Recommendation of Alternative Restaurants

Introduction: Business Problem

In this project we will create a recommender system for alternate restaurants in **Mumbai, Maharashtra, India**. The target audience of this project will be any food delivery apps like **Zomato, UberEats, FoodPanda, Swiggy** or restaurant table reservation apps like **EazyDiner, Zomato, etc.**

The aim of this project is to provide a user with an alternate restaurant options based on the target restaurant of the user when the target restaurant is not delivering at this time or there is no reservation available when they are using any of the above apps for food delivery or table reservation.

We will use our data science powers to generate a relevant recommendation to the user based on the target restaurant, distance from there, pricing, type of cuisine and average user rating. The smarter the recommendations are the more pleasant the experience will be for the user of the website or application.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- distance to target restaurant in the neighborhood
- type of cuisine, picing and average user rating

Following data sources will be needed to extract/generate the required information:

- number of restaurants and their location in every neighborhood will be obtained using **Foursquare API**
- type of cuisine, pricing and average user rating will be obtained using venues feature of **Foursquare API**

Sample JSON data from FourSquare with the features can look like following:


```
{'meta': {'code': 200, 'requestId': '5c9c6221f594df6afa8a99b0'},
 'response': {'groups': [{'items': [{'reasons': {'count': 0,
       'items': [{'reasonName': 'globalInteractionReason',
         'summary': 'This spot is popular',
         'type': 'general'}]},
      'referralId': 'e-0-4caa0096d971b1f7ccca23e1-0',
      'venue': {'categories': [{'icon': {'prefix': 'https://ss3.4sqi.ne
t/img/categories_v2/food/asian_',
          'suffix': '.png'},
         'id': '4bf58dd8d48988d145941735',
         'name': 'Chinese Restaurant',
         'pluralName': 'Chinese Restaurants',
         'primary': True,
         'shortName': 'Chinese'}],
       'id': '4caa0096d971b1f7ccca23e1',
```

```
      'location': {'address': 'Waudby Road',
       'cc': 'IN',
       'city': 'Mumbai',
       'country': 'India',
       'distance': 253,
       'formattedAddress': ['Waudby Road', 'Mumbai', 'Mahārāshtra', 'I
ndia'],
       'labeledLatLngs': [{'label': 'display',
         'lat': 18.938715239156295,
         'lng': 72.83293313173236}],
       'lat': 18.938715239156295,
       'lng': 72.83293313173236,
       'state': 'Mahārāshtra'},
      'name': 'Royal China',
      'photos': {'count': 0, 'groups': []}}},
    {'reasons': {'count': 0,
      'items': [{'reasonName': 'globalInteractionReason',
        'summary': 'This spot is popular',
        'type': 'general'}]},
     'referralId': 'e-0-507fcb7a498e5d07b1604ea2-1',
     'venue': {'categories': [{'icon': {'prefix': 'https://ss3.4sqi.ne
t/img/categories_v2/food/coffeeshop_',
         'suffix': '.png'},
        'id': '4bf58dd8d48988d1e0931735',
        'name': 'Coffee Shop',
        'pluralName': 'Coffee Shops',
        'primary': True,
        'shortName': 'Coffee Shop'}],
      'id': '507fcb7a498e5d07b1604ea2',
      'location': {'address': '10, Elphinstone Building, Veer Nariman
Rd.',
       'cc': 'IN',
       'city': 'Mumbai',
       'country': 'India',
       'crossStreet': 'Horniman Circle',
       'distance': 746,
       'formattedAddress': ['10, Elphinstone Building, Veer Nariman Rd
. (Horniman Circle)',
        'Mumbai 400001',
        'Mahārāshtra',
        'India'],
       'labeledLatLngs': [{'label': 'display',
         'lat': 18.93218954141338,
         'lng': 72.83395939223185}],
       'lat': 18.93218954141338,
       'lng': 72.83395939223185,
       'postalCode': '400001',
       'state': 'Mahārāshtra'},
      'name': 'Starbucks',
      'photos': {'count': 0, 'groups': []}}}],
   'name': 'recommended',
   'type': 'Recommended Places'}],
 'headerFullLocation': 'Mumbai',
 'headerLocation': 'Mumbai',
 'headerLocationGranularity': 'city',
 'suggestedBounds': {'ne': {'lat': 18.96577112700003,
   'lng': 72.86382750225758},
  'sw': {'lat': 18.911771072999972, 'lng': 72.80684349774242}},
```

```
  'suggestedFilters': {'filters': [{'key': 'openNow', 'name': 'Open now
'}],
   'header': 'Tap to show:'},
  'totalResults': 218}}
```

