

# Capstone Project - The Battle of the Neighborhoods (Week 2)

## Table of contents

- [Introduction: Business Problem](#)
- [Data](#)
- [Methodology](#)
- [Analysis](#)
- [Results and Discussion](#)
- [Conclusion](#)

## 1 Introduction: Business Problem

---

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Italian restaurant** in **Shanghai**, China.

Since there are lots of restaurants in Berlin we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Italian restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

## 2 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)
- number of and distance to Italian restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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

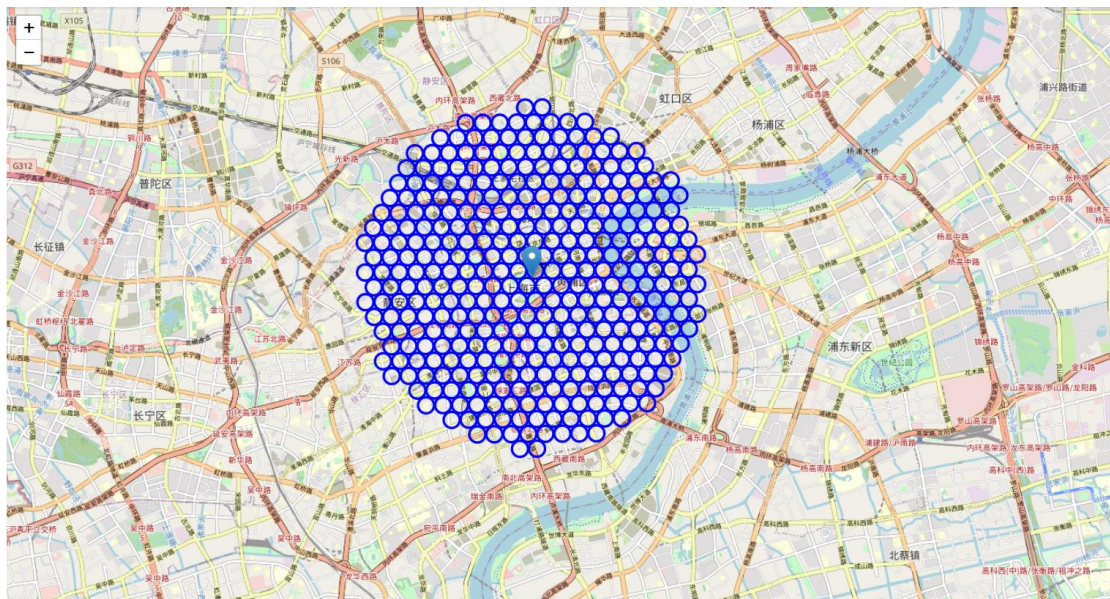
- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Nominatim reverse geocoding**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of Shanghai center will be obtained using **Nominatim geocoding** of well known Shanghai location (Renmin Guangchang)

## Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Shanghai city center.

Let's first find the latitude & longitude of Shanghai city center, using Nominatim.

Create a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**.



	Address	Latitude	Longitude	X	Y	Distance from center
0	中华新路, 静安区, 上海市, 176-0023	31.256585	121.446357	7.872160e+06	1.277142e+07	5992.495307
1	中远两湾城, 远曼路, 宜川路街道, 普陀区, 上海市, 2000065	31.253877	121.444581	7.872760e+06	1.277142e+07	5840.376700
2	M120, 莫干山路, 普陀区, 上海市, 428	31.251170	121.442806	7.873360e+06	1.277142e+07	5747.173218
3	71, 澳门路, 普陀区, 200060	31.248463	121.441032	7.873960e+06	1.277142e+07	5715.767665
4	Habibi, 江宁路, 江宁路街道, 普陀区, 428	31.245756	121.439257	7.874560e+06	1.277142e+07	5747.173218
...	...	...	...	...	...	...
359	沿江风光带, 小东门街道, 上海市, 黄浦区, 上海市, 200011	31.222067	121.502802	7.873360e+06	1.278285e+07	5747.173218
360	中山南路, 小东门街道, 上海市, 黄浦区, 上海市, 200011	31.219362	121.501022	7.873960e+06	1.278285e+07	5715.767665
361	董家渡圣方济各沙勿略堂, 185, 董家渡路, 小东门街道, 上海市, 黄浦区, 上海市, ...	31.216659	121.499242	7.874560e+06	1.278285e+07	5747.173218
362	多稼路, 上海市, 黄浦区, 上海市, 200011	31.213955	121.497462	7.875160e+06	1.278285e+07	5840.376700
363	南浦大桥, 国货路, 半淞园路街道, 上海市, 黄浦区, 200011	31.211251	121.495682	7.875760e+06	1.278285e+07	5992.495307

