Recommending Neighborhoods in Seattle, Washington for Opening a Restaurant

Applied Data Science Capstone Project

Surbhi Bector

Introduction

Aspiring business owners hold a significant amount of risk when evaluating decisions to create the ideal model and strategies to kickstart a successful business. Restaurant owners specifically handle a lot of planning and research, due to the complex nature of the operation, to better their chances for success.

According to Entrepreneur, one of the most challenging parts of starting a restaurant is the pre-planning that goes along with it. Many newly opened restaurants do not profitize in the first year of opening due to failure in planning. One crucial factor in planning a restaurant is the location to open it. Owners have many factors to consider, such as the amount of restaurants near the location, location popularity, and other attractions within this location. Business owners can take advantage of research and data available on any location to then analyze what area would be beneficial for their restaurant. Ideally, owners will want a location that is popular, has complementary attractions, and less competition.

According to Fox Business, one of the top 10 locations to start a restaurant in the U.S. is Seattle, Massachusetts, due to the high amount of annual restaurant sales and high median income level in the city.

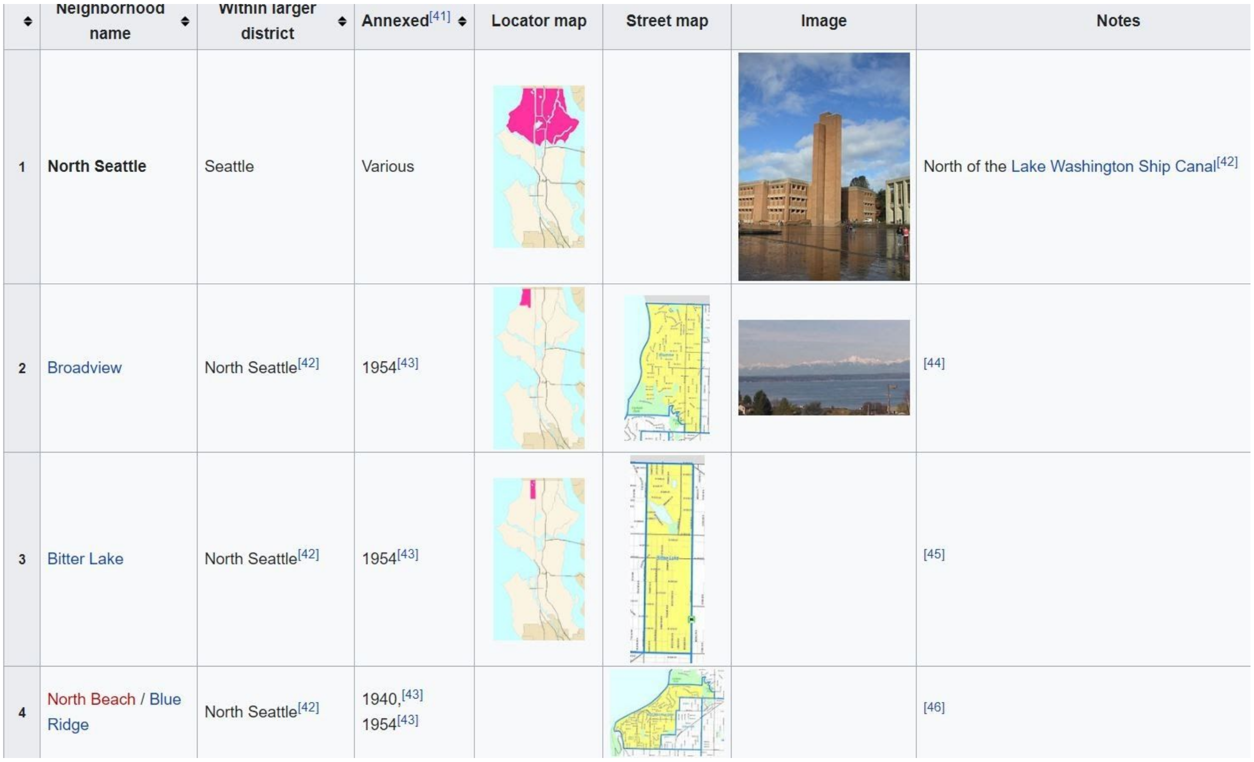
The goal of this project is to recommend the best area in Seattle for aspiring restaurant owners to open a restaurant based on existing data available. By using multiple data mining and machine learning algorithms, clusters in Seattle will be evaluated based on existing attractions. This research will assist new restaurant owners and owners looking to expand franchises by providing advantageous information to plan their restaurant location.

