Battle of the Neighborhoods:
Best Livable Neighborhood of Vancouver

Lydia Langbroek

March, 2021

1. Introduction

1.1 Background

Located in West Canada, Vancouver is a nice city that is home to lots of sightseeing. Aside from being a nice touristic spot, the city is an excellent livable city. According to the results of the Economist Intelligence Unit survey 2019, Vancouver is one of the best livable cities worldwide. Vancouver ranked 6th of the cities that were included within the survey. Factors included in the survey are living standards, crime, infrastructure, access to education and healthcare, and political and economic stability. The city of Vancouver has amazing public transport options. Bus lines, passenger ferries, and sky trains makes it extra convenient to get around without taking the car. Public transit fares are not really expensive. Aside from having great public transport, the city has both great employment opportunities and lots of activities to undertake in your spare time. Its location is perfect because of its close proximity to the ocean and mountains. In addition, the city scores high on safety.

1.2 Problem

Because of the popularity of the city and high demand of dwellings, the property price increases. People move to the city for employment reasons, people who graduated school/ university are likely to stay in the city and look for employment opportunities. Also, people are about to start a family and may be looking to purchase a house. Safety and an affordable living quarter with good amenities nearby are important to consider where to move. Aside from having a good job after graduating school/ university, you want your kids to grow up in a good and safe neighborhood! Therefore, the main question in this research is: What is the best neighborhood to live in?

1.3 Interest

The target audience for this analysis will be young people who graduated university, have a good job in the city, and who are about to start a family life. People who graduated school/ university often stay within the same city because they have a social live and work live there. At the end, I conclude which neighborhood in Vancouver is the best neighborhood to live in, considering the average housing prices in neighborhoods, crime rate, and specific amenities in neighborhoods.

2. Data Collection and Cleaning

2.1 Data Collection

I collect data from different sources. The first data requirement for the analysis is retrieving Vancouver Geo Data. I retrieve the Vancouver Geo Data from Geo Names. This website includes postal code data from various countries worldwide. Aside from postal code data, geospatial data is available. I access the data by the web URL, I download and extract the text file, and I will save the data as a csv file in my documents folder. After collecting postal code and geospatial data, I gather data on housing prices and crime data. The data on housing prices is retrieved from the Vancouver Open Source Portal. The data is available in csv format. Also, the Vancouver crime data is available in csv format at the release of Vancouver Police Department (VPD) crime data website. Lastly, I collect data on amenities in neighborhoods from Foursquare; Foursquare has data available on venues. So, summarizing, I collect the data from the following sources:

Source	Purpose	Web URL
Vancouver Geo Data from Geo Names	Vancouver postal code data, neighborhood data, latitude and longitude data	http://download.geonames.org/export/zip/CA.zip
Property Price Data Vancouver Dwellings	Property price data for each neighborhood	https://opendata.vancouver.ca/explore/dataset/property-tax-report/table/
Vancouver Crime Data from the Vancouver Police Department	Crime numbers	https://geodash.vpd.ca/opendata/
Foursquare Venues Data	Foursquare venue data on amenities in neighborhoods	https://foursquare.com/
Vancouver GeoJSON	Coordinates to create a map with Folium	https://prototype.metabolismofcities.org/cities/vancouver/maps/boundaries/3642/

2.2 Data Cleaning

After data collection, all data requires cleaning. I start with cleaning the Vancouver Geo Data from Geo Names. The data I collected from Geo Names includes postal codes throughout the entire Canada. I select all data from the City of Vancouver. The dataset includes data on postal codes, neighborhoods, state, city, latitude, and longitude.

The property price data from the Vancouver Open Source Portal includes lots of redundant columns. I remove the redundant columns. The only columns we are interested in include zone category, current land value, and property postal code. The zone category is the type of dwelling. Some zone categories are not used for living, such as historic area, industrial, comprehensive development, limited agricultural, and commercial. The remaining categories include one family dwelling, two family dwelling, and multiple family dwelling. I calculate the average price for each neighborhood. Combing the postal code data with the average property price data, I create the final dataset on house prices along with the crime data and Foursquare venue data. I convert the current land value, or property price, to a visually attractive format; I display the prices per thousands (price k). This is by dividing the average property prices for each neighborhood by thousand. All properties in Vancouver Downtown have an enormously high price. This area would not be the best place to buy a property; therefore, I exclude Vancouver Downtown from the analysis (Downtown and Downtown Eastside). Also, UBC has unaffordable prices, so this neighborhood will be excluded as well.

