IBM PROFESSIONAL DATA SCIENCE

CAPSTONE PROJECT REPORT

INTRODUCTION

Chinatown/Penn Quarter is one of Washington DC's most popular neighbourhoods for dining because of its' central location and easy access to many of the city's biggest attractions including the Verizon Center, the National Portrait Gallery & American Art Museum and the International Spy Museum. Chinatown is just a few blocks away and the Washington Convention Center is also within walking distance. The Gallery Place-Chinatown Metro station makes this area easily accessible from across the city. Restaurants in the area offer a wide range of cuisine from contemporary American to Asian Fusion, to Italian or Latin American fare. Chinatown has approximately 20 Chinese and Asian restaurants.

To begin with I am importing all the necessary libraries and installing all the necessary packages. I will be primarily using Pandas, scikit learn, Matplotlib, Geocoders, Foursquare API, Folium

Business Problem – A tourist visiting an unknown city for the first time is often looking for places to eat which are popular in that area but due to unfamiliarity with the set of cuisines and preferences of local people is often confused about which are the best restaurants that he must explore.

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as euclidean-based distance or correlation-based distance. The decision of which similarity measure to use is application-specific.

DATASET USED AND ITS SOURCE

I have primarily worked with Foursquare API to obtain data of all the venues in Chinatown, Washington DC. In addition to that I have used the same data to obtain their longitude and latitude information as well as the categories of venues. I have also used information of likes to obtain reviews of people for the venues

For this assignment, I will be using the Foursquare API to pull the following location data on restaurants in Chinatown, Washington:

1) Venue Name
2) Venue ID
3) Venue Location
4) Venue Category
5) Count of Likes

For data acquisition I am using below methods:

- Get geolocator lat and long coordinates for Chinatown, Washington

- Use Foursquare API to get a list of all venues in Chinatown
- Get venue name, venue ID, location, category, and likes

After data acquisition using foursquare and geolocator, I will do preprocessing on data to remove the categories which are not Food Joints and then categorize them into 5 major food categories based on the data. I will then using histograms create bins to identify five categories of restaurant ratings based on user likes. I will then create a k-means clustering algorithm that will group restaurants into 5 clusters.

METHODOLOGY

In order to determine the popular restaurants in Chinatown, Washington, we will be focusing on finding major restaurant clusters in the area.

Obtain Location

To begin with we will be using geolocator to obtain the location – longitude and latitude coordinates of the Chinatown in Washington DC

OBTAIN NEARBY VENUES

- We will then use the Foursquare API to obtain all the nearby venues to the location coordinates
- The Foursquare API's explore method will return the Location IDs, name, category, and tips information as a json

```
[201]: {'meta': {'code': 200, 'requestId': '5f09a6128a993d2972e04d06'},
        'response': {'suggestedFilters': {'header': 'Tap to show:',
          'filters': [{'name': 'Open now', 'key': 'openNow'},
           {'name': '$-$$$$', 'key': 'price'}]},
         'headerLocation': 'Chinatown',
         'headerFullLocation': 'Chinatown, Washington',
         'headerLocationGranularity': 'neighborhood',
         'totalResults': 115,
         'suggestedBounds': {'ne': {'lat': 38.9048422045, 'lng': -77.01567152474095},
          'sw': {'lat': 38.895842195499995, 'lng': -77.02721450878485}},
         'groups': [{'type': 'Recommended Places',
           'name': 'recommended',
           'items': [{'reasons': {'count': 0,
              'items': [{'summary': 'This spot is popular',
                'type': 'general',
                'reasonName': 'globalInteractionReason'}]},
             'venue': {'id': '58dee3dc54386d4959b23de0',
              'name': 'Poké Papa',
              'location': {'address': '806 H St NW',
               'lat': 38.89951518,
               'lng': -77.02341047,
               'labeledLatLngs': [{'label': 'display',
                 'lat': 38.89951518,
                 'lng': -77.02341047}],
               'distance': 193,
               'postalCode': '20001',
               'cc': 'US',
               'city': 'Washington',
               'state': 'D.C.',
               'country': 'United States',
               'formattedAddress': ['806 H St NW',
                'Washington, D.C. 20001',
                'United States']},
              'categories': [{'id': '5bae9231bedf3950379f89d4',
                'name': 'Poke Place',
                'pluralName': 'Poke Places',
```

READ DATA IN DATAFRAME

- We will read the above data into pandas dataframe which will make it easier to process in the other libraries like scikit and folium
- Clean the data to obtain meaningful Headers
- Read Venue Categories

