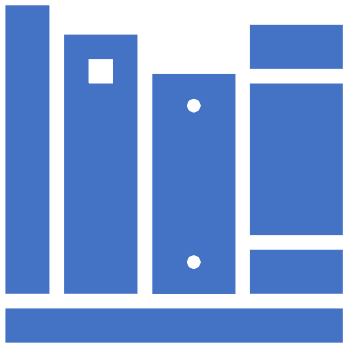


IBM Data Science Capstone Project

The Battle of Neighborhoods: Starting a Chinese Restaurant in Seattle

Jing Yang 03/2021

Seattle, a seaport city on the West Coast of the United States, is full of business opportunities.

Tags of Seattle: #tourismCity #stateParks #education #diversity

Business problem: Where are the best positions to open a Chinese restaurant?

Data Acquisition

1. List of neighborhoods in Seattle from **Wikipedia**: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle

2. Read as dataframe using **BeautifulSoup** and **Requests** packages with Python
Latitude and Longitude data from **geocoder**

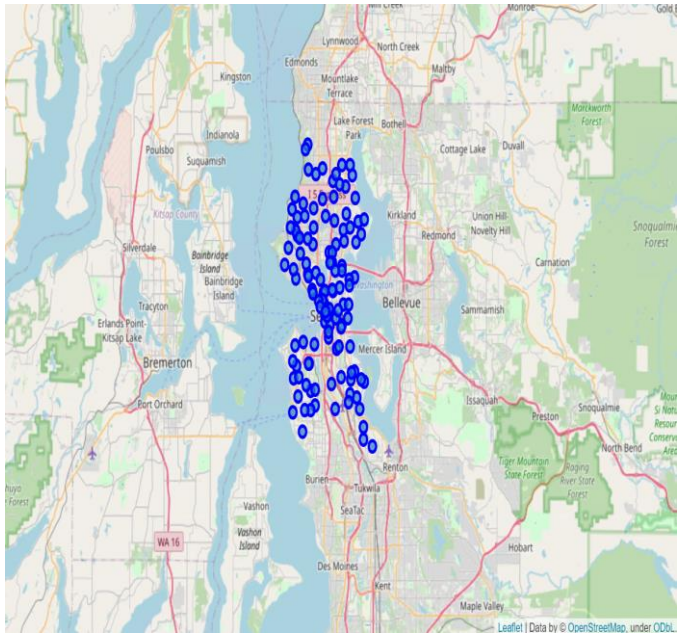
List of districts and neighborhoods [edit]

Despite complications in Seattle's system of neighborhoods and districts, the names and boundaries in the following list are generally accepted and wid used.[37] They are based on the *Seattle City Clerk's Neighborhood Map Atlas*,[38] which in turn is based on a variety of sources, including a 1980 neighborhood map produced by the now-defunct Department of Community Development,[39] Seattle Public Library indexes, a 1984-1986 "Neighborhoc Profiles" feature series in the *Seattle Post-Intelligencer*,[40] numerous park, land use and transportation planning studies, as well as records in the Seattl Municipal Archives.

The following table is largely based on maps from the Seattle City Clerk's Neighborhood Atlas, but also includes designations from other sources.

	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]

	Neighborhood	Latitude	Longitude
0	North Seattle	47.643724	-122.302937
1	Broadview	47.722380	-122.364980
2	Bitter Lake	47.718680	-122.350300
3	North Beach	47.700715	-122.389593
4	Crown Hill	47.695200	-122.374100