To evaluate the safest neighborhood in Vancouver, I gathered Vancouver crime data. I retrieved the data from the Vancouver Police Department crime data website. All types of crime from 2003 till 2021 have been registered. I only use a subset of the data; this is almost the most recent data (all crimes recorded over 2020). I count the number of total crimes in 2020 for all neighborhoods.

In addition, it is required to change the neighborhood names to the right format. The neighborhood name should have the same format as in the Geo JSON data. I correct all neighborhood names in the property price dataset. Most neighborhoods have the correct format in the crime dataset.

Lastly, I retrieve Vancouver Venues data from Foursquare. For each neighborhood, I create a top 5 of most common venues within the neighborhood. I exclude restaurants and more touristic attractions (hotels, bike rentals) from the venues. These venues are less relevant to determine the best neighborhood to live in. The abundance of restaurants in Foursquare venues data could give a distorted picture of reality. The venue categories below include the most daily used venues, such as grocery shopping, leisure activities, public transport, parks, shopping malls, etc.

3. Methodology

After collecting and cleaning the data, I start exploring the data. I describe the distribution of property prices in Vancouver, the count and distribution of most common venues in Vancouver, and the crime numbers per neighborhood in Vancouver. Also, I determine the optimal number of clusters with machine learning. Each cluster has a unique set of top 5 venues, and each neighborhood has a unique set of venues. The neighborhoods with similar venue characteristics in the top 5 venues belong to the same venue cluster. Neighborhoods with different venue characteristics in the top 5 venues belong to a different venue cluster.

Property Prices

The distribution of property prices will be displayed in a histogram. The histogram displays the number of neighborhoods that have a property prices within a certain range.
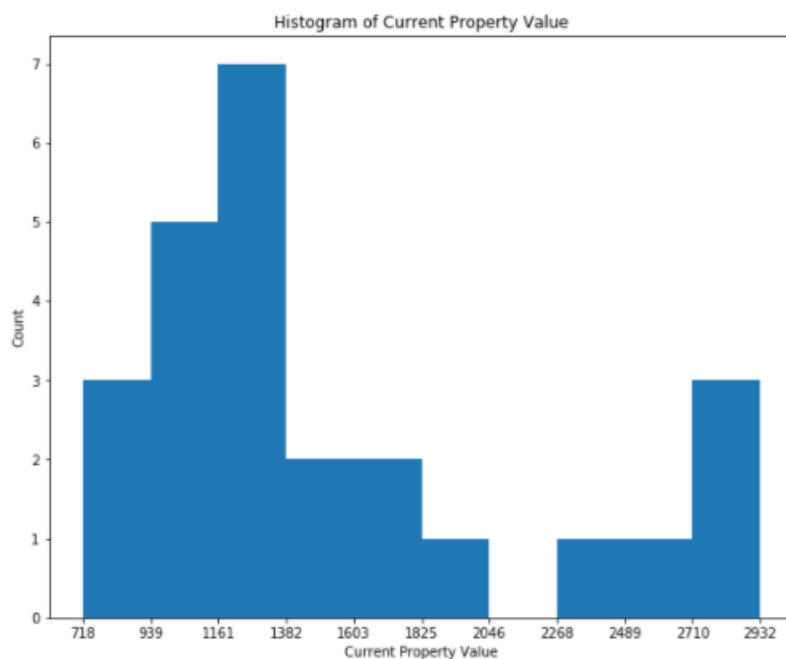
**Foursquare Venues: Most common Venue Types**

This section is a descriptive presentation of the most common venues: the number of venues that are most common in the first place, the number of venues that are most common in the second place, and the number of venues that are most common in the third place. A coffee shop is the most common venue for neighborhoods that are on the first, second, or third position of the top 5 venues. Parks and Yoga Studios are also very common within the top venues.