Methodology

In this project we will direct our efforts on restaurants located within a radius of 3km from the centre of Mumbai city. We are doing this because we will get only 100 nearby venues and if we want details of those venues, we will get only 50 with a free FourSquare account.

In first step we have collected the required data: location, name and type (category) of every nearby venue within 3km from Mumbai centre. But we do not need all the venues as we are looking only for restaurants or food places.

	id	name	categories	lat	lng
0	4caa0096d971b1f7ccca23e1	Royal China	Chinese Restaurant	18.938715	72.832933
1	507fcb7a498e5d07b1604ea2	Starbucks	Coffee Shop	18.932190	72.833959
2	4d15a8e1816af04db4d444c2	Britannia & Co.	Parsi Restaurant	18.934683	72.840183
3	5263e1ba11d265711e8024bf	Town House Cafe	Bar	18.938550	72.833464
4	4cfe22077f2db1f745873bd4	Food for Thought	Café	18.932031	72.831667

Figure 1: Nearby Venues

Second step in our analysis will be filtering out restaurants/food places and getting the venue details like tipCount, visitsCount, tier and rating for each of the filtered out venues. We will then modify the dataframe with one hot encoding for categories and make it ready for k-means clustering.

	id	name	categories	lat	lng
0	4caa0096d971b1f7ccca23e1	Royal China	Chinese Restaurant	18.938715	72.832933
1	507fcb7a498e5d07b1604ea2	Starbucks	Coffee Shop	18.932190	72.833959
2	4d15a8e1816af04db4d444c2	Britannia & Co.	Parsi Restaurant	18.934683	72.840183
3	4cfe22077f2db1f745873bd4	Food for Thought	Café	18.932031	72.831667
4	4b0587d9f964a52023a422e3	Sher-E-Punjab	Indian Restaurant	18.937944	72.837853
5	4c714f4fd7fab1f715d760c9	Cafe Excelsior	Café	18.937701	72.833566
6	4b0587d5f964a52078a322e3	Trishna	Seafood Restaurant	18.928619	72.832356
7	4b0587cef964a52081a222e3	Kyani & Co.	Café	18.944062	72.828576
8	4dc777edcc3ff3b304696c5a	Pratap Lunch Home	Seafood Restaurant	18.933605	72.832854
9	4ba9e19cf964a520ee3b3ae3	Natural's Ice Cream Parlour	Ice Cream Shop	18.934892	72.824222
10	4d691357342b8cfa3170c32c	K Rustoms. Ice Cream	Ice Cream Shop	18.933478	72.824995
11	4b0587cff964a5208ca222e3	Kala Ghoda Café	Café	18.928515	72.832354
12	4c190902d4d9c928e755f029	Taste Of Kerala	Indian Restaurant	18.934205	72.833215
13	4bdac9bb2a3a0f47edf9abb6	Cannon Pav Bhaji	Food Truck	18.940780	72.836007
14	5215d00311d235bf4fe158e9	Dakshin Bar And Kitchen	Seafood Restaurant	18.936489	72.837490
15	4d8229290d5b8cfa68fc4d28	Pizza By The Bay	Pizza Place	18.933516	72.823925
16	536752ce498e2b4b282cc871	Burma Burma	Asian Restaurant	18.929590	72.832043
17	52ee592a498e5dceaa6b6537	La Folie Patisserie	Dessert Shop	18.928504	72.832315

