Introduction:

Pittsburgh is a city in the state of Pennsylvania in the United States and is the county seat of Allegheny County. A population of about 301,048 residents live within the city limits, making it the 66th-largest city in the U.S. In 2015, Pittsburgh was listed among the "eleven most livable cities in the world". In recent years, a lot of technology companies open office in Pittsburgh such as Google, Apple Inc., Bosch, Facebook, Uber, Nokia, Autodesk, Microsoft and IBM. The population is growing because of school, hospital and tech companies. However, when I'm so craving dim sum, I couldn't find any dim sum place in Pittsburgh. The closest one is in Cleveland about two hours drive. With more peoples asking me about the good place to eat dim sum in Pittsburgh, I start to wonder it might be a great opportunity to open a dim sum restaurant in Pittsburgh. In the following discussion, I'm going to find out the best place to open dim sum restaurant in Pittsburgh. This will help restaurant investors make a better business decision.

Data:

Source: Foursquare location data.

Data points: 150 restaurants around three locations (university of Pittsburgh, Square Hill and Shadyside)

Method to get data points:

1. Choose 3 high population area and get latitude and longitude of these places.

```
U of Pitts: 40.44428425 -79.9531832716685
Square Hill: 40.4381258994036 -79.9203284129236
Shadyside: 40.451174 -79.933514
```

2. Use 'search for Venues' API from Foursqaure to search restaurants within 5 miles from these three points.

3. Extract location data from results which include name,id, lat,lng, postalcode etc.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng
0	Prince of India Restaurant	Indian Restaurant	3614 5th Ave	US	Pittsburgh	United States	NaN	543	[3614 5th Ave, Pittsburgh, PA 15213, United St...	[{'label': 'display', 'lat': 40.44163316585947...	40.441633	-79.958575
1	Duranti's Park Plaza Restaurant	Indian Restaurant	128 N Craig St	US	Pittsburgh	United States	NaN	648	[128 N Craig St, Pittsburgh, PA 15213, United ...	[{'label': 'display', 'lat': 40.44813515424374...	40.448135	-79.947444
2	Big Jims Restaurant & Bar	Dive Bar	201 Saline St	US	Pittsburgh	United States	at Alexis St	1941	[201 Saline St (at Alexis St), Pittsburgh, PA ...	[{'label': 'display', 'lat': 40.42704288846097...	40.427043	-79.949695
3	People's Indian Restaurant	Indian Restaurant	5147 Penn Ave	US	Pittsburgh	United States	NaN	2501	[5147 Penn Ave, Pittsburgh, PA 15224, United S...	[{'label': 'display', 'lat': 40.46481627368615...	40.464816	-79.941180
4	Mitchell's Restaurant, Bar & Banquet Center	American Restaurant	304 Ross St	US	Pittsburgh	United States	NaN	3786	[304 Ross St, Pittsburgh, PA 15219, United Sta...	[{'label': 'display', 'lat': 40.43713883620821...	40.437139	-79.996876
5	Restaurant Hana	Sushi Restaurant	4407 Butler St	US	Pittsburgh	United States	NaN	3120	[4407 Butler St, Pittsburgh, PA 15201, United ...	[{'label': 'display', 'lat': 40.471854, 'lng':...	40.471854	-79.959816
6	Savoy Restaurant	Cocktail Bar	2623 Penn Ave	US	Pittsburgh	United States	NaN	2444	[2623 Penn Ave, Pittsburgh, PA 15222, United S...	[{'label': 'display', 'lat': 40.55546680395372...	40.455467	-79.978014
7	fl.2 Restaurant	Restaurant	510 Market St	US	Pittsburgh	United States	NaN	4133	[510 Market St, Pittsburgh, PA 15222, United S...	[{'label': 'display', 'lat': 40.44128365807759...	40.441284	-80.001813

Visualize there 150 data points:

