CMPT 353
Computational Data Science

OSM, Photos, and Tours

Summer 2020 Term
August 13th, 2020

Balijor Dhillon (301298366)
Nhi Le (301323763)
Sayyeda Mussa (301291381)

Project Overview

Open Street Map (OSM) provides open map data to websites, mobile applications, and hardware devices. It is built by a community of mappers that contribute and maintain worldwide data about roads, trails, cafés, railway stations, and much more. In this project, we were provided OSM data for the Vancouver area, and this report would like to do an analysis for the following questions:

1. Do some parts of Vancouver have more Cafe's than others? And are they influenced by restaurant locations?
2. If I was going to choose a hotel (or AirBnb), where should it be? What places have good amenities nearby?
3. Combining question 2 and 3, which winning airbnb is closest to most cafes in Downtown Vancouver? Do we have a winning Airbnb location that tourists must choose to get the best deal?!
4. If I was planning a tour of the city (by walking), where should I go? Are there paths that take me past an interesting variety of things?

This report will consist of different sections outlining each question we would like to look into. Likewise, in each section, we will talk about the data, techniques we used to clean and analyze the data, results we found, and limitations we encountered.

The Data

OSM provides the data in a XML format, and Professor Baker turned this file into a more usable JSON data form: `amenities-vancouver.json.gz.` This dataset has fields for latitude, longitude, amenity type, the name and a tag column (with other useful metadata). Likewise, the following sections will go more into detail about which parts of the dataset was used and what techniques assisted in cleaning and analyzing the data.

Section 1: Cafe Distribution

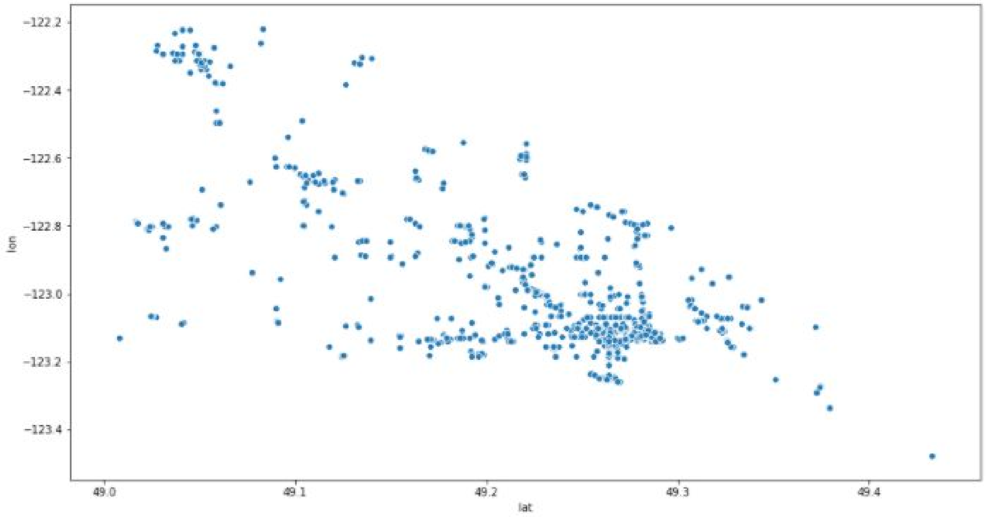
Question Definition:

We wanted to know what the distribution of Cafes were around Vancouver and in the neighbouring cities. Similarly, if some cities have more, is it due to the presence of restaurants around them - are Cafe owners locating their business on a busy street? In order to solve this problem, we did a visual representation of the distribution, conducted a cluster analysis to identify a pattern, and calculated mean and standard deviation of latitude and longitudes to see how far apart are the restaurants and cafes.

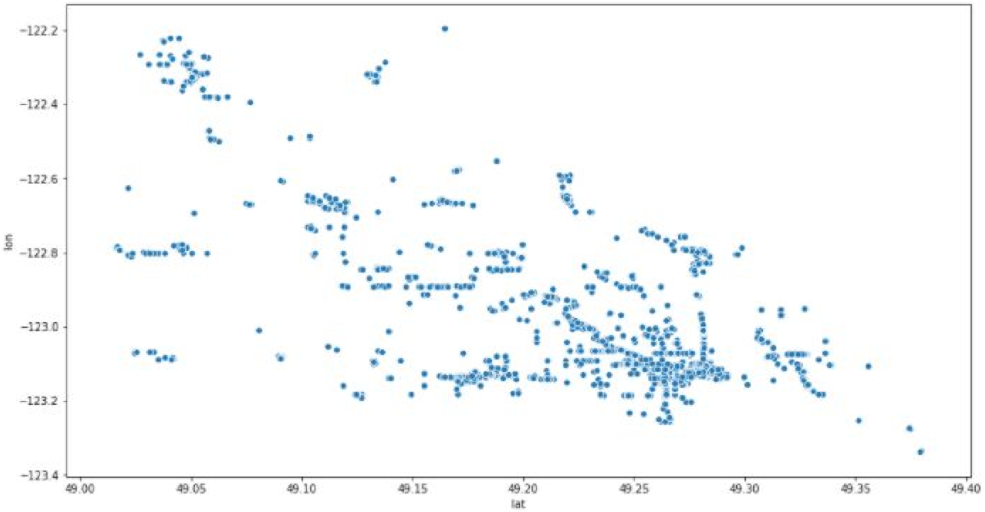
Data analysis/Techniques:

We first read in the big amenities-vancouver.json.gz file and apply a function that extracts all the Cafes, and we drop any missing values.

