Capstone Project – The Battle of Neighbourhoods

Ben Norland

Introduction/Business Problem

This project seeks to find the best location to open a new restaurant in Dublin, Ireland.

A friend of mine is a chef, and is keen to open a new restaurant in Dublin. She is quite flexible in her style of cooking, and is keen to understand what types of restaurant are most popular in the city to inform her plans. She would also like to find an area of the city in which to locate the restaurant. Ideally this should be somewhere with a number of other restaurants to provide footfall, but where her chosen style of restaurant has a chance to stand out from the crowd. She has asked me to look at Foursquare data to provide some recommendations.

Data

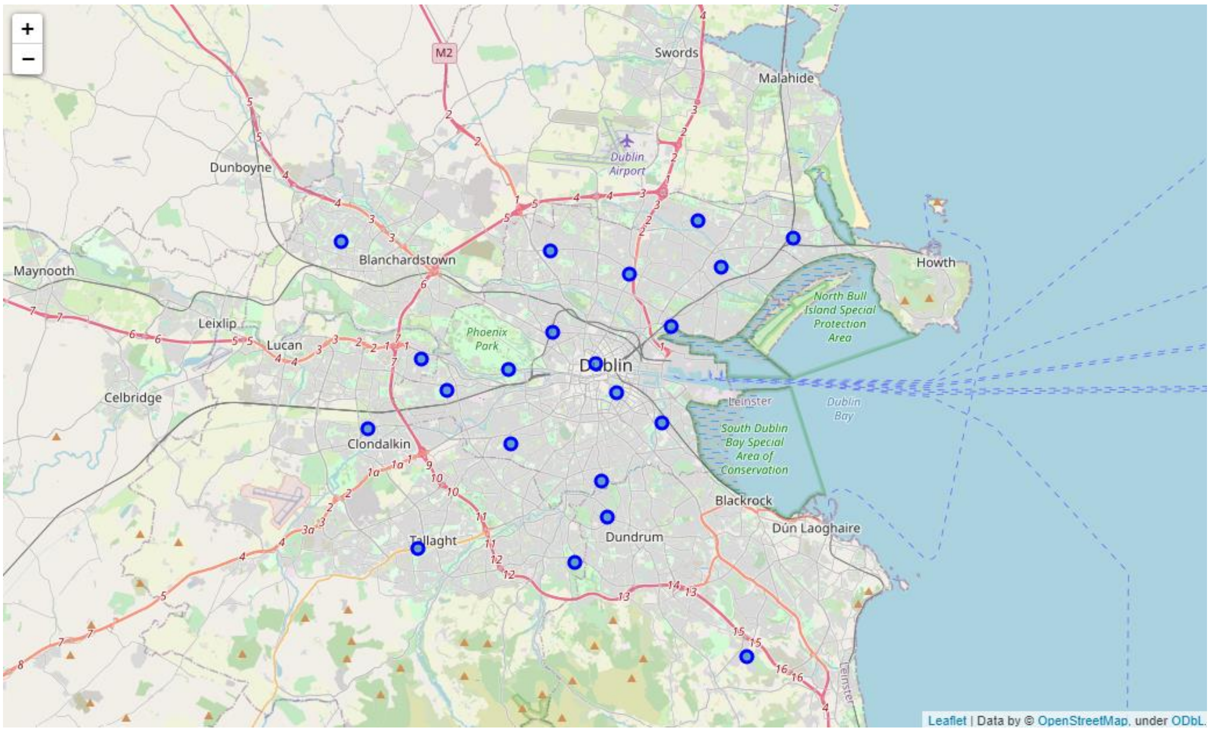
For this analysis, I will use a csv file containing information on Dublin postal districts to define the different areas of the city, and data from the Foursquare API to get information on restaurants and other amenities in each district. Using this data, I will identify:

- The most popular restaurant styles in the city as a whole, informing my friend's choice for her own restaurant
- The areas of the city with the highest density of restaurants
- Among these, districts that have a gap for my friend's chosen cuisine style

I will also cluster the districts based on their most common venue types, to provide alternative areas my friend might look at if it proves difficult to set up the restaurant in her preferred area.