The blue dots represent the 150restaurants. The red dot is the center of three places.
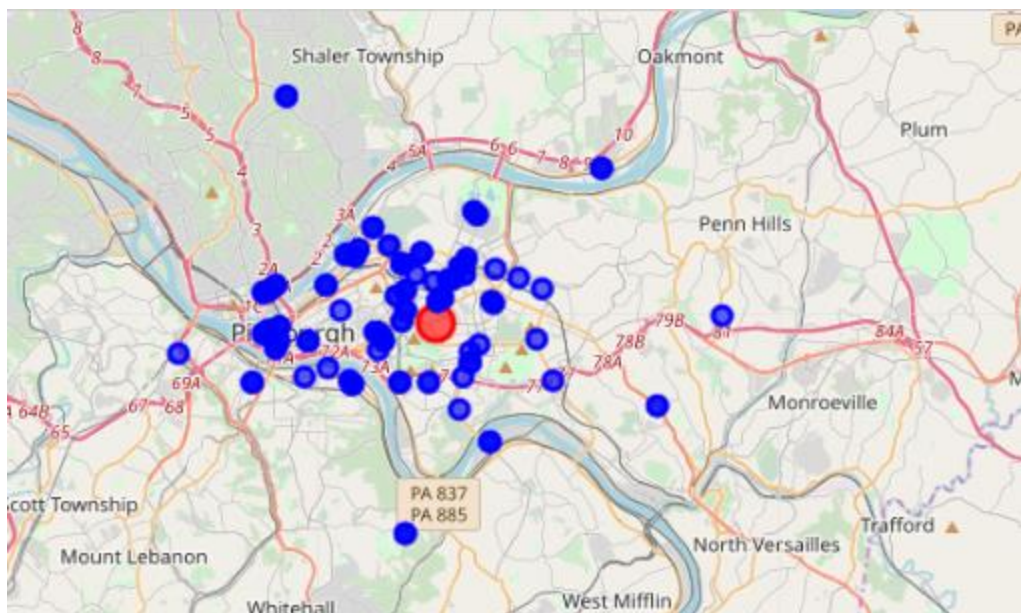
Methodology:

Feature selections: info about these restaurants on Price, Likes, Dislikes, Rating and Tips by 'get details of a venue' API.

One hot encoding: there are 4 levels of price: Cheap, Moderate, Expensive and Very expensive. Any venue that doesn't have Price content was filled by 0. Therefore, one hot encoding on price can generate 5 dummies.

Normalization: To balance the scale, each data point was divided by the max value of each column.

The data frame used in cluster:

|	id	Likes	Dislike	Rating	Tip	0	Cheap	Expensive	Moderate	Very Expensive
0	4ad4c00ef964a52020ee20e3	0.118644	False	0.862069	0.123596	0	0	0	1	0
1	4ad4c00ff964a52037ee20e3	0.000000	False	0.000000	0.000000	0	0	0	1	0
2	4ad7c9bcf964a520b60e21e3	0.344633	False	0.896552	0.393258	0	0	0	1	0
3	4c707b429c6d6dcb6defd27a	0.146893	False	0.804598	0.067416	0	0	0	1	0
4	4ad4c00ef964a5200bee20e3	0.112994	False	0.827586	0.056180	0	1	0	0	0
5	566c88c9498eb66b0019b4b9	0.016949	False	0.000000	0.000000	0	0	0	1	0
6	4daa749693a04642f080f39b	0.242938	False	0.896552	0.303371	0	0	1	0	0
7	59f913efc876c859390b2ae2	0.050847	False	0.896552	0.033708	0	0	0	1	0
8	4f32840b19836c91c7df60c4	0.000000	False	0.000000	0.000000	0	0	0	1	0
9	4f32a31419836c91c7eb2194	0.000000	False	0.000000	0.000000	1	0	0	0	0
10	4c23c4a2136d20a142bee061	0.016949	False	0.620690	0.033708	0	1	0	0	0
11	5400ccc0498ee72e47d61ca7	0.039548	False	0.827586	0.011236	0	0	0	1	0
12	500d9f19e4b099ec37d35b4c	0.000000	False	0.000000	0.000000	0	0	0	1	0
13	4e4e51b8bd4101d0d7a7a1d4	0.000000	False	0.000000	0.000000	0	1	0	0	0
14	4b276cdaf964a5205a8624e3	0.231638	False	0.954023	0.033708	1	0	0	0	0
15	4ad4c00ef964a5201bee20e3	0.000000	False	0.000000	0.000000	0	1	0	0	0
16	4b19b128f964a5201ee223e3	0.033898	False	0.804598	0.134831	0	0	1	0	0
17	4b158b61f964a5207dae23e3	0.146893	False	0.885057	0.191011	0	0	0	1	0
18	5d7d991584f21d0008b51e88	0.005650	False	0.000000	0.000000	0	1	0	0	0
19	4f327efe19836c91c7dd7632	0.000000	False	0.000000	0.000000	1	0	0	0	0

Clusters: Based on these features, 3 clusters would be the best

Results:

Cluster_0(blue): 38 data, price-Expensive, Likes-medium, Rating--medium, tips—middle
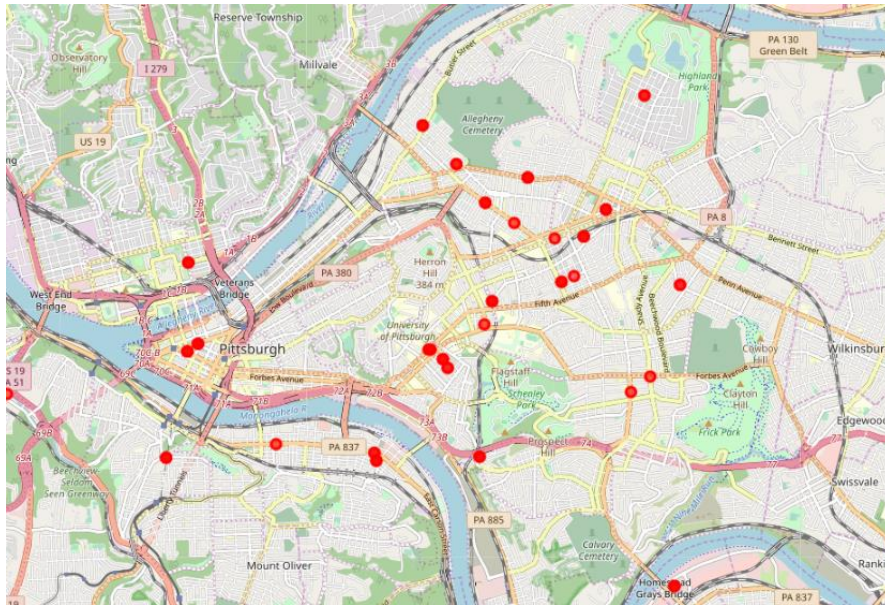
Cluster_1(red): 70 data, price—Moderate, Likes—High, Rating--High, tips—High

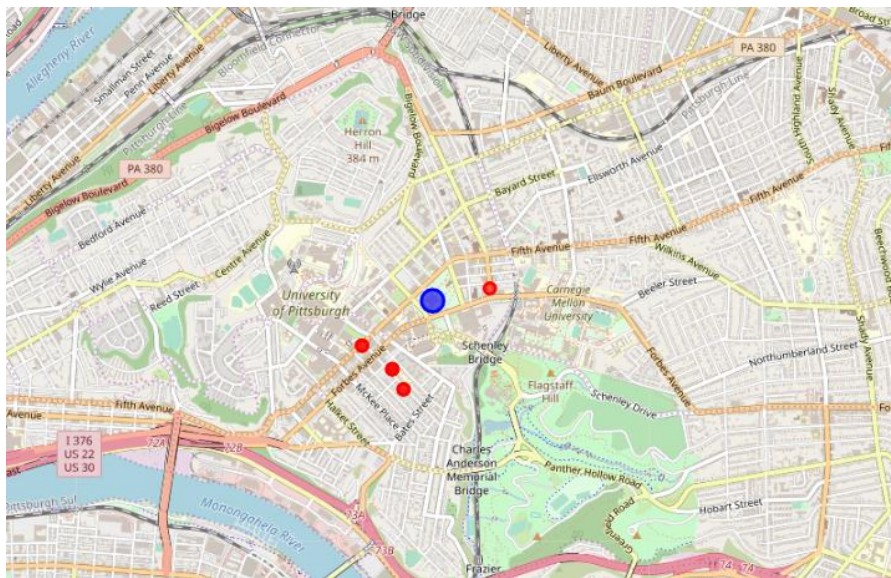Cluster_2(green): 42 data, price—Low, Likes—low, Rating--low, tips—low
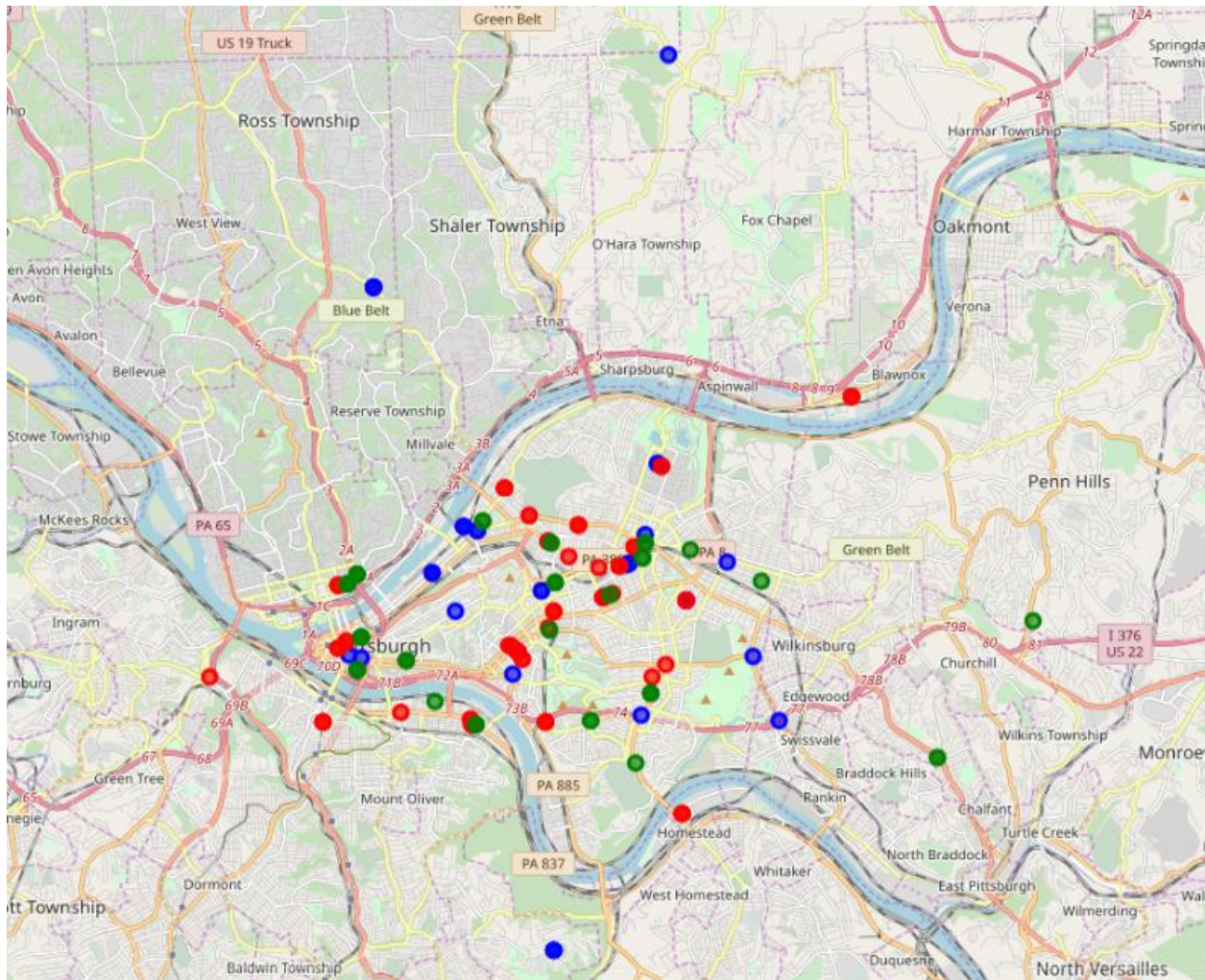
Dim sum restaurant is targeted as cluster_1.