We proceed to visualize **cafe** longitude and latitude on a scatter plot:



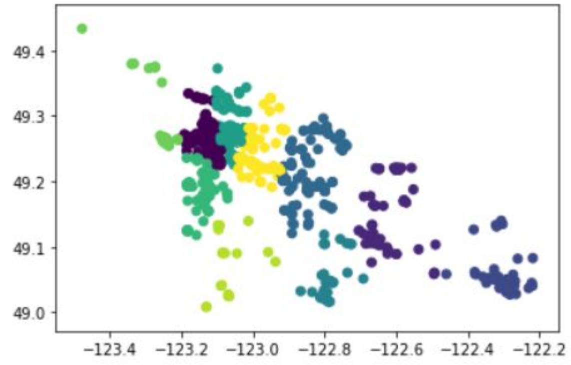
Then, we proceeded to visualize **restaurant** longitude and latitude on a scatter plot:



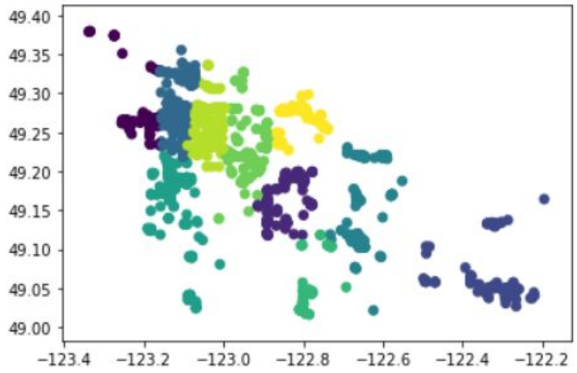
Observing the above two plots, we found that the distribution **for both** was similar in shape and that most points are gathering around the 49.20 - 49.30 latitude range and the (-123.0) - (-123.4) longitude range.

Next, we conducted a K-means clustering based on the longitude and latitude for both amenities:

Cafes:



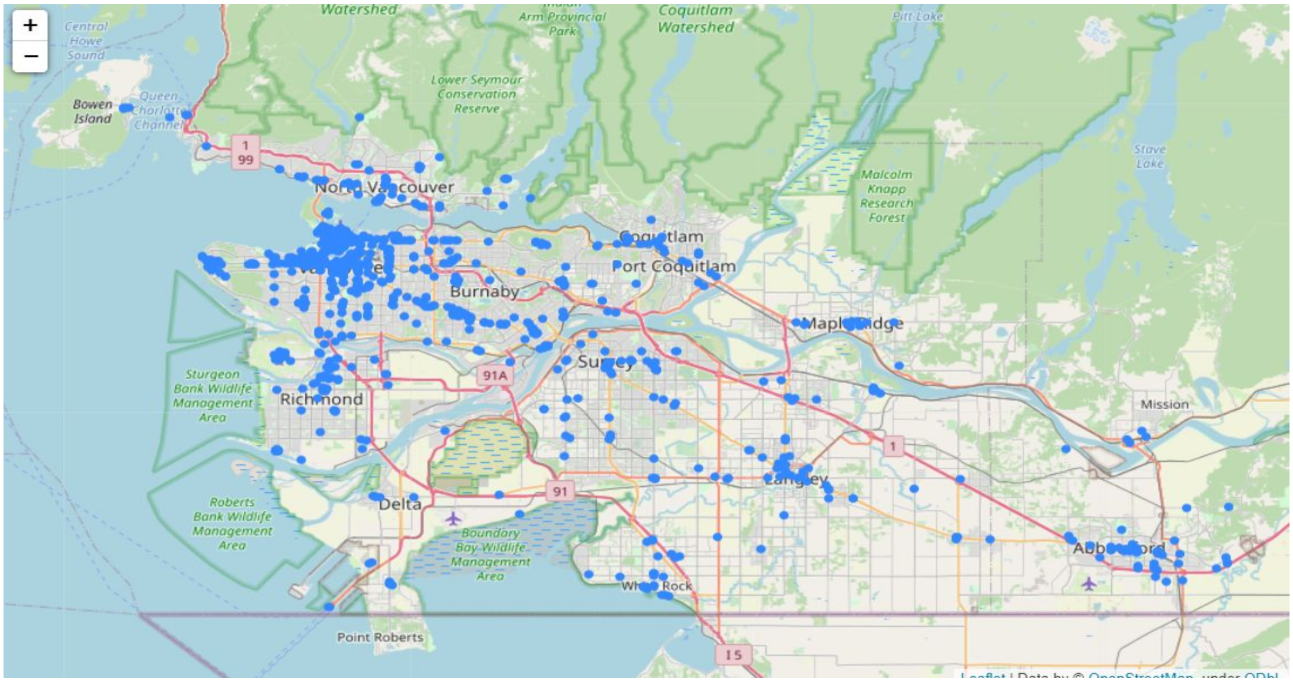
Restaurants:



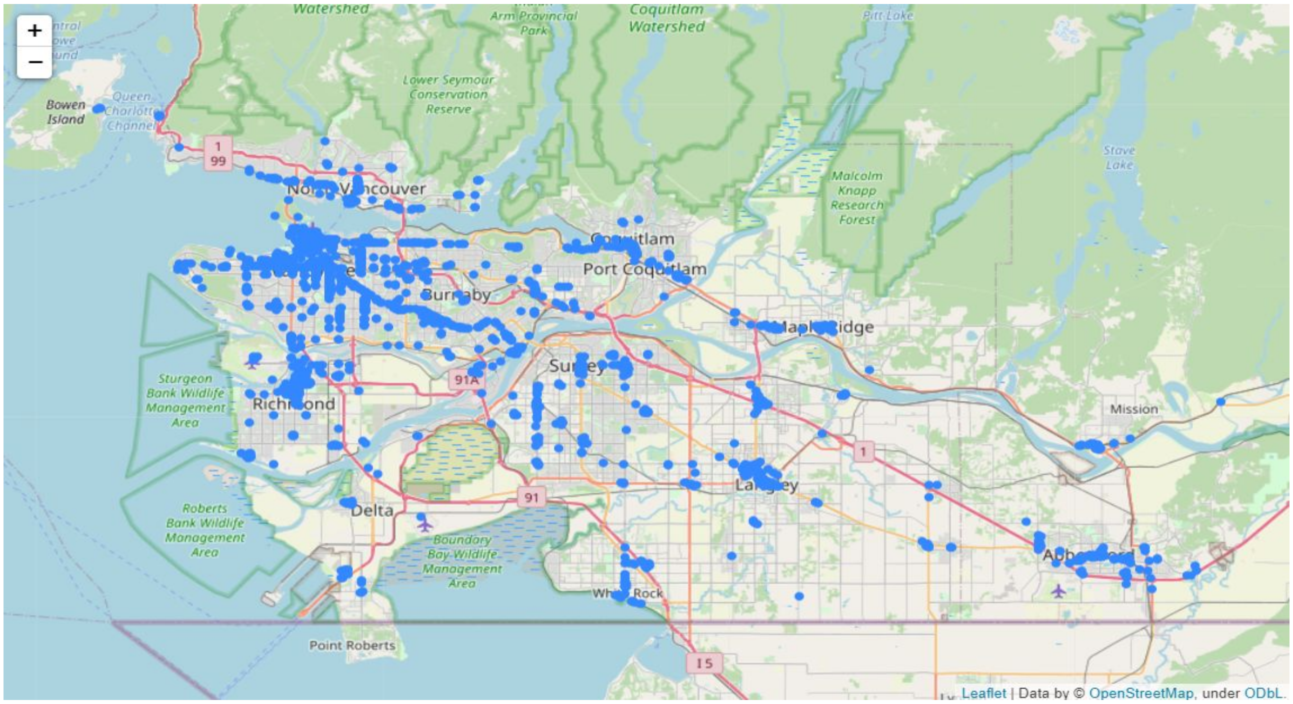
A cluster refers to a collection of data points aggregated together because of certain similarities - again this shows that the distribution is indeed categorized based on the longitude and latitude.

To further visualize the distribution and to be sure that which city had the most, we used the folium package to plot the distribution of both amenities:

Cafes:

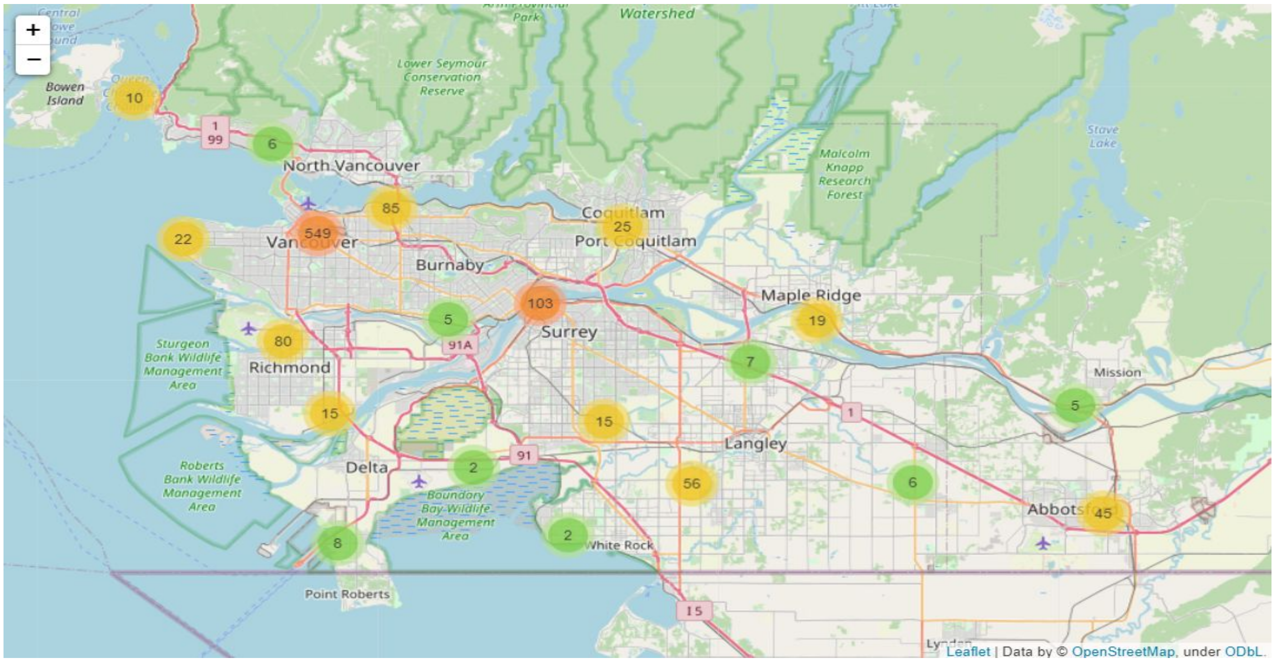


Restaurants:

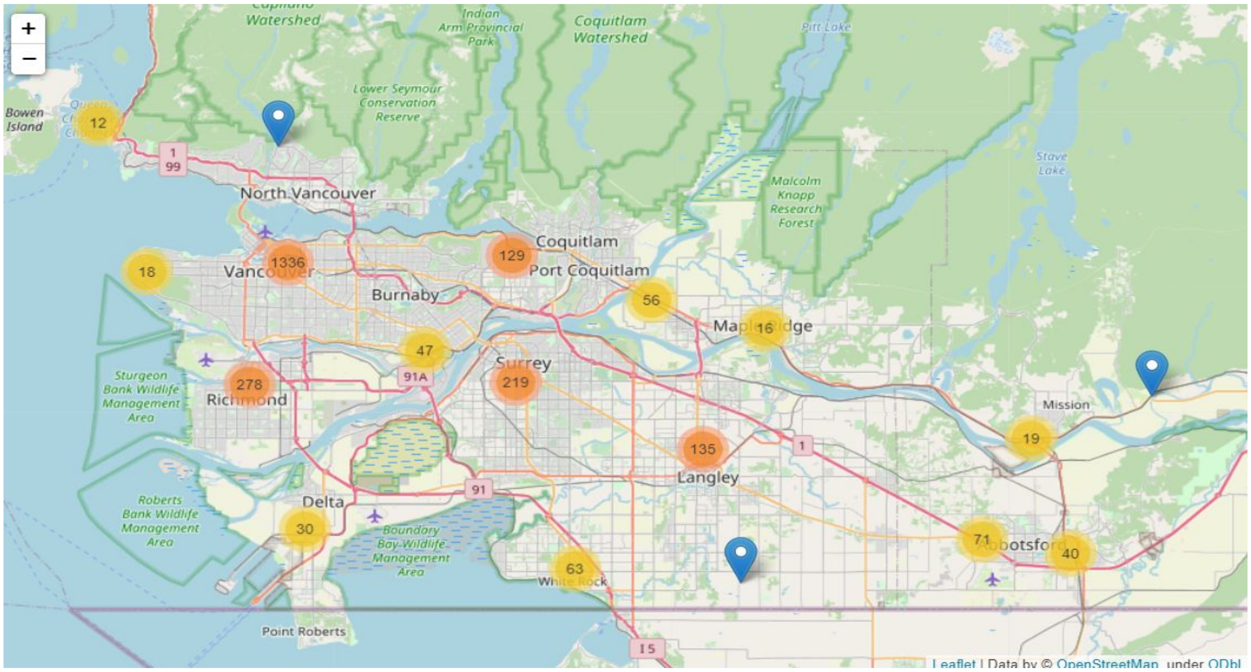

Watershed
Arm Provincial Park
Coquitlam Watershed
Pitt Lake
Bowen Island
Queen Charlotte Channel
1
99
Lower Seymour Conservation Reserve
North Vancouver
Malcolm Knapp Research Forest
Stave Lake
Port Coquitlam
Burnaby
Maple Ridge
Surrey
91A
Sturgeon Bank Wildlife Management Area
Richmond
Mission
Roberts Bank Wildlife Management Area
Delta
91
Langley
Boundary Bay Wildlife Management Area
White Rock
Point Roberts
15
Leaflet | Data by © OpenStreetMap, under ODbL.

Looking at the 2 above maps, both cafes and restaurants seem to have more locations in the downtown Vancouver area (close to english bay and coal harbour area). Similarly, to further go into the gathering pattern, we did a cluster on these two amenities:

Cafes:


Watershed
Bowen Island
10
1
99
6
Lower Seymour Conservation Reserve
North Vancouver
Malcolm Knapp Research Forest
Stave Lake
85
Coquitlam
549
22
Vancouver
25
Port Coquitlam
Burnaby
Maple Ridge
103
5
19
Surrey
80
91A
Sturgeon Bank Wildlife Management Area
Richmond
7
Mission
15
5
15
1
Langley
Roberts Bank Wildlife Management Area
Delta
2
91
56
6
Boundary Bay Wildlife Management Area
45
Abbotsford
8
2
White Rock
Point Roberts
15
Leaflet | Data by © OpenStreetMap, under ODbL.

Restaurants:



We can see from these cluster plots that both cafes and restaurants are mostly located in Vancouver. More than half of cafe locations (count of 549) are in downtown vancouver, and the same goes for restaurants (count of 1336).

Lastly, we calculated the longitude and latitude means and standard deviations for both amenities:

	Longitude mean	**Latitude mean**	**Longitude sd**	**Latitude sd**
Cafes	-123.001527	49.229386	0.217547	0.076407
Restaurants	-122.981881	49.220426	0.215242	0.074321

Longitude/ latitude means and standard deviation for both amenities seem to be close, indicating that cafes and restaurants are not too scattered apart.

Conclusion:

- Both cafes and restaurants are located more in downtown vancouver compared to other cities

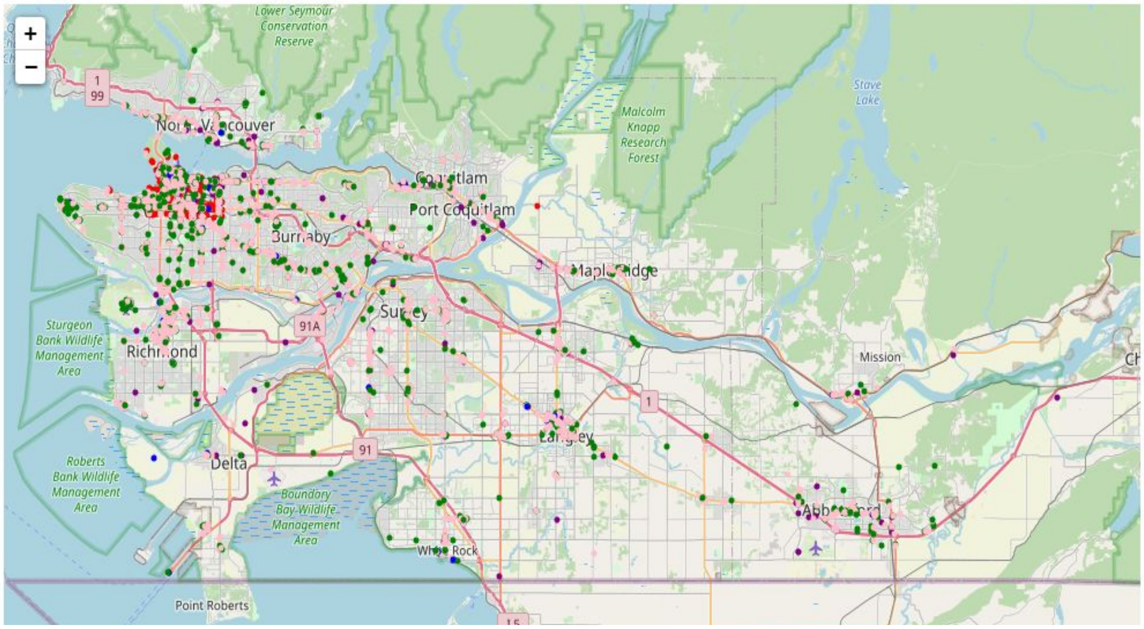
- Cluster analysis reveals the count of each amenity to be higher in the downtown vancouver area
- Statistical data of means and standard deviation on longitude and latitude features reveal that these two amenities are close in proximity
- Hence, we can say that cafe locations are influenced by restaurant locations, and that cafe owners are locating their business on busy roads - potentially due to other factors as well.

