



THE BATTLE OF THE NEIGHBOURHOODS:

WHICH ONE IS MORE
MULTICULTURAL? SOUTHAMPTON VS
PORTSMOUTH

CAPSTONE PROJECT

Vanissa Wanick
vanissa@gmail.com

Table of Contents

1. Description of the problem and a discussion of the background	2
2. Description of the data and how it will be used to solve the problem	2
3. Methodology	3
Step 1 - gather data from each category and exploratory analysis	3
Step 2 - group by category using the mean	4
Step 3 - add to a dataframe.....	6
Step 4 - use k-means to cluster the categories	7
Step 5 - compare the clusters.....	7
4. Results.....	7
Results from data exploration.....	7
Results from k-means - Southampton	7
Results from k-means – Portsmouth	8
5. Discussion.....	9
6. Conclusion	9

1. Description of the problem and a discussion of the background

A multicultural city embraces diversity. For this project, I would like to compare the amount of different “traditional” cuisines in two cities: Portsmouth, UK and Southampton, UK.

Southampton has a population of 254,275 and Portsmouth, 238,137. Both cities are located in the same area of the UK and have similar population size. People tend to travel between those two cities in the weekends or during the week, so this comparison becomes natural. What I want to know is that if I want to have a multicultural experience, where should I go? For example, if I want to eat Lebanese food, which city will give me more options? At the moment, there is no other way of knowing this besides asking people what they think or just looking for the best restaurant using Google. I want to improve the discoverability of other restaurants that offer traditional cuisines around the area of Southampton and Portsmouth, UK. Thus, in summary, the goal of this report/research is to: improve the discoverability of other restaurants that offer traditional cuisines around the area of Southampton and Portsmouth, UK.

Questions: Which city is more multicultural? How many unique cultural restaurants are in Southampton and Portsmouth, UK? If I want to experience a specific cuisine, which city should I go? If a restaurant owner wants to open a business that will attract tourists and commuters, where should they open this restaurant and what are the other restaurants nearby?

By answering these questions, businesses in both areas could increase their discoverability rate and it might be possible that new opportunities could emerge from this search.

Finally, objectives are (at least): to list the current restaurants in each city; to group the restaurants by unique categories; to visualise the area with more variety of cuisines; to compare the results from each city.

2. Description of the data and how it will be used to solve the problem

Data that will be utilised will come from the Foursquare API. The reason for that is that it will be easier to use the queries to look for unique categories and cluster the neighbourhoods according to each category. In foursquare, there is a way to use the function GET request for result, and I will use the ‘categories’ section to group the data. I will use latitude and longitude information from each city as a starting point. The starting point is the train station from each city.

Now that we have in mind the goal of this project and some objectives, we can draw the next stages. These are: 1) analysis of each neighbourhood from each city using the mean of

each category and add it to a dataframe; 2) use k-means to cluster the neighbourhoods; 3) compare the clusters from each city.

3. Methodology

The methodology is composed by 5 parts. Below, each different step.

Step 1 - gather data from each category and exploratory analysis

The starting point is the train station from each city. Here we use the search query for restaurant, to look for restaurants. There are a few categories of restaurants according to foursquare API documentation¹. Limitations: 8000 for radius and 30 as limit of result (due to API constraints). The results were printed as a JSON file. Thus, for visualisation purposes it was necessary to transform it into a dataframe.

It was important to clean the data and only keep the following: venue name, and anything that is associated with location. Then, we could visualise the restaurants nearby. In the image below, we can see the red dot as the train station and the blue dots are all the restaurants nearby (see Figure 1 and Figure 2)

