Applied Data Science Capstone

Final Project Report

Battle of the Neighbourhoods-Asian Delicacies in Seattle

Submitted By:

Snehav Sharma

Introduction/Business Problem

As the number of people migrating to the city of Seattle is at its peak due to its affinity towards tech culture, it has led to a significant increase in the Asian population in the new technology hub of the United States. This has led many businessmen in the food industry to show interest in expanding their franchise or opening new Asian cuisine restaurants in the city. Therefore, the problem to be addressed in this project is to find out the most popular neighbourhood in the city of Seattle to open an Asian Cuisine restaurant that would help the businessmen to make an informed decision.

Data description

Requirements

To find the solution to our business problem it is imperative that we have plenty of data to analyse and recommend the optimal location for an Asian restaurant. To analyse and come up with a solution, we firstly need the latitudes and longitudes for all the neighbourhoods of the city of Seattle, next we would find the restaurants that are there within a radius of around 500 feet of those location in order to establish similarity between the neighbourhoods.

Data Collection

The data used for this project is acquired from the three following sources -

Seattlearea.com - A website named Seattle-area, this is a local website for the city of Seattle which has all the information about the city including neighbourhoods, their zip codes etc. We scraped the text from this website in the form of crucial information for our project, i.e. neighbourhood names and their respective zip codes.

Link - http://seattlearea.com/zip-codes/

Neighborhood Zipcodes

» 98003 - Federal Way
» 98005 - Bellevue
» 98033 - Kirkland
» 98037 - Lynnwood
» 98040 - Mercer Island
» 98052 - Redmond
» 98055 - Renton
» 98101 - Seattle
» 98101 - Downtown
» 98102 - Capital Hill
» 98103 - Greenwood
» 98103 - Freemont
» 98103 - Greenlake

This data was then scraped from the website using the python library Beautiful Soup. Pin codes and neighbourhoods were saved in separate lists first and then a data frame was prepared using those two lists, following is the data frame generated using that –

	PostalCode	Neighborhood
0	98003	Federal Way
1	98005	Bellevue
2	98033	Kirkland
3	98037	Lynnwood
4	98040	Mercer Island

Geopy - Next, I used the python library geopy to extract the coordinates of all the neighbourhoods of the city of Seattle for further processing and data pre-processing. We used the neighbourhood names from the above table to find their co-ordinates using geopy and nominatim application of “ny explorer”. Generating the following addresses for those neighbourhoods-

```
USPS Seattle Network Distribution Center, Federal Way, King County, Washington, USA
Bellevue Place, Capitol Hill, Seattle, King County, Washington, USA
Consulate General of Sweden, Seattle, 5350, Carillon Point, Juanita, Kirkland, King County, Washington, 98033, USA
Seattle Heights, Snohomish County, Washington, 98043, USA
Seattle Public Utilities Water Redistribution Center, 89th Avenue Southeast, Mercer Island, King County, Washington, 98040, USA
Armenian Apostolic Church of Seattle, Woodinville-Redmond Road Northeast, Redmond, King County, Washington, 98052, USA
Renton Avenue South, Rainier Beach, Rainier Valley, Seattle, King County, Washington, 98178, USA
Seattle, Colman Dock, West Edge, Belltown, Seattle, King County, Washington, 98174, USA
Downtown Seattle Transit Tunnel, Marion Street, West Edge, First Hill, Seattle, King County, Washington, 98164, USA
```

Then I extracted the latitudes and longitudes of these addresses using the geopy library, and then these co-ordinates were added to the existing tables in the following format –

	PostalCode	Neighborhood	Latitude	Longitude
0	98003	Federal Way	47.295780	-122.325464
1	98005	Bellevue	47.627073	-122.327626
2	98033	Kirkland	47.654516	-122.205614
3	98037	Lynnwood	47.810375	-122.324020
4	98040	Mercer Island	47.568357	-122.220895
5	98052	Redmond	47.703507	-122.132372
6	98055	Renton	47.510732	-122.260317
7	98101	Seattle	47.602896	-122.339841
8	98101	Downtown	47.604949	-122.333274
9	98102	Capital Hill	14.669678	121.098831
10	98103	Greenwood	47.690981	-122.354877
12	98103	Greenlake	47.676003	-122.322149
13	98104	International District	47.602896	-122.339841

