**Capstone report:**

1. **Introduction where you discuss the business problem and who would be interested in this project.**

**Scenario:**

Kiki got a new job in San Francisco City, so she will move to SF City. The public transportation in SF is great and Kiki has enough time on the commute. Thus, she does not need to stay near the company. However, Kiki needs to take her dogs for a walk in the parks and she also likes to do jogging in the park. Therefore, she prefers a place with most parks nearby.

**Goal:**

Help Kiki find the place with most parks or gardens.

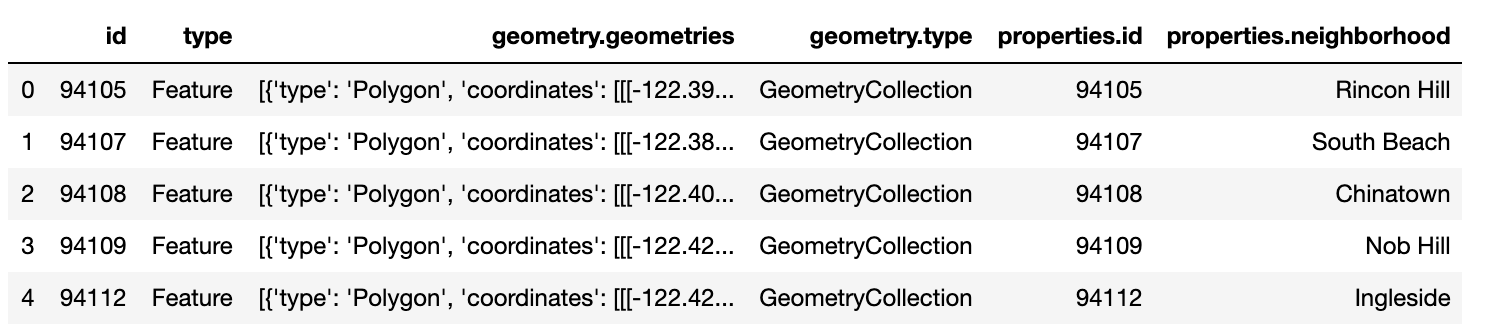
1. **Data where you describe the data that will be used to solve the problem and the source of the data.**

**Source of the dataset:** Json file of the San Francisco City from GitHub.

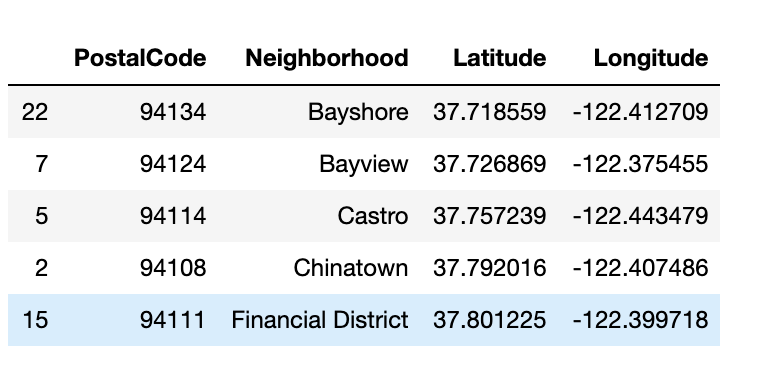
<https://github.com/johan/world.geo.json/blob/master/countries/USA/CA/San%20Francisco.geo.json>