Limitations:
If we had more time, we would have loved to perhaps parse the tag column to get opening hours for each cafe and see which cities had cafes with longer opening hours. Furthermore, are the longer opening hours influenced by restaurant opening hours around these cafes? However, in the following section, we add onto the cafe distribution problem by looking at which airbnb is closest to most cafes in Downtown Vancouver? (we chose downtown because it had the highest count of cafes compared to cities like Surrey, Coquitlam, and Richmond).

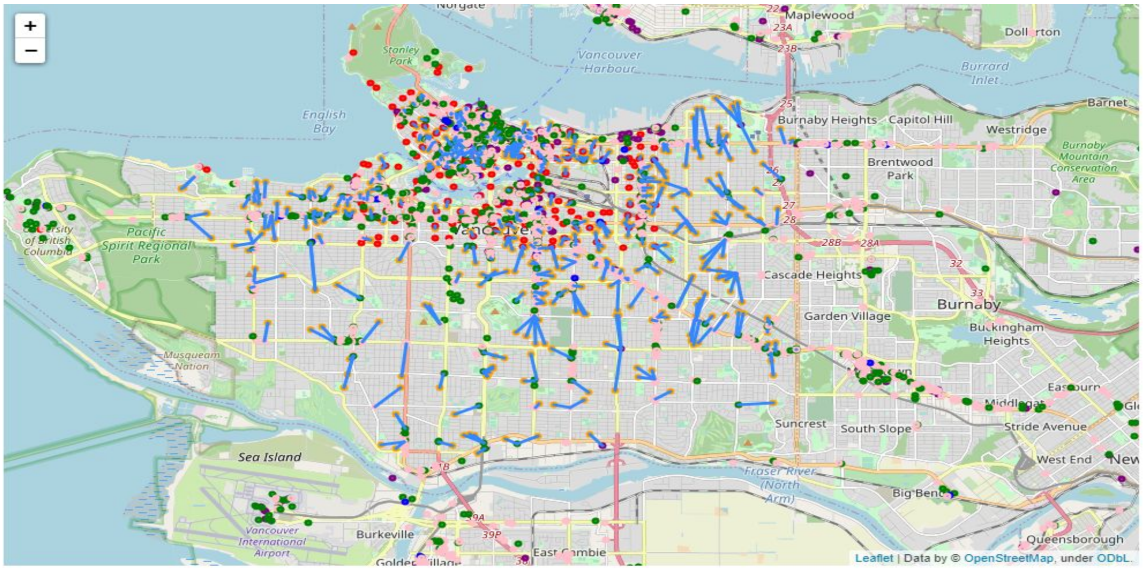
Section 2: Hotel Recommendation + Cafe Proximity

Question Definition:
If someone was to choose a hotel where would it be? And what are some good amenities? To answer this question we must ask what are good amenities? I consider good amenities that tourists would often look for such as outdoor activities, nightlife, and food. Outdoor Activities include bicycle rentals, boat rentals, and parks.Nightlife activities include bars, lounges, nightclubs, stripclubs, pubs,internet cafes, and casinos. Food amenities include bbq, restaurants, ice cream stores, bistro, and cafes. Likewise, keeping in mind that coffee is a big staple in most individuals' lives (especially if you are a tourist), we wanted to further look into which Airbnb in Vancouver has the most cafes near it?

Data analysis/Techniques:
We used the folium package to visualize the data to see where most good amenities were. All nightlife activities are purple dots, outdoor activities are red dots, restaurants are pink dots, cafes are green dots, and all other food amenities are blue dots. As we can see from the figure there are large clusters in the vancouver area, richmond, and langley. We further investigate the vancouver area because it has the most amenities. The Airbnb data has data only located for Vancouver.