364 rows × 6 columns

# Foursquare

Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Italian restaurant' category, as we need info on Italian restaurants in the neighborhood.

```
print('List of all restaurants')
print('-----')
for r in list(restaurants.values())[:10]:
    print(r)
print('...')
print('Total:', len(restaurants))

List of all restaurants
-----
('4d702a88641254813735d893', '阿拉丁新疆风味餐厅', 31.255328843643372, 121.44563742205509, '上海市', '上海市', 290, False, 7872430.057231611, 12771433.988260077)
('50b22bc2e4d0b19f64f07f37a', 'Holiday Inn Express Restaurant', 31.254298459225005, 121.44832683150847, 'Zhabei', '上海市', 334, False, 7872384.447431406, 12771920.297372734)
('4f4a6b3c4d0b19f64f07f37a', 'Da Hiang Dumpling (大响水店)', 31.25408928136432, 121.44837244848153, '中江路1630号', 1630 Zhongjin Rd., '上海市', '上海市', 328, False, 7872416.000903373, 12771946.572191697)
('5527c863498e5d11804df88', '九二二', 31.254150258244554, 121.44504038698314, '上海市', '上海市', 53, False, 7872676.933059199, 12771459.383576192)
('4c0505f86d1a13694ac30f', 'Hongzi Chicken (红子鸡)', 31.249316256574097, 121.4402363277745, '188 Aomen Rd, Putuo', '上海市', 319, False, 7873881.349207895, 12771223.055908475)
('4b45c10f964d520661026e3', 'Nagas (沃歌斯)', 31.24443953867173, 121.43762563400816, '155 Changshou Rd | 长寿路155号 (1/F Paris Printemps | 巴黎春天)', Changshoulu, Putuo, '上海市', 200060, 155, False, 7874914.5531)
('5378b04f0b0e2298a0d1f497', 'Hainuli', 31.24446926584776, 121.4392396400099, '36-38 Ximul Rd, 上海', '上海市', 235, False, 7874778.58747052, 12771537.488326888)
('510cfff6e4d0a7d13a356160', '廖华餐厅(长寿路店) Tsui Wah Restaurant', 31.245469524220994, 121.4378847600011, '中国', 272, False, 7874720.120185865, 12771247.616318382)
('4e5b0d6f999372579f9f4d0', 'The Grandma's (外婆家)', 31.24484723944414, 121.43640442254353, '155 Changshou Rd | 长寿路155号 (5/F Paris Printemps | 巴黎春天)', Changshoulu, Putuo, '上海市', 200060, 224, False, 7874)
('50b10437e4d0e2d5a095d2b7', '金鹰海鲜酒家', 31.242780745654785, 121.43675931348012, '2/F, 1048 Xikang Road | 西藏路1048号金鹰国际大厦2楼 (Xinhuai Road | 近新会路)', '上海市', '上海市', 289, False, 7875263.646351679, 127)
...
Total: 1116

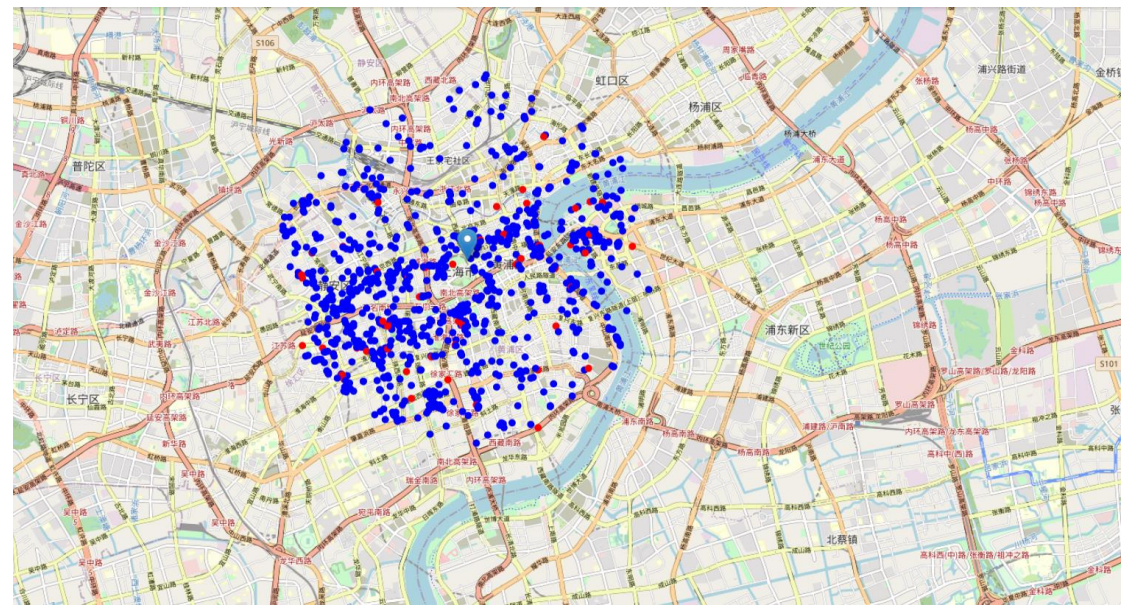
24) print('List of Italian restaurants')
print('-----')
for r in list(italian_restaurants.values())[:10]:
    print(r)
print('...')
print('Total:', len(italian_restaurants))

List of Italian restaurants
-----
('5577a6d4b98e7796e9a5d714', 'SALE & PEPE THE KITCHEN & BAR', 31.235553676191067, 121.43841699661554, '常德路890号1楼-A2室80050HW, 上海', 111, True, 7876345.581455376, 12772264.497506477)