Most Common Venues

Second Most Common Venues

Third Most Common Venues
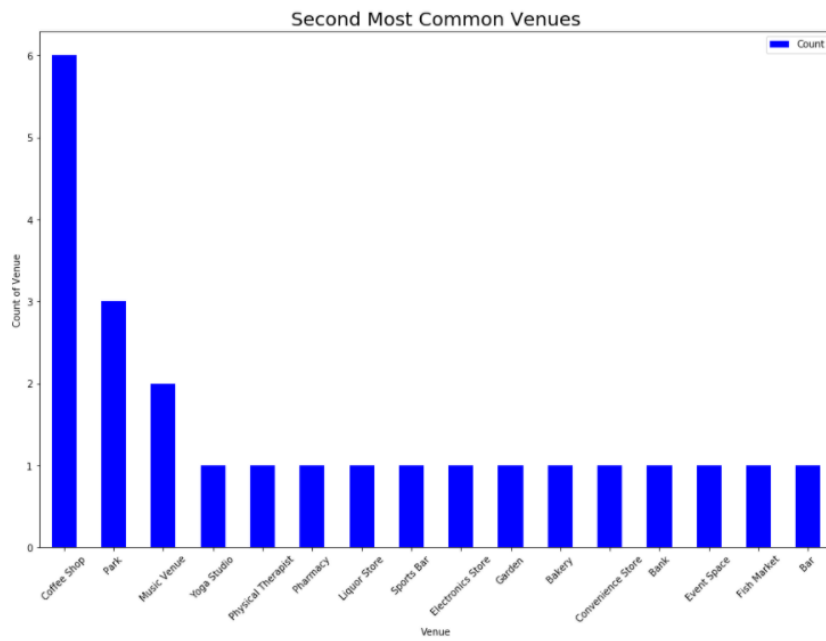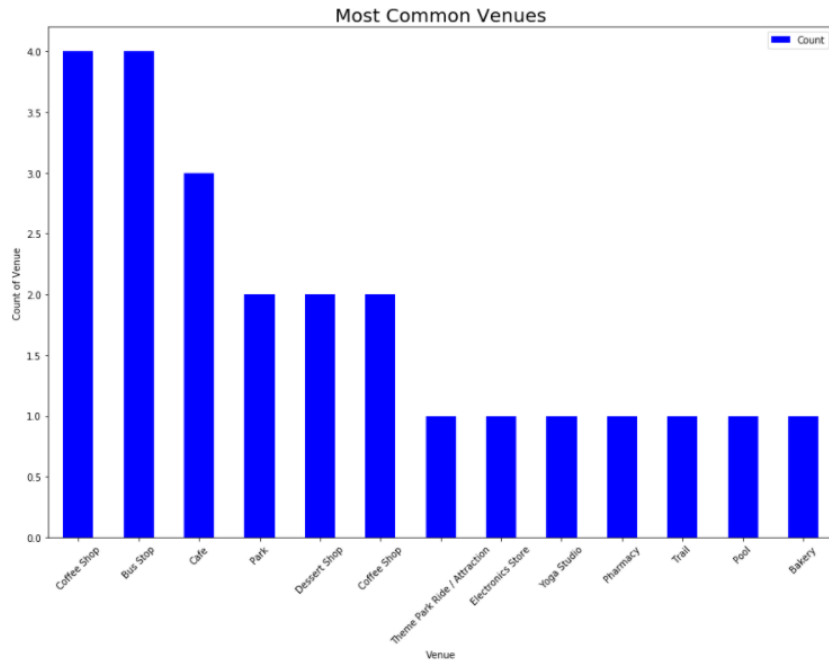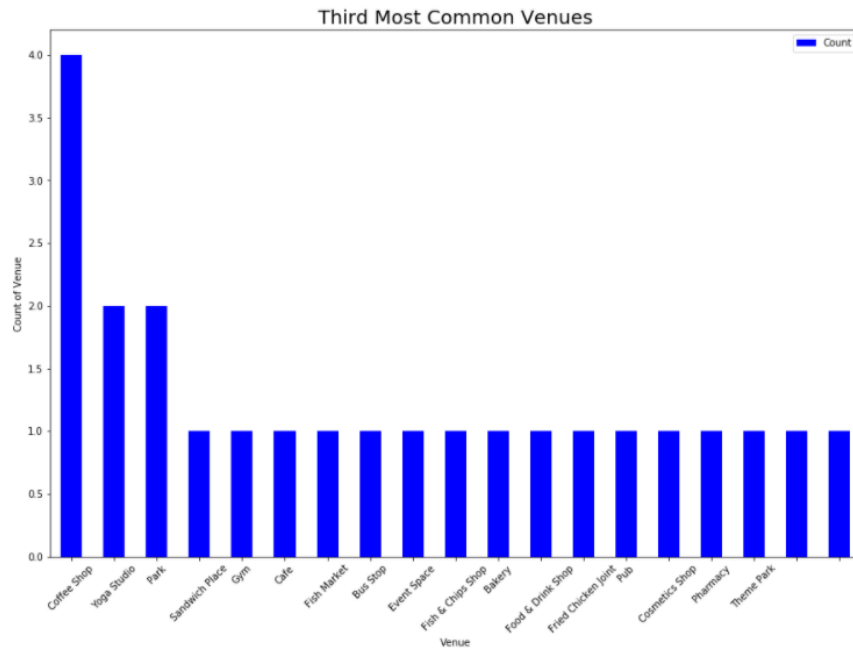
Crime Numbers

