Opening an Italian Restaurant in Seattle, Washington

Jelena Mitrovic
Nov 20, 2019

Introduction

Located on the West Coast of the United States, Seattle is the largest city in the state of Washington and one of the fastest-growing cities since 2013. The reason for this lies in the fact that some of the biggest tech companies have established their business in this city.

Seattle has 127 neighborhoods covering an area of 142.07 sq. mi (367.96 km2) where around 750,000 people live. The population density is around 3151 persons per sq. km and the population growth is around 2.5% every year [1]. Despite the battle the city has concerning the population growth, for people who want to start a new business here, this might be an ideal place.

Before starting a new business, a lot of things should be taken into account. When the idea exists, it requires further steps which include market research, business plan, financial plan, building a team, finding the location, and making it official [2]. The goal of this project is to examine what location would be the most convenient for somebody who wants to open an Italian restaurant in Seattle, Washington. For a business to thrive, the location of the restaurant should be in an area highly populated, where people circulate throughout the day to have lunch, grab a coffee, or have a meeting. Also, for this type of business, it would be ideal to be in an area where other restaurants are. However, it is also important not to be in an area overwhelmed by competitors [3]. Therefore, the following aspects should be considered before making the conclusion: types of the neighborhood, the population in those neighborhoods, and a number of competitors.

Data

In order to determine the most suitable location for a new Italian restaurant, the following data was necessary:

- For Seattle, WA **neighborhood location data**, I used the data provided by Zillow Inc. [4]. I read this data into a pandas dataframe to use for the following steps.
- Once the neighborhood data was in a structured format, I used **Foursquare location data** to extract information about the most common venues, fit them into clusters, and determine the best perspective cluster using k-means clustering.
- For the **population**, I used data provided by the US Census Bureau from the 2010 census [5].

Methodology

The first dataframe I made consisted of four columns: Neighborhoods, Latitude, Longitude, Population, and Coordinates (Figure 1.).

	Neighborhood	Longitude	Latitude	Population	Coordinates
0	Green lake	-122.338530	47.678202	0	{'type': 'MultiPolygon', 'coordinates': [[[[-1...
1	Madison park	-122.285414	47.627222	0	{'type': 'MultiPolygon', 'coordinates': [[[[-1...
2	Meadowbrook	-122.297094	47.707433	3318	{'type': 'MultiPolygon', 'coordinates': [[[[-1...
3	Bryant	-122.285181	47.666595	5223	{'type': 'MultiPolygon', 'coordinates': [[[[-1...
4	Highland park	-122.342096	47.530543	0	{'type': 'MultiPolygon', 'coordinates': [[[[-1...

Figure 1. Dataframe with all neighborhoods, longitude, latitude, and population

Next, using Geopy library, I got the **latitude and longitude of Seattle, WA** so I can show all neighborhoods on the map (Figure 2.)