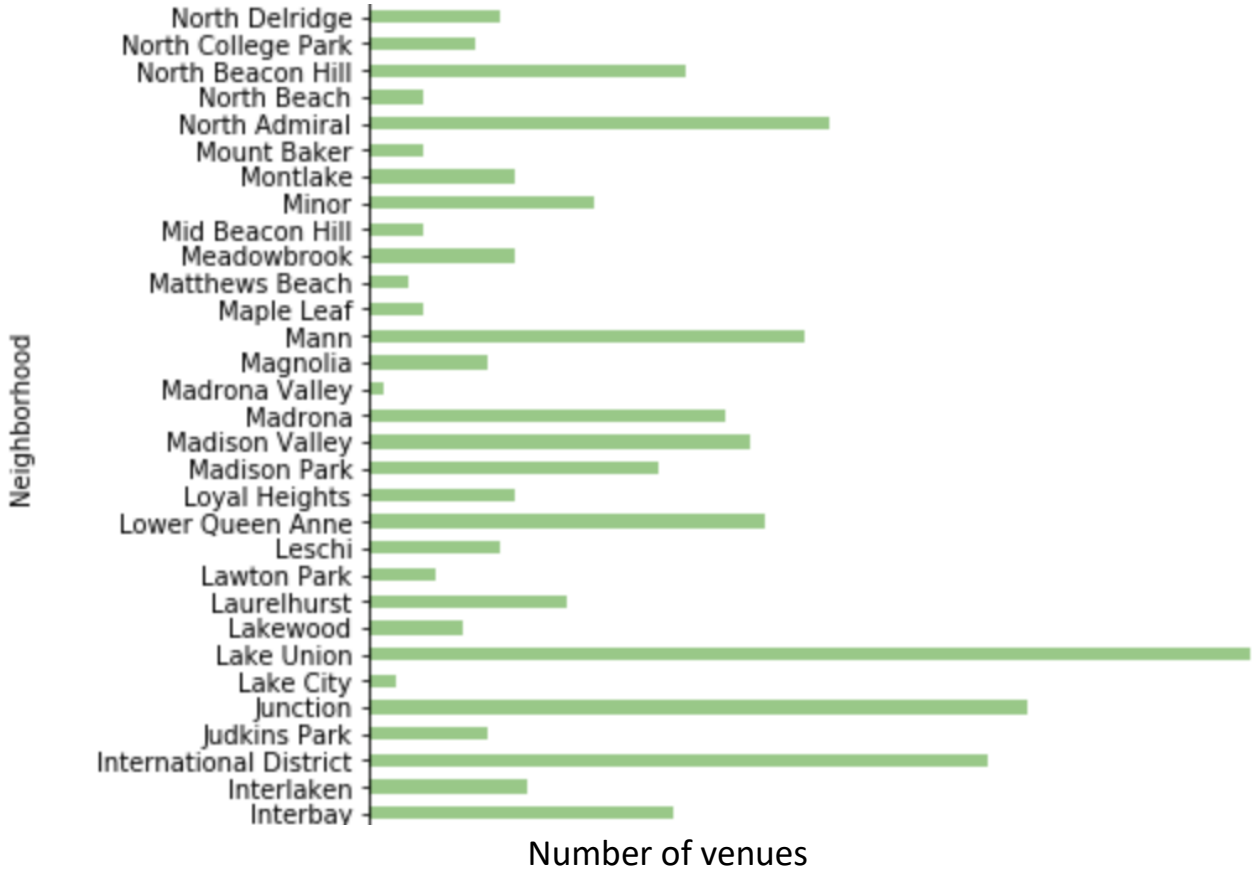
3 All data of venues in each neighborhoods is obtained from **Foursquare API**. There are 127 official neighborhoods and 2970 venues in Seattle

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North Seattle	47.643724	-122.302937	Cafe Lago	47.639698	-122.302256	Italian Restaurant
1	North Seattle	47.643724	-122.302937	Seattle Public Library - Montlake	47.640520	-122.302413	Library
2	North Seattle	47.643724	-122.302937	Montlake Cut	47.647094	-122.304686	Canal
3	North Seattle	47.643724	-122.302937	Fuel Coffee - Montlake	47.639688	-122.302009	Coffee Shop
4	North Seattle	47.643724	-122.302937	Montlake Blvd Market	47.643480	-122.303915	Grocery Store
5	North Seattle	47.643724	-122.302937	Montlake Bicycle Shop	47.639380	-122.302340	Bike Shop
6	North Seattle	47.643724	-122.302937	Traveler Montlake	47.639830	-122.302231	American Restaurant
7	North Seattle	47.643724	-122.302937	Metro Bus Stop #25751	47.644848	-122.304488	Bus Stop
8	North Seattle	47.643724	-122.302937	King County Metro Bus Route 255	47.642409	-122.303858	Bus Line

4 Venue Category is applied with **one-hot encoding** to obtain the frequency of different types of venues

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	venue_ATM	venue_Accessories Store	venue_Acupuncturist	venue_African Restaurant	venue_American Restaurant	.
0	Adams	1716.226920	-4405.953600	1716.135506	-4405.912226	0	0	0	0	0	.
1	Alki Point	237.911772	-611.866435	237.913639	-611.865038	0	0	0	0	0	.
2	Arbor Heights	47.508610	-122.375930	47.510273	-122.376343	0	0	0	0	0	.
3	Atlantic	1047.112220	-2690.649060	1047.163875	-2690.642026	0	0	0	1	0	.

