Identifying suitable neighbourhoods to establish coffee shops in the city of Seattle

1. Introduction/Business Understanding

A client Coffee Inc. is looking to set up operations in the Seattle City, United States, the company would like to open several coffee shops around Seattle and wishes identify which neighbourhoods in Seattle are suitable for this purpose.

The conditions given for a neighbourhoods to be suitable are as follows:

Food and beverage venues (restaurants, bars, pubs etc..) but mainly restaurants must be common in these neighbourhoods, and that there must be a below average number of coffee venues(cafés and coffee shops). This will be necessary in order to avoid excessive competition which will be crucial for survival in the early stages of establishment. Identifying neighbourhoods that meet these requirements will provide a good starting point for further investigation into the specifics of which locations to choose in order to set up the coffee shops, these would be locations that are near groups of restaurants that do not have a large number of coffee shops close by.

Setting up shops in areas with a high number of restaurants and other food venues and beverage venues will be beneficial as the areas are already known as food districts with an established trade.

2. Data to be used

The neighbourhoods of Seattle will be acquired from a Wikipedia web page listing said neighbourhoods, these will be read into a dataframe and the latitude and longitude coordinates for each neighbourhood will added to the dataframe using the python's geoencoder package.

Location data from Foursquare.com will be used in combination with the created dataframe for two purposes:

1. To find nearby coffee shop and café venues for each neighbourhood
2. To find nearby venues of all types for each neighbourhood

Finding the coffee venues for each neighbourhood will allow the number of nearby coffee venues to be determined. From this it can be determined which neighbourhoods have a below average number of nearby coffee venues.

Finding nearby venues of all types will allow the determining of which type of venues are most common in each neighbourhood. Then the neighbourhoods where restaurants and other food and beverage venues are common can be selected from this dataframe.

3. Methodology

3.1 K-means Clustering

K- means clustering is an unsupervised machine learning algorithm that clusters data points in an unlabelled dataset based on data point similarity/dissimilarity. Data points within a cluster will have a high similarity (intra-cluster similarity) and data points between clusters will have a low

similarity(inter-cluster similarity). The algorithm tries to maximise intra-cluster similarity and minimise inter-cluster similarity.

K-means is used here to explore which neighbourhoods are similar to each other. K-means is used under the assumption that there will be neighbourhoods that have a lot of food and beverage venues, enough to be common. This is a simple way grouping neighbourhoods by similarity of venues. Neighbourhoods that are within the same cluster will be similar to each other.

In this case k-means will is used to cluster neighbourhoods based on their 10 most common venues grouped by venue category(bakery, brewery, convenience store etc.). K-means will separate the neighbourhoods into clusters based on what venues that occur most commonly in those neighbourhoods. The clusters with a high prevalence of food and beverage venues(with a focus on restaurants) with be selected for the next stage in the analysis.