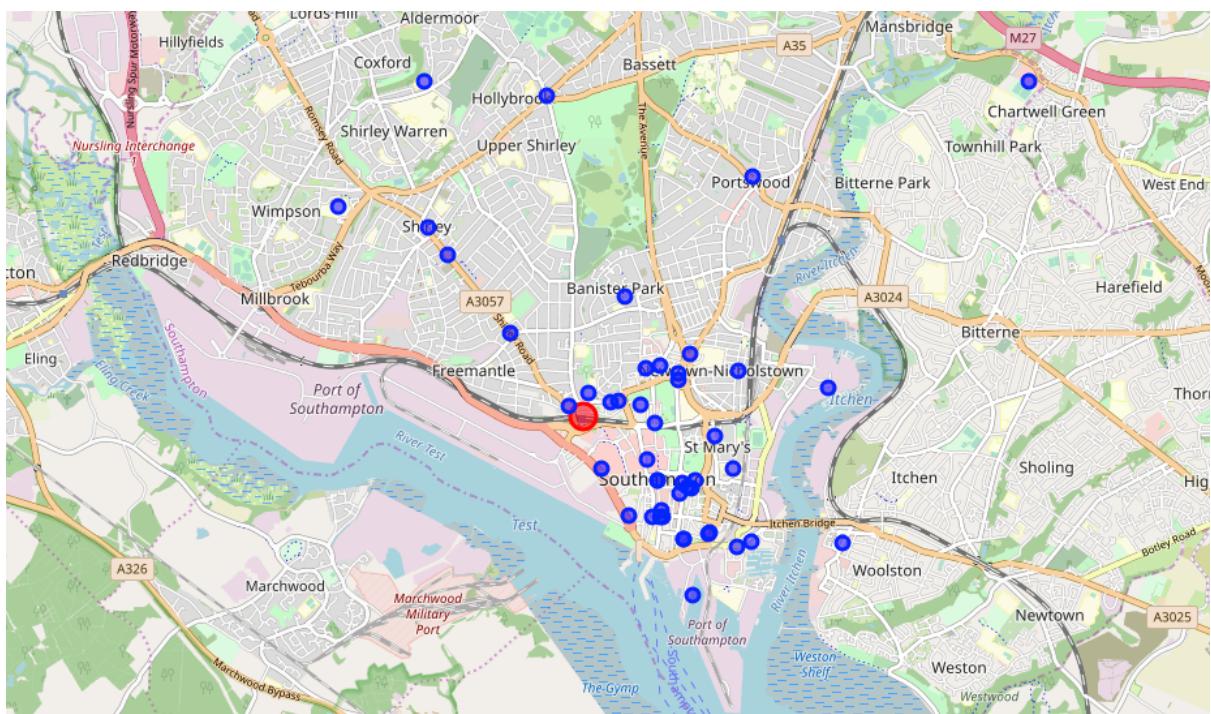


Figure 1 Restaurants near Southampton central

¹ <https://developer.foursquare.com/docs/resources/categories>

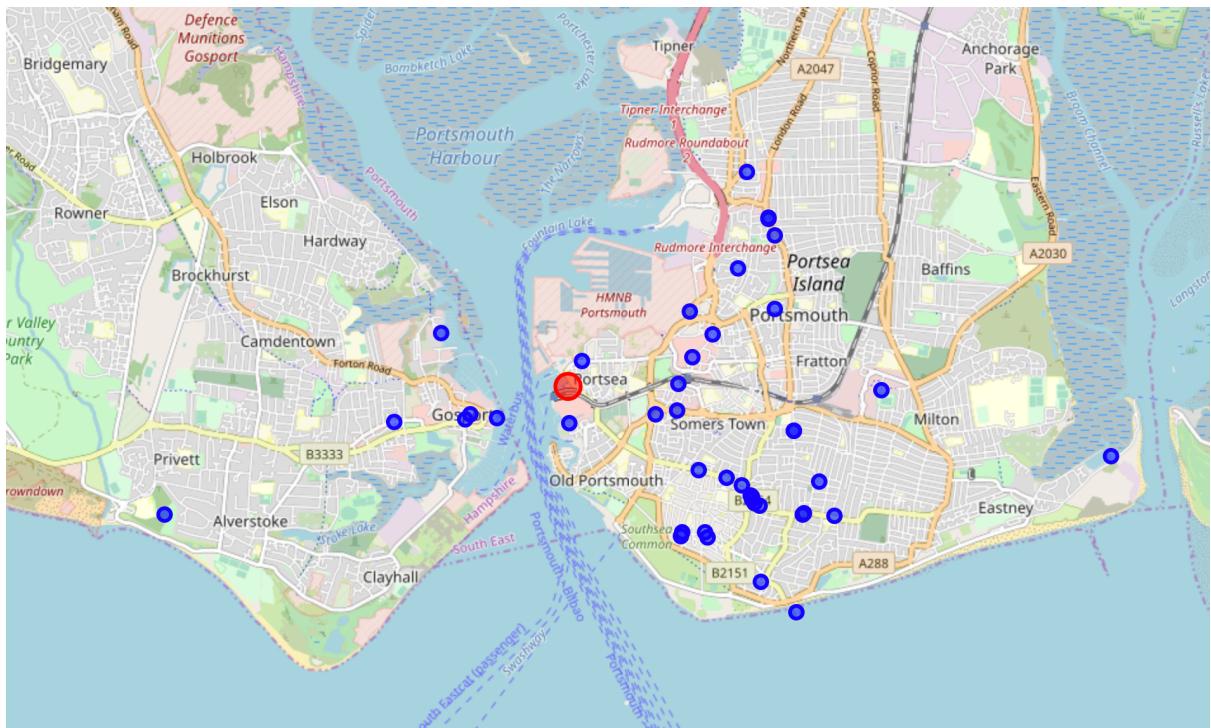


Figure 2 Restaurants near Portsmouth Harbour

Step 2 - group by category using the mean

By grouping the number of restaurants around the initial location, it was possible to cluster them through a visual plugin (see Figure 3 and Figure 4). With this, it is also possible to see the concentration of restaurants near the locations. For example, Southampton has more restaurants (11) near the station than Portsmouth. By doing a query and looking for unique categories, it is possible to know that there are 23 unique categories for Southampton and 22 unique categories for Portsmouth.

The battle of the neighbourhoods: which one is more multicultural? Southampton vs Portsmouth