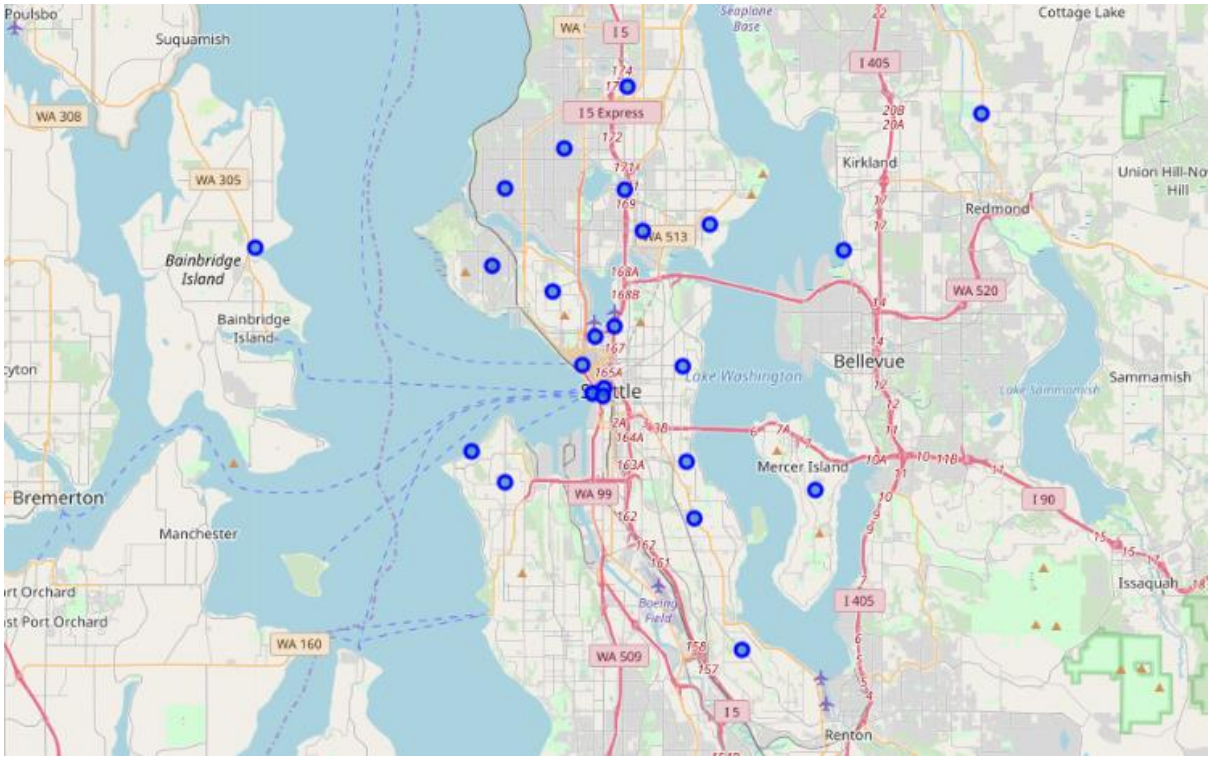
Foursqaure API - I used this API to find out all the popular food/leisure locations within a 500 feet radius of each neighbourhood's location to further narrow down our search for the best neighbourhood to open an Asian cuisine restaurant. After calling the venue information from this API, the following data frame was generated –

