The Battle of Neighborhoods in Seattle, WA

Rodion Dovhan

August 17, 2020

1. Introduction

1.1. Background

Seattle, a city on Puget Sound in the Pacific Northwest, is surrounded by water, mountains, and evergreen forests, and contains thousands of acres of parkland. Washington State's largest city, it's home to a large tech industry, with Microsoft and Amazon headquartered in its metropolitan area. Seattle Metro Area population is about 4 million people. In July 2013, Seattle was the fastest-growing major city in the United States and remained in the top five in May 2015 with an annual growth rate of 2.1%. In July 2016, Seattle was again the fastest-growing major U.S. city, with a 3.1% annual growth rate.

1.2. Problem

In this project we will try to find an optimal location for a restaurant in Seattle, Washington. Specifically, it will be targeted on those interested in opening an Italian restaurant.

We will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Italian restaurants in vicinity. We may take in consideration socioeconomic data of population living in the areas to draw a final decision.

1.3. Interest

This would be of interest to anyone who is looking to open a new Italian Restaurant or wants to get an insight into the business in the area.

2. Data

2.1. Data Sources

We are going to be using this data set with Seattle Neighborhoods by Zip Codes from this page 'http://www.agingkingcounty.org/wp-content/uploads/sites/185/2016/09/SubRegZipCityNeighborhood.pdf'. It is a pdf file containing excel spreadsheet "Sub-Regional, City and Neighborhood Designations by Zip Code".

We will use Foursquare API to later extract data for venues in the corresponding zip codes. Also, we are going to retrieve different statistical data by zip code from zip-codes.com database using their API. Both return data in JSON format.

2.2. Data Preprocessing

I used Adobe Acrobat to extract Seattle Neighborhoods by Zip Codes file into an excel file from the pdf. We will be using data only for the City of Seattle and its adjacent suburbs taken from the third column of the table which is sorted by zip code.

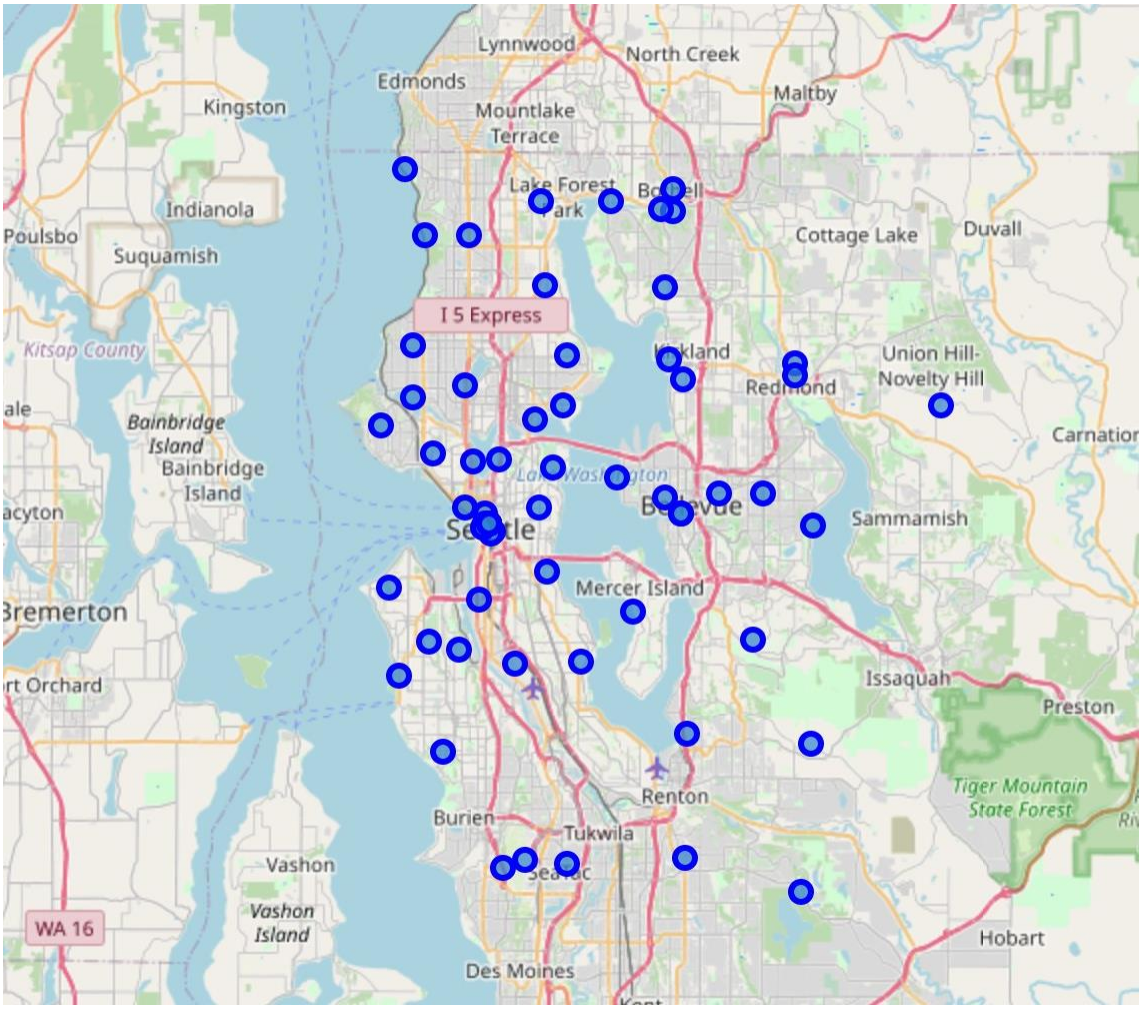
We are using Foursquare API to gather data for 100 venues in radius of 500 meters from each zip code geographic coordinates.