Let's check out the top 10 neighborhoods with the most crime in Vancouver. As you can see, Vancouver Downtown has the largest crime number, followed by West End and Mount Pleasant.

Neighborhoods with Most Crime 2020
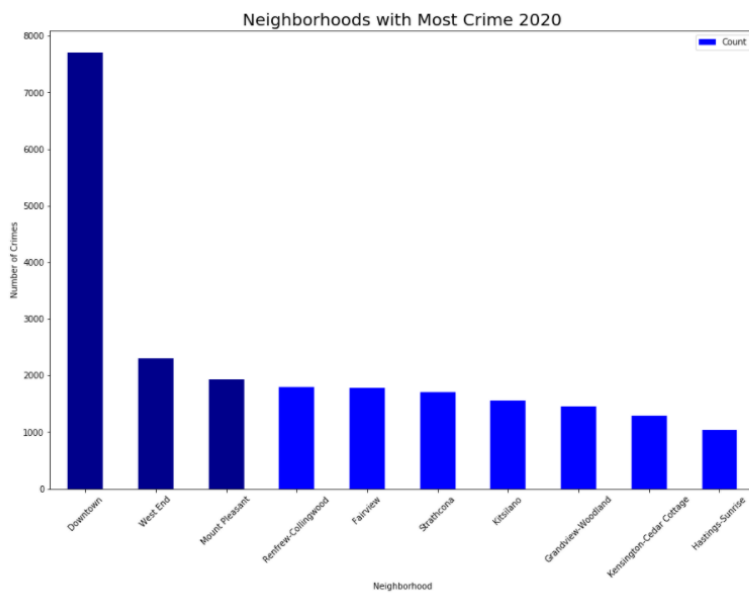
Foursquare Venues: Optimal Clusters

I use machine learning to get the optimal number of clusters. I create a scree plot to determine the cluster maximum. The elbow point in the plot is the optimal number of clusters. The elbow is at k equals 3, so the optimal number of clusters equals 3. This means that there are 3 different clusters with same top 5 venues characteristics. The figure below displays the scree plot.
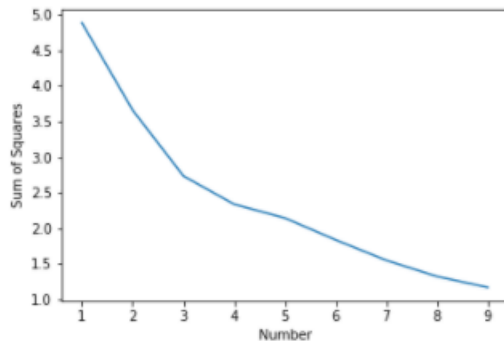

Foursquare Venues: Clusters

How are the clusters distributed over Vancouver? The map shows the 3 clusters of venues data in Vancouver. The 3 clusters are presented by red dots, purple dots, and a green dot.

