

# **Rent and Venues Data Analysis of Vancouver**

Anh Minh Hieu Nguyen

May 28<sup>th</sup>, 2021

## **1. Introduction**

Vancouver is a coastal seaport city on the mainland of British Columbia, Canada. Vancouver life is special because of its setting, surrounded by the sea and the mountains. It has a diverse, metropolitan area, yet you are always close to nature. In addition, Vancouver has a diverse population with nearly 40 percent of the people living in the city are immigrants. Therefore, it ranked second in the world's best cities to live in 2019 (based on Wiki) and second as the world's most expensive cities live in 2020 just after Hong Kong (based on Insider).

I recently moved to Vancouver from Toronto and the process to find a place to rent here took me the longest period of time so I decided to use Vancouver in my project. I think it is not just me but many other people having the same problem when they need to move to a brand-new city that they have never visited before. Depending on whether they have car or not, people will have different approaches and requirements. If the answer is yes, they can easily find a cheaper place in the outskirts area. On the other hand, they feel necessity for a place surrounded by many venues and public transportation is a must. However, it is difficult to obtain information that will guide newcomers to choose a suitable place.

When we consider all these problems, we can create a map and information chart showing the average rent in Vancouver and each area is clustered according to the venue's density.

## **2. Data acquisition and cleaning**

### **2.1 Data sources**

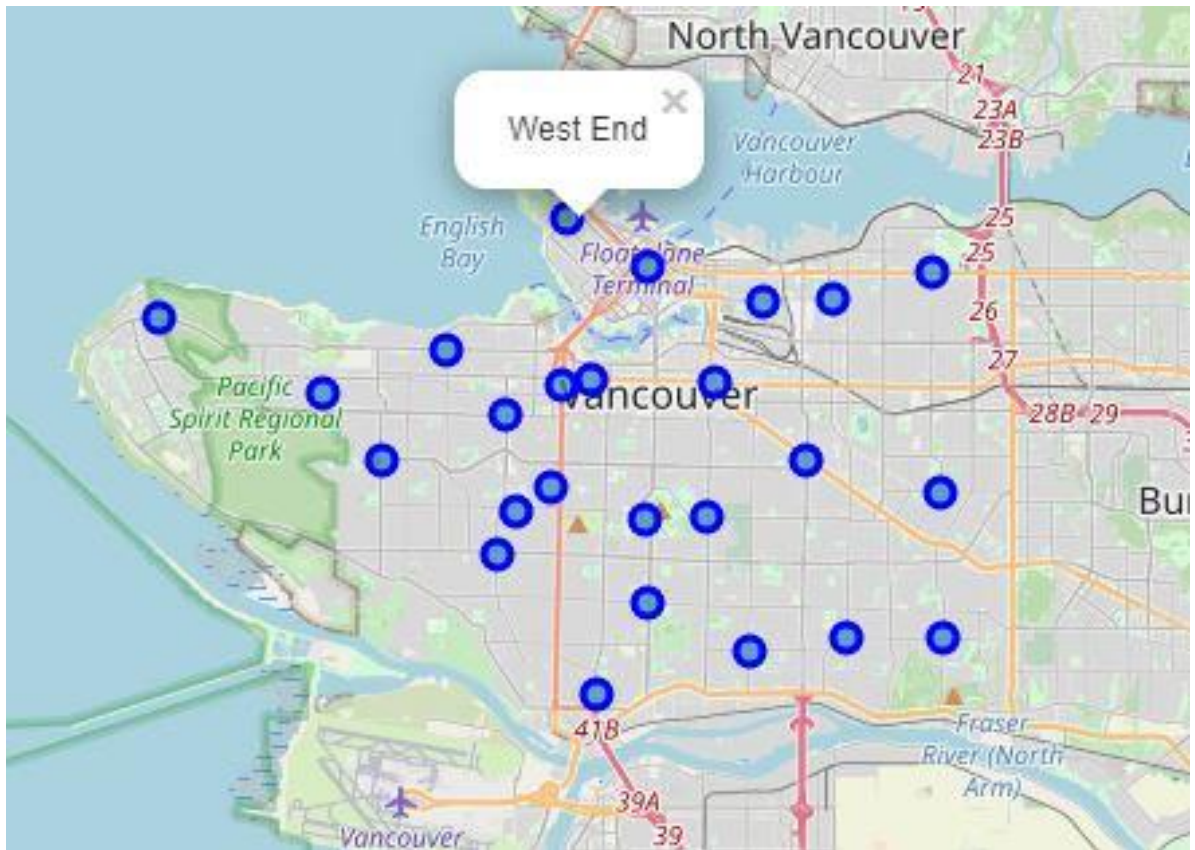
For this project, we will focus on the rent for **1 bedroom apartment**. Here is a list of data after considering the problem:

