**Location Selection of City Catering Shops report**

# Introduction/Business Problem

A Client wants to open a restaurant in the area of Pudong, Shanghai. The client is new to this city and is ont bound to a specific kind of cuisine.

He want to know:

1. What kind of cuisine would attract a lot of customers
2. Which area is suitable for operating his restaurant

I will give him recommendation in this project.

# Data Analysis/Solution section

Three kinds of data are necessary to answer the business problem:

1. The location of neighborhood and restaurant with detail information in it

2. The type of restaurant Category (e.g. Fast Food Restaurant or Steakhouse)

3. The Shanghai geospatial analysis data generated by QGIS.

In the Case of the Foursquare API, the relevant data will be extracted from the 'search' and 'query' endpoint.

**Part 1**

Through three dimensions "Price Tier", "Rating" and "Cost Performance" of a restaurant, to compares different category and screens out the catering types that can open a restaurant.

According to Shanghai catering data obtained via Foursquare API , include three fields: "Price Tier", "Rating" and "Rating Signals". Therefore, we can calculate the "Cost Performance" by indicators: "Price Tier", "Rating". On this basis, we can choose catering types.

**Part 2**

Based on part 1 analysis, choose a restaurant catering type. Divide Shanghai into grid space, and evaluate the spatial index with Python to get the location of catering location. The following four indicators can be used:

Population density index

Road density index

Restaurants number index

These indicators is obtained by using QGIS for geospatial analysis.

Evaluation methods:

Population Density Index The Higher the Score, the Better

Road Density Index The Higher the Score, the Better

Restaurants Index The Higher the Score, the Better

Comprehensive Index = Population Density Index \* 0.4 + Catering Heat Index \* 0.3 + Road Density Index \* 0.2

The population density and road density are calculated by using the population grid map of Shanghai as the basic data. The Restaurants density are calculated by combining the longitude and latitude of catering shops with the catering data of Shanghai. After labeling the four indicators, the weighted average is used to get the comprehensive indicators, and the top10 area is selected as the candidate area by comprehensive ranking. Bokeh can be used to draw a square map of Shanghai according to longitude and latitude, and the alternative areas can be marked by color and size.

# Methodology

### Get Pudong, Shanghai neighborhoods data from

### http://www.shjcdj.cn/djWeb/zhuanti/dangjianfuwu/detail.html

### It contains the Boroughs of Shanghai, and neighborhoods in each Borough w.

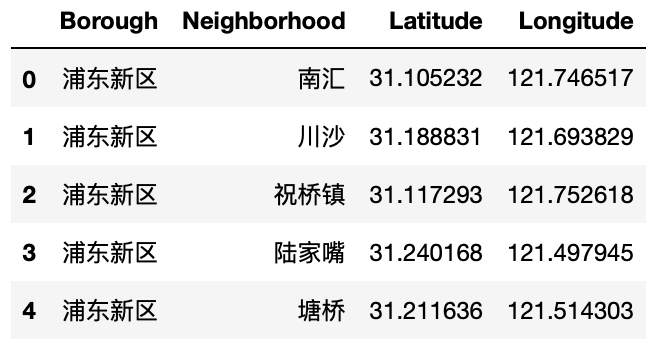


### Data cleansing: Process it (could be used by geolocator)



### Assign the "Latitude, Longitude" to each Neighborhood via geolocator

|  |
| --- |
| for borough, neighborhood\_name in zip( neighborhoods['区县'], neighborhoods['社区乡镇']): # Assign the "Latitude, Longitude" to each Neighborhood via geolocator  address= borough + neighborhood\_name  geolocator = Nominatim(user\_agent="Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.1 (KHTML, like Gecko) Chrome/22.0.1207.1 Safari/537.1")  location = geolocator.geocode(address)  neighborhood\_lat = location.latitude  neighborhood\_lon = location.longitude    pudong\_data = pudong\_data.append({'Borough': borough,  'Neighborhood': neighborhood\_name,  'Latitude': neighborhood\_lat,  'Longitude': neighborhood\_lon}, ignore\_index=True)  pudong\_data.head() #print('The geograpical coordinate of {} are {}, {}.'.format(address,latitude, longitude)) |