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Federal Way	47.295780	-122.325464	CrossFit Federal Way	47.298880	-122.323581	Gym
1	Federal Way	47.295780	-122.325464	Gymnastics Unlimited	47.296597	-122.319283	Gymnastics Gym
2	Bellevue	47.627073	-122.327626	Barjot	47.625701	-122.326748	Café
3	Bellevue	47.627073	-122.327626	Top Pot Doughnuts	47.624630	-122.325470	Donut Shop
4	Bellevue	47.627073	-122.327626	Harry's Fine Foods	47.624402	-122.326771	Restaurant
5	Bellevue	47.627073	-122.327626	Sol Liquor Lounge	47.624605	-122.325434	Cocktail Bar
6	Bellevue	47.627073	-122.327626	Single Shot	47.624688	-122.325431	American Restaurant
7	Bellevue	47.627073	-122.327626	The Lookout	47.626724	-122.326752	Bar
8	Bellevue	47.627073	-122.327626	Silver Cloud Inn Seattle - Lake Union	47.629610	-122.329098	Hotel
9	Bellevue	47.627073	-122.327626	Pro Sports Club	47.623538	-122.329503	Gym
10	Bellevue	47.627073	-122.327626	I Love Sushi	47.628939	-122.331844	Sushi Restaurant

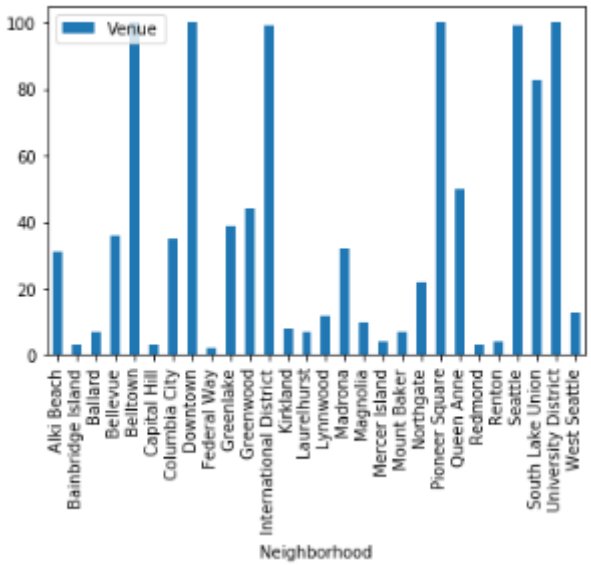
Methodology

Exploratory Analysis

After extracting data from different sources and combining them into a single table, it becomes important to be curious enough to explore it and look at its features. Exploration is very important to generate insightful observations to get initial ideas about using it to solve our problem. The following map was generated based on neighbourhoods –



Also, after analysing the number of restaurants per neighbourhood, it was found out that the neighbourhood of university district had the most number of popular venues was University district, as can be seen from the following bar plot –