[203]:

	venue.name	venue.id	venue.categories	venue.location.lat	venue.location.lng
0	Poké Papa	58dee3dc54386d4959b23de0	Poke Place	38.899515	-77.023410
1	SoulCycle Mount Vernon	55db1c30498e410a92af9ed0	Cycle Studio	38.901930	-77.020469
2	CAVA	54231977498eff43caedb3d5	Mediterranean Restaurant	38.899971	-77.022314
3	RPM Italian	575ae890498e04099a6d1c67	Italian Restaurant	38.902220	-77.020940
4	La Colombe Coffee Roasters	5648c247498e93b42e8897c8	Coffee Shop	38.901054	-77.020102
5	Sixth & I Historic Synagogue	4a9f0416f964a5203a3c20e3	Synagogue	38.900884	-77.020171
6	Nando's	4a4eb693f964a520f4ae1fe3	Portuguese Restaurant	38.900370	-77.021942
7	Dolcezza Gelato	552d5a73498e337582e351bf	Ice Cream Shop	38.900201	-77.024218
8	Taco Bamba	5ac710302b984427e9f885c7	Taco Place	38.901064	-77.022682
9	Chaia	5c34e803dab4b1002c5c4653	Taco Place	38.901015	-77.020934
10	BIBIBOP Asian Grill	5942c26fdec1d635b48cf514	Asian Restaurant	38.898887	-77.021876
11	Zaytinya	3fd66200f964a520d5f11ee3	Mediterranean Restaurant	38.898811	-77.023645
12	Legal Sea Foods	4a561b68f964a52072b41fe3	Seafood Restaurant	38.898795	-77.021853
13	Apple Carnegie Library	5cd4ef3b8b98fd002cb4e11d	Electronics Store	38.902510	-77.022970
14	Capital One Arena	446d8fa5f964a52067331fe3	Basketball Stadium	38.898177	-77.020816
15	Shouk	572a235c498ee06f5eaceb9d	Vegetarian / Vegan Restaurant	38.902737	-77.020600
16	Boqueria	5bdb1cb08fb09e002df5c82e	Tapas Restaurant	38.899416	-77.023755
17	Daikaya	50905502d63e87c2d3448e35	Ramen Restaurant	38.898538	-77.019643
18	Flywheel Sports	5769cf9b498e5ff0eb1e47ed	Cycle Studio	38.900081	-77.024215
19	HipCityVeg	57599792498e1b7150fe95f1	Vegetarian / Vegan Restaurant	38.898953	-77.022145
20	Goethe-Institut	4acbda0bf964a52008c820e3	General College & University	38.900683	-77.021978
21	CHOPT	4a8840cbf964a5208f0520e3	Salad Place	38.899251	-77.022113

REFINE THE DATA

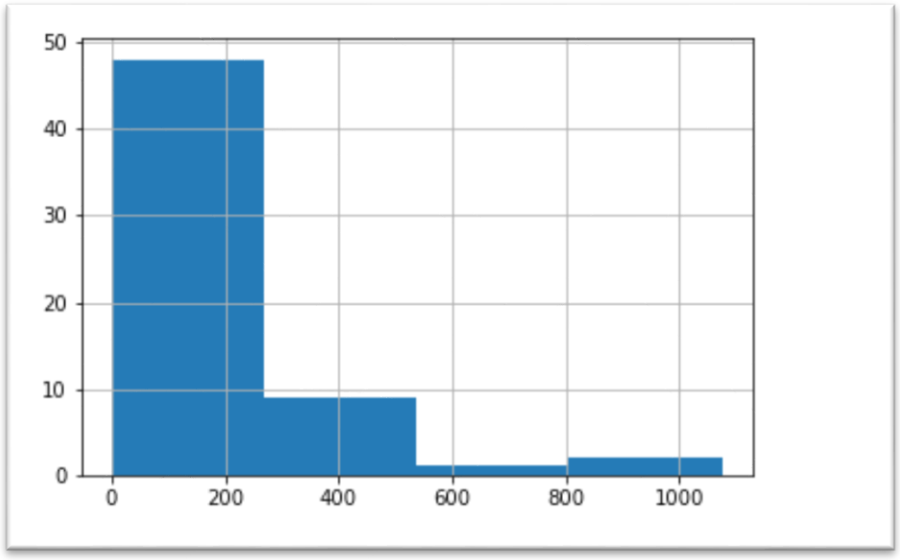
- For cleaning the data we will obtain all the venue categories
- After checking all the categories we well remove the categories that are not food joints from our list
- This will reduce our target set to those venues which are eateries

