**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.

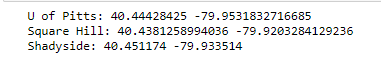
**Data:**

Source: Foursquare location data.

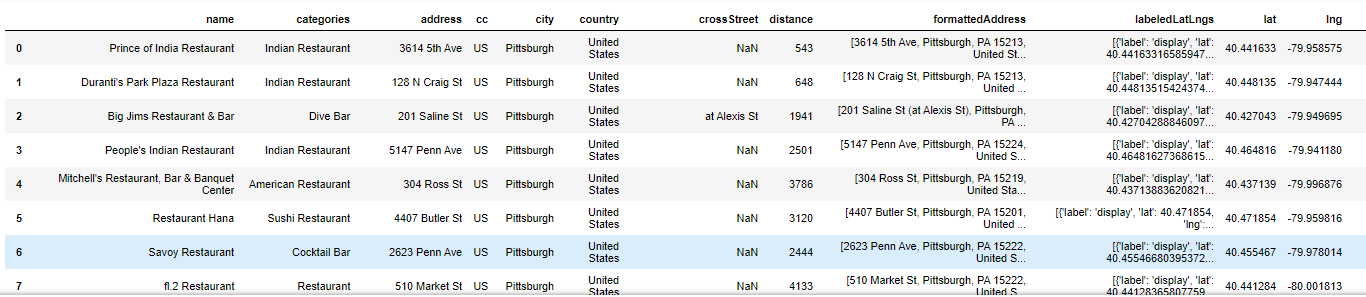
Data points: 150 restaurants around three locations (university of Pittsburgh, Square Hill and Shadyside)

Method to get data points:

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

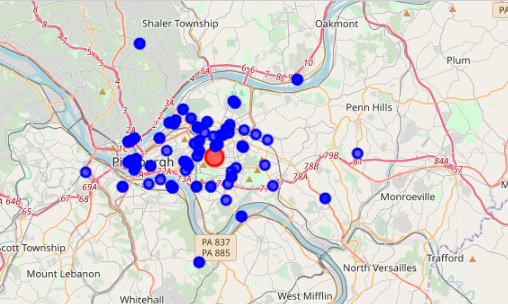


1. Use ‘search for Venues’ API from Foursqaure to search restaurants within 5 miles from these three points.
2. Extract location data from results which include name,id, lat,lng, postalcode etc.



Visualize there 150 data points:

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



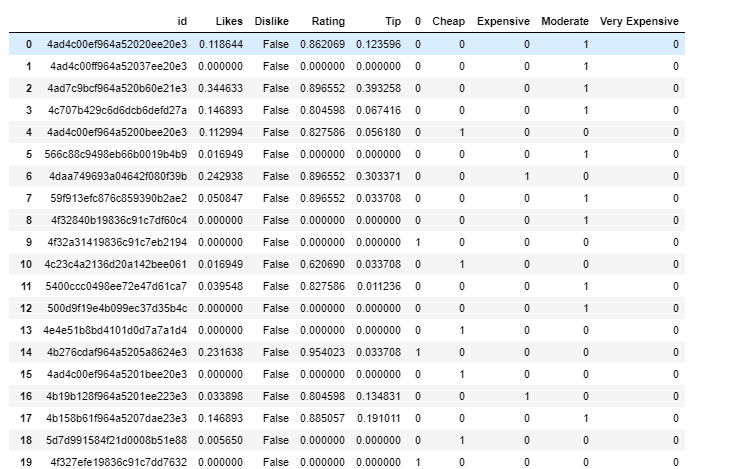
**Methodology:**

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:



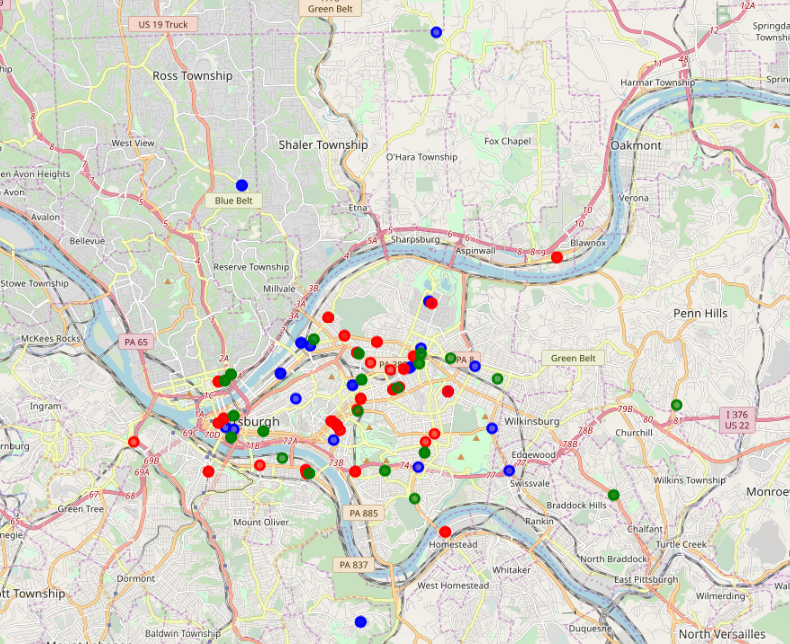
Clusters: Based on these features, 3 clusters would be the best