We also use zip-codes.com and their API to obtain geographic coordinates for our zip codes, as well as average income and population by zip code later.

3. Methodology

3.1. Visualizing Seattle and its neighborhoods

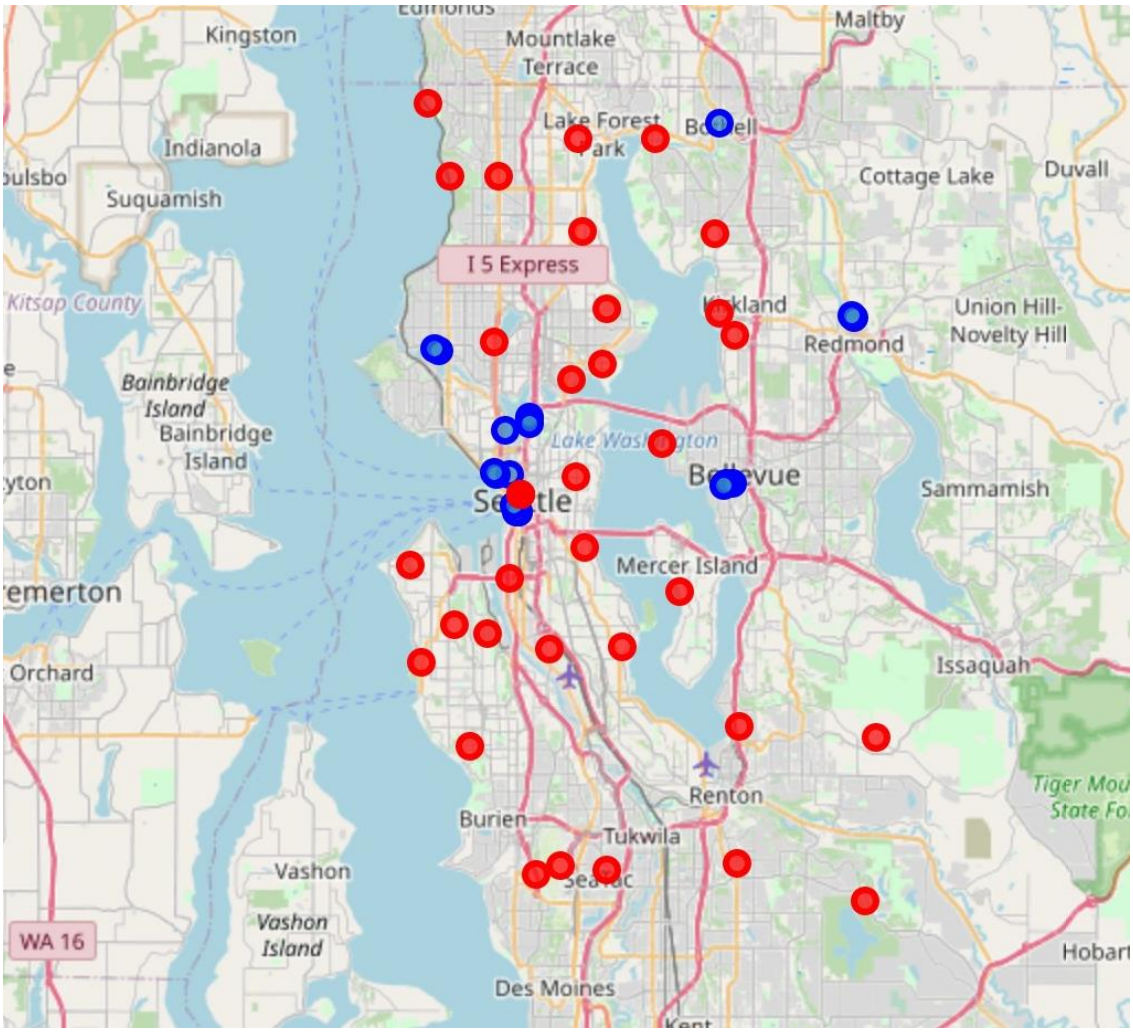
Having gathered the neighborhood-zip code data and respective geographic coordinates for all points we proceed to visualizing them on a map. Each blue point represents a zip code.