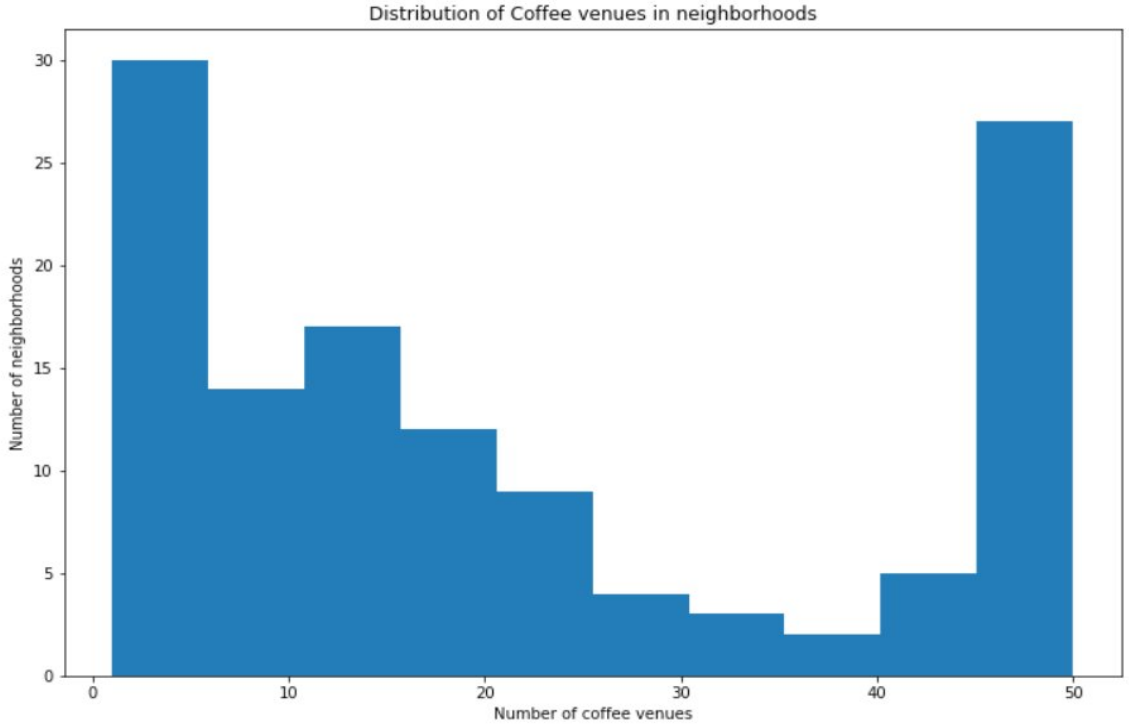
3.2 Comparative analysis

The neighbourhoods in the selected clusters will be compared to the dataframe containing the neighbourhoods with below average number of coffee venues, neighbourhoods that occur in both dataframes will be both in suitable clusters and have a below average number of coffee venues, from this pool of neighbourhoods the suitable neighbourhoods can be selected.

4. Results

4.1 Neighbourhoods with coffee venue count below the median

Initially the idea was to find neighbourhoods with a below average number of coffee venues but looking at the data distribution of coffee venue count for each neighbourhood, the data is quite asymmetrical meaning there are a lot of extreme values in the data which can skew the mean, when data distribution is asymmetrical it is generally better to use median instead the mean which will be the case here. The median number of coffee venues in this case will be 16.



4.2 K-means Clustering

The analysis with k-means clustering yielded the following clusters:

4.2.1 Cluster 1

This cluster seems to have a fair amount of food and beverage venues particularly Asian restaurants, but also some cafés, but looks like it it might contains suitable neighbourhoods. This Cluster will be select for further analysis

Cluster Labels	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	Atlantic	Vietnamese Restaurant	Coffee Shop	Chinese Restaurant	Park	Asian Restaurant	Grocery Store	Pizza Place	BBQ Joint	Thrift / Vintage Store	Performing Arts Venue
0	Beacon Hill	Coffee Shop	Pub	Bakery	Light Rail Station	Brewery	Fast Food Restaurant	Pizza Place	Playground	Pharmacy	Taco Place
0	Bitter Lake	Sushi Restaurant	Hotel	Fast Food Restaurant	Hobby Shop	Bakery	Automotive Shop	Thai Restaurant	Soccer Field	Athletics & Sports	Asian Restaurant
0	Brighton	Vietnamese Restaurant	Convenience Store	Thai Restaurant	Café	Pub	Asian Restaurant	Bank	Fast Food Restaurant	Taco Place	Chinese Restaurant
0	Broadmoor	Garden	Café	Park	Bar	Bank	Spa	Trail	Gas Station	Sushi Restaurant	Dry Cleaner

4.2.2 Cluster 2

This cluster some have convenience stores, recreational venues and open spaces as common venues, there are some food venues but there don't seem to be quite enough, this cluster will not be selected.