### Visualizate Shanghai Pudong with the neighborhoods in it

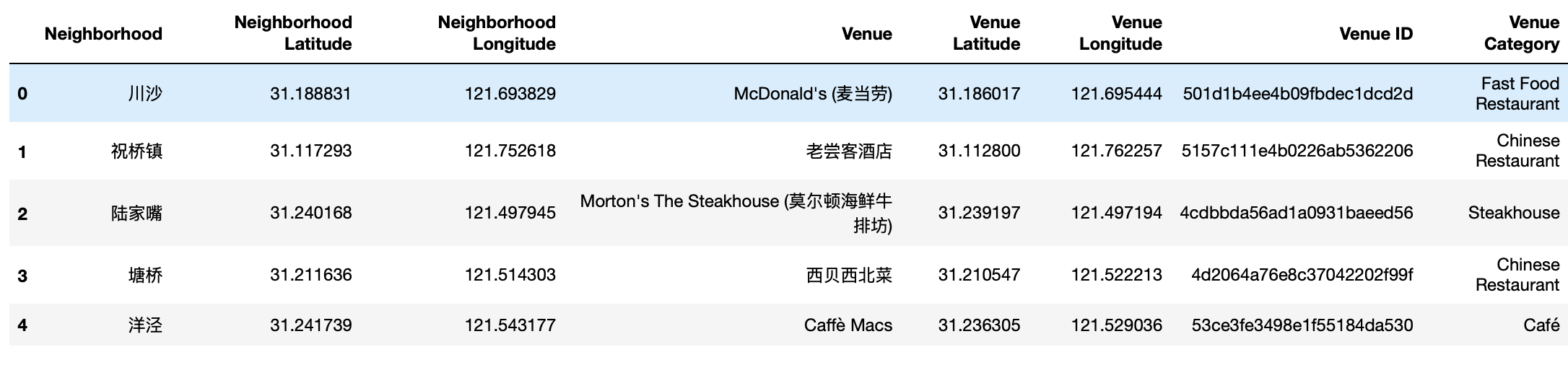
|  |
| --- |
| address\_shanghaipudong = '上海市浦东新区' #Let's get the geographical coordinates of Pudong, Shanghai.  geolocator = Nominatim(user\_agent="Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.1 (KHTML, like Gecko) Chrome/22.0.1207.1 Safari/537.1")  location\_pud = geolocator.geocode(address\_shanghaipudong)  latitude\_pud = location\_pud.latitude  longitude\_pud = location\_pud.longitude  print('The geograpical coordinate of Pudong, Shanghai are {}, {}.'.format(latitude\_pud, longitude\_pud)) |

|  |
| --- |
| map\_pudong = folium.Map(location=[latitude\_pud, longitude\_pud], zoom\_start=11) # Visualizate Shanghai Pudong with the neighborhoods in it  for lat, lng, label in zip(pudong\_data['Latitude'], pudong\_data['Longitude'], pudong\_data['Neighborhood']): # add markers to map  label = folium.Popup(label, parse\_html=True)  folium.CircleMarker(  [lat, lng],  radius=5,  popup=label,  color='blue',  fill=True,  fill\_color='#3186cc',  fill\_opacity=0.7,  parse\_html=False).add\_to(map\_pudong)  map\_pudong |