Let's take look at only cluster_1:

Based the model of this restaurant, university of Pittsburgh is chose as target place.

The blue dot represents u of pitt. The red dots represent 4 places in cluster 1 that is closest to university of Pittsburgh

The center of 4 places is set as location of Dim sum restaurant (yellow) latitude(40.441503672380264) and longitude(-79.95477134701213):
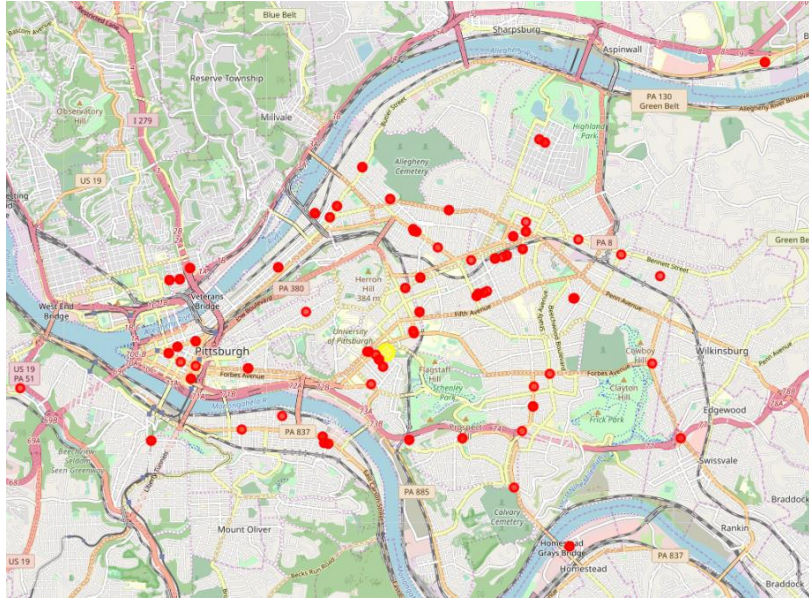
Conclusion:

There are some limitations. First, selected features are not enough to generate an accurate result. Demographics data about Pittsburgh should be considered as features. For example, Dim sum is believed to originate in Guangzhou, China. It tends to attract most Asian people. Therefore, race is an important factor. The average of salary in each district can demonstrate the consumption level. The density of population should also be considered as one feature. Second, in this project, any no value data was filled by 0, which means no data. Each cluster characters were based on average of value in the table. 0 can lower the average, but it doesn't mean the value is low. Therefore, the results might not be accurate. Third, data was collected from Foursquare. It limits how many calls you can access everyday. To improve the accuracy, more data needs to be collected.