('5a7f7b6c8f9f7002c8805a', 'Funkadelic', 31.23184888027929, 121.43555552974364, 163 Yangping Rd (Wuding Rd), Jing'an, '上海市', 200000, 253, True, 7877201.478361639, 12772203.346246048)
('57780770498e61c07c8b349', 'Bar Centrale', 31.230856926080268, 121.43632124241537, '135 Yangping Road (Wuding Road), 上海', '上海市', 328, True, 7877306.880315028, 12772406.71150831)
('50b22bc2e4d0b19f64f07f37a', 'Lucio's', 31.23146631559415, 121.43623515646112, '135 Yangping Road | 延平路135号 (Wuding Road | 武定路), 上海', '上海市', 212, True, 7877211.316893461, 12772336.477111131)
('534d0a34f0b0e2298a0d1f497', 'Dolomites & Hugin Bar', 31.2366340812756, 121.44401370519372, '363 Kandring Rd (Shanank Rd), Jing'an, 上海', 239, True, 7875634.3865754325, 12773101.483384512)
('547a5a7498e52f942e5262', 'Salzeriya', 31.24448793449591, 121.4522882956885, '恒丰路299号4楼 (太阳市)', '上海市', '上海市', 342, True, 7873712.159447037, 12773422.84727059)
('5334893498e7792b2e2a0d', 'Porto Matto', 31.219136222767391, 121.44304838782709, '2/F, 83 Changshu Road | 常熟路83号 (Julu Road | 巨鹿路), 上海', '上海市', 30, True, 7878731.440492073, 12774491.620506946)
('4b5e1eb1344b713feae9c04', 'Bella Napoli', 31.21818159878254, 121.44089305861375, '946 Changle Rd | 长乐路946号2楼 (Wulumuqi Rd | 乌鲁木齐中路), Jing'an, '上海市', 232, True, 7879084.217259768, 12774241.4635436)

[25] print('Restaurants around location')
print('-----')
for i in range(100, 110):
    rs = location_restaurants[i][:10]
    names = ', '.join([r[1] for r in rs])
    print('Restaurants around location {}: {}'.format(i+1, names))

Restaurants around location
-----
Restaurants around location 101: 中家酒楼
Restaurants around location 102: Cloud South (云海肴)
Restaurants around location 103: alia by OM, Nagas (沃歌斯), Sense 8 (晋八仙酒楼), Cloud South (云海肴)
Restaurants around location 104: Little Sheep Hot Pot (小肥羊)
Restaurants around location 105:
Restaurants around location 106: 嘉喜小馆, 德乐创意韩料料理
Restaurants around location 107: Zhen Ding Chicken (镇鼎鸡), Cha's Restaurant (壹餐厅), Dong Lai Shun (东来顺饭店), Wang Jia Sha (王家沙), Taiyuo Teppanyaki (大渔铁板烧), Lapis Yun (丽珑云海), Bellico (德乐创意韩料料理)
Restaurants around location 108: Little Catch, Café Gray Deluxe, Zhen Ding Chicken (镇鼎鸡), Wang Jia Sha (王家沙), Let's Taipei (让爱台北), Curryfi & Taoker, Taiyuo Teppanyaki (大渔铁板烧), Kemi (糯米)
Restaurants around location 109: Little Catch, 朗庭, Kemi (糯米), Curryfi & Taoker, Heat, 和家, Jujube Tree (Vegetarian Life Style) (枣子树), Suno (相扑烤肉)
Restaurants around location 110: 朗庭, Hongtang Small Wonton (弄堂小馄饨), Ruang Thai Kitchen (泰国), 美斯中心店, 厘点港式茶餐厅 Casual Dining
```