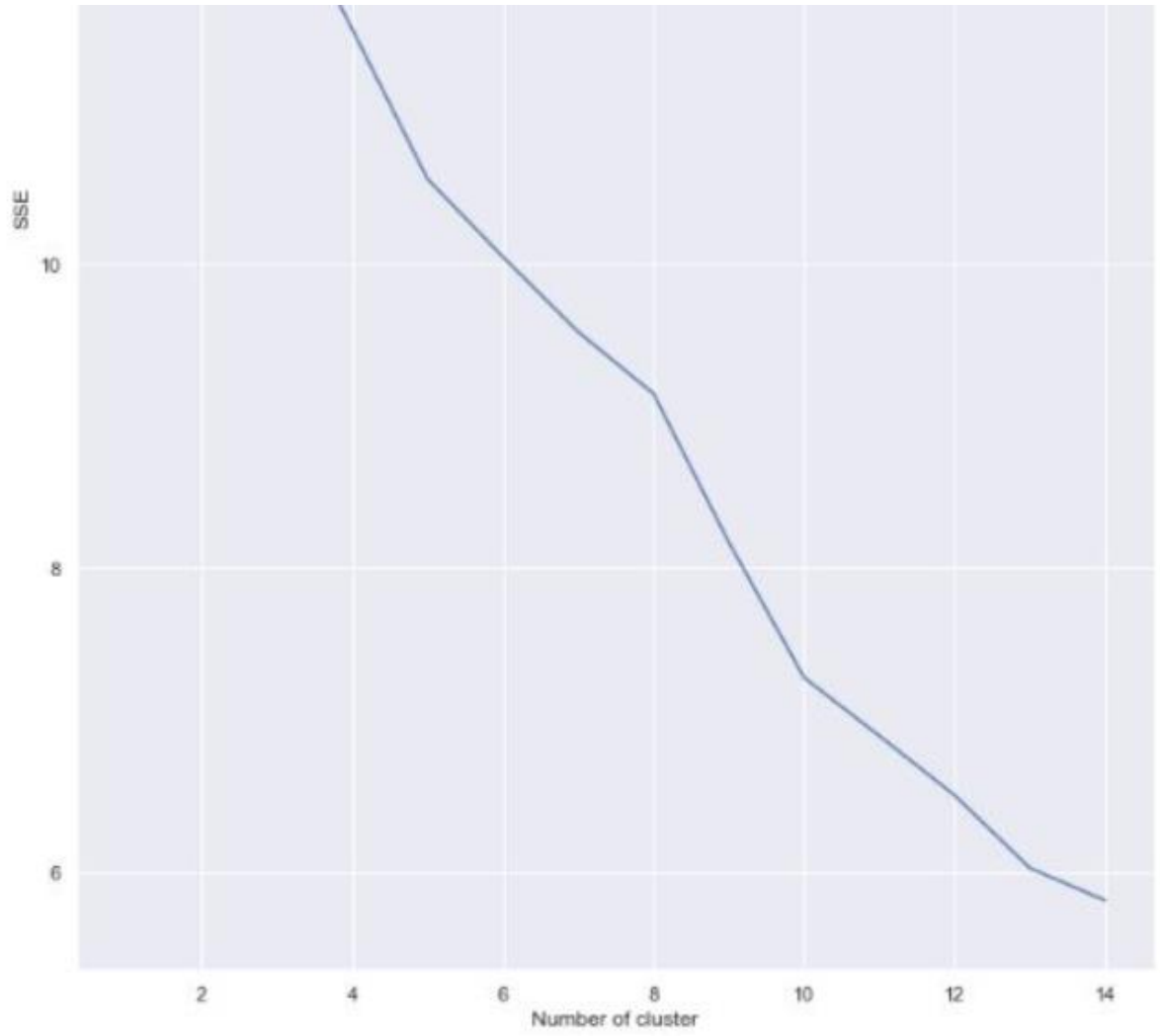
Cluster Analysis

After exploring the data to find that the university district was the most popular neighbourhood having the maximum number of neighbourhoods, so it seemed logical to find out the most frequent

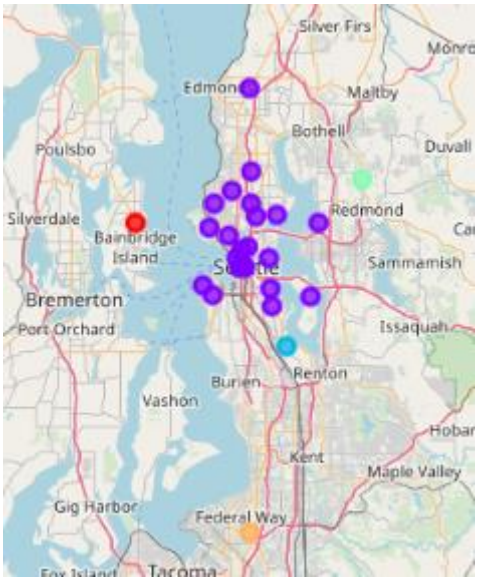
categories of each neighbourhood using the following matrix with each neighbourhood with respect to each venue category –

	Neighborhood	Yoga Studio	ATM	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Video Store	Vietnamese Restaurant	Volleyball Court	Warehouse Store	Waterfront	Weight Loss Center
0	Alki Beach	0.000000	0.00	0.000000	0.000000	0.000000	0.032258	0.000000	0.000000	0.032258	...	0.000000	0.032258	0.032258	0.000000	0.000000	0.000000
1	Bainbridge Island	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Ballard	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.142857	0.000000	0.000000	0.000000	0.000000	0.000000
3	Bellevue	0.027778	0.00	0.000000	0.027778	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Belltown	0.000000	0.00	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000
5	Capital Hill	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Columbia City	0.000000	0.00	0.057143	0.028571	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.057143	0.000000	0.000000	0.000000	0.000000

