Good Locations for Having a New Tea Shop

Chung-Han Lee

2019.08.04

1. Introduction

It is always so hard to have own business, especially when it comes to having a shop at a good location. Bubble tea is a famous beverage and food in Asia. Therefore, I will talk about how shall we find a good location for starting up a bubble tea shop in this project. This project will focus on locations in cities in Germany.

The distribution of venues nearby the bubble tea shops in big german cities will be analysed, and see if the other cities, which haven't had bubble shops yet, are similar to the cities in this venues distribution. If they have high similarity, we can conclude that open a new bubble tea shop in these cities should be a good decision.

This project also can be seen as a good way for thinking how to choose and compare difference places for business.

2. Data

In this project, I will use data from

i. **Wikipedia,** in order to get the rank of cities in population, and its location

ii. **Foursquare**, in which the current running bubble tea shops in Germany, their information and venues near them can easily be found. Besides, we will also fetch venues information of some cities in order to realise if they are potent to host a bubble tea shop.

3. Methodology

At first, I will introduce how do I establish a method to find a place and evaluate it. It can be split into the following steps.

(1) realise what does a good bubble tea shop mean, and what makes them popular in business? In this part, I will collect information of all bubble tea shops in bigger german cities, and see how they popular are. I will use the amount of ratings and reviews as pointer. It will be assumed that more ratings and reviews

mean more customers, which have been visited them, no matter how do they or whether do they like them.

(2) check how many and what kinds of venues are nearby them. This seems to be useful, because a bubble tea shop cannot just stand there alone. For example, you can hardly imagine that a bubble shop in a factory area is easily profitable, but near shopping zone or at walking zone for shopping. They must have a certain relationship with venues nearby them.

(3) analyse the venues in another cities, which don't have bubble shops yet. Then see how these cities are similar to those cities which have bubble shops already. The higher the similarity, the better to be host a bubble shop.

Step 1. create a cities list by population from Wiki:
https://en.wikipedia.org/wiki/List_of_cities_in_Germany_by_population
and wrangle it in a viewable way

	City	Population	latitude	longitude
0	Berlin	3520031	52.51700	13.38300
1	Hamburg	1787408	53.55000	10.00000
2	München	1450381	48.13300	11.56700
3	Köln	1060582	50.93300	6.95000
4	Frankfurt am Main	732688	50.11700	8.68300
5	Stuttgart	623738	48.78300	9.18300
6	Düsseldorf	612178	51.23300	6.78300
7	Dortmund	586181	51.51700	7.46700
8	Essen	582624	51.45000	7.01700
9	Leipzig	560472	51.33300	12.38300
10	Bremen	557464	53.08300	8.80000

figure 1. wrangled german city list

Step 2. search all the bubble tea shops in Germany with important information, such as name, coordinate, city and their ratings and reviews on Foursquare. The number of ratings and reviews are very important, because we see them as a pointer of popularity, as you can see below, the red marked column. Here we will

select the top 15 biggest cities. There are 45 bubble shops selected out, I will only show the head of this data frame here.

	name	categories	address	formattedAddress	cc	city	country	lat	lng	postalCode	state	id	pointer
0	ComeBuy Bubble Tea	Bubble Tea Shop	Rosenthaler Str. 31	[Rosenthaler Str. 31, 10178 Berlin, Deutschland]	DE	Berlin	Deutschland	52.525403	13.403421	10178	Berlin	4ec7c6e0b634b2fd74c12ec9	313
1	ComeBuy Bubble Tea	Bubble Tea Shop	Joachimstaler Str. 14	[Joachimstaler Str. 14, 10719 Berlin, Deutschl...	DE	Berlin	Deutschland	52.502060	13.331109	10719	Berlin	561a8378498ed6032c81078e	40
2	sphere bay	Bubble Tea Shop	Immermannstr. 29	[Immermannstr. 29 (Oststr.), 40210 Düsseldorf,...	DE	Düsseldorf	Deutschland	51.223733	6.787875	40210	Nordrhein-Westfalen	4f5cbb95e4b02628c23716e1	28
3	teamate	Bubble Tea Shop	Immermannstr. 65C	[Immermannstr. 65C (Karlstr.), 40210 Düsseldor...	DE	Düsseldorf	Deutschland	51.221648	6.791776	40210	Nordrhein-Westfalen	57375b8b498ebbf2491e8f81	27
4	Na Na	Bubble Tea Shop	Schweizer Strasse 73	[Schweizer Strasse 73, 60594 Frankfurt am Main...	DE	Frankfurt am Main	Deutschland	50.100432	8.682099	60594	Hessen	4fa8f6b3e4b099106f6163dc	17