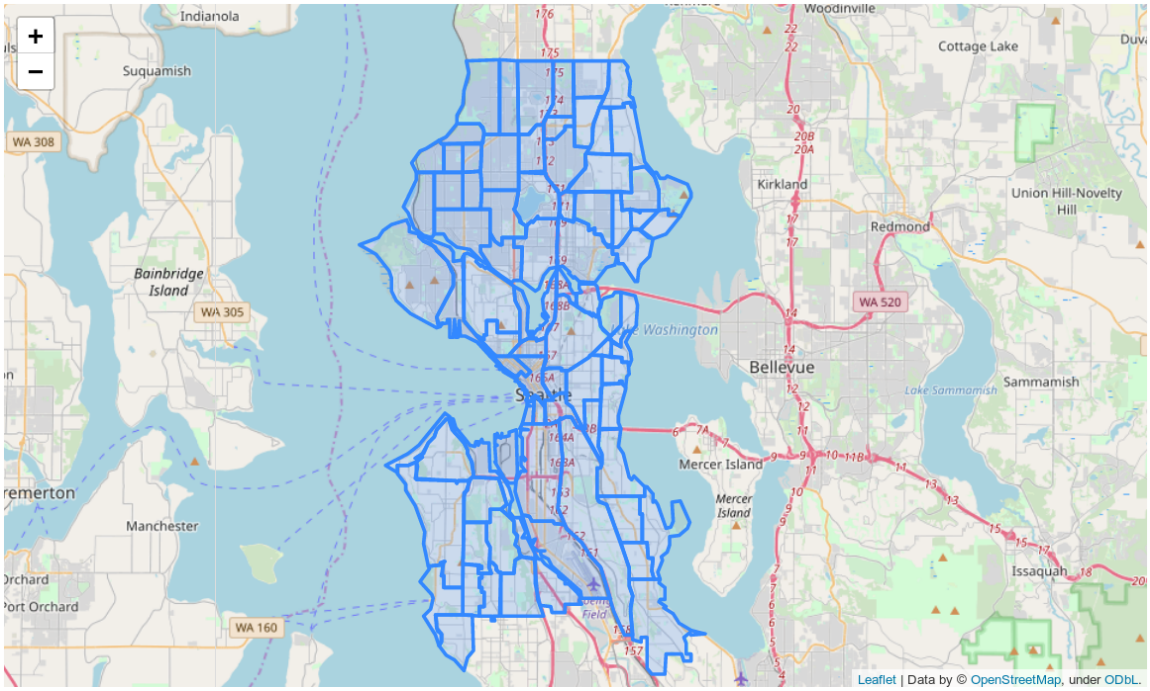

Figure 2. All neighborhoods in Seattle, WA

In order to learn more about the venues located in each neighborhood, I used Foursquare API data to get the **top 100 venues in each neighborhood within a radius of 500 meters**. Then I structured them into pandas dataframe (Figure 3.).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Green lake	47.678202	-122.33853	Green Lake Loop	47.680666	-122.334540	Trail
1	Green lake	47.678202	-122.33853	Green Lake Park	47.680449	-122.332620	Park
2	Green lake	47.678202	-122.33853	El Naranjo	47.681672	-122.343992	Food Truck
3	Green lake	47.678202	-122.33853	Bongos	47.676576	-122.347007	Caribbean Restaurant
4	Green lake	47.678202	-122.33853	Beth's Café	47.682145	-122.344709	Diner

Figure 3. Dataframe with top 100 venues in each neighborhood

Analyzing venues in all neighborhoods, I realized that there were 314 unique categories of venues. Then, I used a one-hot encoding to analyze each neighborhood. Next, I grouped the rows by neighborhood and by taking the mean of the frequency of occurrence of each category. After that, I extracted 5 most common venues in each neighborhood with their frequency. However, for further analysis, I put the top **10 most common venues** into the pandas dataframe (Figure 4.).

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Admiral	Coffee Shop	Pub	Pizza Place	Grocery Store	American Restaurant	Salon / Barbershop	Gift Shop	Thai Restaurant	Bank	Bar
1	Alki	Trail	Beach	Park	Seafood Restaurant	Other Great Outdoors	Burger Joint	Greek Restaurant	Volleyball Court	Post Office	Fast Food Restaurant
2	Arbor heights	Scenic Lookout	Park	Beach	Field	Event Space	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant
3	Atlantic	Park	Bus Station	Coffee Shop	Convenience Store	Garden	Museum	Scenic Lookout	Bakery	Ethiopian Restaurant	Café
4	Ballard	Brewery	Bar	Coffee Shop	New American Restaurant	Sandwich Place	Food Truck	Burger Joint	Mexican Restaurant	Clothing Store	Sushi Restaurant

Figure 4. Neighborhoods with the top 10 most common venues

Results

At this point, my data was ready for cluster analyses. In order to determine the optimal number of clusters, I used The Elbow method. From the figure (Figure 5.) it appears that the optimal number of clusters would be 7 or 8. However, analyzing each cluster I came to the conclusion that the most optimal number would be 3.

