Predicting an optimal location for a Hot Pot restaurant

Caoyu

25 August 2019

1. Introduction: Business Problem

1.1. Background

Chengdu is a UNESCO world food capital.Sichuan hot pot is the most popular and famous food in chengdu.Hotpot in sichuan is famous for its hemp, spicy, fresh and fragrant flavor. It comes from the people and is promoted to the temple. Whether it is menial servants, officials , literati , workers, or young men and women, the range of consumer groups and Consumption per capita are all higher than that of any other place.As a delicacy, hot pot has become the representative cuisine of sichuan and chongqing.

1.2. Problem

Datalization of the location of a hot pot restaurant can help analyze the location of a newly opened hot pot restaurant.Factors such as the number of hotpot restaurants in each block, the distance to the city center, and the data

of nearby shopping malls are available

1.3. Interest

Machine-learning algorithms can help investors choose a suitable address in chengdu to open a new hotpot restaurant

2. Data

Now let's create a grid of area candidates, equaly spaced, centered around city center and within ~16km from TianFu Square. Our neighborhoods will be defined as circular areas with a radius of 400 meters, so our neighborhood centers will be 800 meters apart.
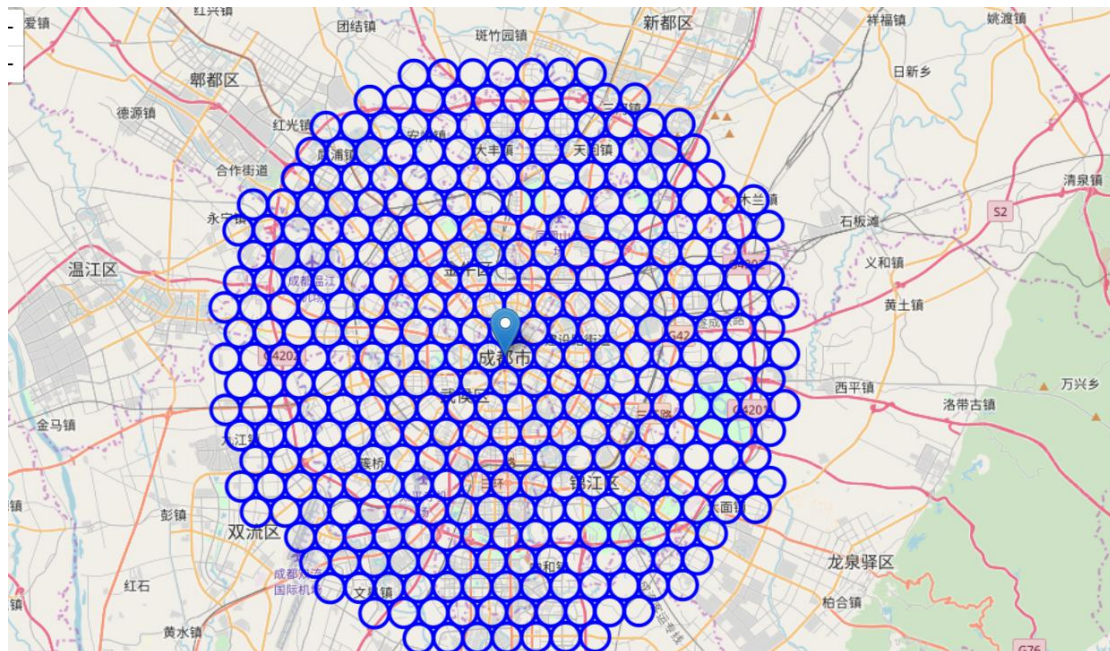
Let's now use BaiDu Maps API to get approximate addresses of those locations and place all this into a Pandas dataframe.

	Address	formatted_address	Latitude	Longitude	X	Y	Distance from center
0		四川省成都市双流区	30.524351	104.014528	405454.853184	3.377303e+06	15979.987484
1		四川省成都市双流区锦江路四段	30.524476	104.031203	407054.853184	3.377303e+06	15574.337867
2		四川省成都市双流区滨河路2段	30.524599	104.047878	408654.853184	3.377303e+06	15325.795249
3		四川省成都市双流区伏龙路1段	30.524720	104.064554	410254.853184	3.377303e+06	15242.047107
4		四川省成都市双流区新通大道	30.524838	104.081229	411854.853184	3.377303e+06	15325.795249
5		四川省成都市双流区新通大道	30.524955	104.097905	413454.853184	3.377303e+06	15574.337867
6		四川省成都市双流区双简路	30.525069	104.114581	415054.853184	3.377303e+06	15979.987484
7		四川省成都市双流区长城路2段-59号	30.536661	103.989386	403054.853184	3.378689e+06	15615.377037
8		四川省成都市双流区	30.536789	104.006063	404654.853184	3.378689e+06	14945.233354
9		四川省成都市双流区荣店路	30.536916	104.022740	406254.853184	3.378689e+06	14422.205102
10		四川省成都市武侯区	30.537040	104.039418	407854.853184	3.378689e+06	14062.716665

Now that we have our location candidates, let's use BaiDu API to get info on restaurants in each neighborhood and these data at local file system (restaurants, hot_pot_restaurants, location_restaurants).