Data

Data on Seattle neighborhoods was initially scraped through wikipedia's page on Seattle, Washington.[1] This page contains a chart of official neighborhoods listed in

[1] https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle

the Seattle area. The data was downloaded and read using pandas and inserted into a data frame.

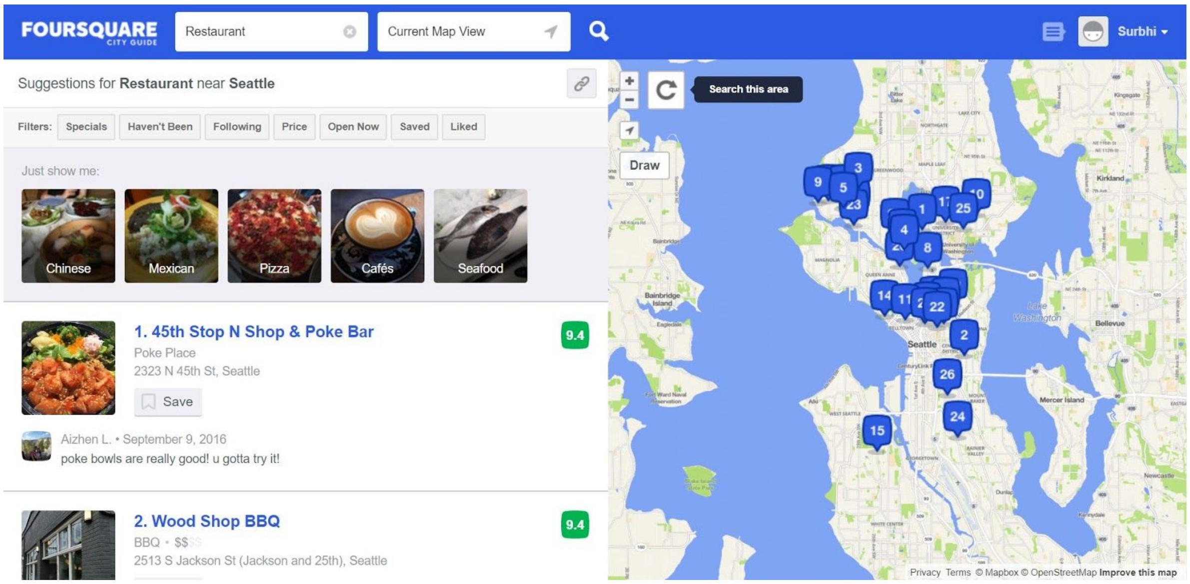


	Neighborhood name	Within larger district	Annexed[41]	Locator map	Street map	Image	Notes
1	**North Seattle**	Seattle	Various				North of the Lake Washington Ship Canal[42]
2	Broadview	North Seattle[42]	1954[43]				[44]
3	Bitter Lake	North Seattle[42]	1954[43]				[45]
4	North Beach / Blue Ridge	North Seattle[42]	1940,[43] 1954[43]				[46]

A new dataframe was created with only the neighborhoods and districts. This new dataframe was also cleaned to remove footnote numbers remaining from the wikipedia page. Additionally, the geopy library was used to identify the latitude and longitude of each neighborhood and coordinates were then added to their respective column and row in the dataframe through a for loop.