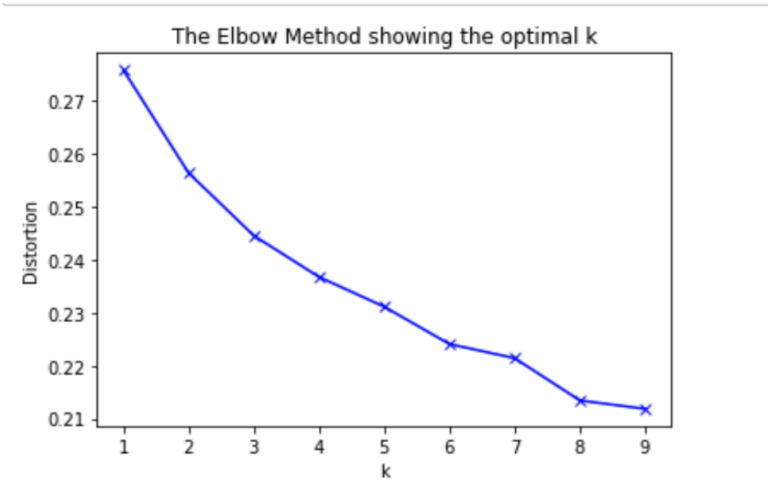

Figure 5. The Elbow Method for Optimal k

Then I created clusters and put data into a new dataframe that includes **the clusters** as well as the **top 10 venues for each neighborhood** (Figure 6.).

	Neighborhood	Longitude	Latitude	Population	Coordinates	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Green lake	-122.338530	47.678202	0	{'type': 'MultiPolygon', 'coordinates': [[[[-1...	0	Park	Pub	Beach	Motorcycle Shop	Diner	Gas Station
1	Madison park	-122.285414	47.627222	0	{'type': 'MultiPolygon', 'coordinates': [[[[-1...	2	Café	Ski Area	Bakery	Surf Spot	Sushi Restaurant	Park
2	Meadowbrook	-122.297094	47.707433	3318	{'type': 'MultiPolygon', 'coordinates': [[[[-1...	2	Pizza Place	Marijuana Dispensary	Pet Store	Pool	Auto Garage	Soccer Field
3	Bryant	-122.285181	47.666595	5223	{'type': 'MultiPolygon', 'coordinates': [[[[-1...	2	Pharmacy	American Restaurant	Coffee Shop	Salon / Barbershop	Café	Chinese Restaurant

Figure 6. Neighborhoods with cluster they belong and top 10 venues

Using the Folium library, I wanted to visualize data, so the following image shows the clusters on the map (Figure 7.).