- I found the latest list of average rent for each neighborhood using [Zumper](#). It is one of the most reliable and easiest free Website to find a place to rent.
- I used **Foursquare API** to get the most common venues of the given neighborhoods.
- I used **Google Maps API geocoding** to get coordinates of each neighborhoods.

## 2.2 Data cleaning

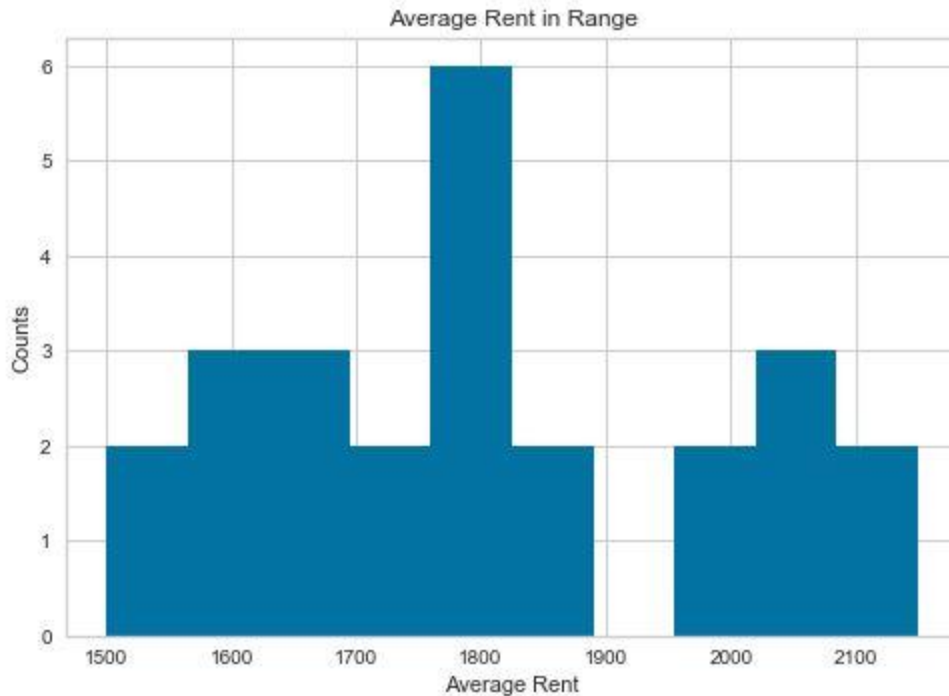
Data of rent and neighborhood was downloaded on [Zumper](#) and scraped using BeautifulSoup. Then, I used Google Maps API Geocoding to the coordinates Latitude and Longitude to merge them into one Data Frame. The rent can be variable using different Websites but I chose to use Zumper to get the latest version. Zumper is also one of the most reliable platforms for renters. Personally, I have used this Website for more than 4 years. However, there are several problems with the dataset.

First of all, Zumper includes most of the areas in Vancouver except Shaughnessy because it is mostly residential areas so there are not many places for rent to get the analysis data. To solve this problem, I searched other platforms and found only one place for rent and it is about \$1,800 for one bedroom apartment in a house so that is the number I used in the dataset. Secondly, there are 2 neighborhoods that does not have any venues around: University Endowment Lands and Sunset. UEL is an unincorporated area that lies to the west of the city of Vancouver and adjacent to the University of British Columbia and the lands associated with that campus. That explained why we can not find any venues within 500 meters. On the other hand, Sunset is the most ethnically diverse neighborhood in Vancouver, BC. It is located in south-east quadrant of the city. It is mainly big residential area and near the Fraser River. However, we can find venues if we increase the radius or change the coordinate, but for the unification with all other neighborhoods, we will exclude Sunset for now. Finally, as I mentioned above, the number of venues can be changed if we used different coordinates and increase the radius but Vancouver is quite a small city with 115 km<sup>2</sup> so I used 500 meters as the radius.