**Results:**

Cluster\_0(blue): 38 data, price-Expensive, Likes-medium, Rating--medium, tips—middle

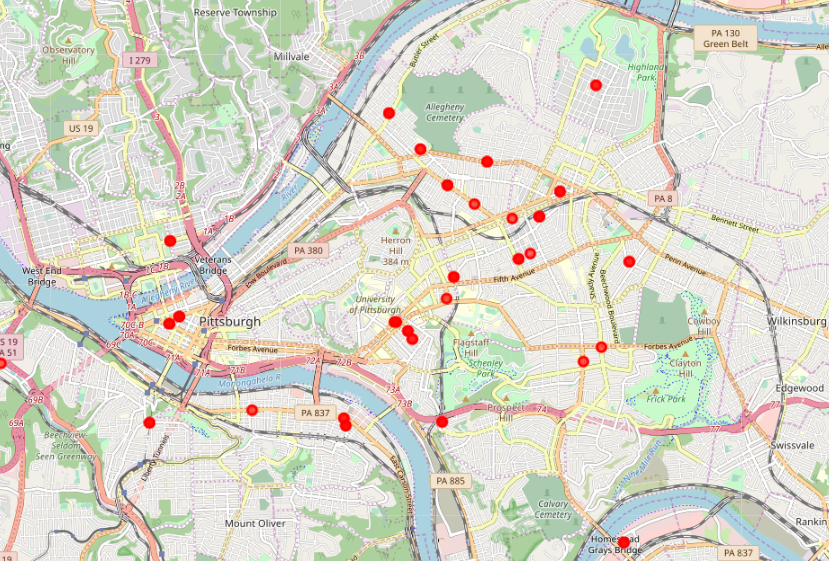
Cluster\_1(red): 70 data, price—Moderate, Likes—High, Rating--High, tips—High

Cluster\_2(green): 42 data, price—Low, Likes—low, Rating--low, tips—low

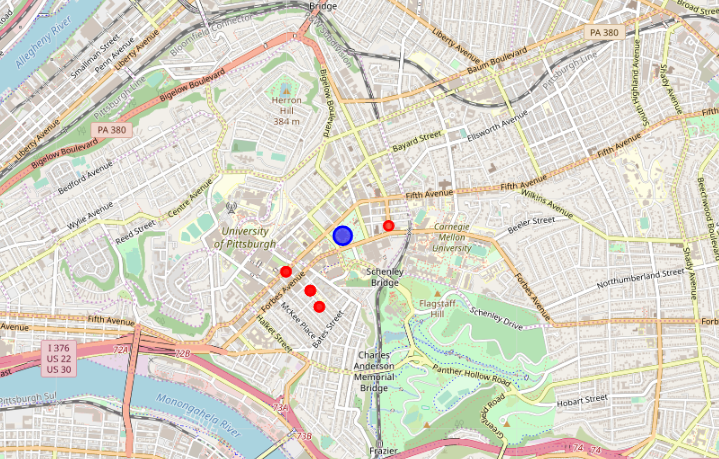


Dim sum restaurant is targeted as cluster\_1.

Let’s take look at only cluster\_1:

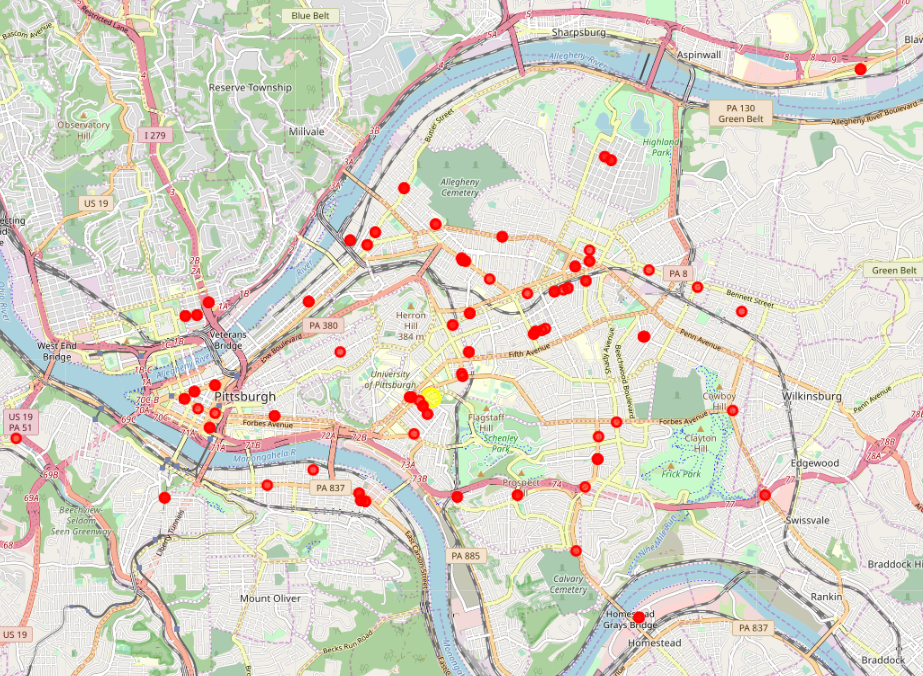


Based the model of this restaurant, university of Pittsburgh is chose as target place.



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.