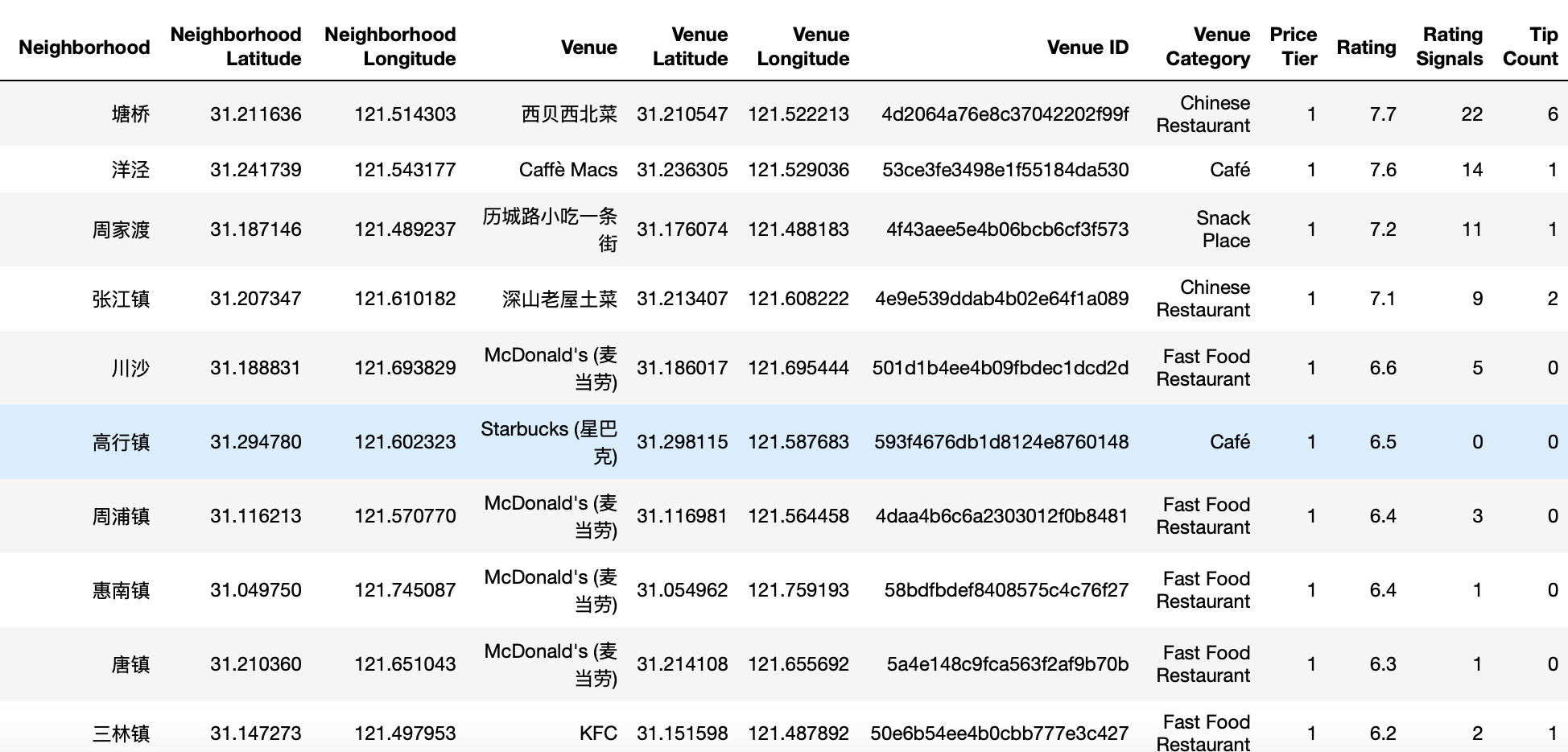
### Search for venues nearby Neighborhood via Foursquare API

|  |
| --- |
| url = 'https://api.foursquare.com/v2/venues/explore?&client\_id={}&client\_secret={}&v={}&ll={},{}&radius={}&limit={}&query=restaurant'.format(  CLIENT\_ID,  CLIENT\_SECRET,  VERSION,  neighborhood\_latitude,  neighborhood\_longitude,  radius,  LIMIT)  results = requests.get(url).json() |



