

Python Capstone Project: New Businesses in Seattle

Introduction/Business Problem

In this project, we will investigate a business scenario using geographical and business data to determine the optimum locations for a new business in Seattle, Washington. They wish to know the best locations for their new premises across the cities districts or neighbourhoods. The businesses are in the leisure market such as restaurants and cinemas.

Clients which may be interested in this scenario include existing chains which are looking to expand their operations into Seattle or new local start-up businesses in Seattle in the leisure or hospitality sectors.

We will use the Foursquare API to examine the Neighbourhoods of Seattle to find the best locations for these businesses, based on the existing businesses and services in those areas. This will allow the business to choose a location to open.

Data






To begin with, we will need a list of the Neighbourhoods/Districts in Seattle.

A list of this kind can be found on Wikipedia at this location:

https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle.

The table looks like this:

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]

We will also require geographic coordinates for the city's districts. We will use the python geopy.geocoders package to return the geographical coordinates (latitude and longitude) for these areas by searching by address.

In order to characterise each neighbourhood, we will use the Foursquare API (<https://foursquare.com/>) to retrieve recent information about these neighbourhoods. Specifically we will use lists of nearby venues in each neighbourhood from Foursquare.

Methodology

Data Import and Clean up

Upon reading the list of Seattle's districts from Wikipedia, we obtain a pandas dataframe that looks like this with **127 rows**:

	Unnamed: 0	Neighborhood name	Within larger district	Annexed[41]	Locator map	Street map	Image	Notes
0	1	North Seattle	Seattle	Various	NaN	NaN	NaN	North of the Lake Washington Ship Canal[42]
1	2	Broadview	North Seattle[42]	1954[43]	NaN	NaN	NaN	[44]
2	3	Bitter Lake	North Seattle[42]	1954[43]	NaN	NaN	NaN	[45]

Some of the Neighbourhood names have references on them, and some of them list two names separated by backslashes (/). We tidy these up and then select just the 'Neighbourhood name' column which we rename 'District'.

Thus we now have a dataframe with a cleaned 'District' column containing each of Seattle's neighbourhoods.

	District
0	North Seattle
1	Broadview
2	Bitter Lake
3	North Beach
4	Crown Hill
5	Greenwood
6	Northgate
7	Haller Lake