figure 2. bubble shops data frame

Step 3. label them in a reasonable way. Here I set **four** classes. The zero means the hottest shops.

Step 4. visualise on a map. The bigger the points are, the hotter they are.

figure 3. hot spot of bubble tea shops in Germany

Step 5. search venues nearby them, in order to know how the spot looks like. Here are 3242 venues searched out in total.

```
In [84]: print(shopnearby_venues.shape)
         shopnearby_venues.head()

         (3242, 7)
Out[84]:
```

	Shop	Shop Latitude	Shop Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ComeBuy Bubble Tea	52.525403	13.403421	BEN RAHIM	52.525168	13.401928	Coffee Shop
1	ComeBuy Bubble Tea	52.525403	13.403421	Die Hackeschen Höfe	52.524094	13.402157	Monument / Landmark
2	ComeBuy Bubble Tea	52.525403	13.403421	Sophiensaele	52.526031	13.401385	Dance Studio
3	ComeBuy Bubble Tea	52.525403	13.403421	Neurotitan	52.524488	13.401809	Art Gallery
4	ComeBuy Bubble Tea	52.525403	13.403421	Casa Camper Boutique Hotel Berlin Mitte	52.525799	13.403818	Hotel

figure 4. venues nearby all the bubble tea shops

Step 6. overview of venues nearby them

	Shop	City	Latitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ComeBuy Bubble Tea	Berlin	52.525403	Clothing Store	Coffee Shop	Hotel	Café	Italian Restaurant	Optical Shop	Ice Cream Shop	Vietnamese Restaurant	Yoga Studio	Boutique
1	ComeBuy Bubble Tea	Berlin	52.502060	Hotel	Clothing Store	Italian Restaurant	Movie Theater	Burger Joint	Café	German Restaurant	Furniture / Home Store	Steakhouse	French Restaurant
2	sphere bay	Düsseldorf	51.223733	Japanese Restaurant	Korean Restaurant	Hotel	Chinese Restaurant	Ramen Restaurant	Café	Grocery Store	Italian Restaurant	Clothing Store	Coffee Shop
3	teamate	Düsseldorf	51.221648	Hotel	Japanese Restaurant	Korean Restaurant	Ramen Restaurant	Chinese Restaurant	Café	Turkish Restaurant	Grocery Store	Coffee Shop	Italian Restaurant
4	Na Na	Frankfurt am Main	50.100432	German Restaurant	Café	Park	Italian Restaurant	Falafel Restaurant	Eastern European Restaurant	Organic Grocery	Supermarket	Sushi Restaurant	Bar
5	Tea Time	Hamburg	53.573097	label	Bakery	Chinese Restaurant	Hotel	Bavarian Restaurant	Italian Restaurant	Steakhouse	Sushi Restaurant	German Restaurant	Spanish Restaurant
6	Yobu	Leipzig	51.344756	label	Hotel	Coffee Shop	Japanese Restaurant	Shopping Mall	Drugstore	Tram Station	Bistro	Hostel	Indian Restaurant

figure 5. sorted venues nearby all the bubble tea shops

Step7. Feed venues and labels to machine learning algorithm, to get the classification label, here we will use decision tree due to its high accuracy so far.

