**Capstone Project - the possible covid-19 outbreak districts in Seattle**

**1. Introduction**

The covid-19 Virus has become a major problem in the world, it is spreading at a fast speed, and can be transmitted thought human to human contact, hence, it is important to find district with lots of places that require human to human contact, such as restaurants. The study is done at a place that I used to live: Seattle.

**2. Data**

The analysis to find the density of different venues in different district in Seattle:

The sources of data are the following to achieve their respective aims:

* I get the information about the district in seattle through the websitle <https://seattle.findwell.com/seattle-neighborhoods/>
* I look up the latitudes and longitudes of the districts through the OpenCage Geocoder API
* I get the location information thought the Foursquare API.

for the sake of simplicity, I copied the information on the website and then pasted it to my excel file, and saved it as the csv format.

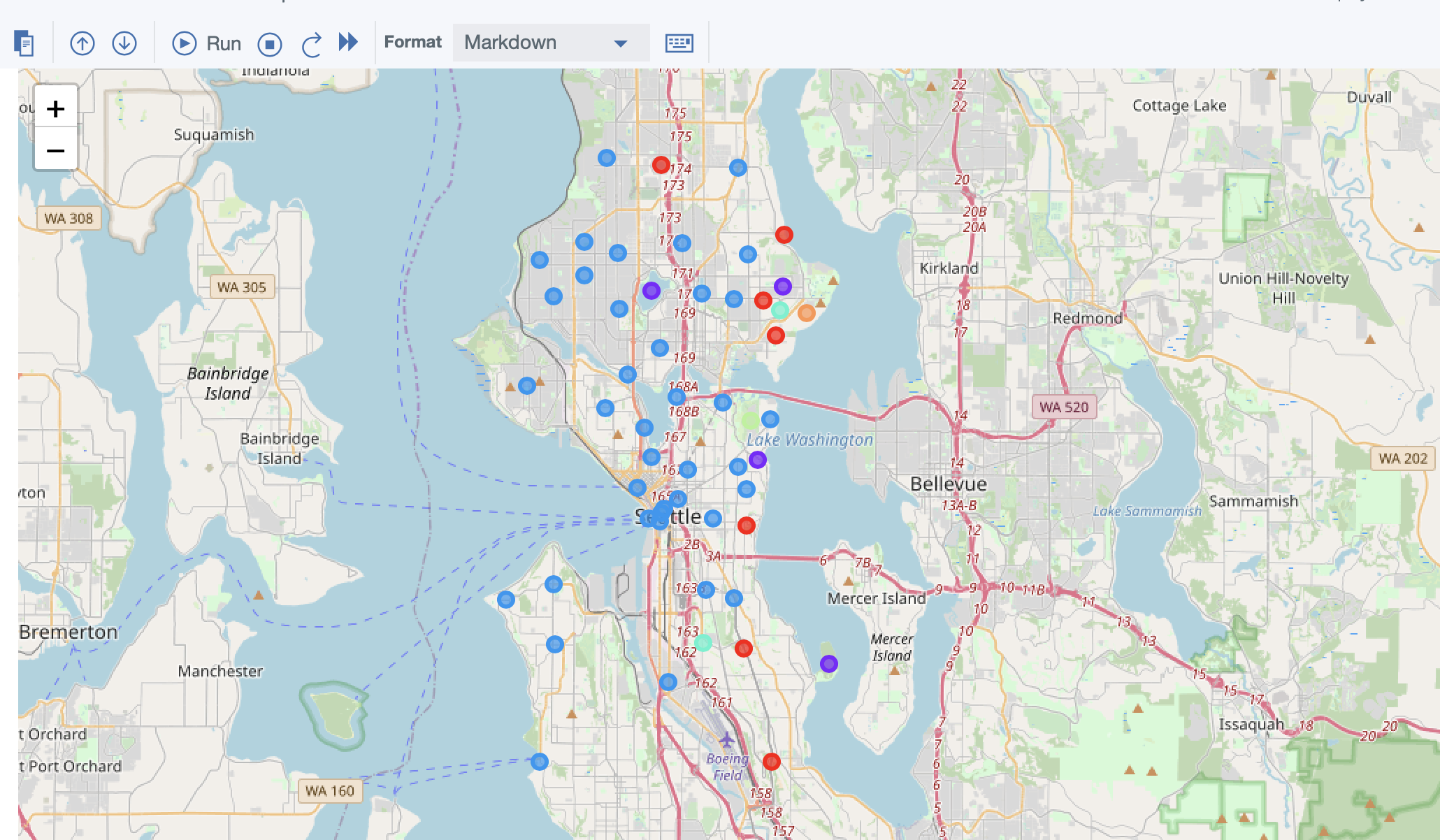
**3. Methodology**

By converting addresses into their equivalent latitude and longitude values. It is possible to use the Foursquare API to explore neighborhoods in Seattle. To explore function to get the most common venue categories in each neighborhood, and K-means clustering algorithm is used to group the neighborhoods into clusters. Finally, the Folium library is used to visualize the neighborhoods in Seattle and their emerging cluster.

The area with the greatest density of huaman to human contact venues is the most vulnerable district.

**4. Analysis**

**Use geopy library to get the latitude and longitude values of SEATTLE**



**5. Results**

The resulting clusters are as follow :

* Cluster 1 - low probility district
* Cluster 2 - low probility district
* Cluster 3 - High probility district
* Cluster 4 - moderate probility district
* Cluster 5 - low probility district
* Cluster 6 - low probility district

From the result, Cluster 3 has the most density of cafe shop and restrants **Therefore, Cluster 3 has the highest potential to have the covid-19 outbreak.**