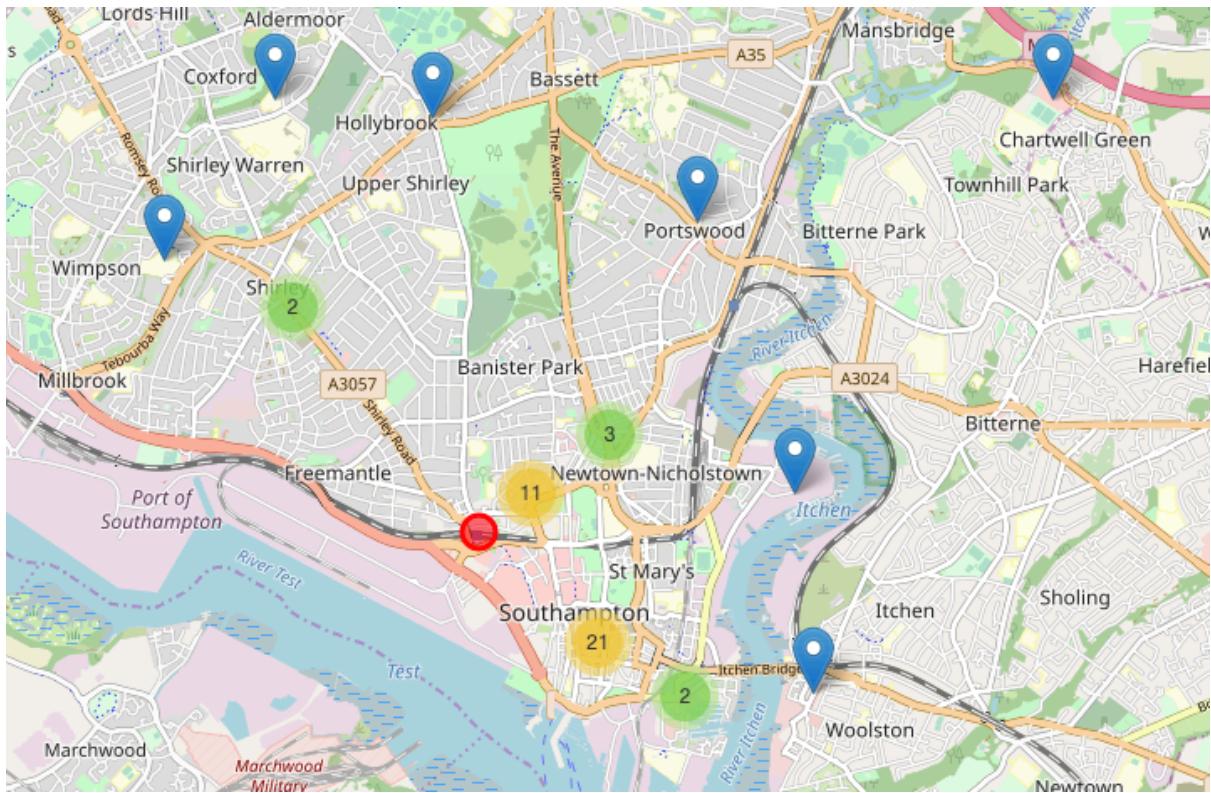


Figure 3 Grouped restaurants in Southampton per distance

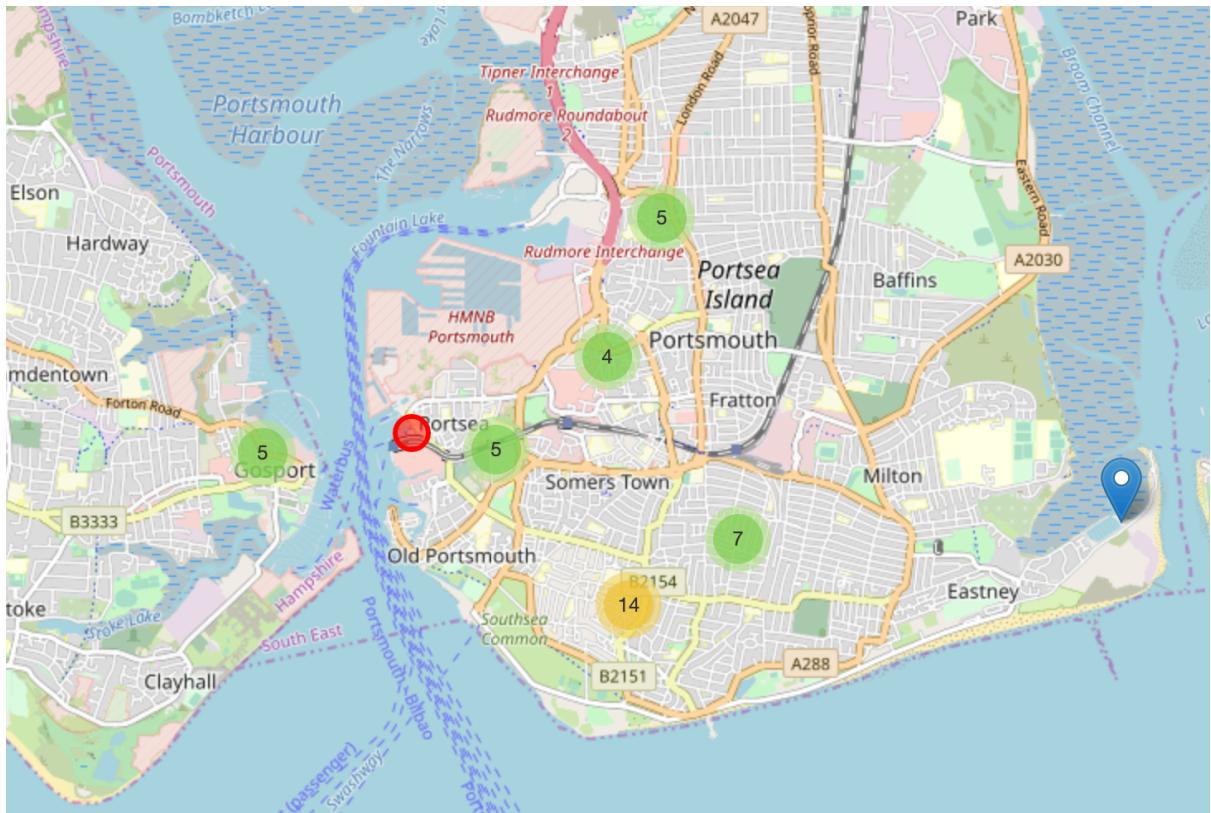


Figure 4 Grouped restaurants in Portsmouth Harbour

The battle of the neighbourhoods: which one is more multicultural? Southampton vs Portsmouth

	categories	lat	lng	distance
17	Thai Restaurant	50.908964	-1.409350	356.0
9	Indian Restaurant	50.909169	-1.409784	386.0
4	Chinese Restaurant	50.907835	-1.410205	418.5
16	Scandinavian Restaurant	50.903382	-1.411634	511.0
3	Cantonese Restaurant	50.911926	-1.403879	852.0
12	Lebanese Restaurant	50.902418	-1.404238	905.0
5	Diner	50.902354	-1.404234	910.0
15	Restaurant	50.907633	-1.408965	1021.6
1	Breakfast Spot	50.906563	-1.400800	1069.0
8	Greek Restaurant	50.900324	-1.403019	1141.0
14	Polish Restaurant	50.912890	-1.399887	1147.0
11	Japanese Restaurant	50.899392	-1.403511	1184.0
0	Afghan Restaurant	50.906039	-1.396637	1234.0
10	Italian Restaurant	50.899683	-1.401814	1251.0
13	Middle Eastern Restaurant	50.911502	-1.393604	1494.0
7	Food	50.897970	-1.397420	1592.0
2	Café	50.918636	-1.377496	3181.0
6	English Restaurant	50.873014	-1.400751	3974.0