EDA


North Delridge
North College Park
North Beacon Hill
North Beach
North Admiral
Mount Baker
Montlake
Minor
Mid Beacon Hill
Meadowbrook
Matthews Beach
Maple Leaf
Mann
Magnolia
Madrona Valley
Madrona
Madison Valley
Madison Park
Loyal Heights
Lower Queen Anne
Leschi
Lawton Park
Laurelhurst
Lakewood
Lake Union
Lake City
Junction
Judkins Park
International District
Interlaken
Interbay
Neighborhood
Number of venues

Methodology

Feature Selection:

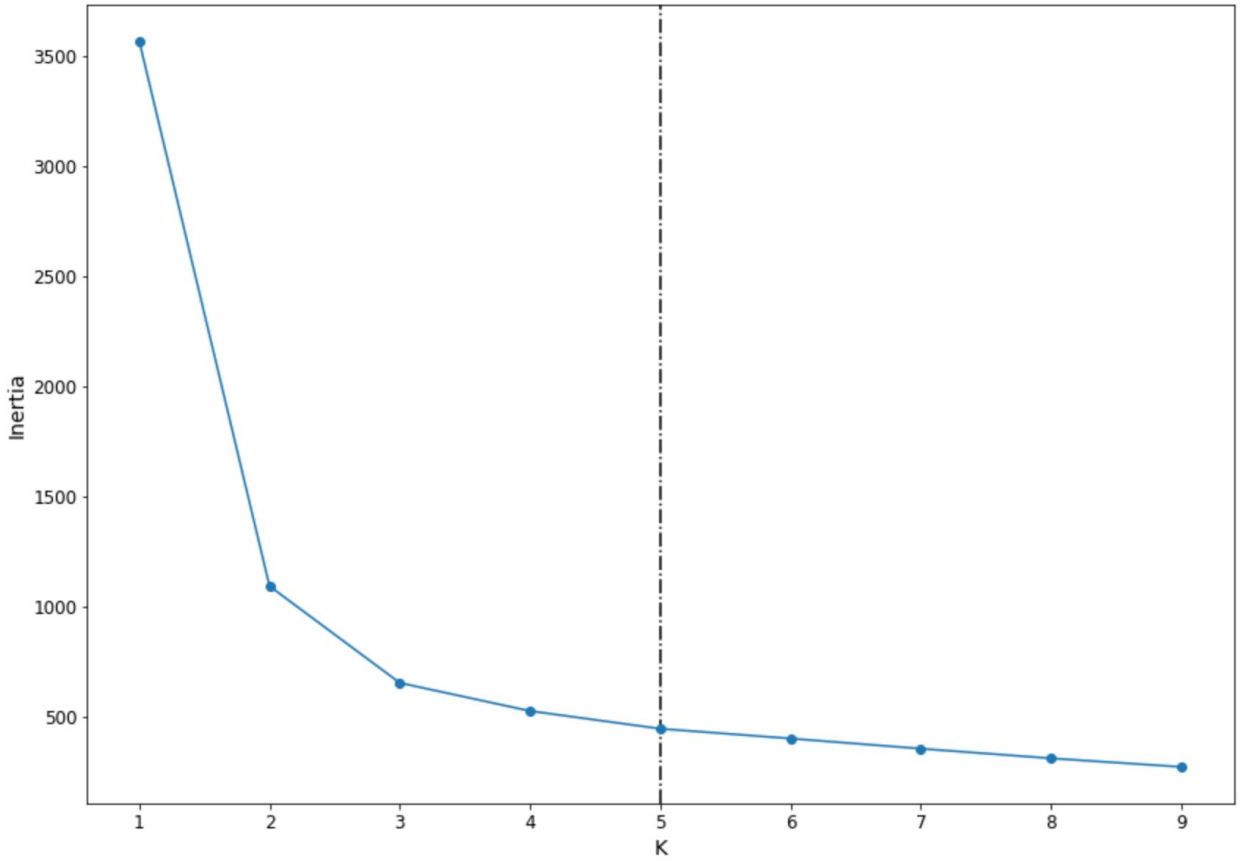
Minus score

Large customer flow: plus score

Methodology

Clustering:

- Based on the selected features
- Exploring number of clusters from 1 to 10
- Elbow Method to select best k
- Final clustering model: best number of clusters



Methodology

Scoring Clusters:

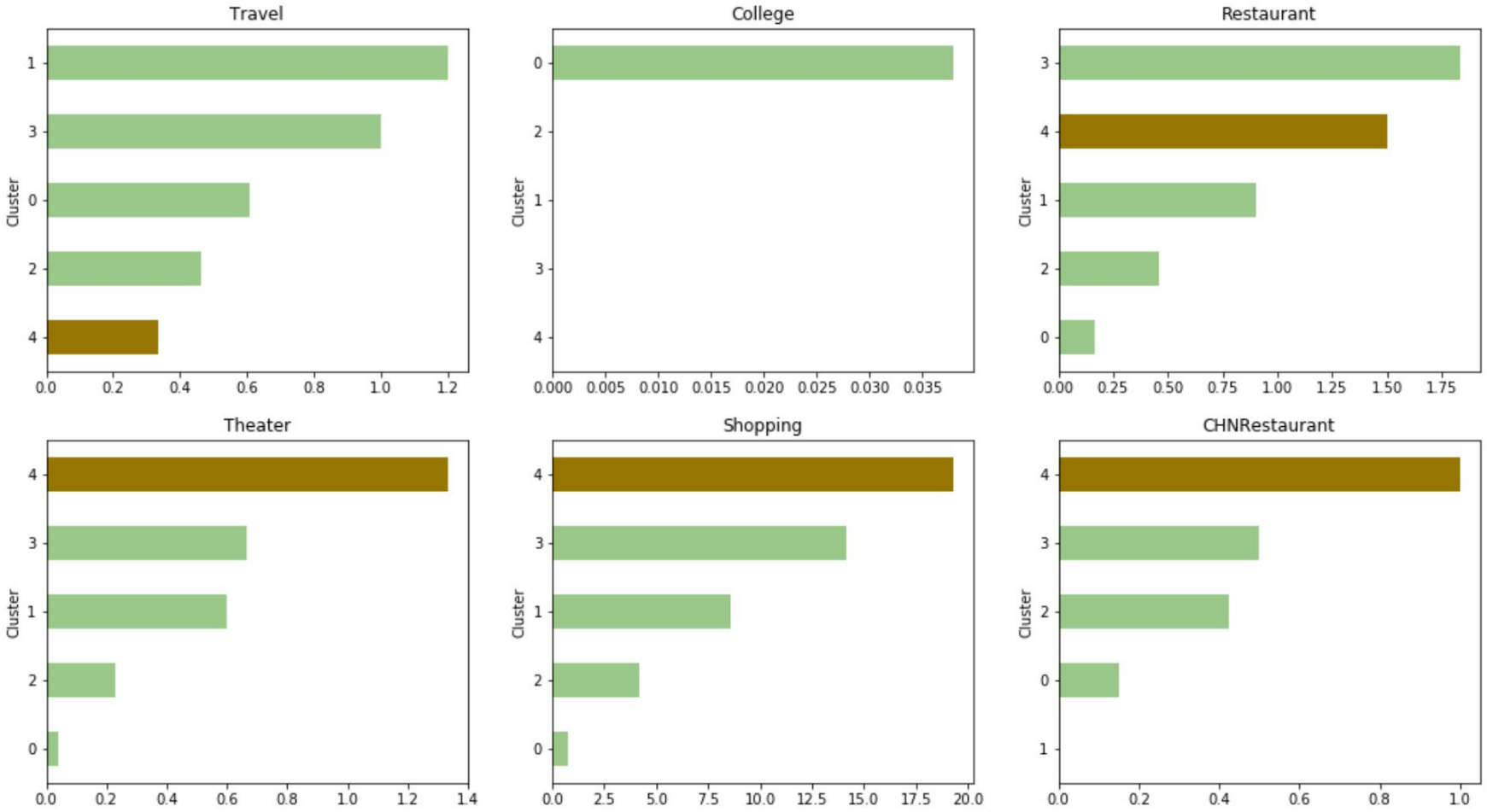
Score = #of restaurants + #of colleges + #of universities + #of shops + #of tourism **– 2*#of Chinese restaurants**

Cluster 4 has the highest score!