Total number of restaurants: 7919

Total number of hot_pot_restaurants: 492

Percentage of Hot Pot restaurants: 6.21%

Average number of restaurants in neighborhood: 21.755494505494507

3. 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: 21.755494505494507

	Address	formatted_address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Hot Pot restaurant
0		四川省成都市双流区	30.524351	104.014528	405454.853184	3.377303e+06	15979.987484	8	3388.639029
1		四川省成都市双流区锦江路四段	30.524476	104.031203	407054.853184	3.377303e+06	15574.337867	0	1815.296410
2		四川省成都市双流区滨河路2段	30.524599	104.047878	408654.853184	3.377303e+06	15325.795249	17	477.208179
3		四川省成都市双流区伏龙路1段	30.524720	104.064554	410254.853184	3.377303e+06	15242.047107	91	373.190451
4		四川省成都市双流区新通大道	30.524838	104.081229	411854.853184	3.377303e+06	15325.795249	2	1080.333151
5		四川省成都市双流区新通大道	30.524955	104.097905	413454.853184	3.377303e+06	15574.337867	0	2671.231484
6		四川省成都市双流区双简路	30.525069	104.114581	415054.853184	3.377303e+06	15979.987484	0	3429.724512
7		四川省成都市双流区长城路2段-59号	30.536661	103.989386	403054.853184	3.378689e+06	15615.377037	0	1937.303086
8		四川省成都市双流区	30.536789	104.006063	404654.853184	3.378689e+06	14945.233354	0	2169.230737
9		四川省成都市双流区荣店路	30.536916	104.022740	406254.853184	3.378689e+06	14422.205102	0	2431.864122

so on average Hot Pot restaurant can be found within ~1500m from every area center candidate. That's fairly close, so we need to filter our areas carefully!


Fig:heat map of restaurant

Fig:heat map of hot pot restaurant

Let's define new, more narrow region of interest, which will include low-restaurant-count parts of South high-tech zone

Not bad - this nicely covers all the pockets of low restaurant density in South high-tech zone. Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates

200m appart).

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Hot Pot restaurant
0	30.550142	104.034076	407354.853184	3.380145e+06	0	1100.503398
1	30.550157	104.036161	407554.853184	3.380145e+06	0	1049.830584
2	30.550173	104.038246	407754.853184	3.380145e+06	2	1035.944391
3	30.550188	104.040331	407954.853184	3.380145e+06	2	1060.291143
4	30.550203	104.042416	408154.853184	3.380145e+06	2	1120.381111
5	30.550219	104.044501	408354.853184	3.380145e+06	4	1210.904769
6	30.550234	104.046586	408554.853184	3.380145e+06	11	1325.642066
7	30.550249	104.048671	408754.853184	3.380145e+06	17	1405.123215
8	30.550264	104.050756	408954.853184	3.380145e+06	24	1239.186899
9	30.550280	104.052841	409154.853184	3.380145e+06	24	1084.802789
10	30.550295	104.054926	409354.853184	3.380145e+06	17	947.633902
11	30.550310	104.057011	409554.853184	3.380145e+06	8	810.185250
12	30.550325	104.059096	409754.853184	3.380145e+06	6	630.186232
13	30.550340	104.061181	409954.853184	3.380145e+06	4	466.764647
14	30.550355	104.063266	410154.853184	3.380145e+06	33	344.389001