Let's now see all the collected restaurants in our area of interest on map, and let's also show Italian restaurants in different color.





# Methodology

In this project we will direct our efforts on detecting areas of Shanghai that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required **data: location and type (category) of every restaurant within 6km from Shanghai center** (RenminGuangchang). We have also **identified Italian restaurants** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Shanghai - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (*and* no Italian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two restaurants in radius of 250 meters**, and we want locations **without Italian restaurants in radius of 800 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

# Analysis

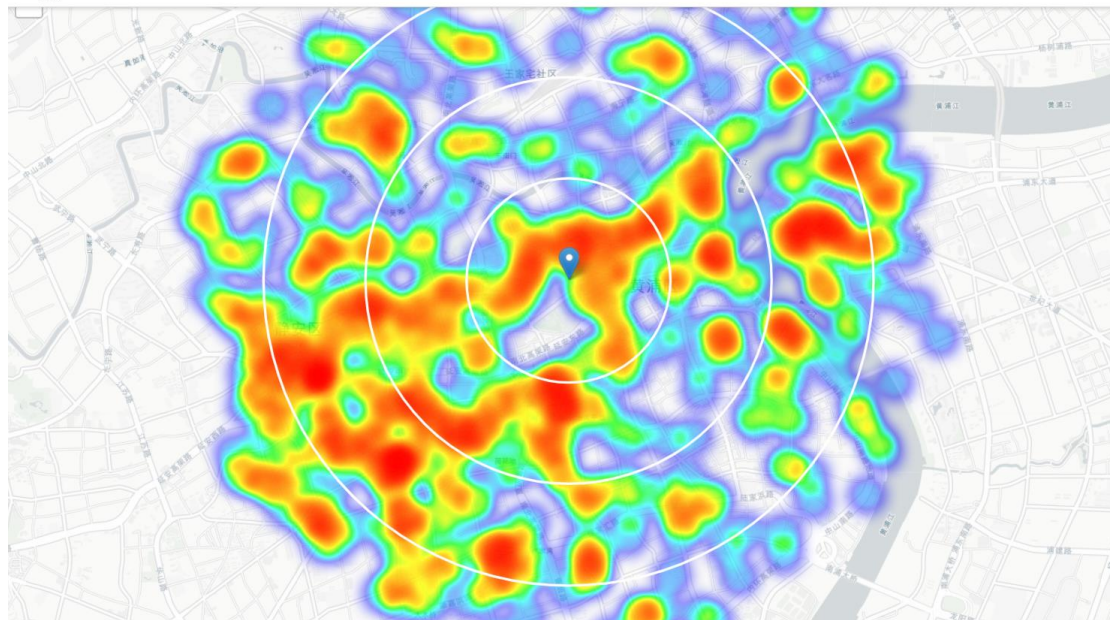
Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **number of restaurants in every area candidate**:

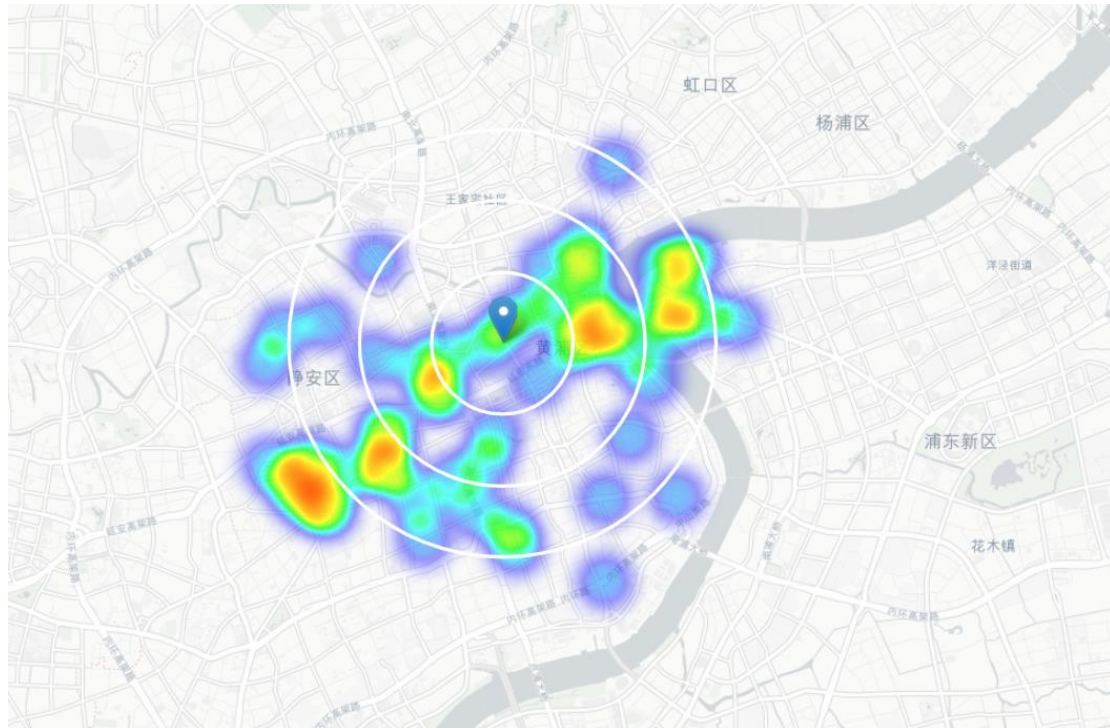
Average number of restaurants in every area with radius=300m: 6.538461538461538									
	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area		
0	中华新路, 静安区, 上海市, 176-0023	31.256585	121.446357	7.872160e+06	1.277142e+07	5992.495307	1		
1	中远两湾城, 远景路, 普陀区, 上海市, 2000065	31.253877	121.444581	7.872760e+06	1.277142e+07	5840.376700	2		
2	M120, 莫干山路, 普陀区, 上海市, 428	31.251170	121.442806	7.873360e+06	1.277142e+07	5747.173218	0		
3	71, 澳门路, 普陀区, 200060	31.248463	121.441032	7.873960e+06	1.277142e+07	5715.767665	0		
4	Habibi, 江宁路, 江宁路街道, 普陀区, 428	31.245756	121.439257	7.874560e+06	1.277142e+07	5747.173218	4		
5	北横通道, 静鼎安邦府邸, 静安区, 200042	31.243049	121.437483	7.875160e+06	1.277142e+07	5840.376700	9		
6	安远路, 江宁路街道, 静安区, 200060	31.240342	121.435709	7.875760e+06	1.277142e+07	5992.495307	5		
7	foot massage heaven, 芷江西路, 芷江西路街道, 王家宅社区, 静安区, ...	31.259325	121.451750	7.871260e+06	1.277194e+07	5855.766389	2		
8	中华新路, 芷江西路街道, 局东村, 王家宅社区, 静安区, 上海市, 200071	31.256618	121.449974	7.871860e+06	1.277194e+07	5604.462508	0		
9	上海长途客运总站, 交通路, 天目西路街道, 静安区, 上海市, 176-0023	31.253910	121.448198	7.872460e+06	1.277194e+07	5408.326913	3		