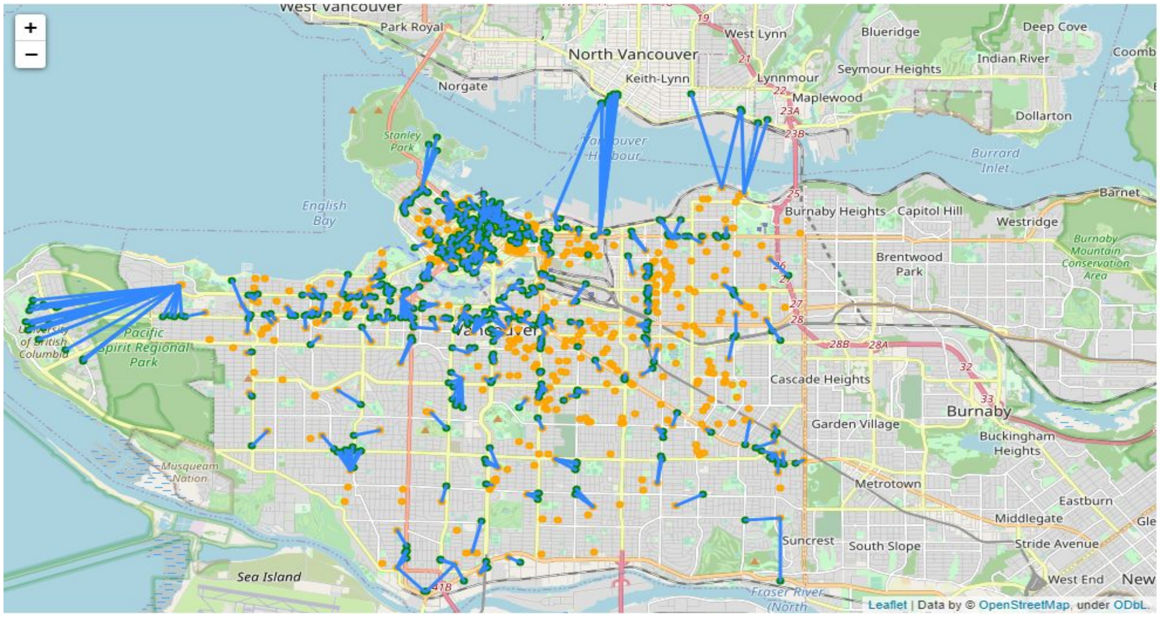

+
−
Lower Seymour Conservation Reserve
North Vancouver
Coquitlam
Port Coquitlam
Burnaby
Maple Ridge
Surrey
Richmond
Delta
Langley
Abbotsford
White Rock
Point Roberts
Stave Lake
Malcolm Knapp Research Forest
Mission
Sturgeon Bank Wildlife Management Area
Roberts Bank Wildlife Management Area
Boundary Bay Wildlife Management Area
1
99
91A
91
15

From the data we used the nearest proximity technique ,using the shapely package and used geopandas for geospatial work, (outlined in project_code notebook) to find the nearest amenities to each Airbnb (Aribnbs are orange color). The Airbnb that were considered had to have at least 90 reviews (the latest review was 2020-01-02, and rating of 80 or higher). With the latest review we know that Airbnb has been used recently, and a rating of 80 or higher is a good rating for a Airbnb place.With at least 90 reviews we know that Airbnb has enough reviews to give a proper average rating.

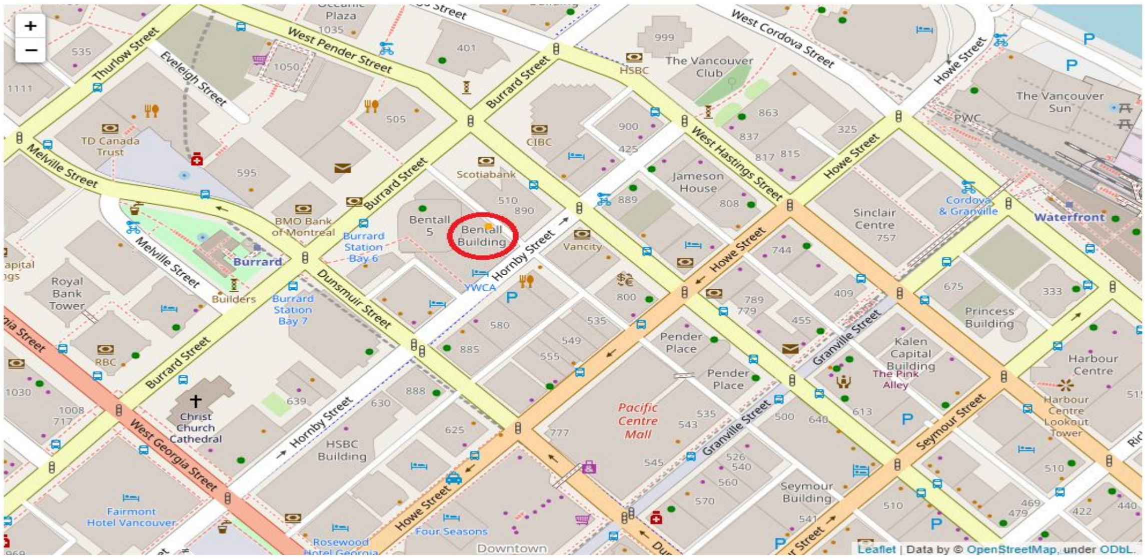


From this we can see that downtown has the most amenities and Airbnb in the downtown area is closest to every type of amenities. The more we move away from downtown are, the diversity of amenities decrease and only restaurants and cafes are left. Hence, people should select a hotel or Airbnb in the downtown area because it has a widest diversity of amenities.

Finally, to find the Airbnb with the most cafes around it we took the cafes and calculated the nearest Airbnb.



Then added the nearest Airbnbs to a dataframe and used the group by function to count the number of cafes that are near to Airbnb. Afterwards, sorted the data frame to show the Airbnb with the most cafes near it and took that id and filtered out the original Airbnb data frame. As a result, the Airbnb with the most cafes near it is located between Burrard Street and Hornby Street. The listing of this Airbnb is called “Private Hotel-Style Suite in a Downtown Boutique High-Rise”.



Conclusion:

The downtown area has the most amenities, and has the most diversity of amenities. Airbnbs that are located outside downtown have longer distances from amenities, also there are less nightlife and outdoor amenities located outside of downtown.
The Airbnb with the most cafes near it is located between Burrard Street and Hornby Street. The listing of this Airbnb is called “Private Hotel-Style Suite in a Downtown Boutique High-Rise”.

Limitations:

We do not have any airbnb data for other cities besides Vancouver. Hence, when looking into the airbnb with the most cafes around it, we would have loved to pick the top rated airbnb in each city such as Richmond, Delta, Surrey or Coquitlam, and inform our readers how many cafes were around it.

