Part 2:

In this part, we are going to help clients to plan a tour of the city by driving.

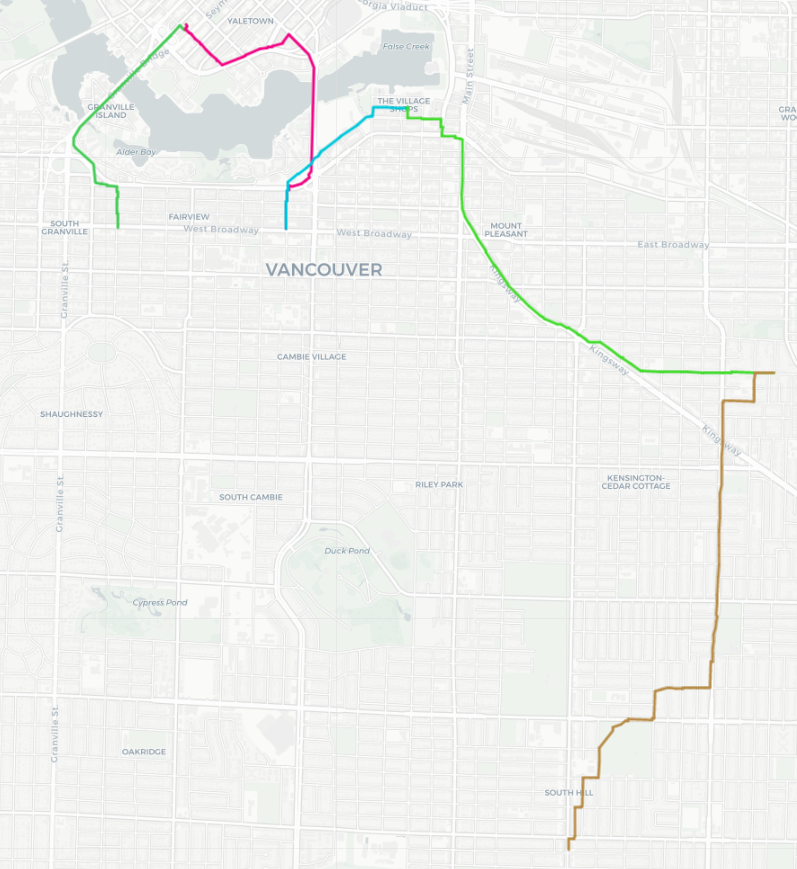
1. Data we use: wiki\_fliter.py

First, we categorized the whole data set into food, facility, service, education, medical, entertainment and transportation. Also, we categorized the set into different cities, like Vancouver, Burnaby and so one. For our example, we assume that the client wants a tour in Vancouver. So, we only use the records which city is equal to Vancouver. Because we need to find some famous places, we need to use Wikipedia data set. Some records contain the Wikipedia entity, some records not. We only use records which contains Wikipedia entity ID.

1. Techniques I used to analyse the data: tour\_plan.py

To check a place is whether famous or not, we are going to check the Wikipedia entity completeness by using the entity ID. So, we need to use wikidata package. We got this idea from the project instruction. We found that the Wikipedia entity more complete, the length of the entity larger. We use member functions of wikidata package to get the length of all entity first. Then, if we have we group the data set by categories. We found the max entity length for each category. We thought we were done. But we found that the famous food is Starbucks. We categorized the cafe into food. We noticed that the chain store like McDonand's or Starbucks got more completed Wikipedia entity. We think our client don’t come to Vancouver for eating fast food. We think they come to Vancouver for local cuisines. So, we excluded record which shows in the data set more than ten times. Then, we got various famous places which are not chain stores. Then, we use the distance function comes from assignment3 to get a shortest path between these famous place. We used osmnx and networkx packages to draw the map.

1. Path Picture:



The intersection of different color lines is the famous place.

1. Limitation:

Our map is too simple, we will include more information in the map, like the address of each point. The map just showing the path. Our clients may unable to find the correct place.