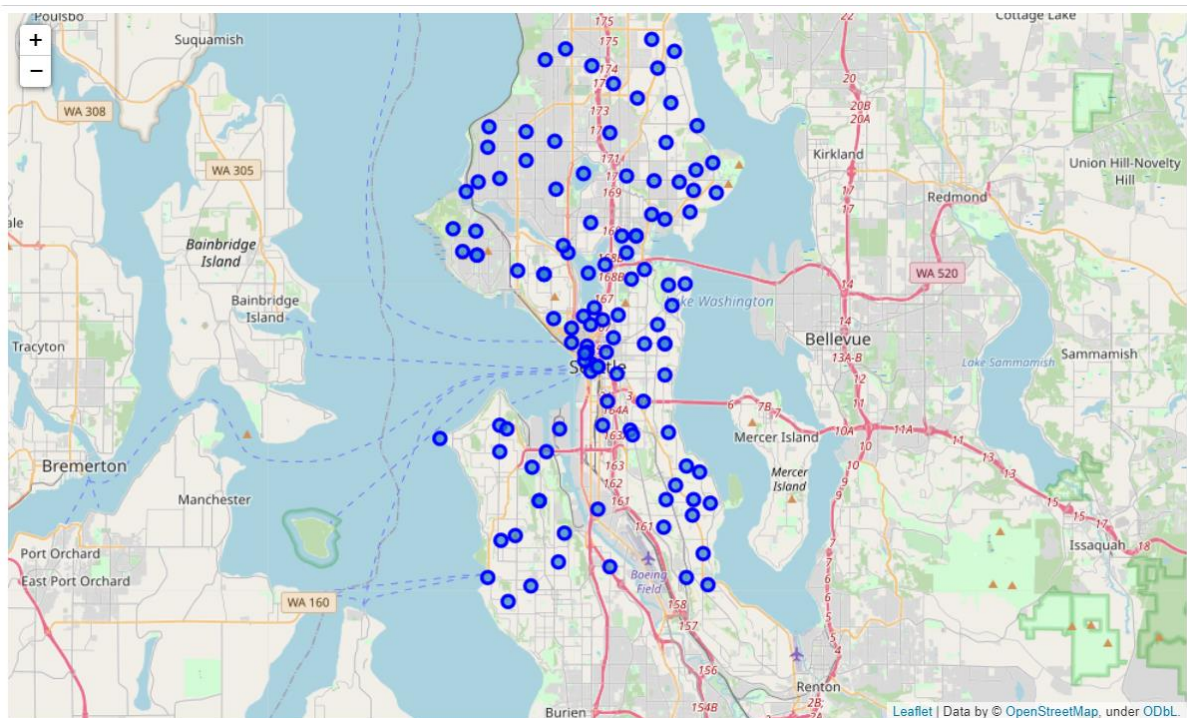
Coordinates

We now need to obtain the geographical coordinates. We can do this using the `geopy.geolocator` package. We call each Neighbourhood address in turn and populate a new dataframe with their latitude and longitude.

	District	Latitude	Longitude
0	North Seattle	47.660773	-122.291497
1	Broadview	47.722320	-122.360407
2	Bitter Lake	47.726236	-122.348764
3	North Beach	47.696210	-122.392362
4	Crown Hill	47.694715	-122.371459
5	Greenwood	47.690981	-122.354877
6	Northgate	47.713153	-122.321231
7	Haller Lake	47.719748	-122.333751

Mapping Seattle

We can now produce a map of Seattle by finding it's coordinates and plotting each neighbourhood onto the map to visualise their distribution.



Foursquare Venues Data

We can now use the Foursquare API to obtain information about the outlets in the vicinity of each neighbourhood centre. We are interested in nearby venues as we can then use this information to decide if there are already too many restaurants in an area to consider opening a new one, for example. The Foursquare API is used to explore each neighbourhood and return the nearby venues. We use a defined function to repeat the process for each neighbourhood and load the data into a new dataframe. The resulting dataframe, listing each neighbourhood and nearby venues looks like this:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North Seattle	47.660773	-122.291497	Burke-Gilman Brewing Company	47.661308	-122.288067	Brewery
1	North Seattle	47.660773	-122.291497	Center for Urban Horticulture	47.657978	-122.290237	College Science Building

We observe 310 unique Venues.

Clustering

In order to do clustering analysis, we need to use one hot encoding to create numerical data that can be analysed in this manner. Clustering is a method of examining how similar or different data points are to other data points. One-hot encoding labels venues in a particular neighbourhood as 1, and shows 0 if there is none of that venue in a particular neighbourhood. Following this process we can group by each neighbourhood and get the average frequency that venues appear in each neighbourhood. The result is a dataframe that looks like this:

	Neighborhood	Zoo Exhibit	ATM	Accessories Store	African Restaurant	Airport Service	R
0	Adams	0.0	0.000000	0.00	0.000000	0.0000	0
1	Alki Point	0.0	0.000000	0.00	0.000000	0.0000	0
2	Arbor Heights	0.0	0.000000	0.00	0.000000	0.0000	0
3	Atlantic	0.0	0.000000	0.00	0.000000	0.0000	0
4	Ballard	0.0	0.000000	0.00	0.000000	0.0000	0
5	Beacon Hill	0.0	0.000000	0.00	0.000000	0.0000	0
6	Belltown	0.0	0.000000	0.00	0.000000	0.0000	0
7	Bitter Lake	0.0	0.038462	0.00	0.000000	0.0000	0

Finally we can extract the top 10 most common venues in each neighbourhood and store in a dataframe like this:

	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	Park	Dog Run	Soccer Field	Zoo	Electronics Store	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair
1	Alki Point	Beach	Lighthouse	Baseball Field	Baseball Stadium	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Farm
2	Arbor Heights	Park	Farm	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Zoo