Step 8. create another data frame, in which the other top 15 cities are included and seen as test set for machine learning. Here is the head of result.

	City	Latitude	label
0	Essen	51.451477	3
1	Bremen	53.534582	3
2	Duisburg	51.431813	3
3	Bochum	51.477398	2
4	Wuppertal	51.255115	1

4. Result

As a result after the set of undiscovered cities labeled by machine learning method, decision tree, we can clearly see the result, as the folium map showed below.
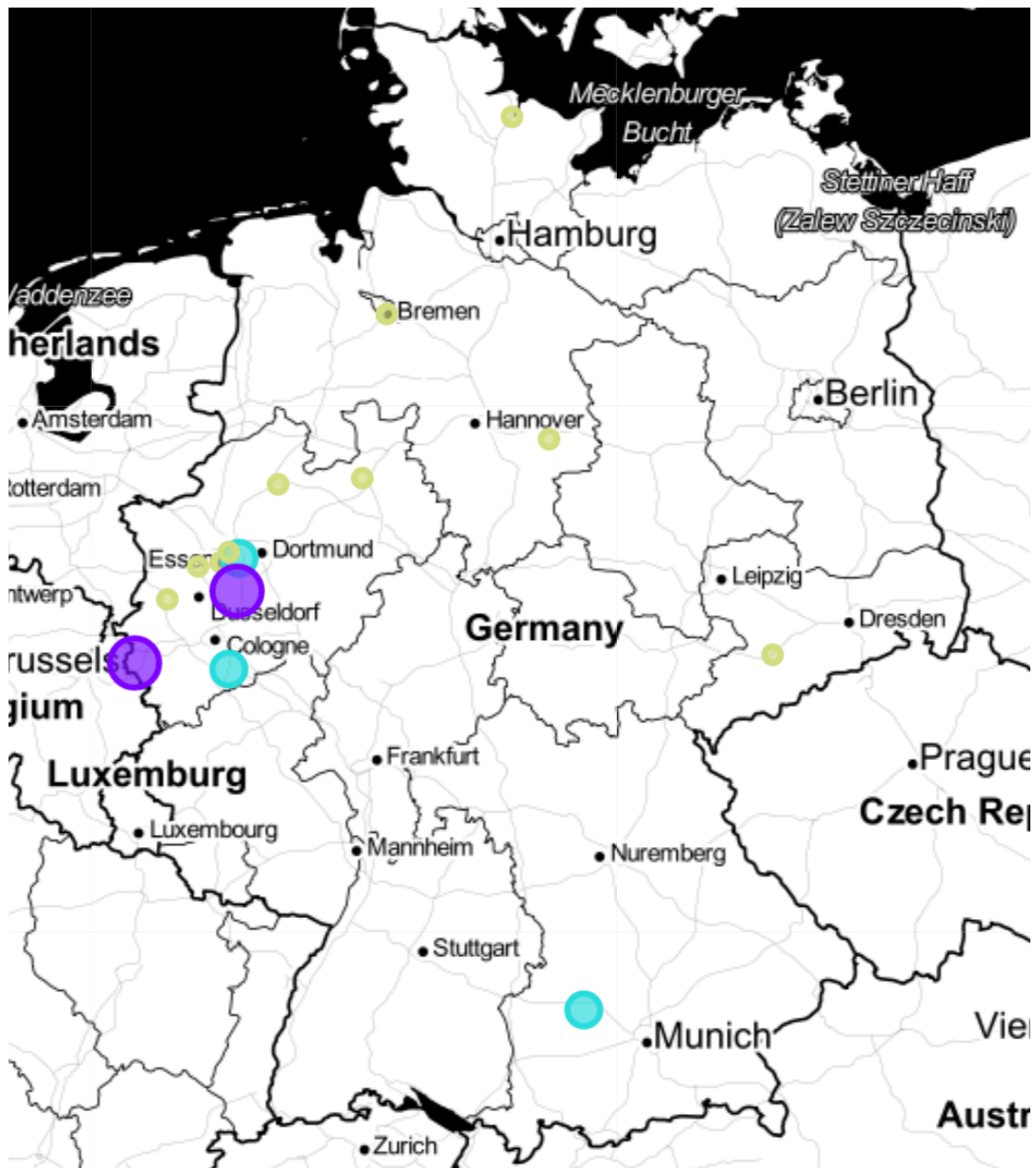
figure 7. predicted hot spots