### Obtain venues detailed information via Foursquare API Premium call

|  |
| --- |
|  |

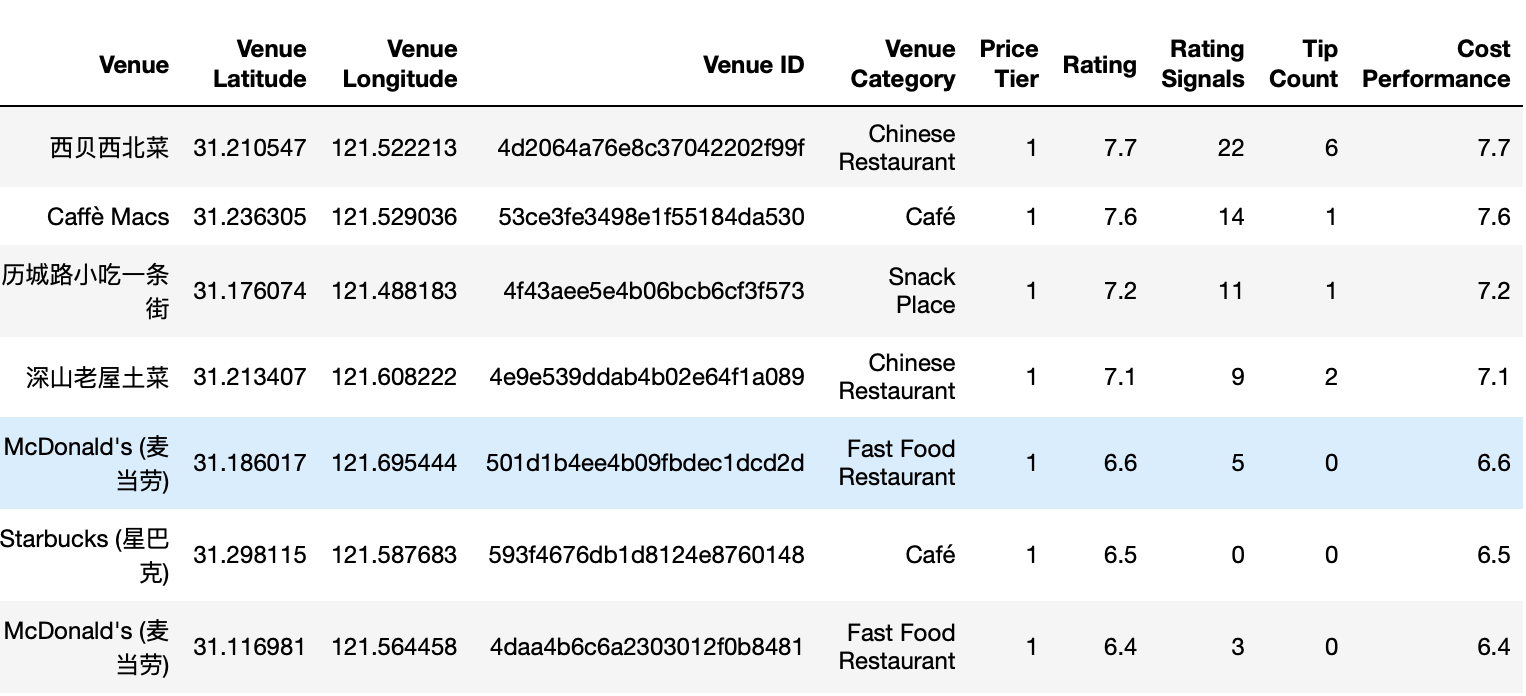


Drop rows with 'NaN' value in column "Rating":

Venue\_dtl\_data.dropna(subset=['Rating'],axis=0,inplace=True)