Figure 2: Food Places after filtering

	Unnamed: 0	Unnamed: 0.1	id	name	categories	lat	lng	tipCount	priceTier	rating	...	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	New American Restaurant	Parsi Restaurant	Pizza Place	Seafood Restaurant	Vegetarian / Vegan Restaurant
0	0	0	4caa0096d971b1f7ccca23e1	Royal China	Chinese Restaurant	18.938715	72.832933	20.0	1.0	8.6	...	0	0	0	0	0	0	0	0	0	0
1	1	1	507fcb7a498e5d07b1604ea2	Starbucks	Coffee Shop	18.932190	72.833959	139.0	1.0	9.1	...	0	0	0	0	0	0	0	0	0	0
2	2	2	4d15a8e1816af04db4d444c2	Britannia & Co.	Parsi Restaurant	18.934683	72.840183	84.0	0.0	8.8	...	0	0	0	0	0	0	1	0	0	0
3	3	3	4cfe22077f2db1f745873bd4	Food for Thought	Café	18.932031	72.831667	24.0	1.0	9.0	...	0	0	0	0	0	0	0	0	0	0
4	4	4	4b0587d9f964a52023a422e3	Sher-E-Punjab	Indian Restaurant	18.937944	72.837853	7.0	2.0	7.9	...	1	0	0	0	0	0	0	0	0	0

Figure 3: One hot encoded data

In third and final step we will focus on clustering the venues using k-means and then giving alternative recommendation based on the cluster and distance from target restaurant. The clustering algorithm used will be K-Means because:

- It is Easy to implement
- With a large number of variables, K-Means may be computationally faster than hierarchical clustering (if K is small).
- K-Means may produce higher clusters than hierarchical clustering

K-Means requires the number of clusters predefined. Elbow method gives us an idea on what a good *k* number of clusters would be based on the sum of squared distance (SSE) between data points and their assigned clusters' centroids. We pick *k* at the spot where SSE starts to flatten out and forming an elbow. We'll use the geyser dataset and evaluate SSE for different values of *k* and see where the curve might form an elbow and flatten out.

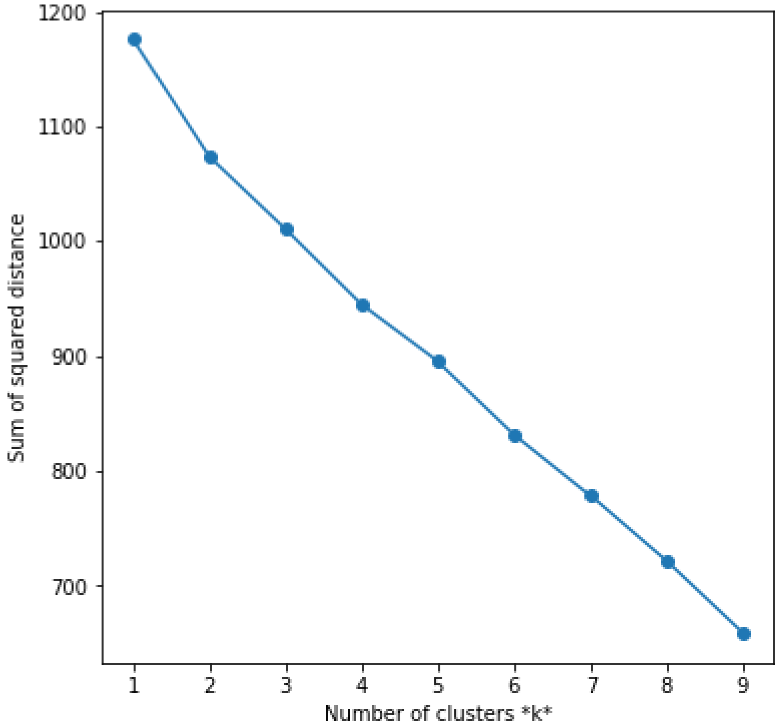


Figure 4: Elbow method results

There is not a very "clear" elbow. But, 4 seems like a reasonable 'k' value.