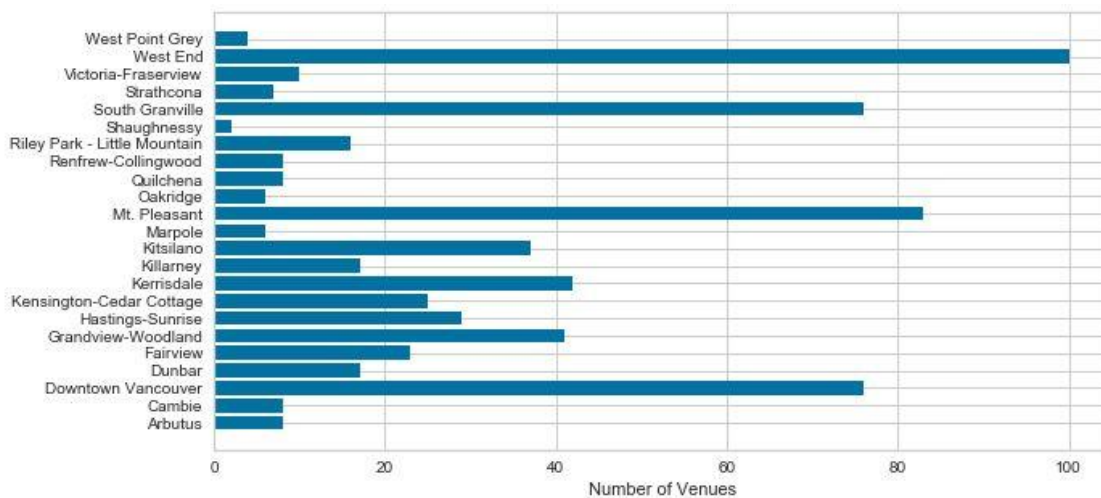


### 3. Methodology

In this project, we will limit our analysis to the main areas of Vancouver and the venues within 500 meters for each neighborhood. After collecting all the data including Neighborhoods, Rents, Latitudes and Longitudes for each neighborhood, I used Foursquare to get all venues within 500 meters. I set the limit of venues as 100 but none of the neighborhood exceeded the limit including their coordinates, however I did not use this information for this project but it is a nice-to-have for future extension or ideas.



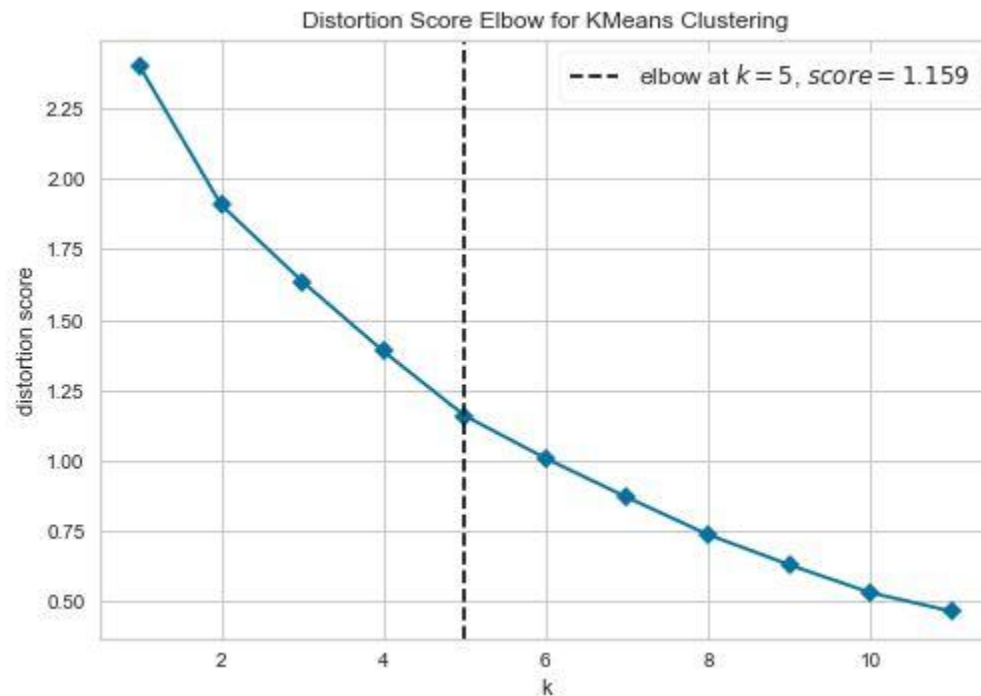
Next, I grouped all the venues together by the neighborhoods to get the number of venues for each of them so it is easier to see which neighborhood have the most number of venues. (West End, Mount Pleasant, South Granville and Downtown Vancouver)