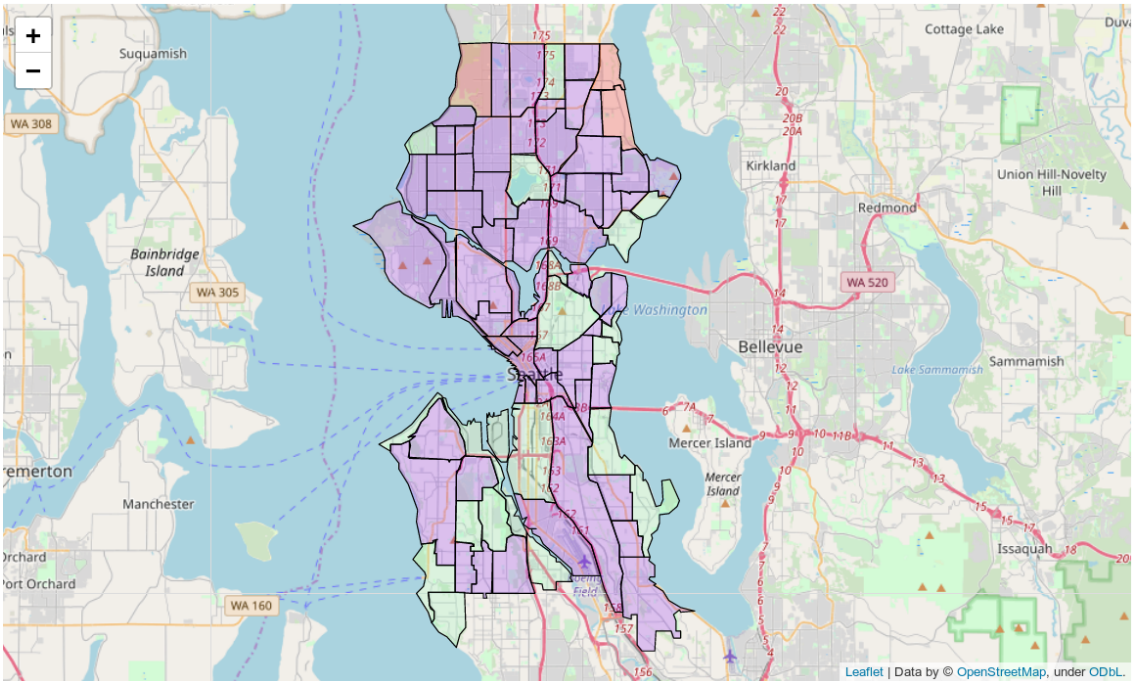

Figure 7. Map of Seattle, WA with all clustered neighborhoods

After examining each cluster, I came to the conclusion that:

- Cluster 1 (green color) represents **parks and scenic lookout**
- Cluster 2 (orange color) represents **outdoor activities**
- Cluster 3 (purple color) represents the **business/residential area** with a lot of coffee shops, convenience stores, restaurants, and bus stops.

The next image shows the above mentioned with markers on top of each neighborhood for easier tracking (Figure 8.).

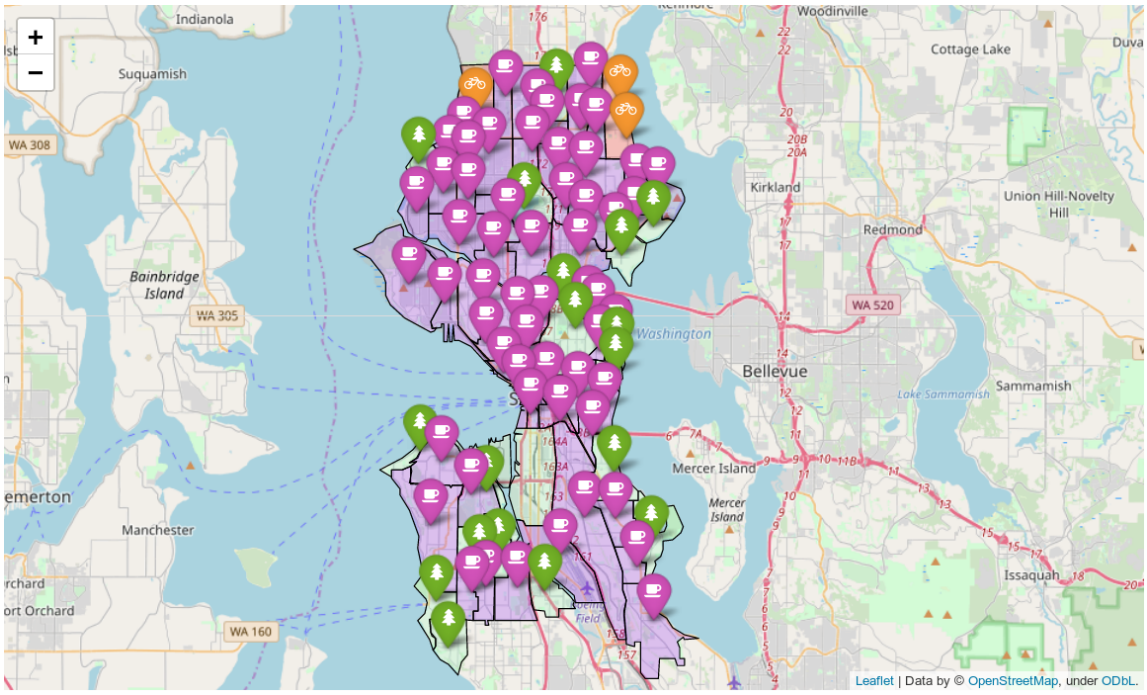

Figure 8. All neighborhoods in Seattle, WA clustered into specific categories

When I got a better insight into the neighborhood's types and realized which neighborhoods would be convenient for opening an Italian restaurant – purple areas, I wanted to see the population in all neighborhoods but with paying attention mostly to ones of interest (Figure 9.).