We can notice that there are a couple of locations where a number of zip codes are tightly clustered. We will keep this on our mind for later so that it doesn't interfere with our results.

3.2. Venues

We gathered data for 100 venues in radius of 500 meters from each zip code geographic coordinates. Then we sorted it to have only Italian Restaurants left. We create a separate dataset from existing data with neighborhoods/zip-codes which do not have any restaurants. The following map shows Italian Restaurants in blue and Neighborhoods without any in red.

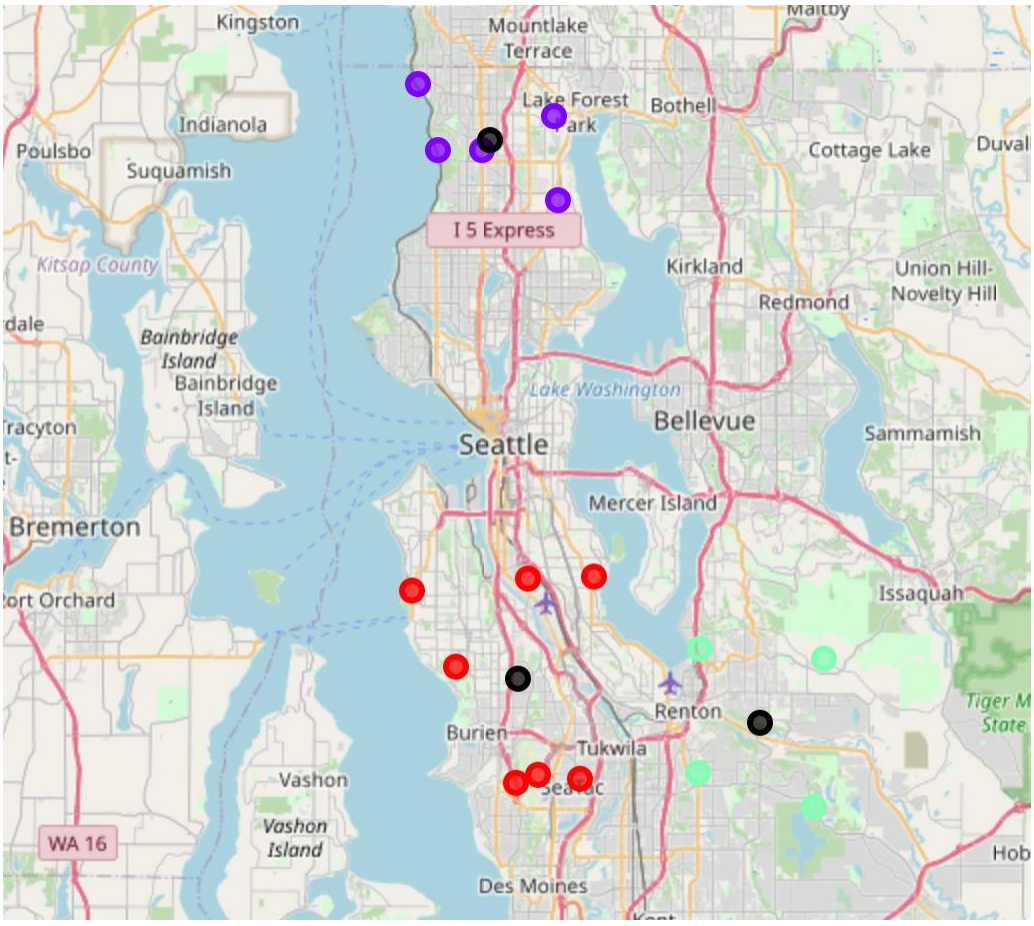


Here we can see that due to the nature of cities division on zip codes we have some overlying red points in very close proximity with existing Italian Restaurants. We decide to remove zips which have an Italian Restaurant in a 4-mile radius to account for that. This leaves us with 16 of those. Here it would be fair to point out that since we are dealing with non-uniform terrain - mainly

with Lake Washington laying in the middle of the area we are looking at, then the method we used to compute distances may not be the best one in comparison with real drive route length.

3.3. Clustering the Neighborhoods

We now want to cluster the neighborhoods without any restaurants as it would be rational to open a restaurant in close proximity to more than one. We are going to use K-Means algorithm in sci-kit learn to cluster our zip code areas without an Italian Restaurant. K being set to 3. I chose this algorithm instead of DBSCAN since as mentioned earlier some areas have a bigger zip code density due to the nature of division itself and the clusters would not make practical sense in this case. The resulting map is shown below with black dots being centers of the clusters, which can be used as reference points for optimal location of a new restaurant.