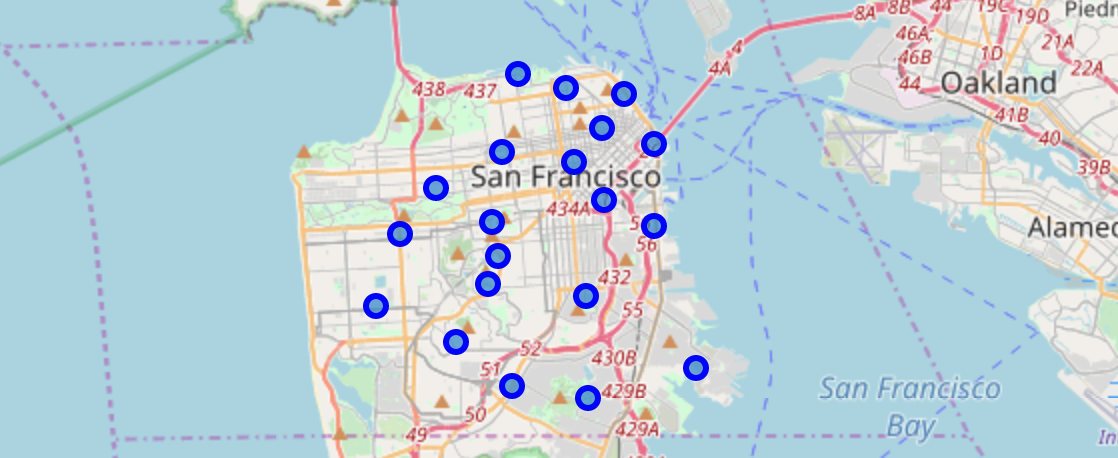
**The dataframe after transformation from the json file:**



**The dataframe after cleaning:**



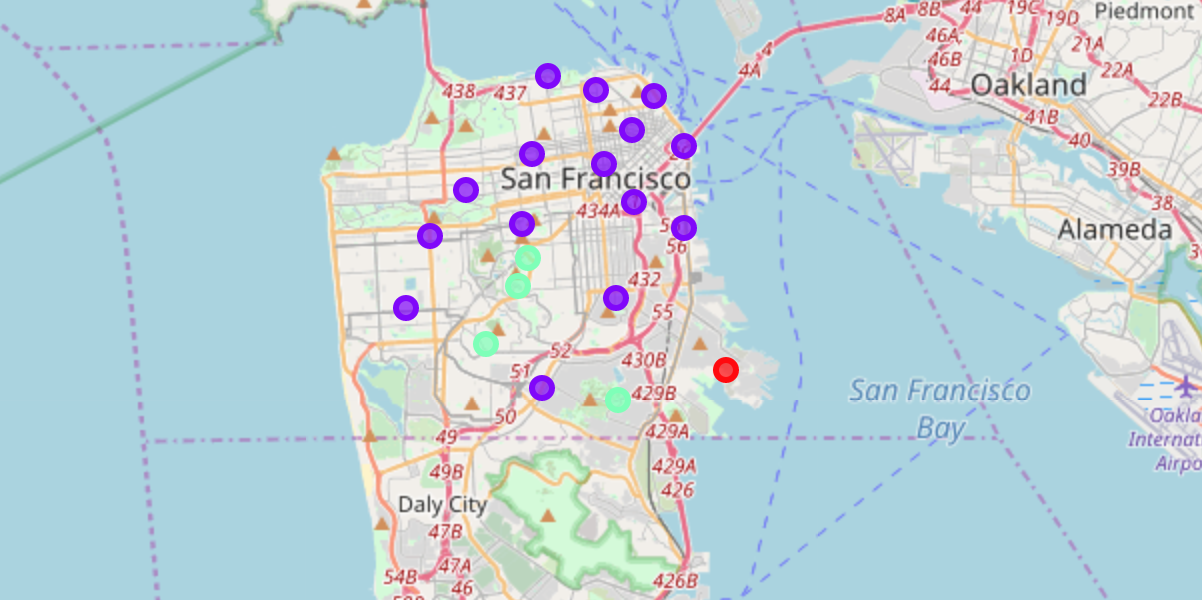
1. **Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.**
2. **Geographical coordinates of SF**