OK. Let us now filter those locations: we're interested only in locations with no more than 5 restaurants in radius of 500 meters, and no Italian restaurants in radius of 1000 meters and see how this looks on a map
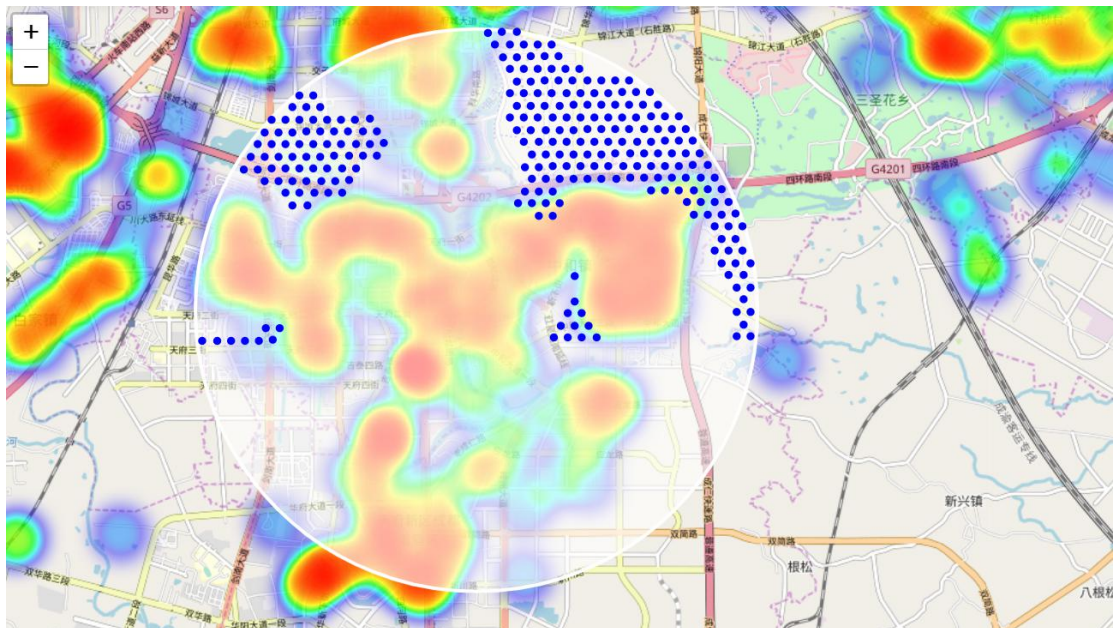
Looking good. We now have a bunch of locations fairly close to High Tech Center, and we know that each of those locations has no more than ten

restaurants in radius of 500m, and no Hot Pot restaurant closer than 1000m. Any of those locations is a potential candidate for a new Hot restaurant, at least based on nearby competition.Let's now show those good locations in a form of heatmap:
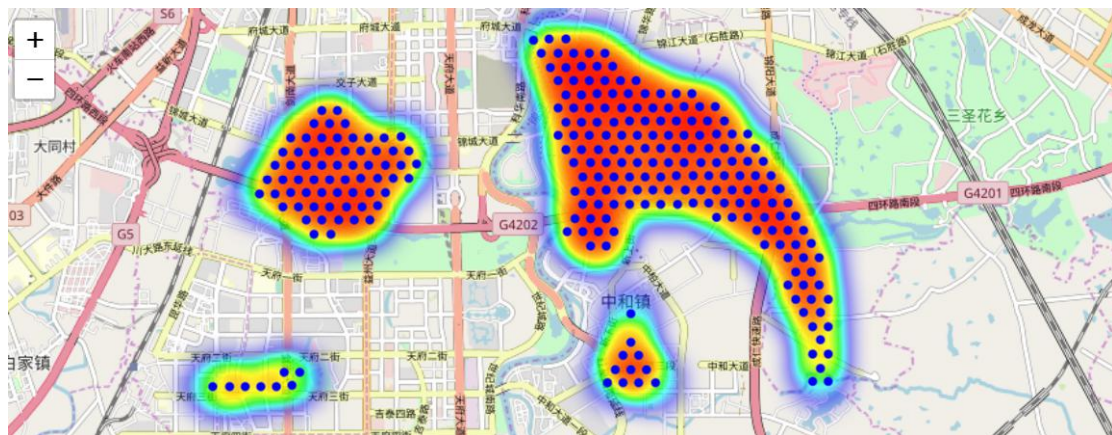

Looking good. What we have now is a clear indication of zones with low number of restaurants in vicinity, and no Hot Pot restaurants at all nearby.

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.

Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.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.

Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:
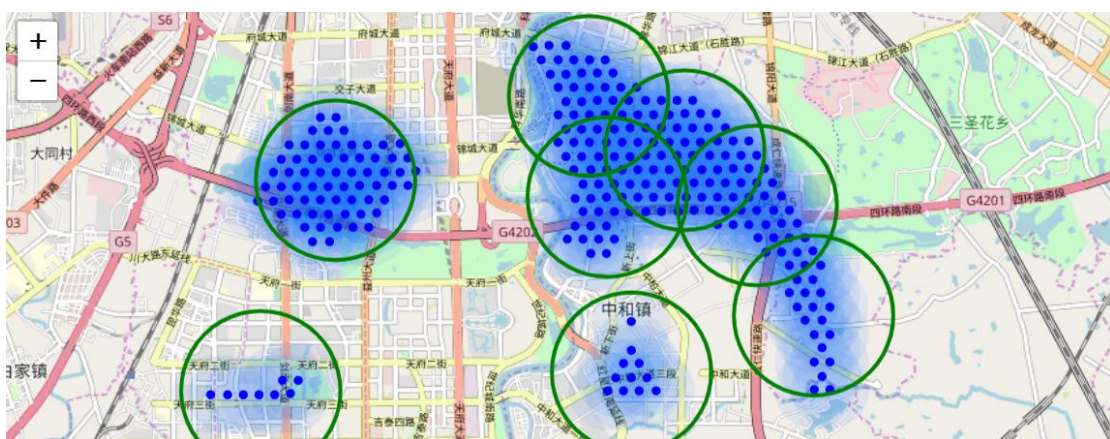
Finaly, let's reverse geocode those candidate area centers to get the addresses which can be presented to stakeholders.

```
========================================================
Addresses of centers of areas recommended for further analysis
========================================================

四川省成都市武侯区锦悦西路                              => 9.8km from TianFuSquare
四川省成都市锦江区康达路                            => 9.9km from TianFuSquare
四川省成都市锦江区                          => 8.9km from TianFuSquare
四川省成都市锦江区琉新路                                  => 12.4km from TianFuSquare
四川省成都市武侯区荣华南路                                => 12.6km from TianFuSquare
四川省成都市双流区中和大道62号                            => 12.4km from TianFuSquare
四川省成都市锦江区G4202(成都绕城高速)                           => 10.2km from TianFuSquare
四川省成都市双流区江家立交                                => 10.9km from TianFuSquare
```

This concludes our analysis. We have created 8 addresses representing centers of zones containin`g locations with low number of restaurants and no Hot Pot restaurants nearby, all zones being fairly close to high-tech district center . Although zones are shown on map with a radius of ~1000 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 high-tech district , which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.
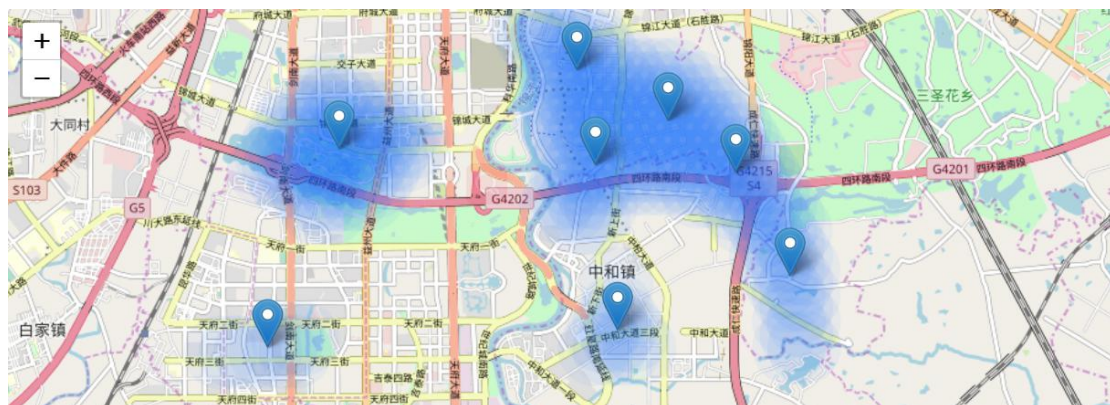
4. Results and Discussion

Our analysis shows that although there is a great number of restaurants in ChengDu (~There are nearly 10,000 restaurants within a 16-kilometer radius of tianfu square), there are pockets of low restaurant density fairly close to high-tech district center. Highest concentration of restaurants was detected north and south west from TianFu square, so we focused our attention to areas south(high-tech district center),which offer a combination of 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. 8x8km from TianFu square) we first created a dense grid of location candidates (spaced 200m appart); those locations were then filtered so that those with more than ten restaurants in radius of 500m and those with an Hot Pot restaurant closer than 1000m 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 baidu api to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 8 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Hot Pot 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 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.

5. Conclusion

Purpose of this project was to identify ChengDu areas close to center with low number of restaurants (particularly Hot Pot restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Hot Pot restaurant. By calculating restaurant density distribution from BaiDu data we have first identified general boroughs that justify further analysis (high-tech district), 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 decission 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 water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.