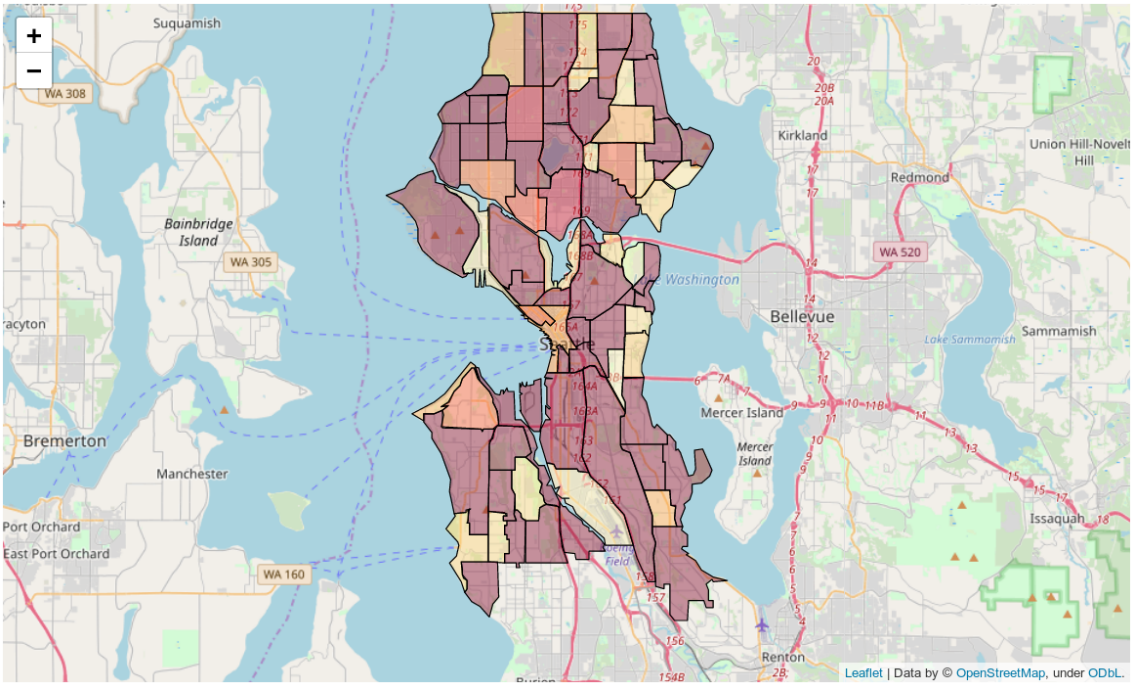

Figure 9. Population in neighborhoods in Seattle, WA

Lastly, I created a bar chart with the frequency of all restaurants in all neighborhoods of business/residential area to determine which area is the most optimal when it comes to the

number of competitors [Figure 10.]. According to the figure, it seems that the neighborhoods can be categorized into 6 different groups concerning the frequency of restaurants: (0-5), (5-10), (10-15), (15-20), (20-25), and (25-30). The top 4 neighborhoods have more than 25 restaurants, so these neighborhoods could be considered as highly competitive. The optimal number of competitors should be somewhere between 10 and 20 restaurants.