Results and Discussion

Here are the 4 clusters:

Cluster 1 - Mainstream Restaurants

In [37]:
```
mumbai_restaurants.loc[mumbai_restaurants['Cluster Labels'] == 0]
```

Out[37]:

	Cluster Labels	Unnamed: 0	Unnamed: 0.1	id	name	categories	lat	lng	tipCount	priceTier	rating
12	0	12	12	4c190902d4d9c928e755f029	Taste Of Kerala	Indian Restaurant	18.934205	72.833215	9.0	2.0	7.9
13	0	13	13	4bdac9bb2a3a0f47edf9abb6	Cannon Pav Bhaji	Food Truck	18.940780	72.836007	14.0	1.0	7.6
14	0	14	14	5215d00311d235bf4fe158e9	Dakshin Bar And Kitchen	Seafood Restaurant	18.936489	72.837490	2.0	3.0	7.6
17	0	17	17	52ee592a498e5dceaa6b6537	La Folie Patisserie	Dessert Shop	18.928504	72.832315	6.0	1.0	8.3
22	0	22	22	507926afe4b0801688753f43	The J	Fast Food Restaurant	18.930374	72.826734	9.0	1.0	8.3
23	0	23	23	4b0587d4f964a52071a322e3	Khyber Restaurant	Indian Restaurant	18.928134	72.831541	30.0	2.0	8.2
24	0	24	24	4b0587dbf964a5206ba422e3	Ayub's	Indian Restaurant	18.928383	72.832951	42.0	2.0	8.1
27	0	27	27	4bcad1c0b6c49c74ab2e9191	Balwas	Indian Restaurant	18.939548	72.828311	8.0	2.0	7.7
28	0	28	28	4b0587d6f964a520c0a322e3	Gaylord Restaurant	Indian Restaurant	18.932948	72.826127	44.0	2.0	8.0
29	0	29	29	4c024e8e0d0e0f479fed009a	Badshah - Falooda, Ice Cream, Syrups	Ice Cream Shop	18.947303	72.833470	57.0	1.0	7.8
30	0	30	30	4eb24b53f5b9448388486154	Gulshan-E-Iran	Middle Eastern Restaurant	18.948118	72.835427	16.0	2.0	7.8
33	0	33	33	516db688d86ce94623b64b39	The Bayview - Hotel Marine Plaza	Italian Restaurant	18.931876	72.823031	4.0	2.0	8.2
35	0	35	35	526a32a711d29613276717b2	Salt Water Cafe	Italian Restaurant	18.933500	72.826300	22.0	2.0	7.8
38	0	38	38	4bea8254b3352d7f7b3655d2	National Hindu	Indian Restaurant	18.935088	72.840371	5.0	2.0	7.5
39	0	39	39	4bc1ccdff8219c74f1c6b310	Volga Paan Wala	Dessert Shop	18.932315	72.832497	4.0	1.0	7.5
40	0	40	40	4bb30edaa32876b0084701fe	Baskin Robbins	Ice Cream Shop	18.933629	72.823944	3.0	1.0	7.8
41	0	41	41	4c029056f56c2d7f28901b66	Piccadilly	Falafel Restaurant	18.921425	72.830936	39.0	1.0	8.4
42	0	42	42	4b1145fff964a520937923e3	Theobroma	Dessert Shop	18.919298	72.829185	77.0	1.0	8.9
43	0	43	43	5674388a498eff76add239fd	Rustico	Vegetarian / Vegan Restaurant	18.929541	72.834630	0.0	2.0	7.5
44	0	44	44	5735f6df498eeb5eb79f56d2	1441 Pizzeria	Pizza Place	18.929081	72.834444	8.0	1.0	7.6
45	0	45	45	4b0587d5f964a5208da322e3	New Martin	Indian Restaurant	18.918624	72.829512	16.0	2.0	8.8
46	0	46	46	4c2eebba452620a135bc1c0f	Cha Bar	Coffee Shop	18.929797	72.827468	8.0	1.0	7.7
48	0	48	48	4b0587d6f964a520aea322e3	Souk	Mediterranean Restaurant	18.922434	72.833934	19.0	2.0	8.0