Cluster 1 represents the largest cluster and the red dots. This cluster includes a variety of stores that are used in everyday life. It mainly includes venues that are related to food, such as coffee shops, pizza places, dessert shops, liquor stores, bakeries, sandwich places, convenience stores, coffee bars, and cafes. In addition, this cluster includes other basic needs of life, such as healthcare products (pharmacy, physical therapist). Bus stops are also very common.

Cluster 2 represents the purple dots in the map. The cluster includes leisure. Parks, playgrounds, pools, or other leisure activities (sports activities) are often found just outside the city center.

Cluster 3 is more of a '*leftover*' cluster. The cluster includes stores you use in everyday life, but it also includes other types of venues that are somewhat related to leisure (yoga studio and dog run). This cluster represents the green dot.
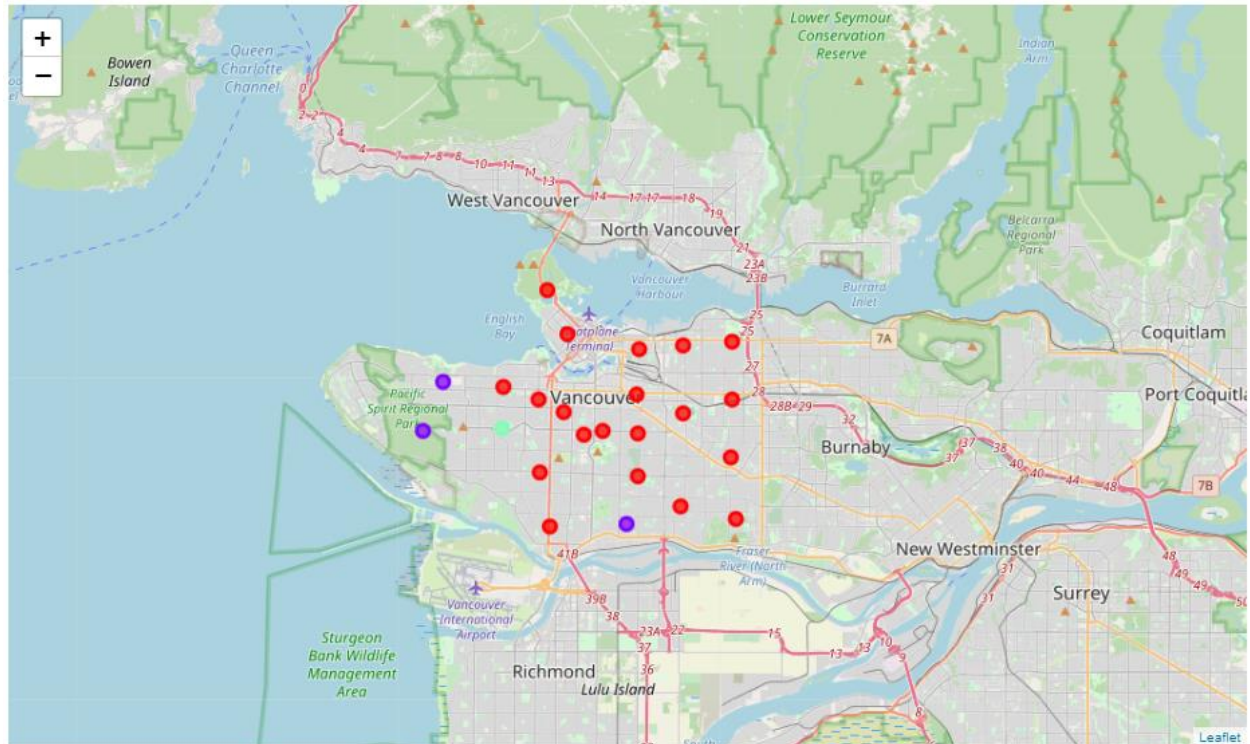
4. Results

