	Discount Store	Event Space	Caribbean Restaurant	Sandwich Place	Seafood Restaurant
3	Fairview	0.170791	Japanese Restaurant	Camera Store	Park	Restaurant	Bakery	Salon / Barbershop	Coffee Shop	Pharmacy	Pet Store	Pizza Place
9	Kitsilano	0.116460	Restaurant	Yoga Studio	Coffee Shop	French Restaurant	Bakery	Beach	Breakfast Spot	Clothing Store	Deli / Bodega	Food Truck
14	Riley Park	0.112924	Café	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Farmers Market	Seafood Restaurant	Coffee Shop	Park	Restaurant	Record Shop	Chinese Restaurant
17	Strathcona	0.467461	Sandwich Place	Restaurant	Park	Café	Gourmet Shop	Brewery	Cheese Shop	Chinese Restaurant	Seafood Restaurant	Food Truck
20	West End	0.182945	Farmers Market	Restaurant	Park	Bookstore	Bakery	Falafel Restaurant	Sandwich Place	Ramen Restaurant	Pub	Convenience Store

Figure 1: *Crime per neighbourhood population visualized using a relative bubble chart in folium.*



	Neighbourhood	Latitude	Longitude	Population	Total Crime	Crime per Population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
7	Kerrisdale	49.220985	-123.159548	13975	1222	0.087442	1	Bus Stop	Golf Course	Park	Café	Supermarket	Gift Shop	Spanish Restaurant
13	Renfrew-Collingwood	49.248577	-123.040179	51530	5315	0.103144	5	Bus Stop	Park	Pizza Place	Bus Station	Bar	Business Service	Deli / Bodega
15	Shaughnessy	49.246305	-123.138405	8430	1094	0.129775	3	Park	Garden	Coffee Shop	Bank	Sandwich Place	Gift Shop	Malay Restaurant
16	South Cambie	49.246464	-123.121603	7970	1110	0.139272	3	Coffee Shop	Park	Garden	Grocery Store	Chinese Restaurant	Sandwich Place	Bank

As seen above, crime is mostly concentrated in downtown locations, mainly Central Business District and Strathcona. Crime rates are the lowest in neighbourhoods such as Victoria-Fraserview and Killarney.

Six clusters are chosen using the k-means algorithm to separate similar neighbourhoods based on venue similarity using the 4-square API. Six clusters are chosen due results producing similar groups throughout. Figure 2 uses a folium map to plot neighbourhoods based on colours for k-means grouping and size for relative crime rates.

Figure 2: *Folium map separating Vancouver neighbourhoods by k-means group and by relative crime rates.*

Group 1 included the following neighbourhoods.

When looking at this group, it seems to be a neighbourhood with casual dining, quick snacks, drinks and treats. These neighbourhoods could be tailed to a younger crowd, or crowds which value this sets of venues the most. In terms of safety, Arbutus Ridge would be the neighbourhood of choice.

Group 2 just contained the neighbourhood Kerrisdale. Other groups such as groups 3 and 5 only contained 1 or 2 neighbourhoods. It was noticed that if these groups were combined, they would share a lot of similar venues. This joined group contains the following neighbourhoods.

These neighbourhoods contain many outdoor venues such as parks, gardens and golf courses. In addition essential venue. This neighbourhood may be more attractive to students. In terms of safety, Kerrisdale is the safest neighbourhood.

Group 3, was the largest group containing 9 neighbourhoods.

	Neighbourhood	Crime per Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Dunbar-Southlands	0.061704	Grocery Store	Café	Gym	Golf Course	Liquor Store	Coffee Shop	Pet Store	Sandwich Place	Restaurant	Pub
5	Hastings-Sunrise	0.107390	Vietnamese Restaurant	Event Space	Park	Indian Restaurant	Sandwich Place	Chinese Restaurant	Portuguese Restaurant	Pharmacy	Convenience Store	Middle Eastern Restaurant
6	Kensington-Cedar Cottage	0.086325	Vietnamese Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Café	Gym / Fitness Center	Seafood Restaurant	Sandwich Place	Pizza Place	Park	Massage Studio
8	Killarney	0.053981	Chinese Restaurant	Bus Stop	Fast Food Restaurant	Farmers Market	Liquor Store	Recreation Center	Sushi Restaurant	Park	Gas Station	Bar
10	Marpole	0.102044	Chinese Restaurant	Sushi Restaurant	Bank	Café	Japanese Restaurant	Indian Restaurant	Grocery Store	Restaurant	Pizza Place	Malay Restaurant
12	Oakridge	0.106907	Sushi Restaurant	Coffee Shop	Fast Food Restaurant	Tea Room	Sporting Goods Shop	Chocolate Shop	Convenience Store	Electronics Store	Cantonese Restaurant	Restaurant
18	Sunset	0.079205	Indian Restaurant	Chinese Restaurant	Bank	Restaurant	Pharmacy	Bakery	Food Truck	Event Space	Discount Store	Dessert Shop
19	Victoria-Fraserview	0.054949	Pizza Place	Bus Station	Sandwich Place	Restaurant	Convenience Store	Bus Stop	Gas Station	Middle Eastern Restaurant	Motorcycle Shop	Fast Food Restaurant
21	West Point Grey	0.088557	Harbor / Marina	Bank	Sushi Restaurant	Beach	Park	Vietnamese Restaurant	Hostel	Sandwich Place	Restaurant	Athletic Spo

It is clear that these neighbourhoods in Vancouver contain many ethnic cuisine options. These neighbours may be tailored towards valuing lots of mid-ranged cuisine options. Citizens who identify as minority may find some of these neighbourhoods interesting as it may contain lots of restaurants with their ethnic cuisine. Victoria-Fraserview is the safest neighbourhood in this group.

Group 4 contains the following neighbourhoods

	Neighbourhood	Crime per Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Central Business District	0.499919	Brewery	Bakery	Ice Cream Shop	Pizza Place	Dessert Shop	Gym	Community Center	Music Venue	Music Store	Movie Theater
4	Grandview-Woodland	0.174910	Coffee Shop	Pizza Place	Indian Restaurant	Brewery	Tapas Restaurant	Scandinavian Restaurant	Mexican Restaurant	Pub	Cuban Restaurant	French Restaurant
11	Mount Pleasant	0.209619	Brewery	Coffee Shop	Sushi Restaurant	Yoga Studio	Grocery Store	Middle Eastern Restaurant	Lounge	Library	Latin American Restaurant	Ice Cream Shop

These neighbourhoods contain many bars, brewery's and snack like shops. A group that seems to have a bustling vibe and tailored towards people who like to go out. It can be seen here that all 3 groups have relatively high crime rates, however Grandview-Woodland is the safest neighbourhood.

Conclusion & Recommendations:

By utilizing crime data and the 4-square API, it was possible to group neighbourhoods utilizing the k-means machine learning algorithm in Vancouver, Canada. These groups were each given a persona that would potentially tailor to certain individuals who are looking to purchase a home in Vancouver. Crime data integration may help the individual decide which neighbourhood in a group they may want to move to. To further refine this model, more data should be used such as

- Full utilization of the census data
- Home purchase prices per neighbourhood
- Further refinement of venue data

These additions to the data model will provide even more insight to home buyers in the Vancouver area. Real estate agencies could use this model to help their customers find the perfect home.