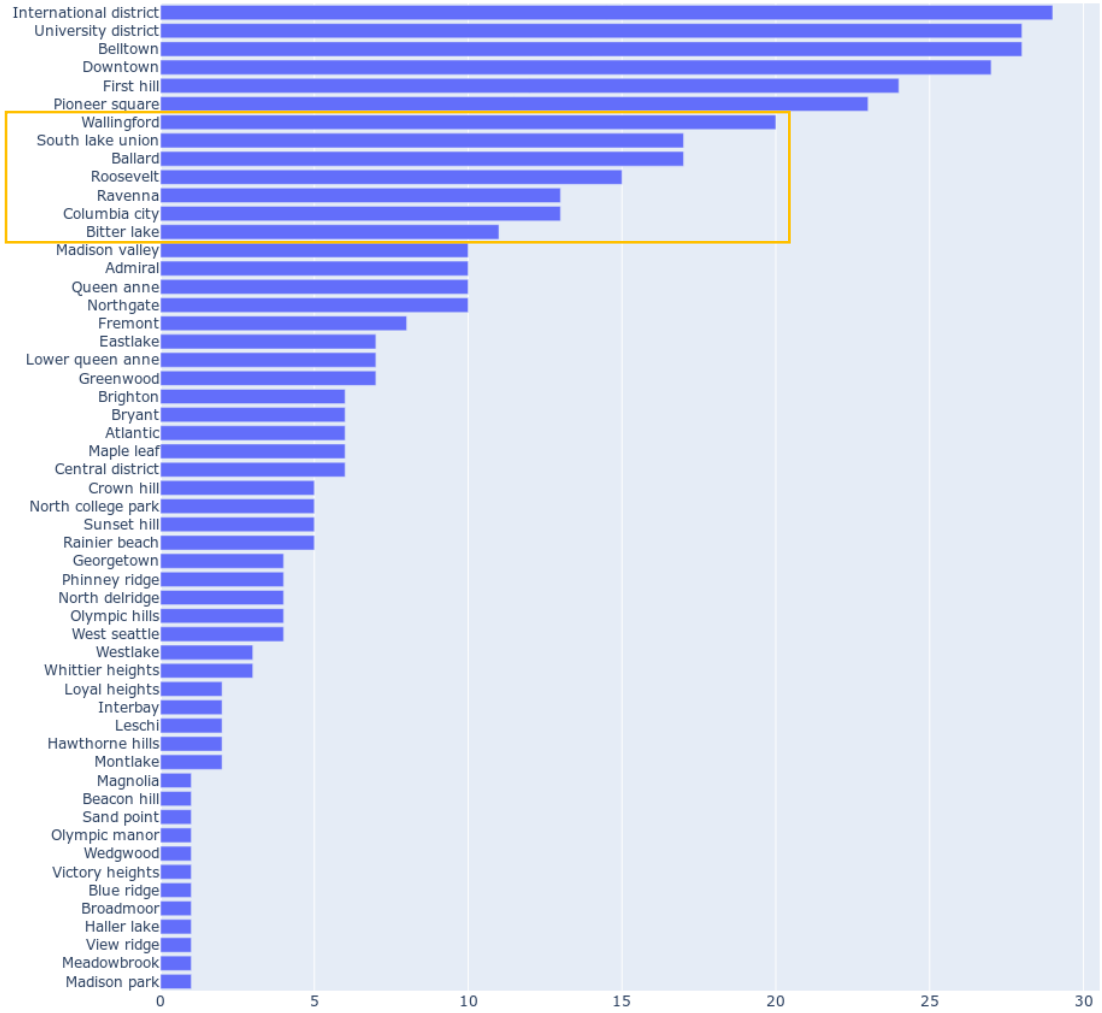

Figure 10. Business/residential neighborhoods with the number of competitors

For easier tracking, I also created a map marked only with the neighborhoods that would be of interest to somebody who wants to open an Italian restaurant (Figure 11.).