Cluster Labels	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
1	Arbor Heights	Convenience Store	Park	Bowling Alley	Bus Stop	Construction & Landscaping	Home Service	Donut Shop	Dry Cleaner	Dumpling Restaurant	Eastern European Restaurant
1	Delridge	Convenience Store	Baseball Field	Park	Food Truck	Gas Station	Residential Building (Apartment / Condo)	Dive Bar	Hardware Store	Bus Line	Zoo
1	Riverview	Convenience Store	Coffee Shop	Park	Baseball Field	Sandwich Place	Dive Bar	Gas Station	Bus Line	Farmers Market	Falafel Restaurant

4.2.3 Cluster 3

This cluster has a variety of restaurants which are widely occurring, there are other common food venues and beverage venues such as bars, however there are also a high number of coffee venues, nevertheless this is a suitable cluster for further analysis.

Cluster Labels	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
2	Adams	Burger Joint	Ice Cream Shop	Bar	Cocktail Bar	Coffee Shop	Mexican Restaurant	Park	Sushi Restaurant	Bakery	New American Restaurant
2	Alki Point	Beach	Coffee Shop	Art Gallery	History Museum	Baseball Stadium	Outdoor Sculpture	Donut Shop	Sandwich Place	Baseball Field	Scenic Lookout
2	Ballard	Coffee Shop	Ice Cream Shop	Bakery	Burger Joint	Thai Restaurant	Vietnamese Restaurant	Cocktail Bar	Gym	Bar	Dessert Shop
2	Belltown	Bar	Coffee Shop	Hotel	Sushi Restaurant	Pizza Place	Sculpture Garden	Bakery	Movie Theater	Gym	Scenic Lookout
2	Broadway	Coffee Shop	Pizza Place	Cocktail Bar	Ramen Restaurant	Mexican Restaurant	Café	Ice Cream Shop	Sandwich Place	Thai Restaurant	Diner
2	Capitol Hill	Coffee Shop	Cocktail Bar	American Restaurant	Bar	Bakery	Italian Restaurant	Thrift / Vintage Store	Garden	Yoga Studio	Indian Restaurant
2	Cascade, Seattle	Coffee Shop	Café	Hotel	Pizza Place	Mexican Restaurant	Bar	Sandwich Place	Bakery	Steakhouse	Restaurant
2	Central Business District	Hotel	Coffee Shop	Seafood Restaurant	American Restaurant	Pizza Place	Bakery	Italian Restaurant	Brewery	Sushi Restaurant	Grocery Store
2	Central Waterfront	Coffee Shop	Brewery	Cocktail Bar	Mexican Restaurant	Bar	Sushi Restaurant	New American Restaurant	Ice Cream Shop	Sandwich Place	Pizza Place
2	Cherry Hill	Coffee Shop	Gay Bar	Burger Joint	Mexican Restaurant	French Restaurant	Donut Shop	Thai Restaurant	Bagel Shop	Ramen Restaurant	Ethiopian Restaurant

4.2.4 Cluster 4

There only one neighbourhood in this cluster and its not suitable, there not enough food/beverage venues

Cluster Labels	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	Windermere	Bus Stop	Greek Restaurant	Park	Bank	Trail	Automotive Shop	Pizza Place	Indie Movie Theater	Beach	Food Truck

4.2.5 Cluster 5

In this cluster open public spaces are common there are some food and beverage venues but there don't seem to be quite enough to meet the requirements

Cluster Labels	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
4	Briarcliff	Park	Playground	Trail	Grocery Store	Pool	Breakfast Spot	Athletics & Sports	Coffee Shop	Video Store	Bus Stop
4	Fauntleroy	Boat or Ferry	Park	Playground	Pier	Pilates Studio	Market	Beach	Bus Station	Scenic Lookout	American Restaurant
4	Gatewood	Park	Pizza Place	Coffee Shop	Bar	Building	Beer Store	Japanese Restaurant	Optical Shop	Gas Station	Bank
4	High Point	Park	Gas Station	Gym / Fitness Center	Storage Facility	Trail	Restaurant	Asian Restaurant	Furniture / Home Store	Food Truck	Rental Car Location
4	Highland Park	Gas Station	Park	Garden	Convenience Store	Vietnamese Restaurant	Trail	Bus Station	Restaurant	Bakery	Coffee Shop
4	Lakewood	Park	Harbor / Marina	Lake	Pet Store	Beach	Organic Grocery	Café	Dog Run	Restaurant	Grocery Store
4	Lawton Park	Park	Trail	Coffee Shop	Brewery	Breakfast Spot	Grocery Store	Greek Restaurant	Bus Stop	Japanese Restaurant	Chiropractor
4	Matthews Beach	Trail	Park	Golf Course	Playground	Locksmith	Zoo	Event Space	Dry Cleaner	Dumpling Restaurant	Eastern European Restaurant
4	North Beach	Beach	Park	Trail	Construction & Landscaping	Coffee Shop	Dog Run	Gluten-free Restaurant	Athletics & Sports	Mexican Restaurant	Playground