	Neighborhood	District	Latitude	Longitude
0	North Seattle	Seattle	47.643724	-122.302937
1	Broadview	North Seattle	47.722380	-122.364980
2	Bitter Lake	North Seattle	47.718680	-122.350300
3	North Beach / Blue Ridge	North Seattle	47.700440	-122.384180
4	Crown Hill	North Seattle	47.695200	-122.374100
5	Greenwood	North Seattle	47.690820	-122.355290
6	Northgate	North Seattle	47.713100	-122.319300
7	Haller Lake	Northgate	47.723200	-122.338700
8	Pinehurst	Northgate	47.718940	-122.314000
9	North College Park (Licton Springs)	Northgate	47.699140	-122.339680
10	Maple Leaf	Northgate	47.700130	-122.317650

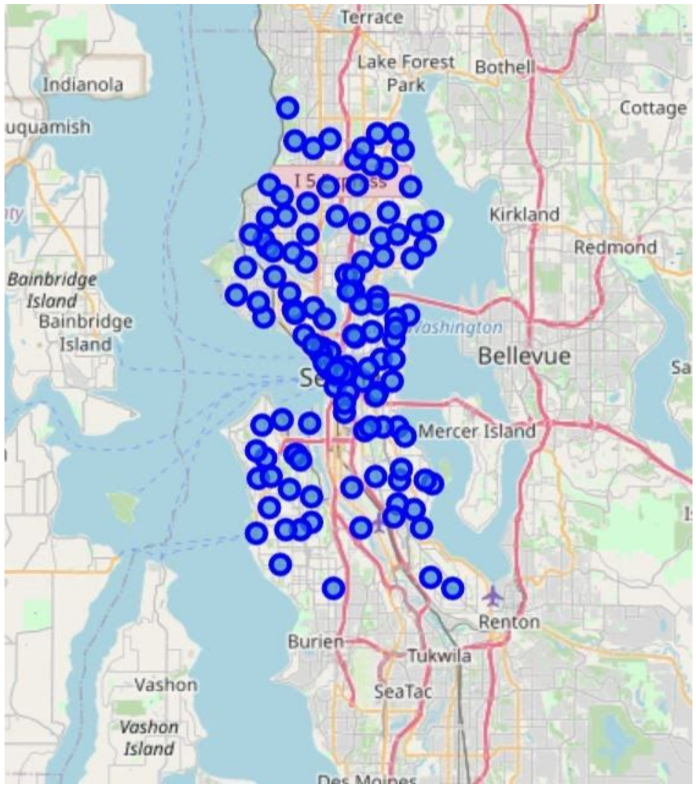
Foursquare API was utilized to extract information on attractions in different neighborhoods in Seattle. The data leveraged by the API included information on venues such as restaurants, theaters, cinemas, malls, etc. It will also be used to categorize venues and obtain their location for visualization purposes.



Methodology

Visualization

The GeoPy library was used to obtain latitude and longitude for neighborhoods in Seattle to visualize the maps of the locations using Folium. Visualization allows us to look at the map of Seattle and get an idea of how spread the neighborhoods are and which areas contain more neighborhoods.



Foursquare API

Foursquare API is a dataset tool used to access accurate data location. By passing the Client ID, Client Secret, and Version credentials and location information, this service provides a list of venues with locations, ID, categories and names for the query entered.

For the purpose of this project, a function *getNearbyVenues* was used to access the data location for nearby businesses and attractions for each neighborhood. The function had parameters: neighborhood, latitude, longitude, and radius, obtained from our data frame. Nearby venues for each of these neighborhoods was appended into a list that was inserted into the dataframe, *seattle_venues*.

Most Common Venues

In order to group all the information in a more understandable way, the dataframe was grouped by categories, which were already classified in the dataframe. The category instances were then averaged to create a frequency. Next, the frequencies were rounded and sorted in descending order to allow us to see the top 10 nearby venues of each neighborhood based on frequency.

K-Means Algorithm

Lastly, to be able to group neighborhoods into areas based on similarities, the K-means algorithm was used to create clusters. This algorithm is an unsupervised learning technique, which is used to find patterns in grouping data. By definition, unsupervised learning is machine learning that uses unlabeled datasets without much human supervision. By looking at the different clusters, we can then conclude which cluster and neighborhoods would be more ideal for a new restaurant.