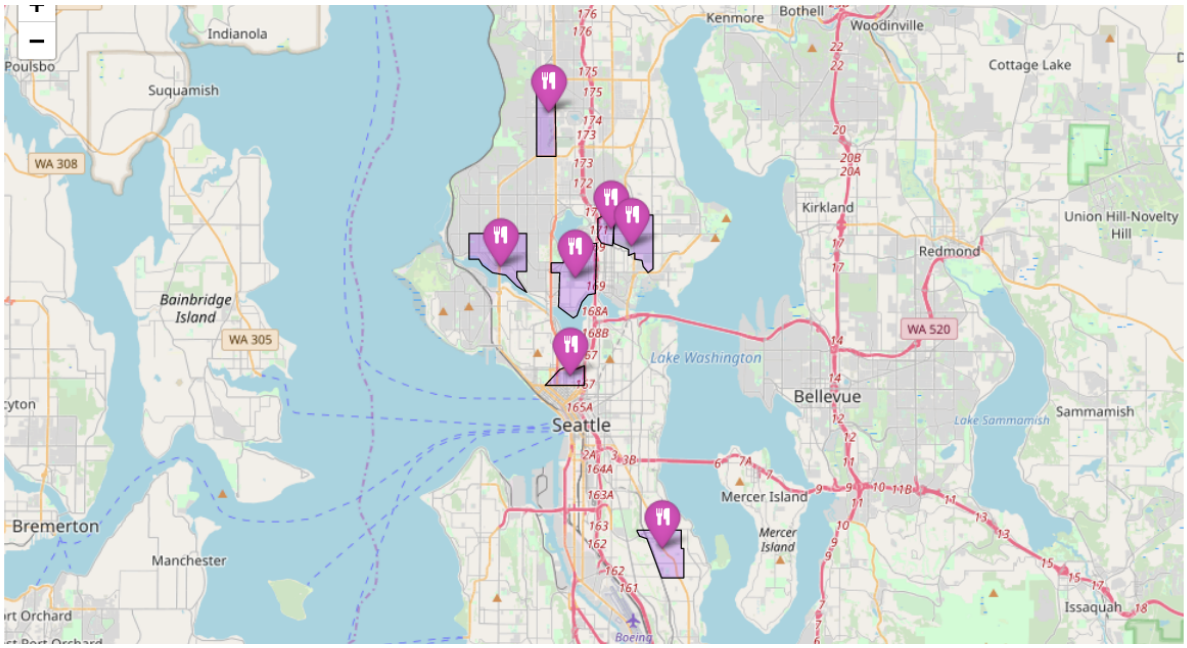

Figure 11. The most convenient neighborhoods for opening an Italian restaurant

Discussion

After obtaining necessary data and structuring it in a proper dataframe, the k-means cluster analysis showed that all neighborhoods in Seattle, WA could be grouped into three separate categories: *parks and scenic lookout, outdoor activities, and business/residential area*. Given the goal of this project, the conclusion was that neighborhoods that belong to business/residential area would be the most suitable for further analyses.

The next analysis showed that highly populated areas were mostly the ones that belong to business/residential areas. The final analysis showed the frequency of restaurants in each neighborhood of business/residential area and it was the final piece in determining which neighborhoods would be the most convenient for opening an Italian restaurant – the ones that are in the category of 10-20 restaurants per neighborhood.

Conclusion

Even though starting a new business requires a lot of steps and considerations before it becomes official, one of the most important things that should be carefully explored is its location. For this project, the goal was to examine the most convenient location for opening an Italian restaurant in Seattle, Washington. The goal was to determine the location based on the type of neighborhood, the number of people who live in that neighborhood, and the number of competitors.

After all analyses were performed, the conclusion is that the most convenient neighborhoods that belong to business/residential area, highly populated, with an optimal number of competitors would be Wallingford, Salt Lake Union, Ballard, Roosevelt, Ravenna, Columbia City, and Bitter Lake.

Although this project narrowed the most suitable neighborhoods for opening an Italian restaurant, in future projects it should be explored even further. Depending on the target customers, in the next project, it would be convenient to use other data such as the structure of the population in a specific neighborhood, business revenues in that area (if available), or public parking areas.

References

[1] Population Of Seattle | Seattle Population 2019. (December 30, 2018). Retrieved from https://theuspopulation.com/seattle-population/

[2] McCreary, Matthew. (n.d.) The Complete, 12-Step Guide to Starting a Business. Retrieved from https://www.entrepreneur.com/article/297899

[3] Why the Location of Your Restaurant is So Important. (n.d.). Retrieved from https://www.tigerchef.com/why-the-location-of-your-business-is.html

[4] Zillow Neighborhood Maps of Seattle in GeoJSON and KML Formats. (December 2, 2013). Retrieved from https://github.com/craic/zillow_seattle_neighborhoods

[5] Overview of Seattle, Washington - Population. (n.d.). Retrieved from https://statisticalatlas.com/place/Washington/Seattle/Overview