:

	name	id	categories	lat	lng
2	CAVA	54231977498eff43caedb3d5	Mediterranean Restaurant	38.899971	-77.022314
3	RPM Italian	575ae890498e04099a6d1c67	Italian Restaurant	38.902220	-77.020940
4	La Colombe Coffee Roasters	5648c247498e93b42e8897c8	Coffee Shop	38.901054	-77.020102
6	Nando's	4a4eb693f964a520f4ae1fe3	Portuguese Restaurant	38.900370	-77.021942
7	Dolcezza Gelato	552d5a73498e337582e351bf	Ice Cream Shop	38.900201	-77.024218
8	Taco Bamba	5ac710302b984427e9f885c7	Taco Place	38.901064	-77.022682
9	Chaia	5c34e803dab4b1002c5c4653	Taco Place	38.901015	-77.020934
10	BIBIBOP Asian Grill	5942c26fdec1d635b48cf514	Asian Restaurant	38.898887	-77.021876
11	Zaytinya	3fd66200f964a520d5f11ee3	Mediterranean Restaurant	38.898811	-77.023645
12	Legal Sea Foods	4a561b68f964a52072b41fe3	Seafood Restaurant	38.898795	-77.021853
15	Shouk	572a235c498ee06f5eaceb9d	Vegetarian / Vegan Restaurant	38.902737	-77.020600
16	Boqueria	5bdb1cb08fb09e002df5c82e	Tapas Restaurant	38.899416	-77.023755

CATEGORIZE THE LIKES DATA TO GET RATINGS

- Based on the number of likes obtained for each restaurant we will categorize them to provide ratings
- In order to do this we will use histograms to bins the likes data
- After observing and analyzing the data we will define the percentiles to create explicit bins
- These bins will then be converted into ratings category
- Eg - <10 percentile – poor, >95 percentile is best



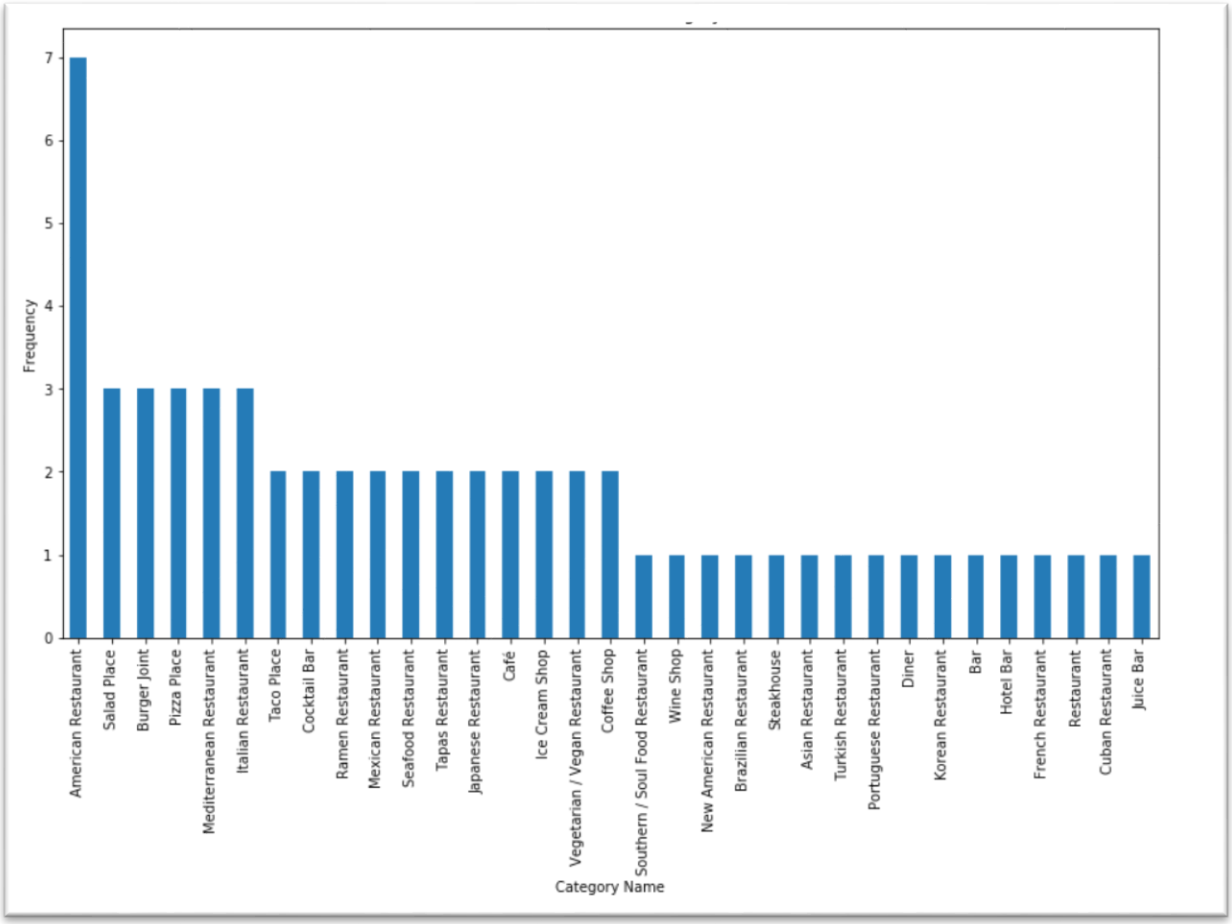
Number of restaurants vs Likes

Percentile	Likes	Category
<10	11	Poor
25	38	Below average
50	116	Above average
75	194	Good
>95	532	Best