4.2.6 Cluster 6

This cluster has only one neighbourhood and its not suitable, there a some restaurants but not enough.

Cluster Labels	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
5	Rainier View	Bakery	Gun Range	Lounge	Zoo	Dry Cleaner	Eastern European Restaurant	Electronics Store	Entertainment Service	Ethiopian Restaurant	Event Space

4.3 Selecting suitable neighbourhoods

Selecting cluster 1 and 3, It can now can be determined which neighbourhoods have coffee venue counts below the median which also fall within suitable clusters, these neighbourhoods will be the pool of candidates from which the suitable neighbourhoods are selected, the selection criteria for a neighbourhood to be suitable are as follows:

1) A neighbourhood must have have at least 2 restaurants in the top 4 most common venues

2) A neighbourhood must have at least 5 food and/or beverage venues(excluding coffee shops and cafés) in the top 10 most common venues

5. Discussion and Conclusion

Applying the above mentioned criteria, 7 suitable neighbourhoods met the requirements:

Number of coffee venues	Cluster Labels	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	0	Columbia City	Park	Pizza Place	Ethiopian Restaurant	Ice Cream Shop	Pub	Soccer Field	Bar	Mexican Restaurant	Bakery	Convenience Store
15	2	Genesee	Coffee Shop	Pizza Place	Gym	Burger Joint	Furniture / Home Store	Mexican Restaurant	Brewery	Asian Restaurant	American Restaurant	BBQ Joint
15	2	Greenwood	Coffee Shop	Mexican Restaurant	Pizza Place	Bar	Mediterranean Restaurant	Greek Restaurant	Chinese Restaurant	Automotive Shop	Spa	Playground
13	0	Holly Park	Vietnamese Restaurant	Grocery Store	Chinese Restaurant	Mexican Restaurant	Convenience Store	Coffee Shop	Pizza Place	Cajun / Creole Restaurant	Bakery	Bank
3	0	Lake City	Mexican Restaurant	Vietnamese Restaurant	Coffee Shop	Bank	Gas Station	Sushi Restaurant	Thai Restaurant	Breakfast Spot	Café	Bakery
4	0	Olympic Hills	Bus Station	Thai Restaurant	Bank	Chinese Restaurant	Gymnastics Gym	Vietnamese Restaurant	Bar	Beer Store	Mexican Restaurant	Gaming Cafe
9	0	Renton Hill	Pizza Place	Vietnamese Restaurant	Coffee Shop	Post Office	Japanese Restaurant	BBQ Joint	Thai Restaurant	Bank	Szechuan Restaurant	Food Truck

Looking these neighbourhoods and their venues on a map, areas with clusters of food/beverage seem to be fairly common, often these clusters are accompanied by a fair amount of coffee venues as it is common place to have coffee venues near food venues such as restaurants, however there are food venue clusters which do not have any coffee venues nearby or have only a small number of nearby coffee venues, it is recommended that these clusters are selected as the primary areas for further investigation into finding suitable locations to establish coffee shops, with secondary areas being locations with clusters of food venues but with higher coffee venue count.

All neighbourhoods with the exception of Columbia city and Greenwood have primary clusters. Lake City and Olympic hills share a primary cluster as they are adjacent neighbourhoods. Holly Park, Columbia City, Greenwood, and Genesee all have secondary clusters for further investigation.