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Italian restaurant
0	中华新路, 静安区, 上海市, 176-0023	31.256585	121.446357	7.872180e+06	1.277142e+07	5992.495307	1	2534.614895
1	中远两湾城, 远置路, 宜川路街道, 普陀区, 上海市, 2000065	31.253877	121.444581	7.872780e+06	1.277142e+07	5840.376700	2	2218.548998
2	M120, 莫干山路, 普陀区, 上海市, 428	31.251170	121.442806	7.873380e+06	1.277142e+07	5747.173218	0	2034.611942
3	71, 澳门路, 普陀区, 200060	31.248463	121.441032	7.873980e+06	1.277142e+07	5715.767665	2	2019.240526
4	Habibi, 江宁路, 江宁路街道, 普陀区, 428	31.245756	121.439257	7.874580e+06	1.277142e+07	5747.173218	5	1975.475700
5	北横通道, 静安安邦府邸, 静安区, 200042	31.243049	121.437483	7.875180e+06	1.277142e+07	5840.376700	8	1456.050953
6	安远路, 江宁路街道, 静安区, 200060	31.240342	121.435709	7.875780e+06	1.277142e+07	5992.495307	7	1028.428175
7	foot massage heaven, 芷江西路, 芷江西路街道, 王家宅社区, 静安区, ...	31.259325	121.451750	7.871280e+06	1.277194e+07	5855.786389	2	2866.214037
8	中华新路, 芷江西路街道, 周东村, 王家宅社区, 静安区, 上海市, 200071	31.258618	121.449974	7.871880e+06	1.277194e+07	5604.462508	0	2373.366691
9	上海长途客运总站, 孔家木桥路, 天目西路街道, 静安区, 上海市, 176-0023	31.253910	121.448198	7.872480e+06	1.277194e+07	5408.326913	3	1941.791954

OK, so **on average Italian restaurant can be found within ~950m** from every area center candidate. That's fairly close, so we need to filter our areas carefully!

Let's crete a map showing **heatmap / density of restaurants** and try to extract some meaningfull info from that. Also, let's show **borders of Shanghai boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from RenminGuangchang.