	name	id	categories	lat	lng	total likes	total likes_cat
2	CAVA	54231977498eff43caedb3d5	Mediterranean Restaurant	38.899971	-77.022314	164	good
3	RPM Italian	575ae890498e04099a6d1c67	Italian Restaurant	38.902220	-77.020940	161	good
4	La Colombe Coffee Roasters	5648c247498e93b42e8897c8	Coffee Shop	38.901054	-77.020102	238	great
6	Nando's	4a4eb693f964a520f4ae1fe3	Portuguese Restaurant	38.900370	-77.021942	355	great
7	Dolcezza Gelato	552d5a73498e337582e351bf	Ice Cream Shop	38.900201	-77.024218	240	great
8	Taco Bamba	5ac710302b984427e9f885c7	Taco Place	38.901064	-77.022682	55	above average
9	Chaia	5c34e803dab4b1002c5c4653	Taco Place	38.901015	-77.020934	14	below average
0	BIBIBOP Asian Grill	5942c26fdec1d635b48cf514	Asian Restaurant	38.898887	-77.021876	23	below average
1	Zaytinya	3fd66200f964a520d5f11ee3	Mediterranean Restaurant	38.898811	-77.023645	1075	best
2	Legal Sea Foods	4a561b68f964a52072b41fe3	Seafood Restaurant	38.898795	-77.021853	161	good
5	Shouk	572a235c498ee06f5eaceb9d	Vegetarian / Vegan Restaurant	38.902737	-77.020600	107	above average
6	Boqueria	5bdb1cb08fb09e002df5c82e	Tapas Restaurant	38.899416	-77.023755	18	below average
7	Daikaya	50905502d63e87c2d3448e35	Ramen Restaurant	38.898538	-77.019643	829	best
9	HipCityVeg	57599792498e1b7150fe95f1	Vegetarian / Vegan Restaurant	38.898953	-77.022145	61	above average
1	CHOPT	4a8840cbf964a5208f0520e3	Salad Place	38.899251	-77.022113	137	good
2	The Capital Burger	5aa6f0eee075507441b7a499	Burger Joint	38.903107	-77.021751	83	above average
8	Farmers & Distillers	57f24558498e2c8382f6adb3	American Restaurant	38.901358	-77.020206	176	good
9	Momiji	4a750244f964a52044e01fe3	Japanese Restaurant	38.899996	-77.019193	131	good
0	Crimson Diner + Whiskey Bar	5952f04c7b2e047f4e3fdf63	Diner	38.899990	-77.020975	58	above average

GROUP THE RESTAURANTS CATEGORY

- As the data may still contains a number or categories too big to encode we will further refine this data to obtain the major categories of restaurants
- To group the data we will use a bar chart analysis to sea major restaurant types and then divide them into broad meaningful labels of restaurants like – Italian, European , Asian, etc
- We will then add the newly added label and drop the previous one



Categories	**New Category**
'American Restaurant', 'Restaurant', 'Diner', 'New American Restaurant', 'Burger Joint'	American
'Hotel Bar', 'Wine Shop', 'Cocktail Bar', 'Bar'	Bars
'Taco Place', 'Brazilian Restaurant', 'Southern / Soul Food Restaurant', 'Mexican Restaurant', 'Cuban Restaurant'	South American
'Korean Restaurant', 'Mediterranean Restaurant', 'Asian Restaurant', 'Vegetarian / Vegan Restaurant', 'Ramen Restaurant', 'Japanese Restaurant', 'Tapas Restaurant'	Middle East and Asian Food
'Italian Restaurant', 'Portuguese Restaurant', 'Pizza Place', 'French Restaurant', 'Turkish Restaurant'	European

'Seafood Restaurant', 'Salad Place', 'Steakhouse', 'Juice Bar', 'Ice Cream Shop', 'Coffee Shop', 'Café'	others