The larger the points, the better location they have. We can see the two biggest conspicuous purple points on the west side of Germany. Let's see who are they.

	City	Population	latitude	longitude	hbf_lat	hbf_lng	label
4	Wuppertal	350046	51.267	7.183	51.255115	7.151224	1
14	Aachen	245885	50.783	6.083	50.767494	6.089211	1
3	Bochum	364742	51.483	7.217	51.477398	7.218545	2
6	Bonn	318809	50.733	7.100	50.732113	7.098196	2
8	Augsburg	286374	48.367	10.900	48.365862	10.888834	2
0	Essen	582624	51.450	7.017	51.451477	7.014468	3
1	Bremen	557464	53.083	8.800	53.534582	8.597749	3
2	Duisburg	491231	51.433	6.767	51.431813	6.776627	3
5	Bielefeld	333090	52.017	8.533	52.028578	8.533019	3
7	Münster	310039	51.967	7.633	51.956334	7.633311	3
9	Gelsenkirchen	260368	51.517	7.100	51.504768	7.102298	3
10	Mönchengladbach	259996	51.200	6.433	51.196720	6.446273	3
11	Braunschweig	251364	52.267	10.517	52.252357	10.538698	3
12	Chemnitz	248645	50.833	12.917	50.837106	12.929267	3
13	Kiel	246306	54.333	10.133	54.313446	10.131014	3

figure 8. label list sorted by class

The two best choices are Wuppertal and Aachen. We also can see the venues distribution near them.

City	Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Wuppertal	1	Hotel	Drugstore	Café	Supermarket	Clothing Store	Ice Cream Shop	Platform	Bakery	Thai Restaurant	Modern European Restaurant
Aachen	1	Bakery	Fast Food Restaurant	Drugstore	Vietnamese Restaurant	Italian Restaurant	Convenience Store	Pizza Place	Bar	Supermarket	Chinese Restaurant

figure 9. venue distribution nearby predicted hot spots

5. Discussion

Although we have brought out the results of how to find a good city to have a bubble tea shop, the research, which I have done here, can still be improved.

(1) low accuracy - as you can see, the accuracy provided by **jaccard** method, was not good.

```
print("DecisionTrees's Accuracy: ", metrics.accuracy_score(y_testset, predTree))
```

```
DecisionTrees's Accuracy:  0.2857142857142857
```

figure 10. result of accuracy by decision tree algorithm

Why does it look like that? The **main reason** should be that we have too many categories but too less samples. In this project, only 45 shops were analysed. **The second point** could be that those characters we selected out are not so obviously related to a successful bubble shop. To be a solution, another characters may have to be considered and involved, such as weather (average temperature), amount of tourists, age distribution of the population in those cities, and so on.

(2) targets are not selected well - bubble tea shops are relatively rare in Germany, nevertheless I adopted the data from Germany over Foursquare directly. As we can see, even the Foursquare doesn't have so many data for this topic. If we can use another provider, such as Google Map, or choose another topics, such as vegetarian restaurant, which are getting more popular, we might have a more accurate results.

6. Conclusion

This project tries to establish a possible method and process to evaluate places, in order to know if they are proper for hosting a new shop or stores. Although the result might be not very accurate, due to the amount of available data, this try is not in vain. This research can be improved if more data sources are provided, more characters are considered, we even can involve another Europe cities if data are not enough. This project will be seen an initial research in this field and I am glad to see that it is analysed more deeply as I suggested.