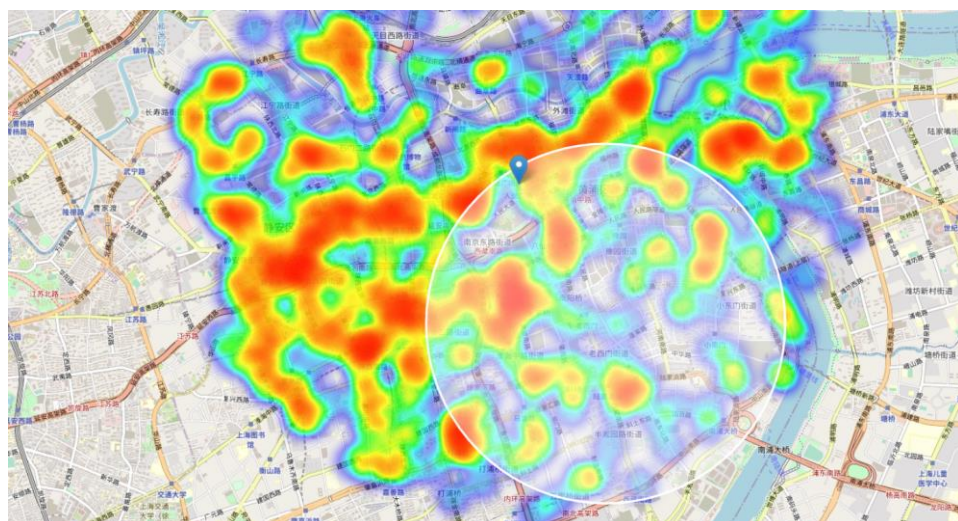




This map is not so 'hot' (Italian restaurants represent a subset of ~6.4% of all restaurants in Shanghai) but it also indicates higher density of existing Italian restaurants directly east and south-west from RenminGuangchang, with closest pockets of **low Italian restaurant density positioned south-east, north-west and north from city center**.

Based on this we will now focus our analysis on areas *south-east, north-west and north from Shaihai center* - we will move the center of our area of interest and reduce it's size to have a radius of **2.0km**. This places our location candidates mostly in boroughs **Jingan and Huangpu** .

Let's define new, more narrow region of interest, which will include low-restaurant-count parts of Huangpu and Jingan closest to Renmin Guangchang.