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new
2	CAVA	54231977498eff43caedb3d5	Mediterranean Restaurant	38.899971	-77.022314	164	good	Middle Eastern and Asian
3	RPM Italian	575ae890498e04099a6d1c67	Italian Restaurant	38.902220	-77.020940	161	good	European
4	La Colombe Coffee Roasters	5648c247498e93b42e8897c8	Coffee Shop	38.901054	-77.020102	238	great	others
6	Nando's	4a4eb693f964a520f4ae1fe3	Portuguese Restaurant	38.900370	-77.021942	355	great	European
7	Dolcezza Gelato	552d5a73498e337582e351bf	Ice Cream Shop	38.900201	-77.024218	240	great	others
8	Taco Bamba	5ac710302b984427e9f885c7	Taco Place	38.901064	-77.022682	55	above average	Mexican South American
9	Chaia	5c34e803dab4b1002c5c4653	Taco Place	38.901015	-77.020934	14	below average	Mexican South American
10	BIBIBOP Asian Grill	5942c26fdec1d635b48cf514	Asian Restaurant	38.898887	-77.021876	23	below average	Middle Eastern and Asian
11	Zaytinya	3fd66200f964a520d5f11ee3	Mediterranean Restaurant	38.898811	-77.023645	1075	best	Middle Eastern and Asian
12	Legal Sea Foods	4a561b68f964a52072b41fe3	Seafood Restaurant	38.898795	-77.021853	161	good	others
15	Shouk	572a235c498ee06f5eaceb9d	Vegetarian / Vegan Restaurant	38.902737	-77.020600	107	above average	Middle Eastern and Asian
16	Boqueria	5bdb1cb08fb09e002df5c82e	Tapas Restaurant	38.899416	-77.023755	18	below average	Middle Eastern and Asian
17	Daikaya	50905502d63e87c2d3448e35	Ramen Restaurant	38.898538	-77.019643	829	best	Middle Eastern and Asian
19	HipCityVeg	57599792498e1b7150fe95f1	Vegetarian / Vegan Restaurant	38.898953	-77.022145	61	above average	Middle Eastern and Asian
21	CHOPT	4a8840cbf964a5208f0520e3	Salad Place	38.899251	-77.022113	137	good	others
22	The Capital Burger	5aa6f0eee075507441b7a499	Burger Joint	38.903107	-77.021751	83	above average	American
28	Farmers & Distillers	57f24558498e2c8382f6adb3	American Restaurant	38.901358	-77.020206	176	good	American
29	Momiji	4a750244f964a52044e01fe3	Japanese Restaurant	38.899996	-77.019193	131	good	Middle Eastern and Asian
30	Crimson Diner + Whiskey Bar	5952f04c7b2e047f4e3fdf63	Diner	38.899990	-77.020975	58	above average	American
32	Del Frisco's Double Eagle Steakhouse	54171297498e3d2a9d876590	Steakhouse	38.900750	-77.024588	134	good	others
33	Izakaya	580d476738fa69fff46fc06f	Japanese Restaurant	38.898577	-77.019865	39	above average	Middle Eastern and Asian
34	Fig & Olive	55781818498e506f812dda67	Mediterranean Restaurant	38.900270	-77.024530	217	great	Middle Eastern and Asian

CONVERTING CATEGORICAL VALUES TO CODES

- We will use hot encoding methodology to encode the categorical variables
- add neighborhood column back to dataframe
- move neighborhood column to the first column

Since clustering algorithms including kmeans use distance-based measurements to determine the similarity between data points, it's recommended to standardize the data to have a mean of zero and a standard deviation of one since almost always the features in any dataset would have different units of measurements such as age vs income.

CLUSTERING THE DATA

- Since this is an unsupervised data we will use k-means clustering to obtain clusters
- We used elbow technique to obtain the best K
- We have selected 5 clusters for this data

The way kmeans algorithm works is as follows:

-Specify number of clusters *K*.

- Initialize centroids by first shuffling the dataset and then randomly selecting *K* data points for the centroids without replacement.

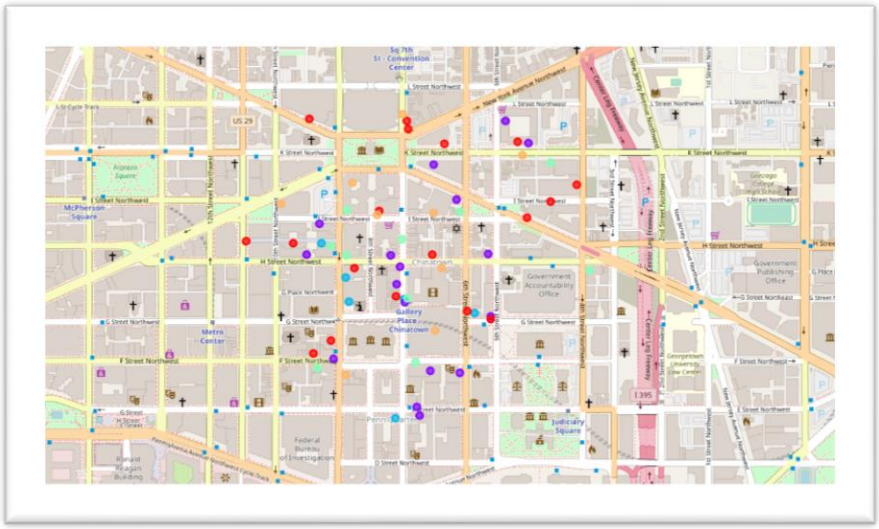
- Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

- Compute the sum of the squared distance between data points and all centroids.