Figure 5 Top categories of restaurants in Southampton ordered by distance

	categories	lat	lng	distance
18	Seafood Restaurant	50.794419	-1.106634	350.000000
4	Department Store	50.800040	-1.090082	1213.000000
22	Winery	50.802148	-1.124063	1312.000000
21	Wine Bar	50.804035	-1.090426	1363.000000
3	Chinese Restaurant	50.795314	-1.091644	1516.250000
7	Greek Restaurant	50.784693	-1.088026	1952.000000
15	Mediterranean Restaurant	50.787947	-1.082008	2051.000000
14	Malay Restaurant	50.804248	-1.078895	2104.000000
13	Lebanese Restaurant	50.787354	-1.080876	2153.000000
20	Turkish Restaurant	50.785926	-1.083122	2158.500000
11	Japanese Restaurant	50.799260	-1.083330	2198.000000
16	Middle Eastern Restaurant	50.790246	-1.075737	2365.000000
5	English Restaurant	50.784018	-1.080661	2400.000000
0	Asian Restaurant	50.810489	-1.078856	2440.000000
12	Korean Restaurant	50.789471	-1.072915	2553.000000
6	French Restaurant	50.780903	-1.080822	2607.000000
9	Indian Restaurant	50.789473	-1.071518	2706.750000
1	Breakfast Spot	50.797063	-1.147405	2883.333333
19	Soccer Stadium	50.797236	-1.064455	2985.000000
2	Café	50.811013	-1.080133	3119.333333
17	Restaurant	50.813015	-1.114875	3181.428571
10	Italian Restaurant	50.788722	-1.130276	4997.000000
8	Hotel	50.720495	-1.109804	8581.000000

Figure 6 Top categories of restaurants in Portsmouth ordered by distance

Step 3 - add to a dataframe

By adding to the dataframe then we can use to cluster them. There will be 2 dataframes, one for each city. These are ordered by the top categories. By doing this, we can check which venues are closer to the starting point and which categories they are related to (see Figure 5 and Figure 6). This was done separately (per city).

Step 4 - use k-means to cluster the categories

Then we cluster the categories of restaurants according to the distance. A number of 5 should be sufficient. By running iterations of the code, then it is possible to find the following tables in the results section.

Step 5 - compare the clusters

After that there will be a discussion around the comparison between both cities using the main aspects mentioned here.