****

1. **API of San Francisco (Use Foursquare) + Analyze Each Neighborhood**

****

1. **Cluster Neighborhoods**

****

1. **Examine Clusters**

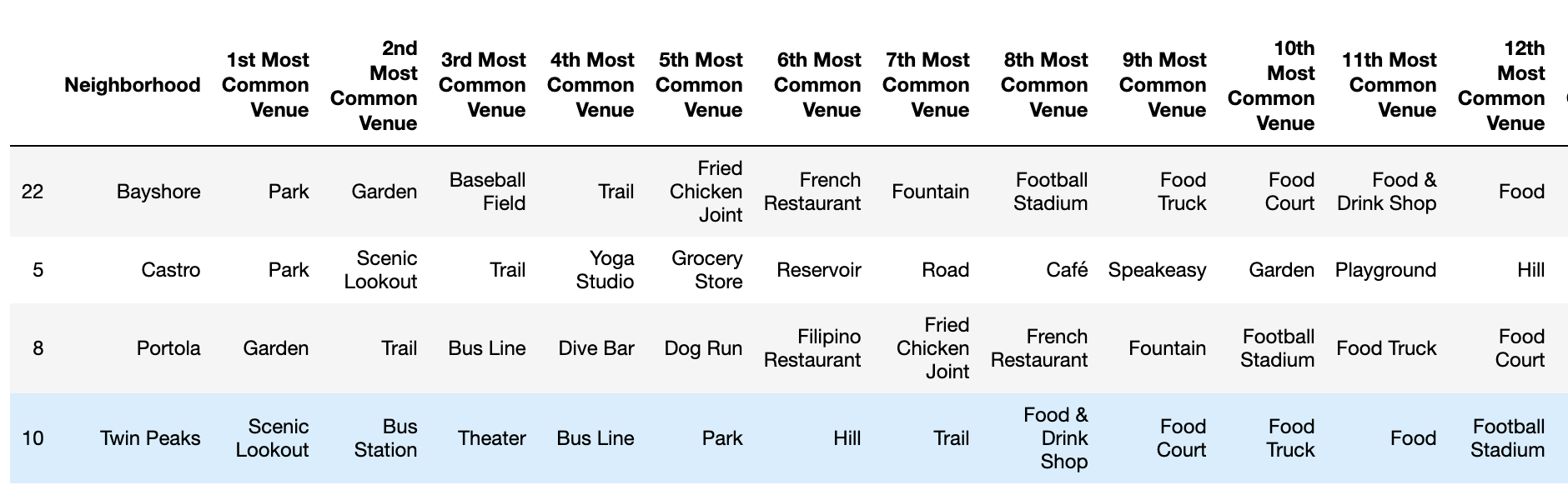
**Cluster 0:**

****

**Cluster 1:**

****

**Cluster 2:**

****

1. **Results section where you discuss the results.**

After trying 3,4 and 5 clusters, I found that three segments are the idealist. Then I examined the clusters. I found that Cluster 0 is surrounded with art, shops and then restaurants. Cluster 1 is surrounded with lots of restaurants and entertainment places. Cluster 2 is surrounded with gardens, lookouts and parks.

However, Cluster 0 and 2 are relatively far away from the downtown. The Cluster 1 are near the center of the downtown and has most neighborhoods.

1. **Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.**

Based on the result, we can see that the neighborhoods in Cluster 2 are the most ideal one for Kiki. Although it is far from the downtown of the San Francisco City, it is not a matter for Kiki. Kiki does not care about the length of the commute. The neighborhoods included in Cluster 2: Bayshore, Castro, Portola, Twin Peaks.

1. **Conclusion section where you conclude the report.**

According to the preference of each group of people, they can find the most suitable neighborhood for them in the San Francisco City. For Kiki’s specific situation, the Cluster 2 is the most ideal place to live. Although there are four neighborhoods altogether in this cluster, they have similar characteristics. However, these four neighborhoods still have some little difference. For example, if Kiki would like a place near the bus station, the Twin Peaks in the Cluster 2 is the best choice.