The results show two separate maps of Vancouver. The first map, created with Folium, shows the property price per neighborhood and the venue clusters. Downtown is excluded from the analysis because of very high property prices. Neighborhood 'Dunbar Southlands' did not have data available. The darker the neighborhood, the higher the property price. Neighborhoods in the west have a higher property price. However, more venues with leisure activities are located in the west.
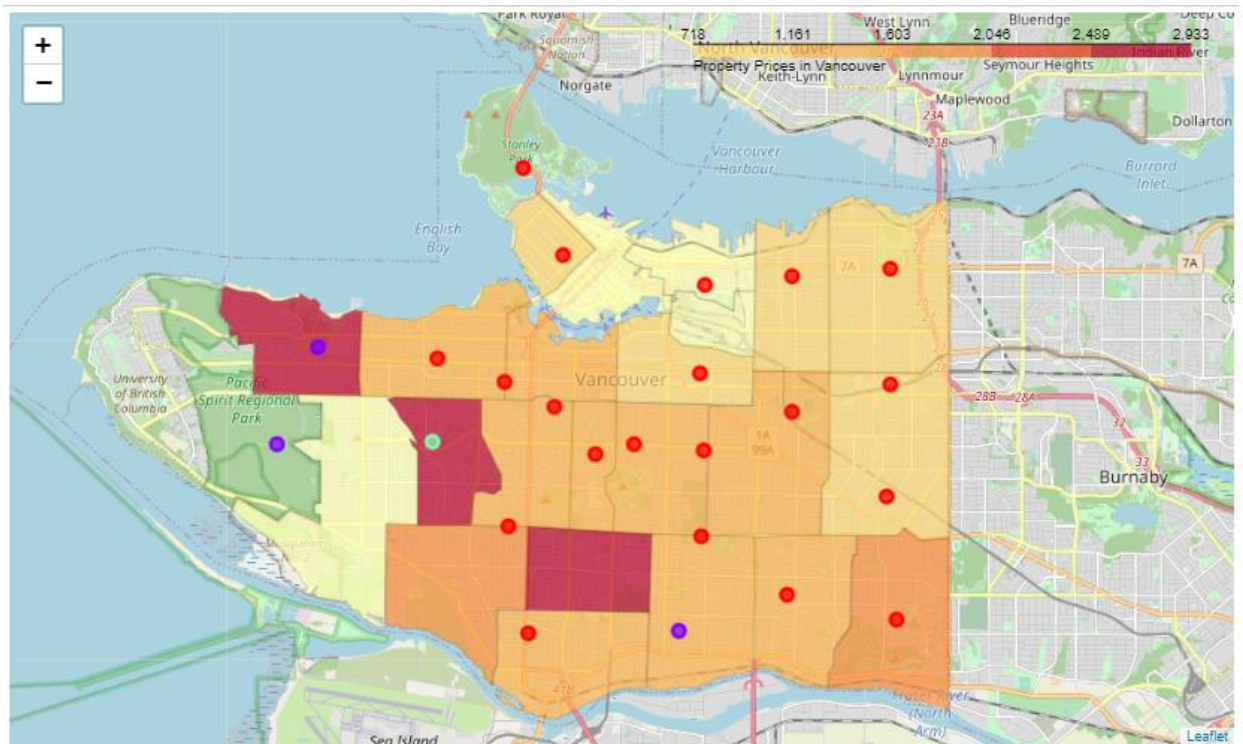
The second map displays the crime rate in each neighborhood in Vancouver along with the clusters. The darker the color of the neighborhood, the higher the crime number. As the map shows, the crime numbers are highest around Vancouver downtown. Neighborhood Renfrew-Collingwood has a higher crime number as well.
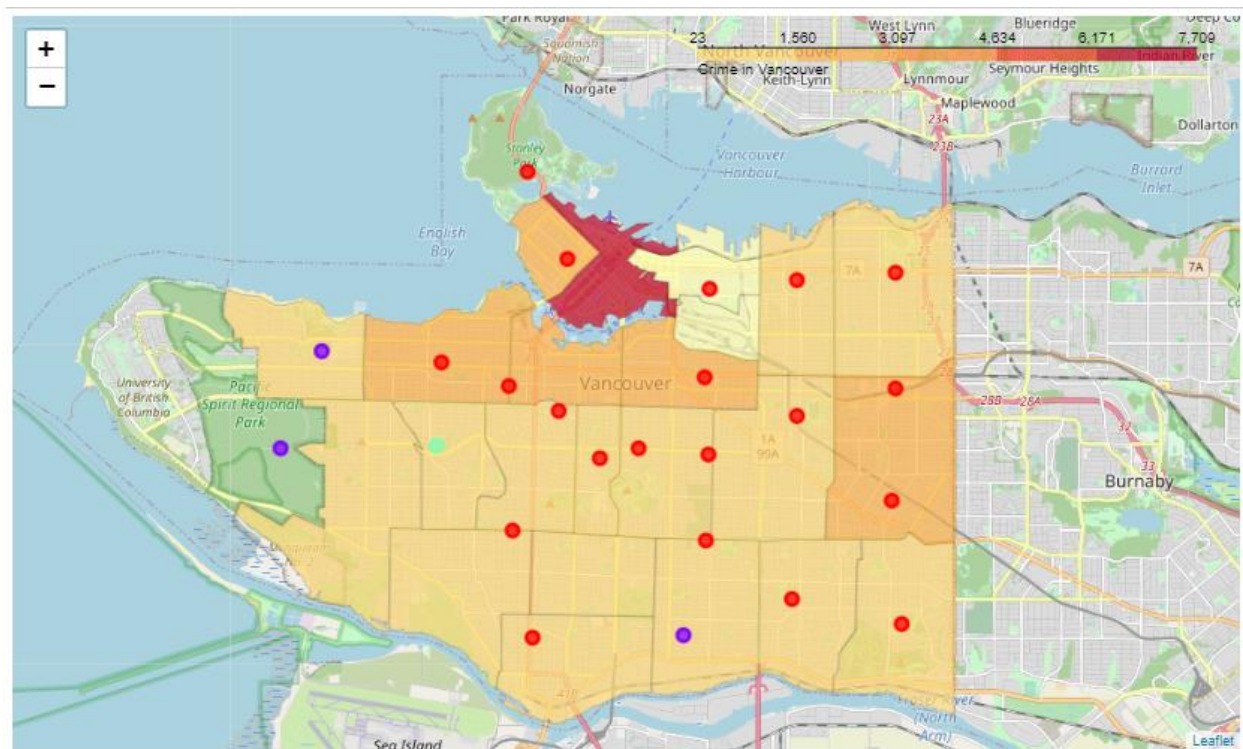
5. Discussion