4. Results

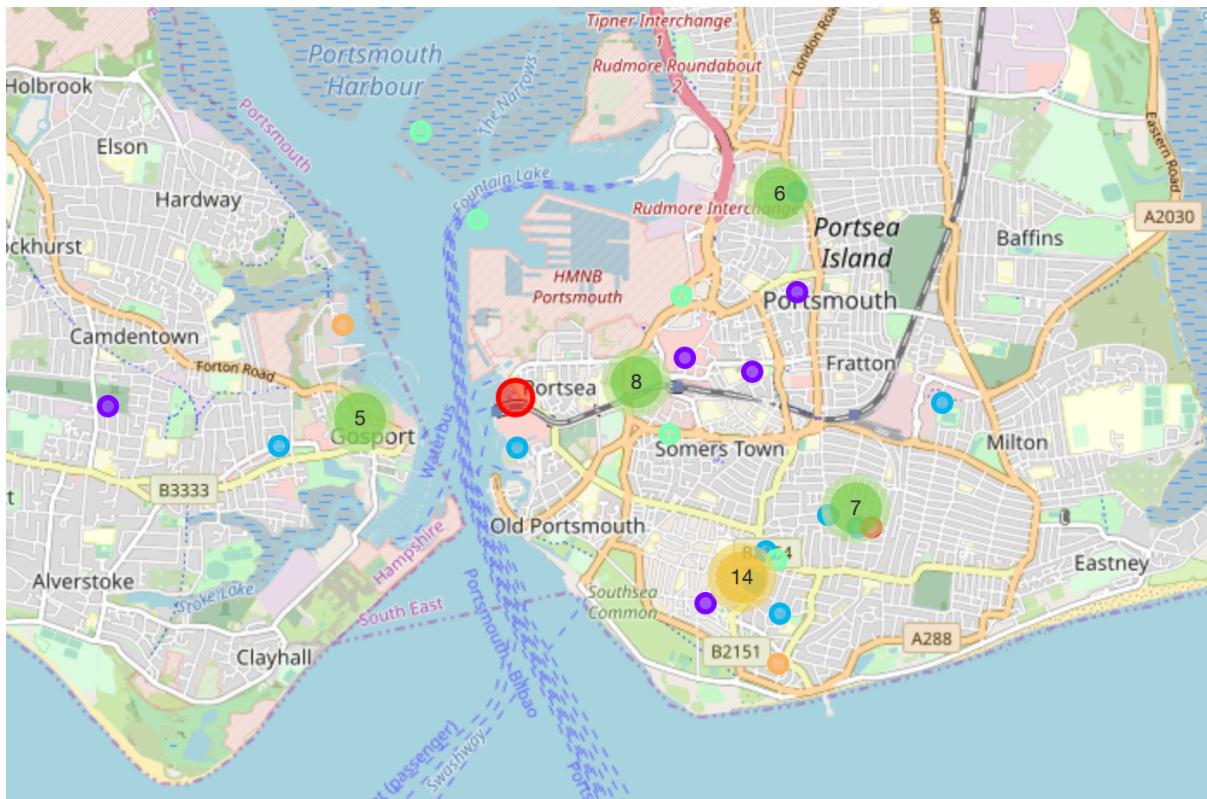
Results from data exploration

There is more variety of restaurants in Southampton than Portsmouth. Also, there are more restaurants near the train station than in Portsmouth.

Results from k-means - Southampton

The figures below show each different cluster from the k-means for the Southampton area. It is possible to see that certain categories tend to be surrounded by specific ones. For example, Greek restaurants are closer to English restaurants.