Section 3: City Tour/Route Recommendations

Question Definition:

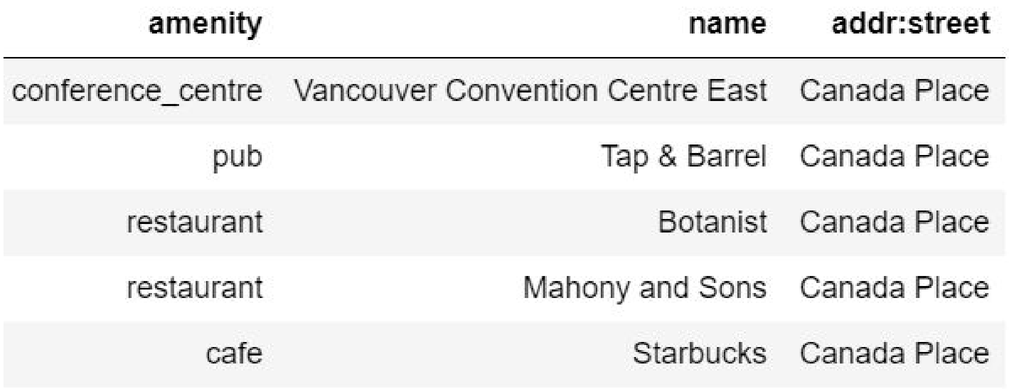
If I was planning a tour of the city by walking, where should I go? Are there paths that take me past an interesting variety of things? Below will be some snippets that will help tourists when they plan to come to Vancouver BC.

Data analysis/Techniques:

Since the amenities-vancouver.json data is too complete it has many unnecessary amenities recorded, the amenities are grouped into categories so that onew we want to look into/or are interesting to tourists can be extracted with ease. For example, fast food, restaurants and bistros belong to the ‘food’ category, and clock, marketplace and conference_centre fall into the ‘interesting places’ category.

Since tags are stored under dictionary, the variable is splitted into multiple columns using pd.Series and filtered so that the most used (>1000 records) tags are saved for further usage.

Name and tags are stored in a way that sometimes it is not easy to actually find a place with a name we always know. Having Canada Place as an example:



amenity	name	addr:street
conference_centre	Vancouver Convention Centre East	Canada Place
pub	Tap & Barrel	Canada Place
restaurant	Botanist	Canada Place
restaurant	Mahony and Sons	Canada Place
cafe	Starbucks	Canada Place

From the figure, we can see that the keyword ‘Canada Place’ is shown on the tag ‘addr:street’ with 5 resulting records and the ‘Canada Place’ that we informally refer to is the Vancouver Convention Centre East.

Since the ‘interesting places’ category consists of only 56 candidates, we adapted 2 additional datasets including Parks from City of Vancouver Open Data Portal and the Visitor Centres Listing from HelloBC.

We focus on travelling around downtown Vancouver by walking. As skytrain stations are the centers of transportation around Metro Vancouver, we selected the 4 skytrain stations downtown including: Waterfront, Burrard, Granville and Stadium - Chinatown stations as the initial point.

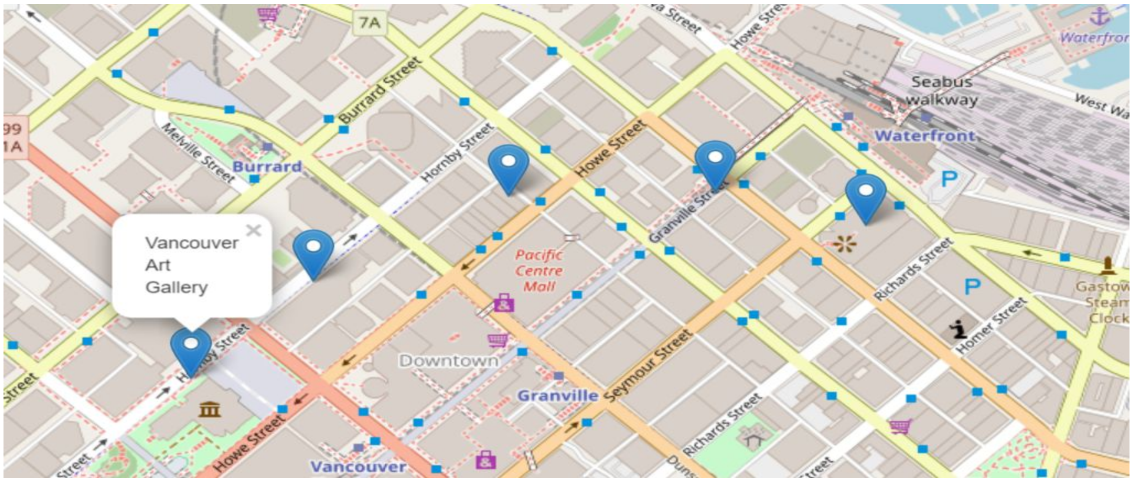
We deployed the Shapely library to find the first nearest places to the 4 stations. The next step involves finding the next nearest places from the output recommended places. The findings are implemented and added 4 times with the resulting 5 recommended points from each station.

Lastly , we visualized the locations and added name tags to maps to make the place identifiable for the tourists to see, using Folium library.

Results/Conclusion:

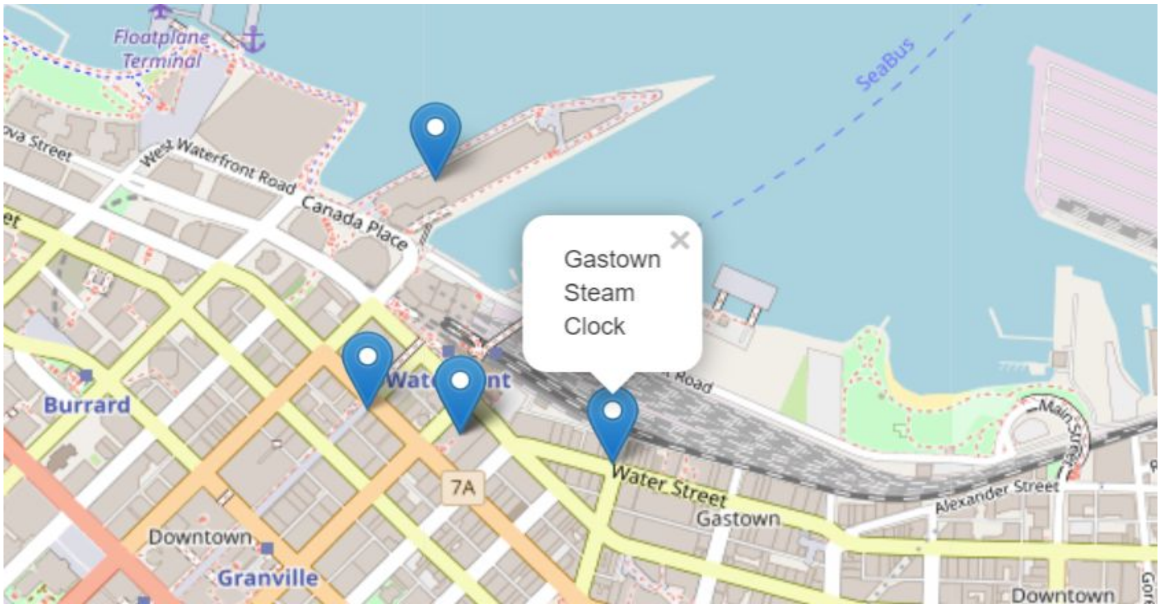
The resulting recommendations from Burrard and Graville stations are identical with the route consisting of:

Vancouver Art Gallery → Bill Reid Gallery of Northwest Coast Art → LeSoleil Fine Art Gallery → Birks Clock → Vancouver City Passport



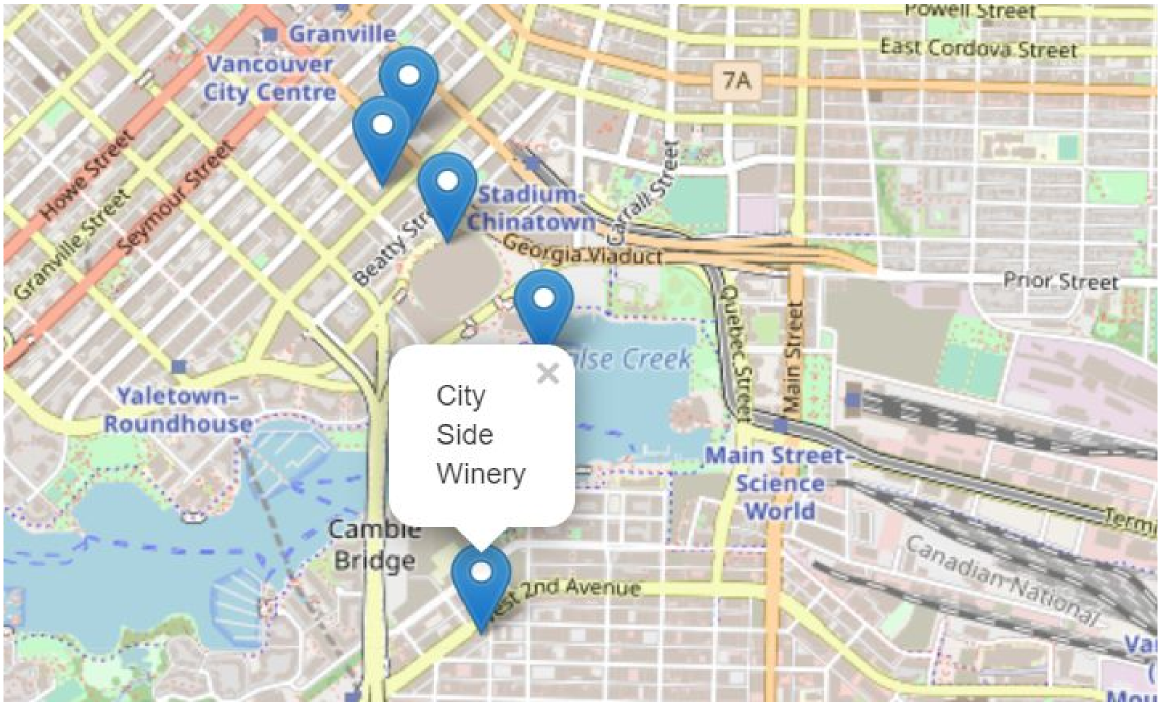
<u>From Waterfront Station:</u>

Vancouver Convention Centre East → Birks Clock → Vancouver Lookout → Gastown Steam Clock



<u>From Stadium - Chinatown station:</u>

Downtown Farmers Market → The Post at 750→ BC Sports Hall of Fame → Harbour Convention Centre → City Side Winery



<u>Limitation</u>:
The route plan with a real geological path at the time since I was not able to find an appropriate library. Additionally, the visualization should need geotagged pictures attached.

Project Experience Summary

Balijor Dhillon

OSM Data - CMPT 353 Final Project **July 2020 - August 2020**

- Utilized folium package to add details to map visualization
- Used pandas to group the multiple amenities into different classes
- Cleaned airbnb and osm data and used neighbouring technique to see which amenities were closeby
- Designed and organized report to ensure analysis of each question was explained thoroughly with visualization

Nhi Le

OSM Data - CMPT 353 Final Project **July 2020 - August 2020**

- Utilized the folium package to ensure route recommendation visuals were clear and informative
- Utilized shapely package to assist with map visualization and proximity techniques tp plan city tour route recommendation
- Adapted and cleaned 2 additional datasets including Parks from City of Vancouver Open Data Portal and the Visitor Centres Listing from HelloBC
- Designed and organized report to ensure analysis of each question was explained thoroughly with visualization

Sayyeda Mussa

OSM Data - CMPT 353 Final Project **July 2020 - August 2020**

- Optimized workflow by splitting work efficiently and assisting with question ideation
- Conducted data visualization for cafe and restaurant distribution using folium package
- Applied clustering techniques to further visualize distribution
- Designed and organized report to ensure analysis of each question was explained thoroughly with visualization