**The Location of A New Restaurant In Kirkland**

**I. Description of The Problem And Discussion of The Background**

Kirkland is a city in western Washington with a population of about 93,000 according to 2019 census report. It is adjacent to Bellevue and Redmond, where many big companies such as Expedia, and Microsoft locate. And is about 20 minutes drive from Seattle downtown. Kirkland has beautiful waterfront areas, numerous parks, and several art galleries. It has attracted more and more people to move in and live there. According to World Population Review, in the past decade, its population has grown by 88.47% and is continuing to grow at a current rate of 1.32% each year. Hence, there is a growing need for leisure, such as entertainment and dining.

#### This report will use Python segmenting and clustering techniques to fetch venue data from Foursquare to locate an area which will be ideal for someone to open a new restaurant in Kirkland. It will not consider other factors that generally impact such decisions, like cuisine preferences and average spending of targeted customers, rent cost, etc..

**II. Data Used in The Report**

#### In this report, I will use neighborhood information from kirklandwa.org for clustering and segmentation, and the venue information from Foursquare to analyze each neighborhood to find the ideal area to open this new restaurant.

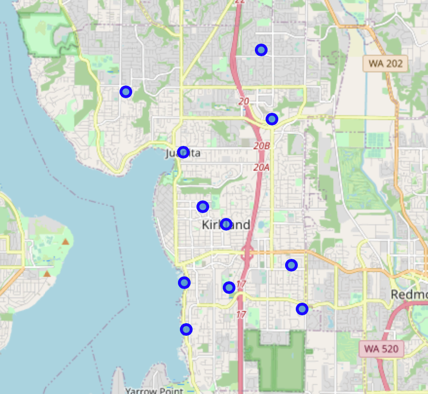
**III. Methodology**

#### *1. Create A Dataframe With Neighborhood Information*

#### Due to the fact that Kirkland is a small city, there isn't any readily usable open source data online, I will have to create a dataframe with the neighborhood data obtained from kirklandwa.org, using pandas.

Then I created a second dataframe of coordinates to combine it with the first one, using Geocoder Python package. In this process, I encountered several issues. One of them was that Geocoder was not able to obtain the geographical coordinates of two neighborhoods of North Rose Hill and South Rose Hill. Initially, I decided to remove these two rows of data and continued with the analysis. However, the second issue I encountered could not be ignored. The geographical coordinates of most of the neighborhood obtained by geocoder were off. When the map was generated with Matplotlib, I could see that only three neighborhoods were pinned correctly in Kirkland, all the rest were either in Seattle or the east coast of the United States. To solve this issue, I added the city of Kirkland between neighborhood name and the state abbreviation to increase the geographical accuracy of where those neighborhoods are located and I added North Rose Hill and South Rose Hill back to see if this time it will generate the coordinates. The amended neighborhood column seemed to work. However, by looking closely at the coordinates, Highlands neighborhood’s figures still didn’t look right. Then I mapped out the neighborhoods again to confirm the accuracy of the coordinates, in this process I found two more neighborhoods’ figures were also off. Therefore, I had to manually replace those numbers with the correct ones found in Google Search Engine. Please see *figure 1* and *figure 2* below for the final dataframe of Kirkland neighborhoods and the map.

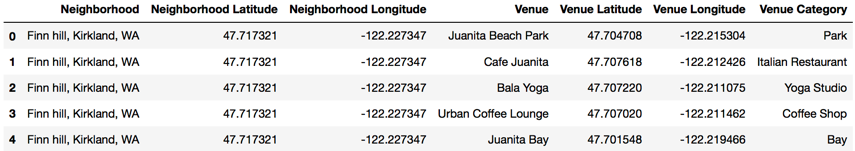
*Figure 1 Figure 2*

**

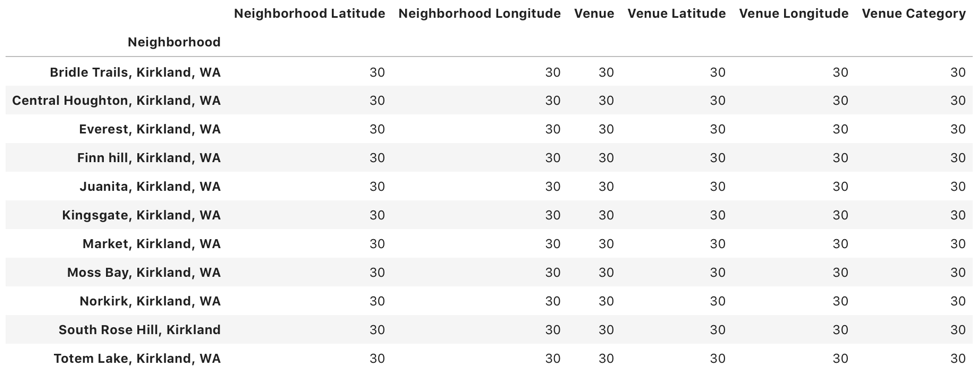
#### *2. Analyze Neighborhoods in Kirkland*

Now that we have the correct coordinates of each neighborhood in Kirkland, we can start our exploratory analysis.