Results

Finding the frequency using the most common venues algorithm allows us to look at specific types of attractions present in existing neighborhoods. For example, in the following figure we see the neighborhood, Adams, has more burger joints and bars as they are listed with the top two frequencies.

```
----Adams----
               venue  freq
0       Burger Joint  0.08
1                Bar  0.05
2    Thai Restaurant  0.05
3             Bakery  0.05
4        Coffee Shop  0.05
5        Candy Store  0.03
6   Tapas Restaurant  0.03
7                Gym  0.03
8          Bookstore  0.03
9        Supermarket  0.03
```

After inserting this information into the neighborhood_venues_sorted dataframe, it is easier to identify the most common venues in order for each neighborhood.

	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	Adams	Burger Joint	Bakery	Bar	Thai Restaurant	Coffee Shop	Indian Restaurant	Ice Cream Shop	Sri Lankan Restaurant	Restaurant	Pub
1	Alki Point	Scenic Lookout	Park	Convenience Store	Women's Store	Field	Exhibit	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market
2	Arbor Heights	Spa	Women's Store	Filipino Restaurant	Exhibit	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field
3	Atlantic	Coffee Shop	Vietnamese Restaurant	Bakery	Pharmacy	Seafood Restaurant	Sandwich Place	Bank	BBQ Joint	Park	Skate Park
4	Ballard	Mexican Restaurant	Cocktail Bar	Bakery	Italian Restaurant	Sushi Restaurant	Ice Cream Shop	Burger Joint	New American Restaurant	Thai Restaurant	Dessert Shop

Finally, the neighborhoods are grouped into 5 clusters with similarities. Therefore, based on the clusters, we can identify a particular cluster where the owner should start their new restaurant.

Cluster 1

In [22]:
```
seattle_final.loc[seattle_final['Cluster Labels'] == 0, seattle_final.columns[[1] + list(range(5, seattle_final.shape[1]))]]
```

Out[22]:

	District	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	Seattle	Bus Stop	Coffee Shop	Grocery Store	Park	Library	American Restaurant	Italian Restaurant	Canal	Trail	Bike Shop
2	North Seattle	Marijuana Dispensary	Automotive Shop	Convenience Store	Hardware Store	Food Truck	Fast Food Restaurant	Furniture / Home Store	Steakhouse	Gas Station	Sushi Restaurant
4	North Seattle	Pizza Place	Coffee Shop	Sports Bar	Greek Restaurant	Playground	Pet Store	Rock Club	Park	Sandwich Place	Bus Station
5	North Seattle	Coffee Shop	Mexican Restaurant	Bar	Bookstore	Spa	Pizza Place	Indian Restaurant	Comic Shop	Steakhouse	Sports Bar
6	North Seattle	Arts & Crafts Store	Yoga Studio	Jewelry Store	Sporting Goods Shop	Financial or Legal Service	Supermarket	Taco Place	Greek Restaurant	Pet Service	Kids Store
7	Northgate	Soccer Field	Chinese Restaurant	Vietnamese Restaurant	Bakery	Pool	Tennis Court	Gym	Juice Bar	Track	Athletics & Sports
8	Northgate	Pizza Place	Coffee Shop	Indian Restaurant	Bakery	Grocery Store	Liquor Store	Park	Rock Club	Sandwich Place	Smoke Shop
9	Northgate	Movie Theater	Asian Restaurant	Sandwich Place	Chinese Restaurant	Tanning Salon	Martial Arts School	Bus Station	Thai Restaurant	Coffee Shop	Shipping Store

Cluster 2

In [23]:
```
seattle_final.loc[seattle_final['Cluster Labels'] == 1.0, seattle_final.columns[[1] + list(range(5, seattle_final.shape[1]))]]
```

Out[23]:

	District	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
1	North Seattle	Trail	Concert Hall	Women's Store	Filipino Restaurant	Exhibit	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant
11	North Seattle	Trail	Women's Store	Ethiopian Restaurant	Exhibit	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field