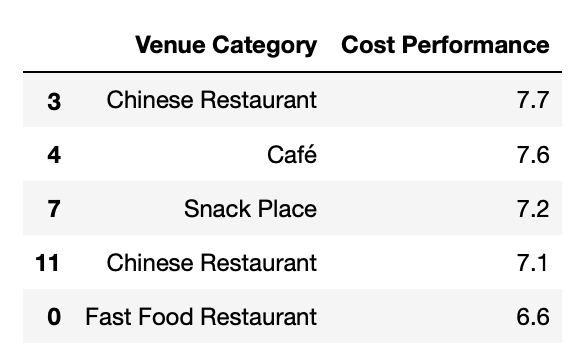
### Evaluation:

Cost Performance= Rating/ Price Tier, I used it to weight the cuisine to be selected as the restaurant category.

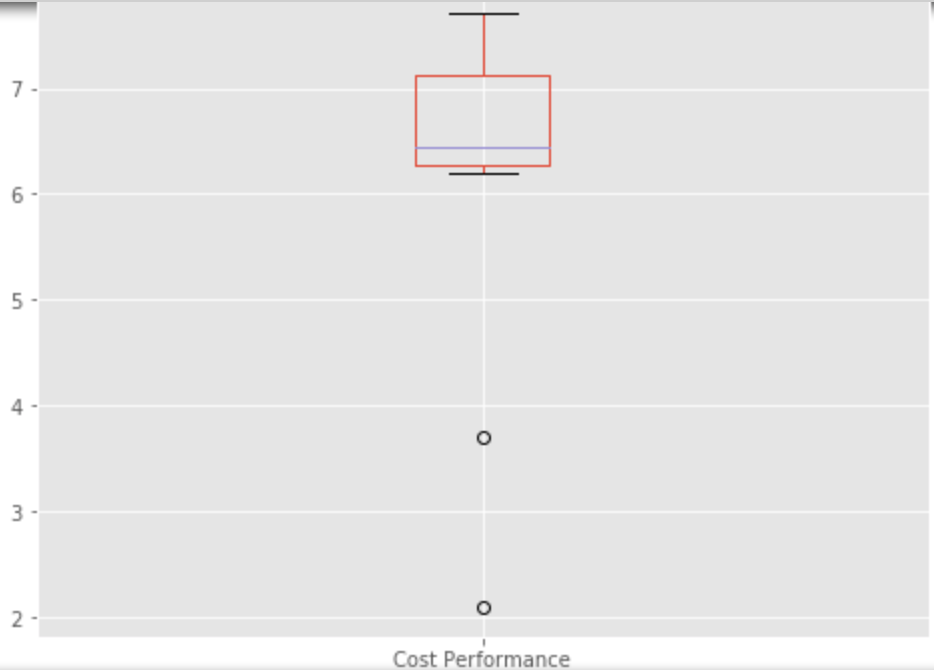
So “Chinese Restaurant” is most “Cost Performance”.

### Data Visualization

Visualizing Data using Matplotlib



|  |
| --- |
| dvcolumn\_names = ['Venue Category','Cost Performance']  DV\_data = pd.DataFrame(columns=dvcolumn\_names)  DV\_data['Venue Category'] = Venue\_dtl\_data['Venue Category']  DV\_data['Cost Performance']=Venue\_dtl\_data['Cost Performance']  DV\_data.head() |
| DV\_data.plot(kind='box', figsize=(8, 6))  plt.title('Box plot of Venue Category for Cost Performance')  plt.show() |



As per analysis according to Cost Perfomance, I recommond to "Chinese Restaurant" as catering shop type.