Now we can use the data in the above table to make clusters of these neighbourhoods using k-means, taking k=5, after this was implemented using the value k based on the following elbow graph –



It gave out 5 clusters of neighbourhoods with common attributes of category of neighbourhoods. Each of which is depicted in the following map –



Then these clusters were analysed on the bases of their venue categories and it was found that in cluster number 4 seen in the following table had majority of Asian restaurants and was situated near downtown and university area with the neighbourhood names as the University district being the most optimal one having mostly Asian restaurant in the top 10 most popular venues in that neighbourhood –

	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue
1	98005	Bellevue	47.627073	-122.327626	1	Bar	Hotel	Seafood Restaurant	Park	Restaurant	Harbor / Marina	Coffee Shop	Yoga Studio	Music Venue	Sushi Restaurant
2	98033	Kirkland	47.654516	-122.205614	1	Coffee Shop	Wine Bar	Hotel	French Restaurant	American Restaurant	Restaurant	Seafood Restaurant	Harbor / Marina	Falafel Restaurant	Electronics Store
3	98037	Lynnwood	47.810375	-122.324020	1	Korean Restaurant	Mexican Restaurant	Latin American Restaurant	Bar	Donut Shop	BBQ Joint	Pharmacy	Greek Restaurant	Japanese Restaurant	Warehouse Store
4	98040	Mercer Island	47.568357	-122.220895	1	Park	Stadium	Playground	Track Stadium	Eastern European Restaurant	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant
7	98101	Seattle	47.602896	-122.339841	1	Seafood Restaurant	Coffee Shop	Gift Shop	Boat or Ferry	Park	New American Restaurant	Theme Park Ride / Attraction	Pier	Bar	Café
8	98101	Downtown	47.604949	-122.333274	1	Hotel	Coffee Shop	New American Restaurant	Seafood Restaurant	Cocktail Bar	Café	Korean Restaurant	Italian Restaurant	Scenic Lookout	Breakfast Spot
9	98102	Capital Hill	14.669678	121.098831	1	Dim Sum Restaurant	Pizza Place	Basketball Court	Women's Store	Electronics Store	Food Truck	Food & Drink Shop	Food	Fish & Chips Shop	Fast Food Restaurant
10	98103	Greenwood	47.690981	-122.354877	1	Coffee Shop	Bar	Mexican Restaurant	Pizza Place	Theater	Spa	Mediterranean Restaurant	Cheese Shop	Bridal Shop	Brewery
12	98103	Greenlake	47.676003	-122.322149	1	Coffee Shop	Spa	Vegetarian / Vegan Restaurant	Burger Joint	Yoga Studio	Electronics Store	Frozen Yogurt Shop	Thai Restaurant	Salon / Barbershop	Organic Grocery
13	98104	International District	47.602896	-122.339841	1	Seafood Restaurant	Coffee Shop	Gift Shop	Boat or Ferry	Park	New American Restaurant	Theme Park Ride / Attraction	Pier	Bar	Café
14	98104	Pioneer Square	47.602139	-122.333927	1	Coffee Shop	Cocktail Bar	Café	Hotel	Seafood Restaurant	Donut Shop	Italian Restaurant	Salad Place	Indian Restaurant	Gift Shop
15	98105	University District	47.661298	-122.313152	1	Coffee Shop	Chinese Restaurant	Sandwich Place	Bubble Tea Shop	Café	Korean Restaurant	Indian Restaurant	Thrift / Vintage Store	Asian Restaurant	Mexican Restaurant

Results and Conclusions

The result of the project was that if businessman wants to open an Asian restaurant in the city of Seattle, the optimal location for that is the neighbourhood of **University district** due to the fact that it was in the cluster having number of Asian restaurants and had the most number of Asian Restaurants in its top 10 most common venue list. It seems reasonable as there seems a large cluster of international students and population in that area. Therefore, that might be the reason for them to have the most number of popular venues too due to a majority of youth population in that area.