I also grouped all the venues by venue categories to get top 10 most common venues for each neighborhood.

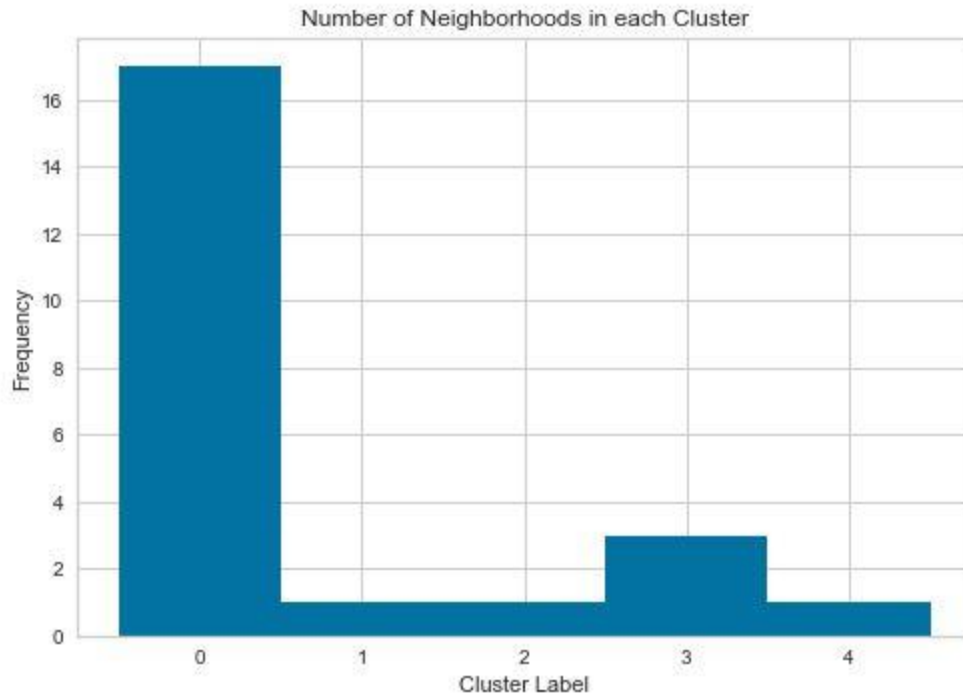
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arbutus	Grocery Store	Chinese Restaurant	Ice Cream Shop	Pet Store	Business Service	Café	Gym	Yoga Studio	Fish & Chips Shop	Fast Food Restaurant
1	Cambie	Garden	Park	Plaza	Tennis Court	Seafood Restaurant	Ethiopian Restaurant	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm
2	Downtown Vancouver	Hotel	Café	Seafood Restaurant	Restaurant	Coffee Shop	Concert Hall	Electronics Store	Clothing Store	Taco Place	New American Restaurant
3	Dunbar	Pharmacy	Sushi Restaurant	Bank	Grocery Store	Diner	Park	Pub	Coffee Shop	Sandwich Place	Café
4	Fairview	Coffee Shop	Park	Paper / Office Supplies Store	Sandwich Place	Korean Restaurant	Nail Salon	Japanese Restaurant	Camera Store	Restaurant	BBQ Joint

Based on this information, I created clusters of all neighborhoods. I used K-Means algorithm for this problem and the Elbow method to decide the number of clusters should be used which was 5.