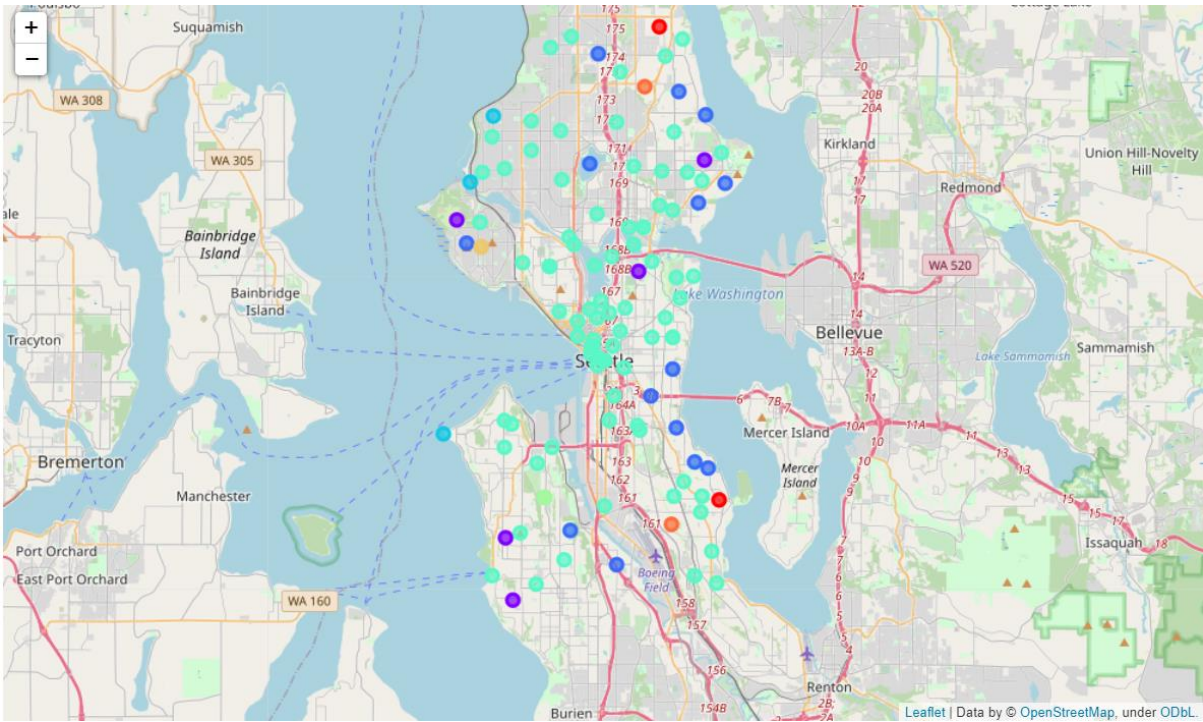
We can now carry our K-means clustering on the neighbourhoods to determine which neighbourhoods have similar top venues and what the top venues are in each area.

Results

Following the clustering we have a similar dataframe with each neighbourhood assigned to one of 8 cluster labelled 0-7.

	Neighborhood	Latitude	Longitude	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	7th Most Common Venue
0	North Seattle	47.660773	-122.291497	4	Burger Joint	Donut Shop	Salon / Barbershop	Golf Driving Range	Park	Coffee Shop	Chinese Restaurant
1	Broadview	47.722320	-122.360407	4	Thai Restaurant	Sushi Restaurant	Convenience Store	Pizza Place	Antique Shop	Food Truck	Video Store
2	Bitter Lake	47.726236	-122.348764	4	Fast Food Restaurant	Hotpot Restaurant	Thrift / Vintage Store	Thai Restaurant	Bank	Szechuan Restaurant	Mobile Phone Shop

We can produce a map of the neighbourhoods and mark the clustering assigned by colour.



We can print out lists of Neighbourhoods and top venues in each cluster, for example in this format:

	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
14	Olympic Hills	Middle Eastern Restaurant	Farmers Market	Zoo	Fast Food Restaurant	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Farm
91	Seward Park	Beach	Music Venue	Bus Station	Farmers Market	Zoo	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Farm

Discussion

Our clients are restaurants and cinema business looking for new locations. Let's look at the clusters and select good locations for them.

Cluster 1:

The first cluster contains two neighbourhoods. While there are some food outlets, including the most common venue in Olympic Hills, there is not much variety and fast food outlets are common. This neighbourhood appears to be served well with food outlets already.

	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
14	Olympic Hills	Middle Eastern Restaurant	Farmers Market	Zoo	Fast Food Restaurant	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Farm
91	Seward Park	Beach	Music Venue	Bus Station	Farmers Market	Zoo	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Farm

Cluster 2:

Parks dominate cluster 2. There are some Ethiopian and Falafel restaurants, but could be demand for other type of restaurants to cater for park users.

	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
17	View Ridge	Park	Construction & Landscaping	Farm	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant
53	Interlaken	Park	Playground	Scenic Lookout	Zoo	Falafel Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop
82	South End	Park	Trail	Dry Cleaner	Electronics Store	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair
107	Gatewood	Park	Gym	Pet Store	Coffee Shop	Filipino Restaurant	Financial or Legal Service	Electronics Store	Ethiopian Restaurant	Event Space	Exhibit
109	Arbor Heights	Park	Farm	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Zoo

Cluster 3:

Cluster three is similar to cluster 2 and is dominated by parks and lakes. There is some more variety of restaurants but again there could be a market for more choices.