- Assign each data point to the closest cluster (centroid).

- Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

-



RESULTS

The result of the above analysis are –

Cluster 1:

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
8	Taco Bamba	5ac710302b984427e9f885c7	Taco Place	38.901064	-77.022682	55	above average	Mexican South American	0
15	Shouk	572a235c498ee06f5eaceb9d	Vegetarian / Vegan Restaurant	38.902737	-77.020600	107	above average	Middle Eastern and Asian	0
19	HipCityVeg	57599792498e1b7150fe95f1	Vegetarian / Vegan Restaurant	38.898953	-77.022145	61	above average	Middle Eastern and Asian	0
22	The Capital Burger	5aa6f0eee075507441b7a499	Burger Joint	38.903107	-77.021751	83	above average	American	0
30	Crimson Diner + Whiskey Bar	5952f04c7b2e047f4e3fdf63	Diner	38.899990	-77.020975	58	above average	American	0
33	Izakaya	580d476738fa69fff46fc06f	Japanese Restaurant	38.898577	-77.019865	39	above average	Middle Eastern and Asian	0
36	Kinship	567b3262498e1d6a13fc352c	New American Restaurant	38.903305	-77.021797	80	above average	American	0
49	Free State	586c359f5e56b4471f7a4d26	Bar	38.898380	-77.019125	41	above average	bars	0
53	Texas de Brazil	5575f88b498e6a2031ed0eef	Brazilian Restaurant	38.900900	-77.017948	49	above average	Mexican South American	0
54	Five Guys	4a50ee92f964a52044b01fe3	Burger Joint	38.899643	-77.023466	77	above average	American	0
57	Centrolina	555fa470498efbaa5d7894a5	Italian Restaurant	38.900266	-77.025417	79	above average	European	0
60	Alta Strada	56ecad4f498e640591fe0207	Italian Restaurant	38.902788	-77.018256	35	above average	European	0
81	Ella's Wood-Fired Pizza	4a85fb08f964a520eaff1fe3	Pizza Place	38.897845	-77.024218	115	above average	European	0
82	Ottoman Taverna	5723657e498e11c97d95f9e9	Turkish Restaurant	38.901729	-77.016373	72	above average	European	0
84	Succotash	59b71bcb1499466c7ae06b13	Southern / Soul Food Restaurant	38.897526	-77.024780	81	above average	Mexican South American	0
88	Fruitive	5661d923498e05208d892307	Juice Bar	38.900321	-77.026929	63	above average	others	0
93	Marriott Marquis Lobby Bar	5387dcd0498eed7dcda7a3b8	Hotel Bar	38.903361	-77.024918	37	above average	bars	0
95	Eye Street Cellars	53617ffc498e43d4e75f29a2	Wine Shop	38.901304	-77.017201	7	poor	bars	0

Cluster 2:

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
2	CAVA	54231977498eff43caedb3d5	Mediterranean Restaurant	38.899971	-77.022314	164	good	Middle Eastern and Asian	1
3	RPM Italian	575ae890498e04099a6d1c67	Italian Restaurant	38.902220	-77.020940	161	good	European	1
12	Legal Sea Foods	4a561b68f964a52072b41fe3	Seafood Restaurant	38.898795	-77.021853	161	good	others	1
21	CHOPT	4a8840cbf964a5208f0520e3	Salad Place	38.899251	-77.022113	137	good	others	1
28	Farmers & Distillers	57f24558498e2c8382f6adb3	American Restaurant	38.901358	-77.020206	176	good	American	1
29	Momiji	4a750244f964a52044e01fe3	Japanese Restaurant	38.899996	-77.019193	131	good	Middle Eastern and Asian	1
32	Del Frisco's Double Eagle Steakhouse	54171297498e3d2a9d876590	Steakhouse	38.900750	-77.024588	134	good	others	1
41	Capital One Café	5ba1ab18de3bbf002c709dad	Café	38.899690	-77.022013	7	poor	others	1
45	Bantam King	575b2c62498e538453bc141f	Ramen Restaurant	38.898452	-77.019117	130	good	Middle Eastern and Asian	1
55	Denson Liquor Bar	546d6c82498e1cad95af4a5e	Cocktail Bar	38.897047	-77.020107	159	good	bars	1
56	sweetgreen	50326b53e4b0bf3e6b821f75	Salad Place	38.903307	-77.018647	121	good	others	1
58	Compass Coffee	574d9a03cd10a7141a8a4cd3	Coffee Shop	38.897099	-77.021048	187	good	others	1
66	DBGB Kitchen and Bar	53ce74e7498ee536a98f1940	French Restaurant	38.899978	-77.025007	141	good	European	1
83	The Smith	586ea12f6119f44c2591e0ae	American Restaurant	38.897401	-77.024148	120	good	American	1
94	Luke's Lobster	4db64690ec8e571a35e13e24	Seafood Restaurant	38.896192	-77.021467	165	good	others	1
96	sweetgreen	5d98c4c2c7f86100083f20f7	Salad Place	38.895981	-77.021389	1	poor	others	1
98	Mandu	4d39f3890edc594184d58dec	Korean Restaurant	38.902758	-77.017921	117	good	Middle Eastern and Asian	1