Cluster	Score	Travel	Restaurant	Theater	Shopping	College	CHNRestaurant
4	20.500000	0.333333	1.500000	1.333333	19.333333	0.000000	1.000000
3	16.666667	1.000000	1.833333	0.666667	14.166667	0.000000	0.500000
1	11.300000	1.200000	0.900000	0.600000	8.600000	0.000000	0.000000
2	4.500000	0.461538	0.461538	0.230769	4.192308	0.000000	0.423077
0	1.265823	0.607595	0.164557	0.037975	0.721519	0.037975	0.151899

Results

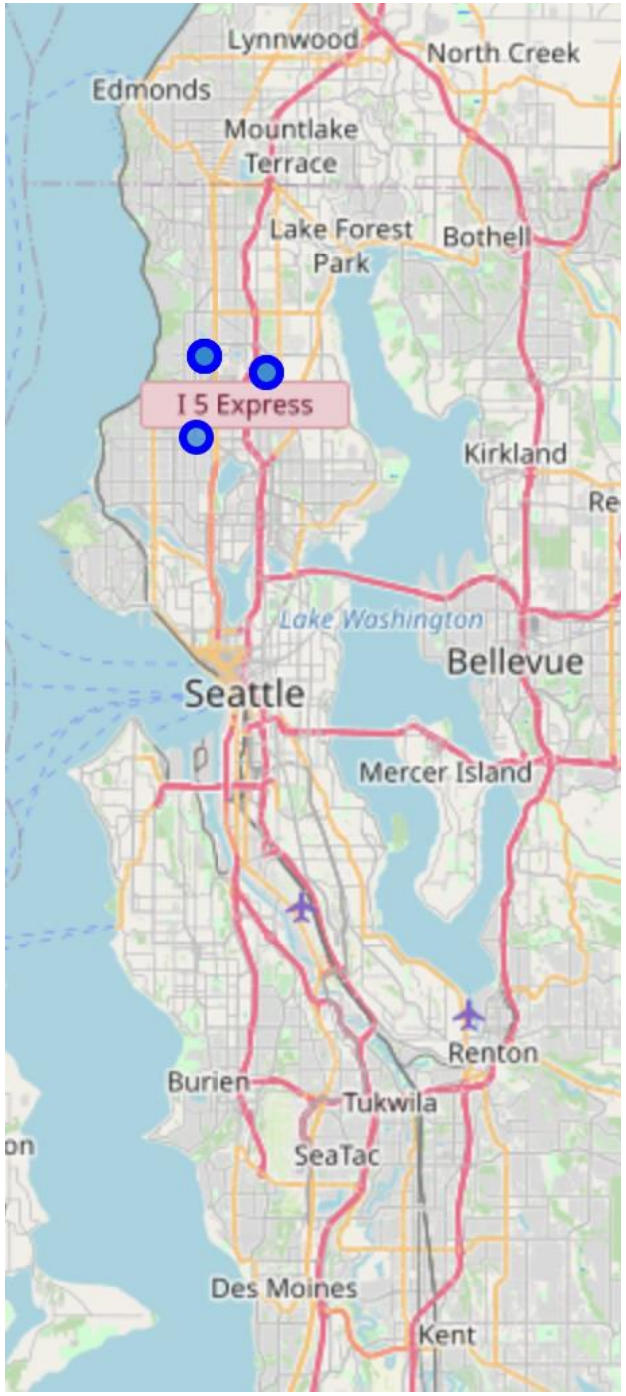
Check the distribution of venues in Cluster 4:



Results

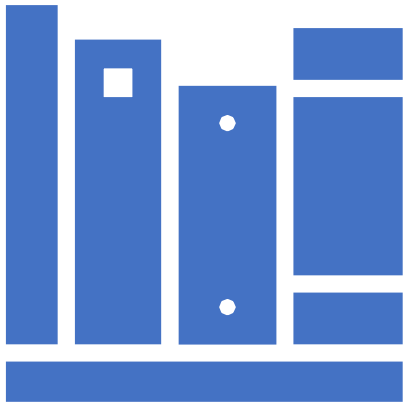
Top 6 neighborhoods in Cluster 4:

	Neighborhood	Latitude	Longitude
0	Bitter Lake	47.71868	-122.35030
1	Northgate	47.71310	-122.31930
2	Bitter Lake	47.71868	-122.35030
3	Bitter Lake	47.71868	-122.35030
4	Greenwood	47.69082	-122.35529
5	Northgate	47.71310	-122.31930



Conclusion & Discussion

- The goal is to find positions of neighborhoods to open a Chinese restaurant in Seattle
- We acquire the detailed data of neighborhoods and venues of Seattle
- We select 6 key features that are more directly related to a Chinese restaurant's position
- Trough k-means clustering with selected features, we segment the neighborhoods into 5 parts
- We score the clusters and choose the highest scored cluster




Thank you!