2.1 Obtain venue data for each neighborhood from Foursquare



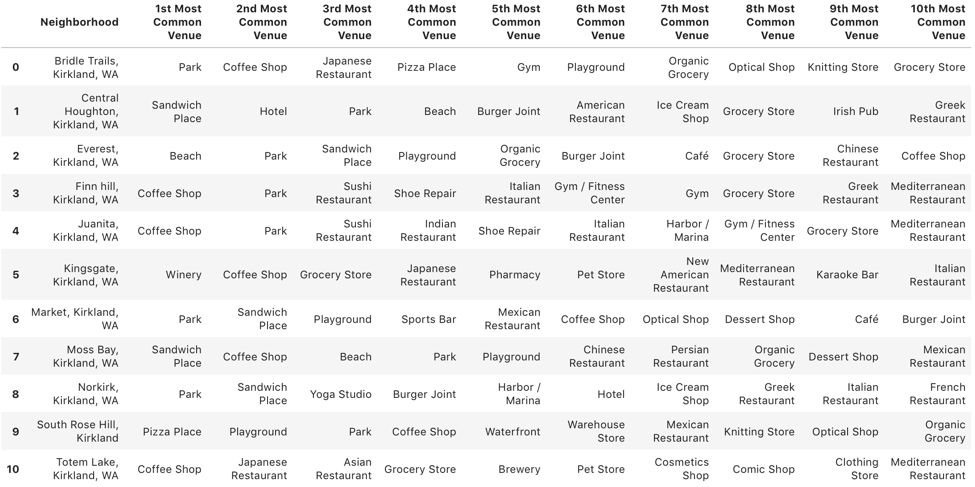
2.2 Get venue counts for each neighborhood



From this step, I found out that only 11 neighborhoods has venue data from Foursquare. So I will use these 11 neighborhoods to continue with the following analysis.

2.3 Analyze each neighborhood

In this step, I found out the total amount of unique venue types in Kirkland. Then assigned venue types to each neighborhood to create a new dataframe. From this dataframe, I then grouped the counts for each venue type by neighborhood to further get the means of each venue type for each neighborhood. By doing this, I was able to create a new dataframe with top 10 most popular venues for each neighborhood.

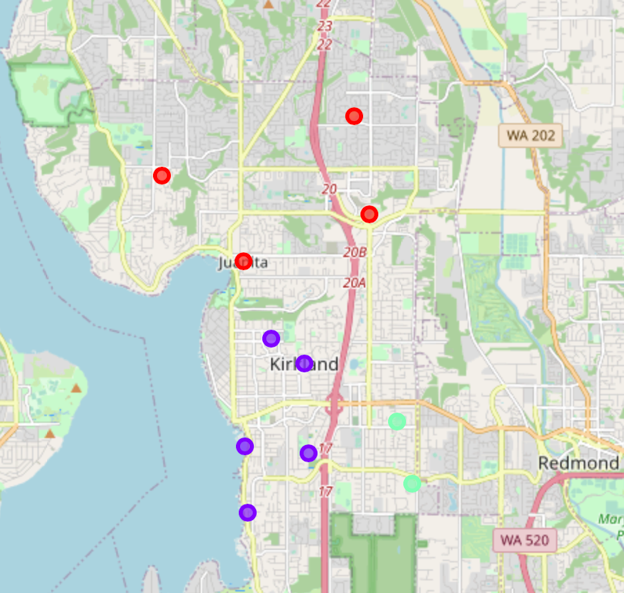


2.4 Cluster neighborhoods

In this step, I tried randomly clustering neighborhoods with k equals to 3, 4, 5 and found that when k equals to 3, neighborhoods were distributed more evenly into groups.

Then I inserted the Cluster Labels as a new column into the dataframe. While I was merging this dataframe with the original one named “df”, a new issue came up. Because the one with top 10 venues only has 11 neighborhoods, when merged, there are three rows that contains NaN values, which automatically turned cluster labels values into float datatype, which resulted in errors when generating the map of clustered neighborhoods. To solve this issue, I went back and dropped those three neighborhoods that doesn’t have any venue information. Then did the merge again.

Please see below the map of the clustered neighborhoods.



2.5 Examine each cluster

Once the neighborhoods were clustered, I then examine them to find out the unique characteristics of each one. Please see the top 10 most common venues of each cluster below.

*Cluster 1*



*Cluster 2*



*Cluster 3*



**IV. Results**

#### From the above analysis of neighborhoods in kirkland, each cluster has its unique characteristics. Cluster 1 is popular for family leisure and dining. Cluster 2 is popular for family leisure and fastfood. Cluster 3 is popular for family leisure, fast food and errands.

**V. Discussion**

#### All three clusters have parks to attract families to spend time in the area, especially during weekends and holidays. But each cluster also has other unique characteristics to offer. Cluster 1 has more table service dining options, which is a go-to area for meals. This guarantees a certain amount of potential customers for any new restaurants as it helps the exposure of the existence of a new restaurant. Cluster 2 seems to have more fast food options. Similar to Cluster 1, Cluster 3 also has more fast food option than table service restaurants.

**VI. Conclusion**

#### Based on each clusters' characteristics in Kirkland, it's ideal to open the new restaurant in cluster 1 area.