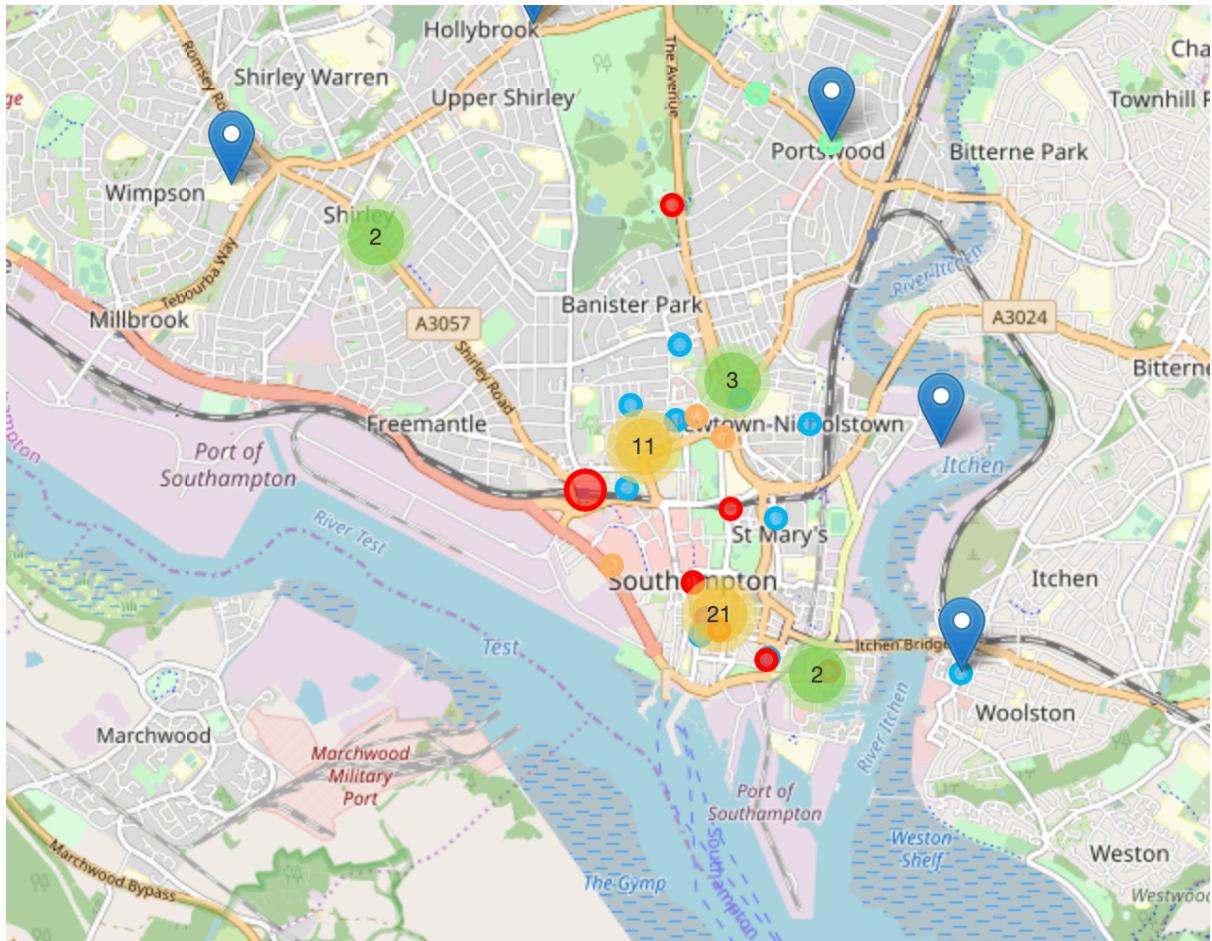
categories		categories		categories	
16	Lebanese Restaurant	5	Chinese Restaurant	21	Scandinavian Restaurant
7	Diner	17	Mediterranean Restaurant	4	Cantonese Restaurant
2	Breakfast Spot	19	Polish Restaurant	12	Hotel
14	Italian Restaurant	15	Japanese Restaurant	3	Café
9	Food	0	Afghan Restaurant	23	Thai Restaurant
22	Steakhouse	11	Gastropub	8	English Restaurant
11	Greek Restaurant	20	Restaurant	6	Coffee Shop
8	Bar	1	Cafe	23	Thai Restaurant



Results from k-means – Portsmouth

The figures below show each different cluster from the k-means for the Portsmouth area. It is possible to see that certain categories tend to be surrounded by specific ones as well. For example, Malay restaurants are closer to Chinese, Italian, Mediterranean and Korean restaurants. Interestingly, Lebanese and Japanese restaurants are not surrounded by any other category.

categories		categories		categories		categories	
4	Department Store					18	Seafood Restaurant
22	Winery					7	Greek Restaurant
15	Mediterranean Restaurant	21	Wine Bar			14	Malay Restaurant
20	Turkish Restaurant	3	Chinese Restaurant			16	Middle Eastern Restaurant
5	English Restaurant	19	Soccer Stadium			0	Asian Restaurant
1	Breakfast Spot	10	Italian Restaurant			12	Korean Restaurant
2	Café	8	Hotel	13	Lebanese Restaurant	6	French Restaurant
						9	Indian Restaurant
						17	Restaurant
						11	Japanese Restaurant



5. Discussion

Although Southampton has more different categories of restaurants, both cities have similar numbers (23). Therefore, the most important aspect might be the distance between the restaurants and the initial point (train station). In the future I will use another location to compare.

There are other things that might need improvement, such as cleaning the restaurant category. This category needs more exploration. But let's revisit the main question of this report. If I want to have a multicultural experience, where should I go? I can go to both places but if I look for a Chinese restaurant in Southampton, there is a chance I can find other cuisines. And if I look for a Greek restaurant in Portsmouth, there is a chance I can find other cuisines nearby.

6. Conclusion

So which city is better? It is not a yes/no answer. Because they both have similar unique values, there is a chance you might have a multicultural experience in both cities. Now it is a matter on which cuisine to choose. Due to the machine learning algorithm developed, it is

possible that if we go to a Chinese restaurant in Southampton, there is a chance I can find other cuisines. But if we look for French restaurants in Portsmouth, then there is little chance to see other restaurants nearby. Of course, this is a report developed for the capstone project of coursera and there are small flaws, but it is good to know that we can answer specific questions by data science.