# Suitable for operating this restaurant

The following four indicators can be used:

1. Population density index
2. Road density index
3. Restorants nunber index

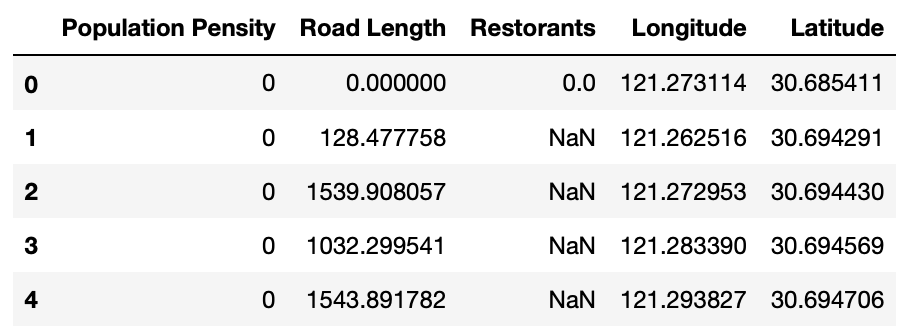
These indicators need is obtained by using QGIS for geospatial analysis.

Evaluation methods: Population Density Index The Higher the Score, the Better;Road Density Index The Higher the Score, the Better; Restaurants Index The Higher the Score, the Better

Comprehensive Index = Population Density Index \* 0.4 + Catering Heat Index \* 0.3 + Road Density Index \* 0.2

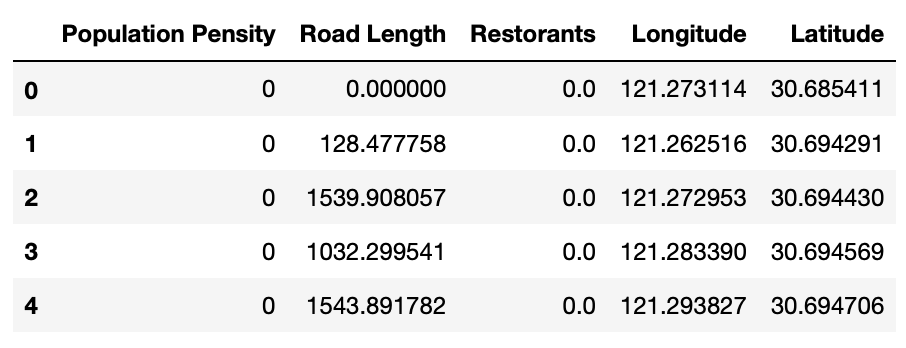
After labeling the four indicators, the weighted average is used to get the comprehensive indicators, and the top10 area is selected as the candidate area by comprehensive ranking.

### Data generated by QGIS to csv format :“Population Pensity”, “Road Length”, “Restaurants”, “Longitude, Latitude”



### Data Standardization

Process "NaN" values. replace it with 0



1. Calculate Population Density Index

df\_data['Population Nor'] = (df\_data['Population Pensity']-df\_data['Population Pensity'].min())/(df\_data['Population Pensity'].max()-df\_data['Population Pensity'].min())

1. Calculate Road Density Index

df\_data['RoadLenght Nor'] = (df\_data['Road Length']-df\_data['Road Length'].min())/(df\_data['Road Length'].max()-df\_data['Road Length'].min())

1. Calculate Restorants Index

df\_data['Restorants Nor'] = (df\_data['Restorants']-df\_data['Restorants'].min())/(df\_data['Restorants'].max()-df\_data['Restorants'].min())