Cluster 2 - Cheap Chinese Places

In [38]:
```
mumbai_restaurants.loc[mumbai_restaurants['Cluster Labels'] == 1]
```

Out[38]:

	Cluster Labels	Unnamed: 0	Unnamed: 0.1	id	name	categories	lat	lng	tipCount	priceTier	rating
0	1	0	0	4caa0096d971b1f7ccca23e1	Royal China	Chinese Restaurant	18.938715	72.832933	20.0	1.0	8.6
19	1	19	19	50aa6378e4b0034097bc6c14	Golden Dragon	Chinese Restaurant	18.928479	72.832493	4.0	1.0	8.2

Cluster 3 - Cafes

In [39]:
```
mumbai_restaurants.loc[mumbai_restaurants['Cluster Labels'] == 2]
```

Out[39]:

	Cluster Labels	Unnamed: 0	Unnamed: 0.1	id	name	categories	lat	lng	tipCount	priceTier	rating
1	2	1	1	507fcb7a498e5d07b1604ea2	Starbucks	Coffee Shop	18.932190	72.833959	139.0	1.0	9.1
2	2	2	2	4d15a8e1816af04db4d444c2	Britannia & Co.	Parsi Restaurant	18.934683	72.840183	84.0	0.0	8.8
3	2	3	3	4cfe22077f2db1f745873bd4	Food for Thought	Café	18.932031	72.831667	24.0	1.0	9.0
4	2	4	4	4b0587d9f964a52023a422e3	Sher-E-Punjab	Indian Restaurant	18.937944	72.837853	7.0	2.0	7.9
5	2	5	5	4c714f4fd7fab1f715d760c9	Cafe Excelsior	Café	18.937701	72.833566	20.0	1.0	7.8
6	2	6	6	4b0587d5f964a52078a322e3	Trishna	Seafood Restaurant	18.928619	72.832356	54.0	3.0	9.1
7	2	7	7	4b0587cef964a52081a222e3	Kyani & Co.	Café	18.944062	72.828576	98.0	1.0	8.4
8	2	8	8	4dc777edcc3ff3b304696c5a	Pratap Lunch Home	Seafood Restaurant	18.933605	72.832854	13.0	3.0	8.0
9	2	9	9	4ba9e19cf964a520ee3b3ae3	Natural's Ice Cream Parlour	Ice Cream Shop	18.934892	72.824222	22.0	1.0	8.9
10	2	10	10	4d691357342b8cfa3170c32c	K Rustoms. Ice Cream	Ice Cream Shop	18.933478	72.824995	64.0	1.0	8.8
11	2	11	11	4b0587cff964a5208ca222e3	Kala Ghoda Café	Café	18.928515	72.832354	55.0	1.0	8.7
15	2	15	15	4d8229290d5b8cfa68fc4d28	Pizza By The Bay	Pizza Place	18.933516	72.823925	100.0	1.0	8.5
18	2	18	18	4e8c15697ee6beb5b5efb7e4	Shree Thaker Bhojnalay	Indian Restaurant	18.951217	72.828326	15.0	2.0	9.0
20	2	20	20	4b9c68ddf964a5209a6636e3	Café Mondegar	Café	18.924219	72.832106	176.0	2.0	9.1
21	2	21	21	4b0587d8f964a52006a422e3	Bhagat Tarachand Restaurant	Indian Restaurant	18.951802	72.830486	48.0	2.0	8.7
25	2	25	25	56742fd9498eee4a3320853a	145 Kala Ghoda	Café	18.928088	72.831844	9.0	1.0	8.1
26	2	26	26	50caded2e4b06f07c9b15053	PizzaExpress Colaba	Pizza Place	18.923899	72.833452	46.0	1.0	8.7
32	2	32	32	56b9ee80498edc99be627f0e	Le15 Cafe	Café	18.923709	72.832735	10.0	1.0	8.5
34	2	34	34	4ce68ea5f1c6236a571f54f0	Leopold Café	Café	18.922828	72.831709	248.0	2.0	8.6