Sample data:

	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
7	Haller Lake	Lake	Playground	Park	Coffee Shop	Dance Studio	Fire Station	Falafel Restaurant	Event Space	Fish & Chips Shop	Exhibit
12	Matthews Beach	Park	Trail	Playground	Zoo	Falafel Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop
13	Meadowbrook	Lake	Pizza Place	Park	Coffee Shop	Pool	Financial or Legal Service	Fair	Ethiopian Restaurant	Fire Station	Event Space

Cluster 4:

This cluster contains 3 neighbourhoods. While there are some food outlets, including the most common venue in Seaview, the most common venues here are recreational. These neighbourhoods could be suitable for new restaurants. This could also be a good location for cinemas.

	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
3	North Beach	Beach	Park	Scenic Lookout	Electronics Store	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant
101	Alki Point	Beach	Lighthouse	Baseball Field	Baseball Stadium	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Farm
104	Seaview	Seafood Restaurant	Beach	Scenic Lookout	Harbor / Marina	Surf Spot	Snack Place	Bar	Caribbean Restaurant	Outdoor Supply Store	Cuban Restaurant

Cluster 5:

Cluster 5 is the largest cluster with a wide variety of neighbourhoods and venues. Many of these neighbourhoods have a variety of restaurant types in the top 10 venues. Therefore it is suggested that these area are well served by food outlets and new business could struggle to compete with established premises in this area.

Sample data:

	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
0	North Seattle	Burger Joint	Donut Shop	Salon / Barbershop	Golf Driving Range	Park	Coffee Shop	Chinese Restaurant	Gym	College Science Building
1	Broadview	Thai Restaurant	Sushi Restaurant	Convenience Store	Pizza Place	Antique Shop	Food Truck	Video Store	Furniture / Home Store	Beer Bar
2	Bitter Lake	Fast Food Restaurant	Hotpot Restaurant	Thrift / Vintage Store	Thai Restaurant	Bank	Szechuan Restaurant	Mobile Phone Shop	Fabric Shop	Café
4	Crown Hill	Coffee Shop	Fast Food	Bus Station	Taco Place	Sandwich	Burger	Grocery	Pet Store	Sports Bar

Cluster 6:

Cluster 6 is essentially made up of one neighbourhood. There are a number of restaurant categories in the top 10 venues here, and as such would not be recommended as a location for new restaurants.

	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
110	Delridge	Gas Station	Pet Service	Park	Pizza Place	Restaurant	Trail	Vietnamese Restaurant	Zoo	Dumpling Restaurant	Electronics Store
111	North Delridge	Gas Station	Pet Service	Park	Pizza Place	Restaurant	Trail	Vietnamese Restaurant	Zoo	Dumpling Restaurant	Electronics Store
115	South Delridge	Gas Station	Pet Service	Park	Pizza Place	Restaurant	Trail	Vietnamese Restaurant	Zoo	Dumpling Restaurant	Electronics Store

Cluster 7:

Cluster 7 is essentially made up of one neighbourhood. There are no restaurant categories in the top 10 venues here, and there a number of event/exhibit spaces. Therefore this could be a good location for new restaurants. This could also be a good location for cinemas.

	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
38	Magnolia	Video Store	Pool	Bus Stop	Locksmith	Farm	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair
41	Southeast Magnolia	Video Store	Pool	Bus Stop	Locksmith	Farm	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair

Cluster 8:

Cluster 8 has two neighbourhood with few restaurants. Therefore this could be a good location for new restaurants.

	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
15	Victory Heights	Playground	Marijuana Dispensary	Zoo	Farm	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant
95	Holly Park	Playground	Zoo	Farmers Market	Event Space	Exhibit	Eye Doctor	Fabric Shop	Fair	Falafel Restaurant	Farm

Clusters 4, 7 and 8 contains significant event or exhibit spaces and are suggested as good locations for possible cinema businesses.

Conclusion

To conclude, in this report, we have examined the neighbourhoods of Seattle in order to examine area for new business to locate. We sourced neighbourhood names from Wikipedia and used geopy and Foursquare to obtain geographic coordinates and information on nearby venues respectively. We have clustered the neighbourhoods using the top 10 venues in each area. We suggest that clusters 2,3,4,7 and 8 could be good locations for new restaurants and that clusters 4,7 and 8 would be good locations for new cinema businesses.