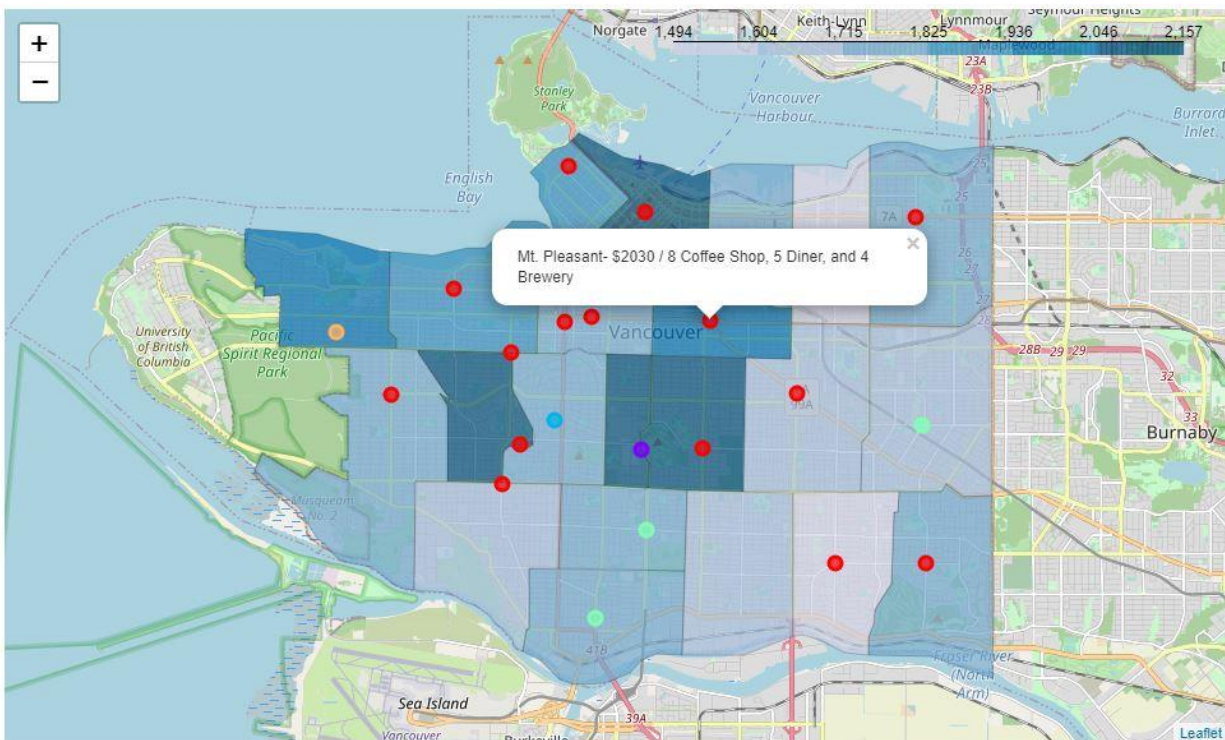
Now, we can see the number of neighborhoods in each cluster.





Finally, I presented Vancouver map including the average rent, top 3 venues and area's name for each neighborhood.

## 4. Results and Discussion



Our map shows that the multiple social venues are mostly near downtown. The rent in the middle area and west area are clearly more expensive than the east area.

As I mentioned before, we can try different approach by changing the coordinate and the radius for finding venue. However, Vancouver is not too big of a city to increase the range of radius so I used 500 meters which i think is good enough. After getting all the venues around the neighborhood within 500 meters, I grouped them together by the Neighborhood to choose out the 10th most common venues. I used K-Means algorithm as part of this clustering project and I also used Elbow method to find the optimum  $k = 5$ . I also tried to set  $k=2$  or  $k=3$  but West Point Grey was included in Cluster 0 which I do not think should be grouped together because it has a landscaping.

## **5. Conclusion**

As a result, with this information, it is easier for us to minimize the options in choosing neighborhood to look into for more details. Of course, the rent can vary and the venues are limited but they can help us to gain some basic information and what to expect for each areas when we are newcomers.