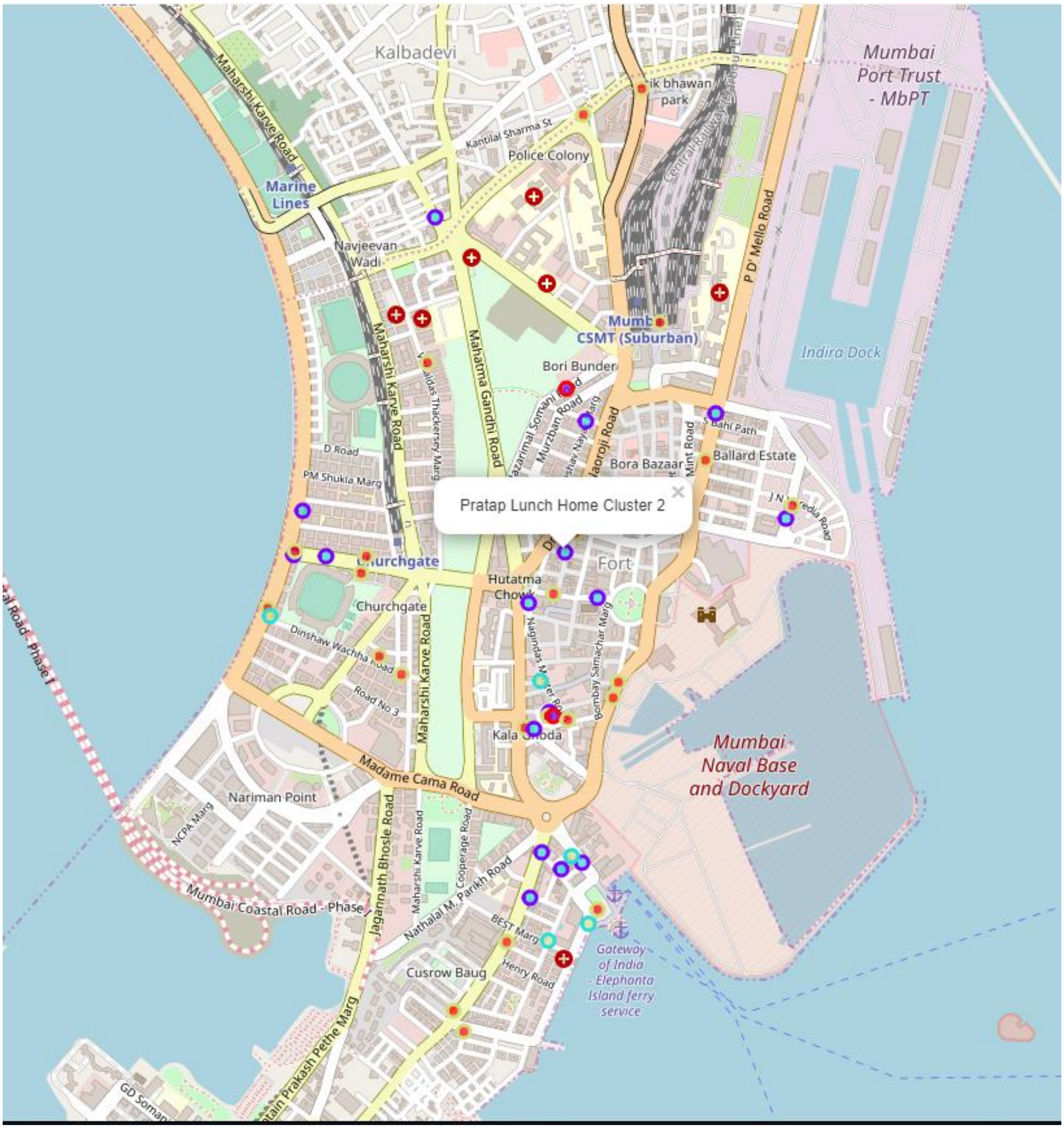
Cluster 4 - Premium Asian Restaurants

In [40]:
```
mumbai_restaurants.loc[mumbai_restaurants['Cluster Labels'] == 3]
```