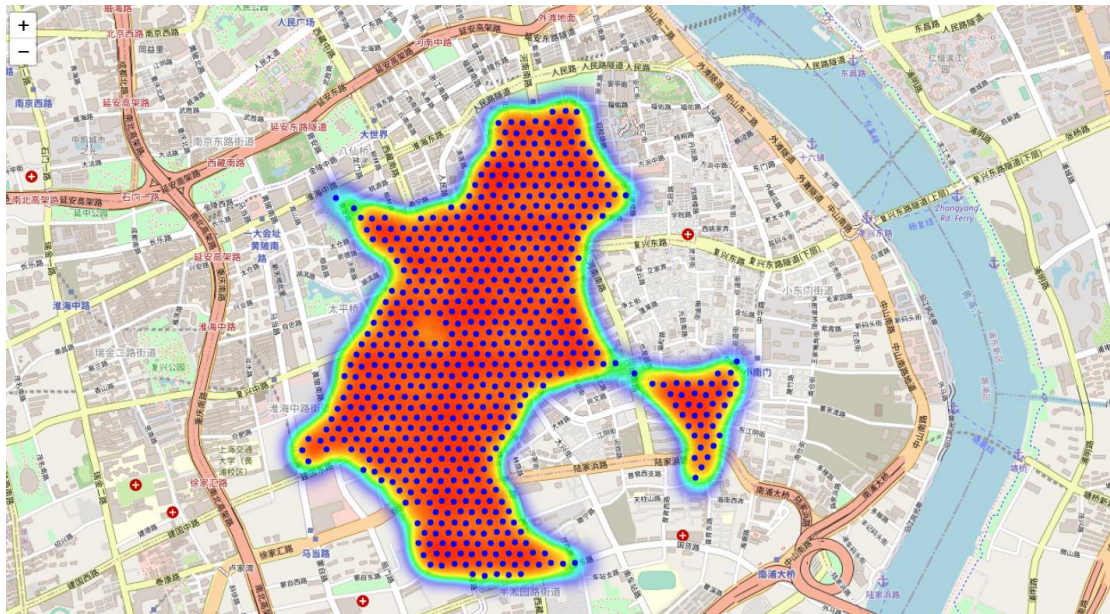


Now let's calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Italian restaurant** and then **filter** those locations: we're interested only in **locations with no more than 2 restaurants in radius of 250 meters**, and **no Italian restaurants in radius of 800 meters**.

```
Locations with no more than two restaurants nearby: 1205  
Locations with no Italian restaurants within 800m: 698  
Locations with both conditions met: 646
```

We now have a bunch of locations fairly close to Renmin Guangchang (mostly in Huanpu and Jingan), and we know that each of those locations has no more than two restaurants in radius of 250m, and no Italian restaurant closer than 800m. Any of those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition.

Let's now show those good locations in a form of heatmap:

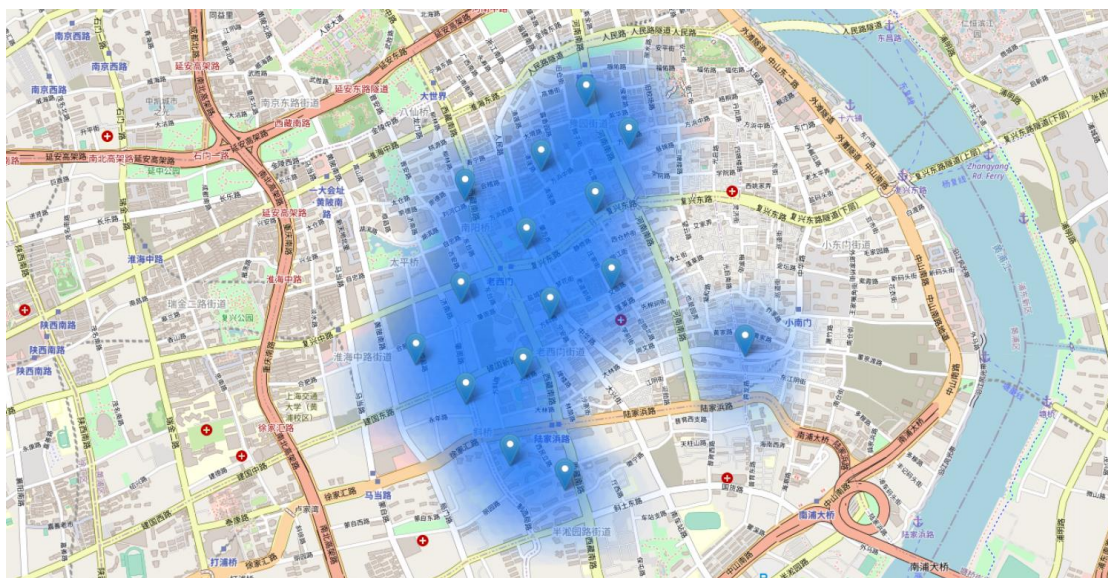


Let us now cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.



Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby, all zones being fairly close to city center (all less than 4km from Alexanderplatz, and about half of those less than 2km from Renmin Guangchang). Although zones are shown on map with a radius of ~200 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located in Huangpu boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.



## Results and Discussion

---



Our analysis shows that although there is a great number of restaurants in Berlin (~1116 in our initial area of interest which was 12x12km around Renmin Guangchang), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and west from Alexanderplatz, so we focused our attention to areas south-east and east, corresponding to boroughs Huangpu and Jingan which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics *and* a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approx. 4x4km south-east from Renmin Guangchang) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Italian restaurant closer than 800m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Berlin center but not crowded with existing restaurants (particularly Italian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area.

Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

## Conclusion

---

Purpose of this project was to identify Shanghai areas close to center with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Huangpu and Jingan), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or Huangpu River), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.