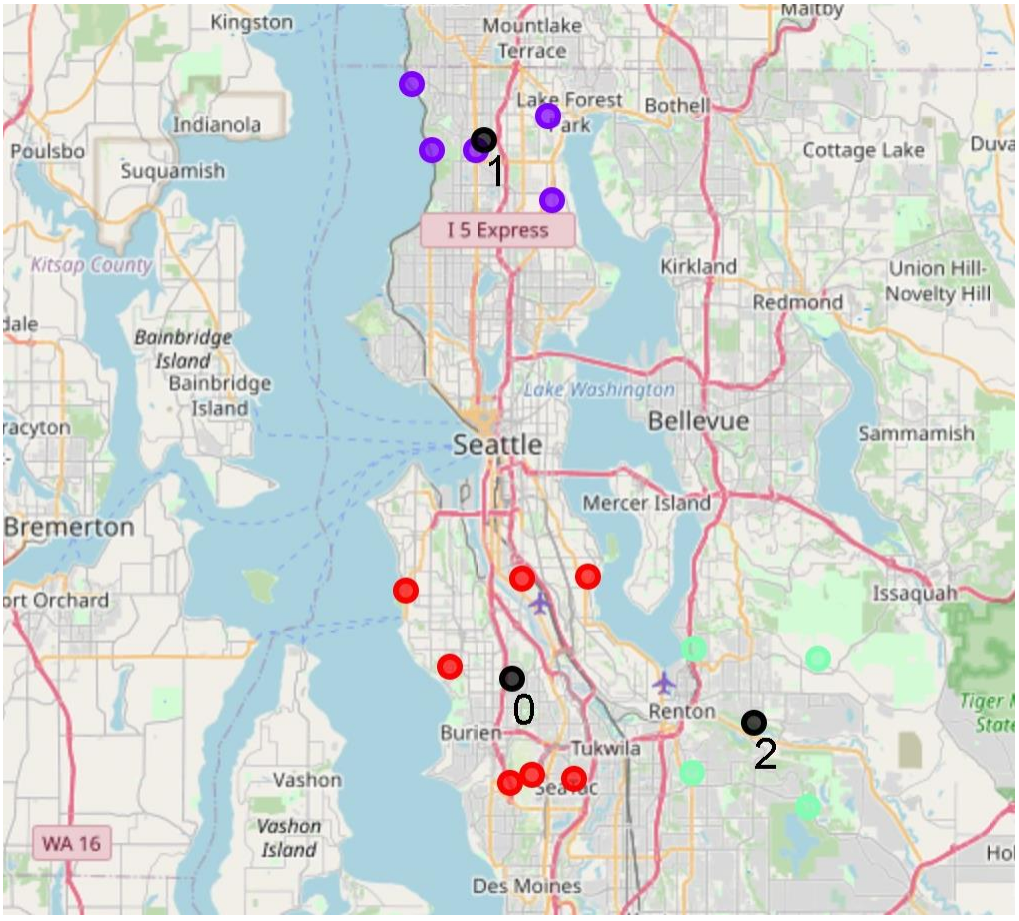


3.4. Adding Socioeconomic Data

We also gather some socioeconomic statistics in the zip codes comprising the clusters. More specifically, we look at income per household and total population per zip code. We then calculate average income and total population in each cluster and combine all data in one table.

	Cluster	Latitude	Longitude	Average Income	Total Population
0	0	47.494956	-122.320450	69683.666667	138918
1	1	47.743946	-122.339707	81131.500000	133444
2	2	47.474792	-122.154365	87052.000000	130794

Here we can see that the cluster in Renton area has the highest average income, while the cluster north of Seattle is right behind. Also, the total populations in the area do not differ much.



4. Results and Discussion

Our analysis shows that although there are a lot of restaurants in Seattle Area there are not that many Italian Restaurants specifically. Moreover, a lot of them are highly concentrated in several neighborhoods like Seattle Downtown, Bellevue Downtown, or even South Lake Union and Ballard. This leaves out a big area in the south free and a smaller one in the north.

We narrowed down our attention on the areas which do not have any Italian Restaurants in a ~4-mile vicinity radius. Basically, this left us with neighborhoods where we would not have any competition whatsoever. We used K-Means clustering algorithm with 3 clusters, which was deemed reasonable in this case. In general, areas nearby the centers of the clusters may be used as optimal locations for a restaurant.

In order to get some insight into which one of them could be the better one we decided to take a little look at the socioeconomic demographic available in the areas. Looking at the latter one might see Renton or Shoreline areas as more preferable, but not necessarily.

5. Conclusion

Purpose of this project was to identify Seattle areas with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major

zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.