Part 3:

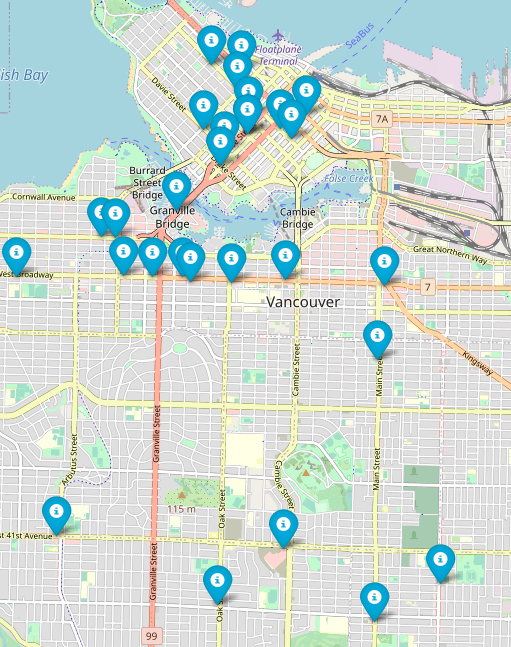
In this part, we are going to help clients to choose a hotel or AirBnb.

1. Data we use:

We assume that the client wants to choose a hotel or AirBnb in Vancouver. We categorized the set into different cities. We only use data points which locate in Vancouver. We assume that a good hotel or airBnb is surrounded by many restaurants. So, we excluded all data records which are not restaurant.

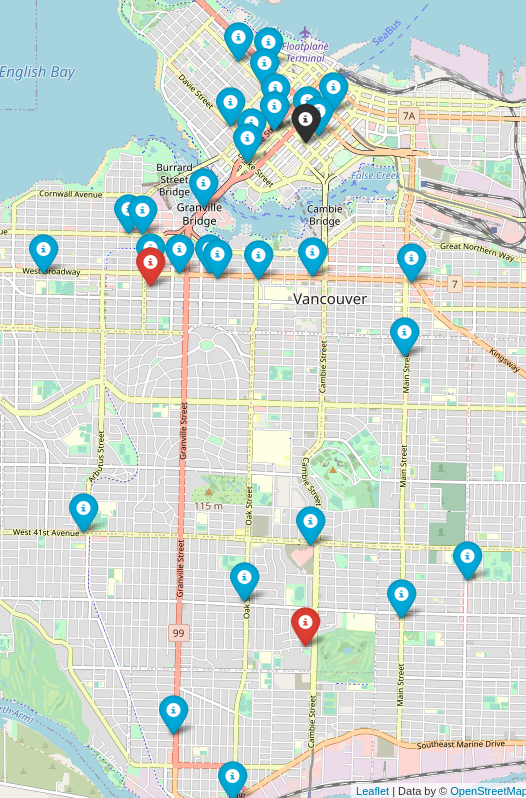
1. Techniques I used to analyse the data:

When we were going to analyze the data set, we found that there are no hotel or airBnb data point in the data set. We thought the data set contains those point. So, we change our goal of this section. Instead finding a actual hotel of airBnb, we planed to find a location which is surrounded by many restaurants. We created an map and show the all restaurants on the map first.



After seeing the map, an idea came to our mind is clustering all the data point into several groups. The center point of each group is surrounded by many restaurants. We used the K-Means algorithm to cluster those points. In K-Means algorithm, choose a appropriate k value is very important. By doing some research on Google, we found the elbow method. Then, we used the elbow method to get the k value. We used the data set and the k value to train the K-Means model.

1. Result we got:



We got three new points. One of them is black. Two of them are red. These three points are centers points of three different cluster group. We recommend the black point to our clients because the black points is the center point of the largest cluster group.

1. Limitation:

We can’t help our client to find a actual hotel or airBnb. We can only help them to find a good location. A hotel or airBnb nearby this location is a good choice.

Accomplish Statement of Yangxin Ma:

I created a tour path of a city and I found some good locations of hotels or airBnb. For creating a path on the map, I learned some knowledge of using Folium, Osmnx, and Networkx packages to create a map and a path by searching on Google. For clustering data points, I learned how to use the sklearn package which is very useful. The sklearn package contains many models. My tasks is distributed by the project leader. I learned how to work in a group. A group project is not like a individual project. Our task is relative. I used some functions from his code. This project is a great experience for me to learn how to cooperate with other people.