Out[40]:

	Cluster Labels	Unnamed: 0	Unnamed: 0.1	id	name	categories	lat	lng	tipCount	priceTier	rating
16	3	16	16	536752ce498e2b4b282cc871	Burma Burma	Asian Restaurant	18.929590	72.832043	21.0	2.0	8.2
31	3	31	31	4d31c8065c2db60cbb6ebb6c	The Table	New American Restaurant	18.924099	72.833097	48.0	4.0	8.5
36	3	36	36	4b41e0cff964a520b8c925e3	Wasabi by Morimoto	Japanese Restaurant	18.922012	72.833663	29.0	4.0	8.7
37	3	37	37	4b3504a2f964a520be2925e3	The Oriental Blossom, Marine Plaza	Asian Restaurant	18.931605	72.823121	5.0	2.0	8.1
47	3	47	47	58b6d8957d0f6d1bd9e7edb8	Pa Pa Ya (Colaba)	Asian Restaurant	18.921456	72.832337	3.0	2.0	8.2

Here is the map of clustered restaurants:



To recommend an alternative restaurant, we first obtain the cluster number of the restaurant and present the rest of the clusters sorted in a certain way. E.g. A common logical way is to sort by the distance from the original restaurant.

We would use geopy.distance from geodesic to calculate the distance between 2 points (with latitude and longitude).

Using the restaurant Khyber Restaurant as an example showing the top 5: (First row is the original restaurant hence a distance of 0)

	index	Cluster Labels	Unnamed: 0	Unnamed: 0.1	id	name	categories	lat	lng	tipCount	priceTier	rating	distance
5	23	0	23	23	4b0587d4f964a52071a322e3	Khyber Restaurant	Indian Restaurant	18.928134	72.831541	30.0	2.0	8.2	0.000000
3	17	0	17	17	52ee592a498e5dceaa6b6537	La Folie Patisserie	Dessert Shop	18.928504	72.832315	6.0	1.0	8.3	0.091260
6	24	0	24	24	4b0587dbf964a5206ba422e3	Ayub's	Indian Restaurant	18.928383	72.832951	42.0	2.0	8.1	0.151055
19	44	0	44	44	5735f6df498eeb5eb79f56d2	1441 Pizzeria	Pizza Place	18.929081	72.834444	8.0	1.0	7.6	0.323307
18	43	0	43	43	5674388a498eff76add239fd	Rustico	Vegetarian / Vegan Restaurant	18.929541	72.834630	0.0	2.0	7.5	0.360799
21	46	0	46	46	4c2eebba452620a135bc1c0f	Cha Bar	Coffee Shop	18.929797	72.827468	8.0	1.0	7.7	0.466862

We can also sort by distance from your current geo location or rating etc.

Discussion

Cluster 2 (Cheap Chinese Places) and 4 (Premium Asian Restaurants) look like the restaurants are clustered together based on the categories and price tier. Cluster 1 and 3 look like they can use some more features to differentiate the restaurants within them. We can get a more accurate picture of how busy a restaurant is from the "visit counts" of the paid FourSquare account.

Some ideas of collecting more features include:

- Get a more accurate picture of how busy a restaurant is from the "visit counts" of the paid FourSquare account.
- Use image recognition services to see what kinds of dishes people take photos of
- Use natural language processing to see what dishes the restaurants offer in the menu posted

The data set size is also smaller due to the limitation of the FourSquare free account. It can be solved by using the paid account. The process would be exactly the same.

Conclusion

We can see that machine learning can help with finding things that are "similar" to each other. It can help to make smarter recommendations that humans can't since we can base the recommendations on a large amount of data and features. We have demonstrated an algorithm that can help map or restaurant booking applications to provide strategically calculated recommendations of alternative restaurants.