Neighborhoods located more to the west have a higher property price. These neighborhoods have reasonably low crime rates and these neighborhoods are home to venues with more leisure activities. Due to the high property price, these neighborhoods might not be optimal to live.

A good option would be the South of Vancouver. Crime rates are lower, you have close proximity to venues with leisure activities, and prices are not as high as in the neighborhoods located in the west. Neighborhood 'Sunset' or neighborhood 'Marpole' could be an option to live in!

However, property prices in the north eastern part are lower. Crime rates in neighborhood Grandview-Woodland and Hastings-Sunrise are rather low compared to other neighborhoods centered around Vancouver Downtown. In these neighborhoods, property prices are also lower. The only drawback is that it is that the neighborhoods are not located near a cluster with leisure activities. Neighborhood Renfrew-Collingwood has lower property prices, but a reasonably high crime rate. Renfrew-Collingwood is less safe, and therefore, this neighborhood would not be optimal to live in.

Considering the crime numbers, property prices, and venues, I compare 4 optional neighborhoods to live in: neighborhood Marpole, neighborhood Grandview-Woodland, neighborhood Sunset, and neighborhood Hastings-Sunrise.
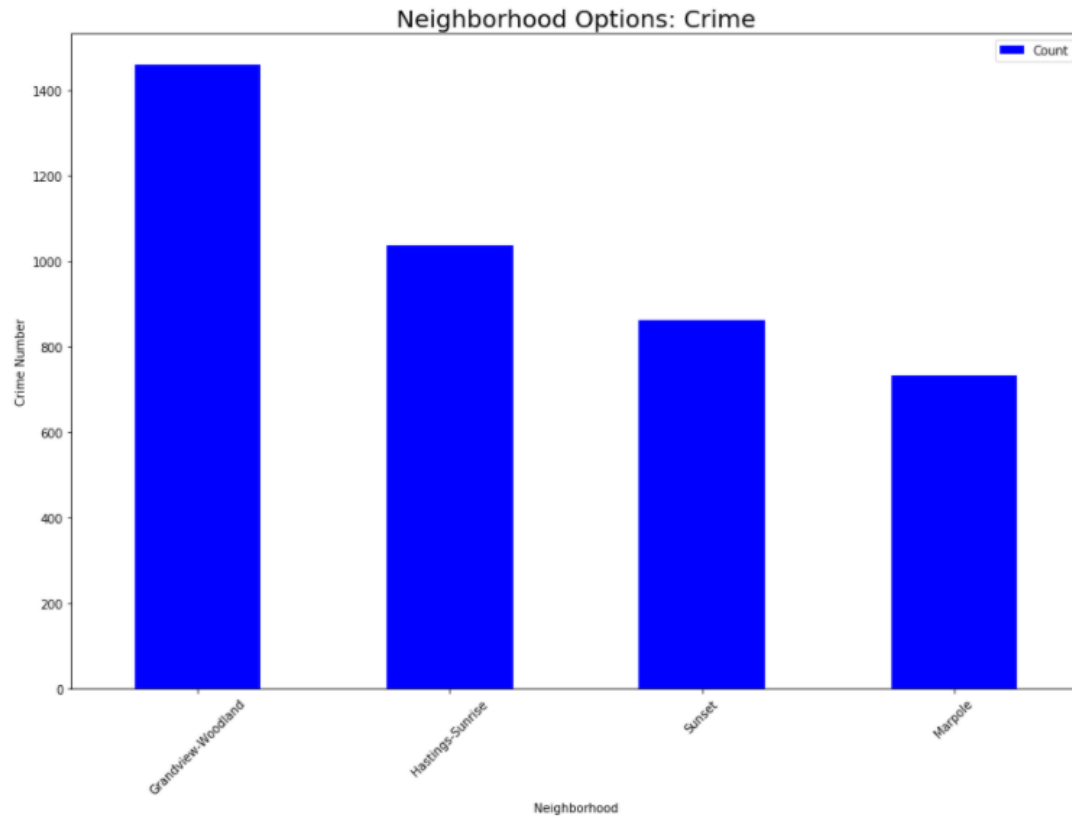
The Foursquare Venues data shows the top 5 most common venues for each neighborhood.

	Postal_Code	Neighborhood	State	City	Latitude	Longitude	Property Prices	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	V5K	Hastings-Sunrise	British Columbia	Vancouver	49.2807	-123.0397	1052.207532	0	Theme Park Ride / Attraction	Event Space	Theme Park	Amphitheater	Stadium
1	V5L	Grandview-Woodland	British Columbia	Vancouver	49.2795	-123.0667	746.260426	0	Café	Coffee Shop	Convenience Store	Fried Chicken Joint	Pizza Place
9	V5W	Sunset	British Columbia	Vancouver	49.2327	-123.0917	1195.994199	0	Bus Stop	Coffee Shop	Fried Chicken Joint	Dessert Shop	Sandwich Place
10	V5X	Marpole	British Columbia	Vancouver	49.2156	-123.0979	1353.811116	1	Electronics Store	Park	Yoga Studio	Cosmetics Shop	Dog Run

6. Conclusion

What is the best neighborhood to live in? I conclude that Neighborhood '*Sunset*' is the best neighborhood to live in. Compared to other neighborhoods, it has a relatively low crime rate. Neighborhood Marpole has a lower crime rate, but the property prices are quite high for this neighborhood. Neighborhood Sunset has some food venues nearby, there is a park nearby, and there is a bus stop nearby.