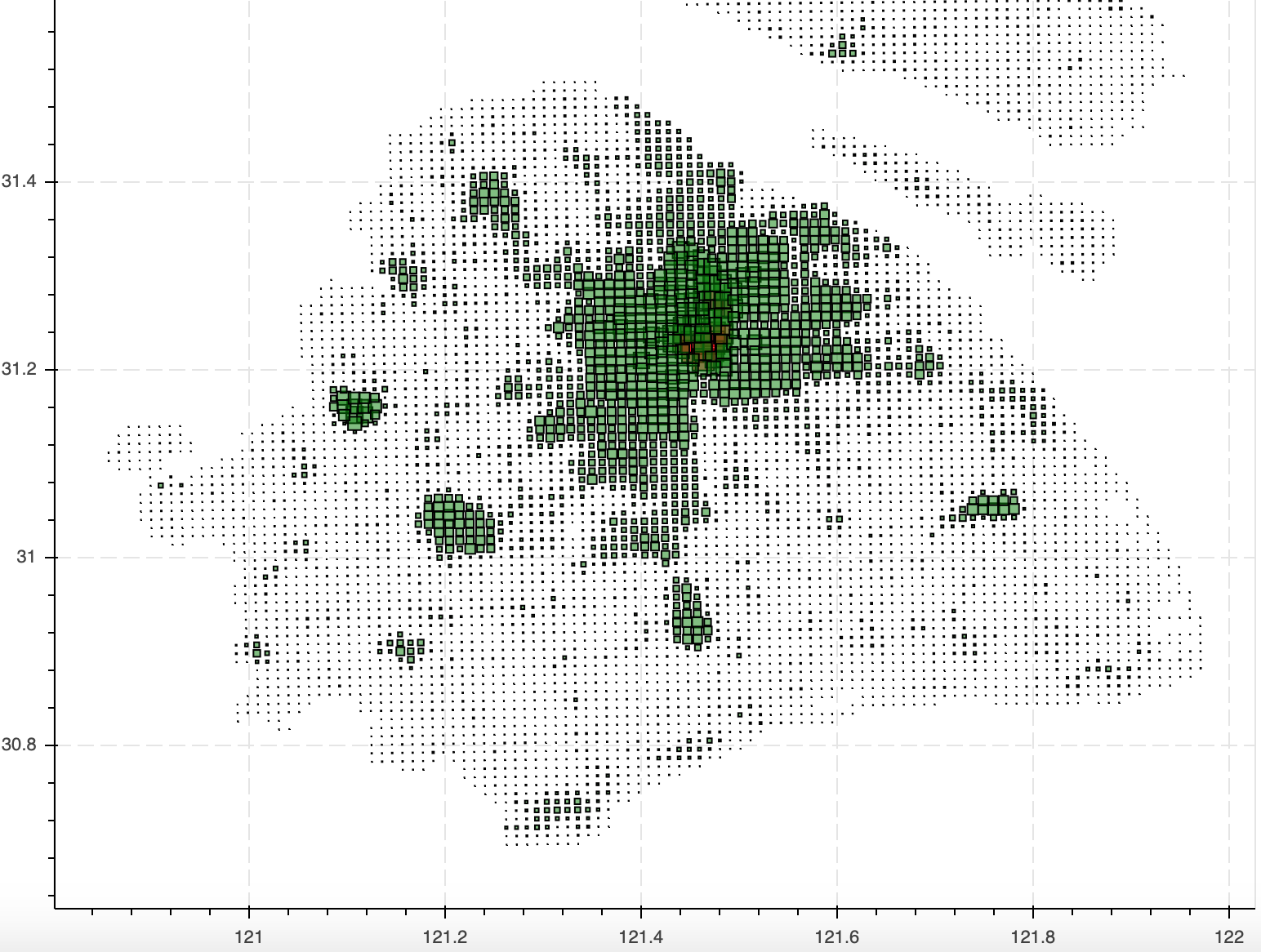
1. Comprehensive Index

df\_data['Score'] = df\_data['Population Nor']\*0.4+df\_data['RoadLenght Nor']\*0.2+df\_data['Restorants Nor']\*0.3

1. Filter out the top 10 higher score

df\_score = df\_data[['Longitude','Latitude','Score']].sort\_values('Score',ascending =False)

1. Figure out the top 10 candidates via Bokeh



# Result

1. What kind of cuisine would attract a lot of customers

Chinese restaurant

1. Which area is suitable for operating his restaurant

Top 10 candidates locations have been marked in red in the above chart

# Conclusion

* For this assignment location data from the Foursquare API was gathered and was analyzed using multiple data science related packages in Python
* The number of data points for this analysis was limited. Despite that, it possible to get a good overview of the business metrics for specific restaurant types in the area of San Francisco