Cluster 3

```
In [24]: seattle_final.loc[seattle_final['Cluster Labels'] == 2, seattle_final.columns[[1] + list(range(5, seattle_final.shape[1]))]]
```

Out[24]:

	District	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
18	North Seattle	Park	Bus Stop	Women's Store	Field	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Filipino Restaurant
42	Magnolia	Bus Stop	Women's Store	Filipino Restaurant	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field	Financial or Legal Service

Cluster 4

```
In [25]: seattle_final.loc[seattle_final['Cluster Labels'] == 3, seattle_final.columns[[1] + list(range(5, seattle_final.shape[1]))]]
```

Out[25]:

	District	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
3	North Seattle	Park	Café	Photography Studio	Garden Center	Women's Store	Field	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market
13	Lake City	Park	Trail	Playground	Field	Event Space	Exhibit	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market
16	Lake City	Garden	Marijuana Dispensary	Beach	Playground	Park	Women's Store	Field	Eye Doctor	Fabric Shop	Falafel Restaurant
17	North Seattle	Park	Supermarket	Coffee Shop	Video Store	Steakhouse	Flower Shop	Exhibit	Eye Doctor	Fabric Shop	Falafel Restaurant
24	Windermere / North Seattle	Trail	Arts & Crafts Store	Nature Preserve	Playground	Park	Women's Store	Exhibit	Eye Doctor	Fabric Shop	Falafel Restaurant
28	North Seattle	Park	Seafood Restaurant	Boat or Ferry	Café	Food Truck	Gourmet Shop	Diner	Pizza Place	Bus Line	Tapas Restaurant
31	North Seattle	Park	Spa	Mexican Restaurant	General Travel	Gas Station	Music Store	Grocery Store	Karaoke Bar	Pharmacy	Coffee Shop
40	Central Seattle	Harbor / Marina	Park	Burger Joint	Scenic Lookout	Bus Station	Bus Line	Seafood Restaurant	Sporting Goods Shop	Financial or Legal Service	Fish & Chips Shop

Cluster 5

```
In [26]: seattle_final.loc[seattle_final['Cluster Labels'] == 4, seattle_final.columns[[1] + list(range(5, seattle_final.shape[1]))]]
```

Out[26]:

	District	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
76	Downtown	Park	Women's Store	Filipino Restaurant	Eye Doctor	Fabric Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field	Financial or Legal Service

Discussion

The information gathered helps us use common knowledge and outside information to determine the advantages and disadvantages of each cluster.
We see a lot of differences in both size and common venues between each cluster.

According to the article "Tips on Where to Locate Your Restaurant", restaurant owners should find a location that has high traffic. While it is important for there to be other restaurants in the area to gain target customers, the location should also not have too many competing restaurants. Additionally, the neighborhoods in the clusters should have complementary venues more common, such as movie theaters or malls. In other words, restaurants profitize more when there are other attractions that customers can spend time at before or after they eat.

Looking at the results we found, clusters one and four have more traffic and seem to be more advantageous for a restaurant owner to look at compared to two, three, and five. Looking further into clusters one and four, cluster four seems to have more green area with the most common nearby venues being parks, gardens, and trails in those neighborhoods. In cluster one there seems to be more entertainment such as cafes, movie theaters, and other restaurants as the most common venues.

Thus, it would be most advantageous for a new restaurant owner to open a restaurant up in cluster one of Seattle as there are more attractions to complement a restaurant and high traffic levels in that location.

Conclusion

Using the information obtained from this project, future business owners can strategically make decisions on where to open up a new restaurant. By using this data gathered in the code and research from the internet, we were able to recommend a condensed area in Seattle, cluster one, as it had beneficial characteristics. Furthermore, with more detailed information, such as the type of restaurant opening, the owner can make a more condensed location by looking at specific neighborhoods in cluster one. The owner would similarly pick a neighborhood with complementing venues. However, he or she would want to choose the neighborhood with a similar demographic yet no restaurants with the same idea present. Additionally, they would take into account other external factors such as availability and rent.