Cluster 3:

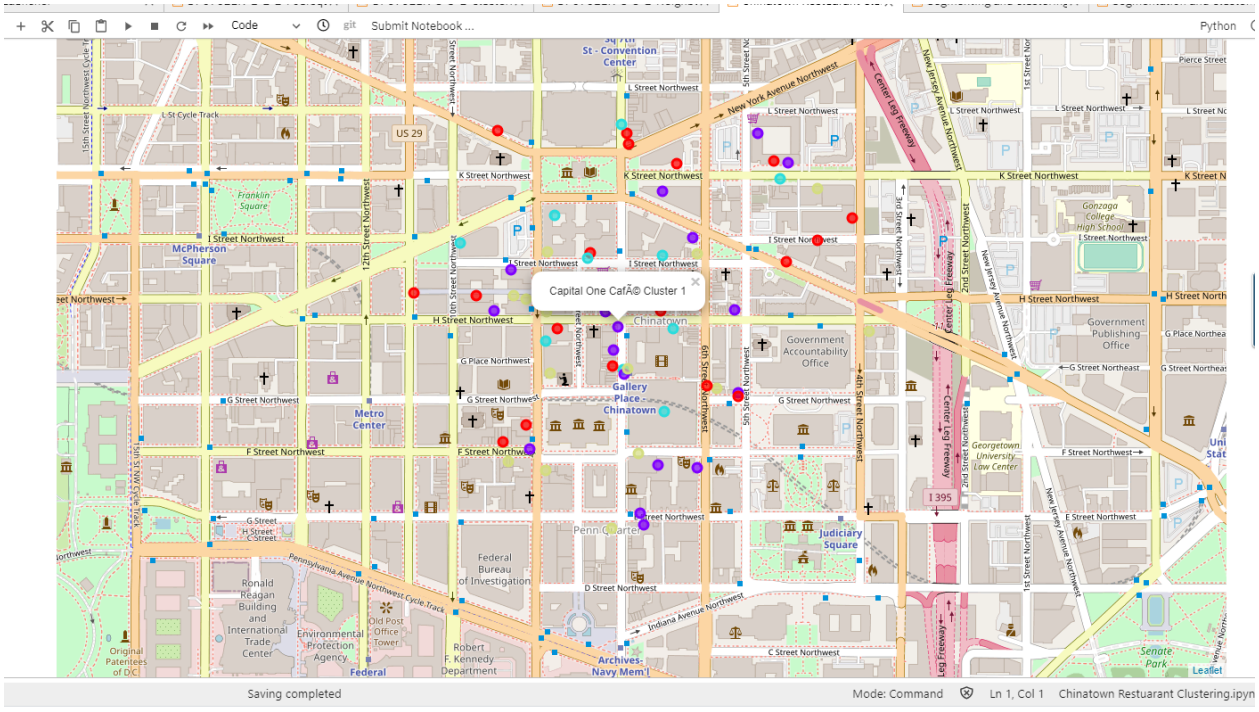
	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
10	BIBIBOP Asian Grill	5942c26fdec1d635b48cf514	Asian Restaurant	38.898887	-77.021876	23	below average	Middle Eastern and Asian	2
11	Zaytinya	3fd66200f964a520d5f11ee3	Mediterranean Restaurant	38.898811	-77.023645	1075	best	Middle Eastern and Asian	2
16	Boqueria	5bdb1cb08fb09e002df5c82e	Tapas Restaurant	38.899416	-77.023755	18	below average	Middle Eastern and Asian	2
17	Daikaya	50905502d63e87c2d3448e35	Ramen Restaurant	38.898538	-77.019643	829	best	Middle Eastern and Asian	2
34	Fig & Olive	55781818498e506f812dda67	Mediterranean Restaurant	38.900270	-77.024530	217	great	Middle Eastern and Asian	2
86	Jaleo	4a6e4e04f964a52074d41fe3	Tapas Restaurant	38.895914	-77.022164	493	great	Middle Eastern and Asian	2

Cluster 4:

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
4	La Colombe Coffee Roasters	5648c247498e93b42e8897c8	Coffee Shop	38.901054	-77.020102	238	great	others	3
6	Nando's	4a4eb693f964a520f4ae1fe3	Portuguese Restaurant	38.900370	-77.021942	355	great	European	3
7	Dolcezza Gelato	552d5a73498e337582e351bf	Ice Cream Shop	38.900201	-77.024218	240	great	others	3
43	Clyde's of Gallery Place	454a14f8f964a5209e3c1fe3	American Restaurant	38.898905	-77.021783	522	great	American	3
51	A Baked Joint	5581a9a8498e9053810aa260	Café	38.902280	-77.017215	436	great	others	3
62	City Tap House Penn Quarter	528fe1a5498eba71d4957543	American Restaurant	38.901091	-77.023701	400	great	American	3
71	Rosa Mexicano	41ca0a80f964a520971e1fe3	Mexican Restaurant	38.897319	-77.021527	288	great	Mexican South American	3
78	Pi Pizzeria	4e562f4cc65be3cea4621714	Pizza Place	38.897168	-77.024658	478	great	European	3
79	Wiseguy NY Pizza	509db390e4b01b9e4932c44f	Pizza Place	38.899608	-77.015963	300	great	European	3
80	Cuba Libre	4bd22974a8b3a593474d675f	Cuban Restaurant	38.899974	-77.023736	283	great	Mexican South American	3

Cluster 5:

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
9	Chaia	5c34e803dab4b1002c5c4653	Taco Place	38.901015	-77.020934	14	below average	Mexican South American	4
35	Acela Club	4baaaeb4f964a5206f7f3ae3	American Restaurant	38.898102	-77.020897	23	below average	American	4
40	Rita's Italian Ice & Frozen Custard	4c6daf4865eda09341614fd0	Ice Cream Shop	38.899659	-77.020675	25	below average	others	4
47	Morris	5a72342e6c08d128d729c78b	Cocktail Bar	38.903482	-77.021937	21	below average	bars	4
65	Renaissance Downtown Concierge Lounge	4e89a1a39a52f9bbd0e733dd	American Restaurant	38.901774	-77.023538	26	below average	American	4
75	Poca Madre	5b2c4e88fe63bd002c3bd625	Mexican Restaurant	38.900970	-77.022739	12	below average	Mexican South American	4
76	Shake Shack	51c22f467dd2142513d69372	Burger Joint	38.896995	-77.023762	729	best	American	4
91	Estuary	5c848408fd16bb00395f63d1	Restaurant	38.901258	-77.025806	12	below average	American	4
92	Busboys and Poets	5b8f5ad8464d65002c541ae9	American Restaurant	38.902457	-77.018107	23	below average	American	4



DISCUSSION SECTION

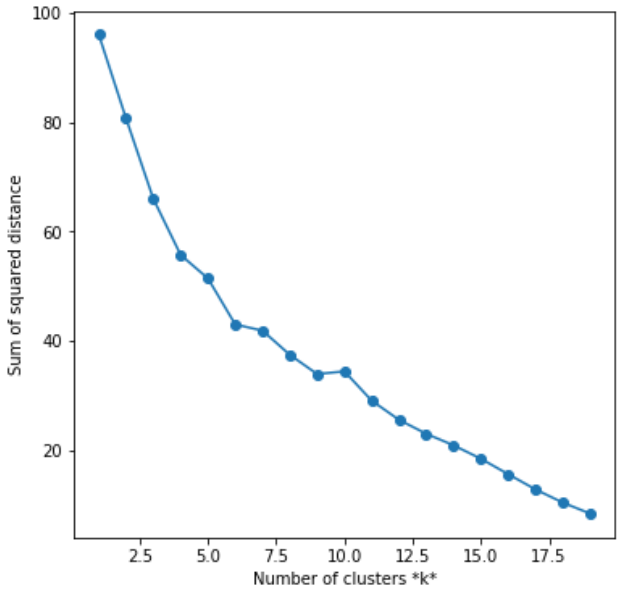
Based on the results obtained on cluster as 5, we observed that the likes and restaurant category have been a determining factor for the categorization of the restaurant.

No locational data was used to test clustering hence we see that all 5 clusters are speeded over the place

Contrary to supervised learning where we have the ground truth to evaluate the model's performance, clustering analysis doesn't have a solid evaluation metric that we can use to evaluate the outcome of different clustering algorithms. Moreover, since kmeans requires k as an input and doesn't learn it from data, there is no right answer in terms of the number of clusters that we should have in any problem.

Using the elbow method we can see that for the given data the ideal number of clusters is not available , so we tested the data for k=5, 9, 11

```
# Plot sse against k
plt.figure(figsize=(6, 6))
plt.plot(list_k, sse, '-o')
plt.xlabel(r'Number of clusters *k*')
plt.ylabel('Sum of squared distance');
```



Hence, I felt we needed more data to create the clusters with required accuracy.

CONCLUSION

Below are the major clusters that I obtained –

1) Above Average Restaurant
2) Good Restaurant
3) Middle Eastern and Asian restaurant with best or great rating
4) Great Restaurants
5) Below Average or poor American restaurant