Methodology

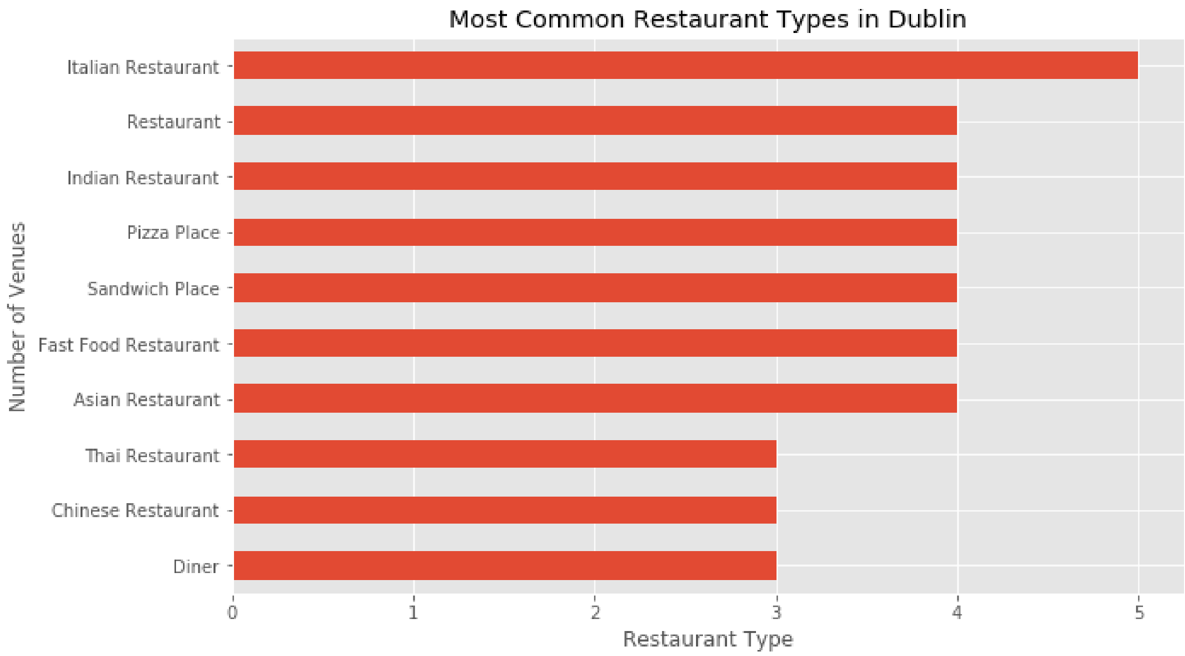
The first step in the process is to bring in location data for the different districts in Dublin. This is in the form of a csv file. We can use this to create an initial map of the city, to get a sense of the various districts, as shown below.



We then use the latitude and longitude of each district to query the Foursquare API and pull venue data. This contains the name, category, and location of all venues within a 500m radius of the district. By analysing the resulting category information, we're able to create a subset of data focusing on restaurants only, a sample of which is shown below.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	D1	53.35014	-6.266155	The Black Sheep	53.349708	-6.269066	Gastropub
1	D1	53.35014	-6.266155	BoCo	53.351353	-6.269949	Pizza Place
2	D1	53.35014	-6.266155	Aobaba	53.348801	-6.268864	Vietnamese Restaurant
3	D1	53.35014	-6.266155	Arisu	53.348922	-6.268964	Korean Restaurant
4	D1	53.35014	-6.266155	Kingfisher Restaurant	53.351663	-6.263265	Diner
5	D1	53.35014	-6.266155	Boojum	53.347371	-6.265438	Burrito Place
6	D1	53.35014	-6.266155	Musashi Noodles & Sushi Bar	53.346968	-6.268147	Sushi Restaurant
7	D1	53.35014	-6.266155	The Winding Stair	53.346596	-6.263784	Restaurant
8	D1	53.35014	-6.266155	Il Fornaio	53.346773	-6.263193	Italian Restaurant
9	D1	53.35014	-6.266155	El Patron	53.350662	-6.270850	Mexican Restaurant
10	D1	53.35014	-6.266155	Chapter One	53.354249	-6.263890	Restaurant

With our dataset limited to restaurants, we're able to address the first key question posed by our friend: which are the most common types of restaurants in the city. We find this by grouping the restaurant dataset by category, and produce the chart below, which shows that Italian restaurants are the most common, followed by Indian, Pizza, Fast Food, and Asian.



Given that no single cuisine has saturated the market, our friend indicates that she will set up an Italian restaurant, and we move on to consider which district might be the best location.

We can use the dataset to identify which districts already have Italian restaurants, and which have a high concentration of restaurants in general, giving us a shortlist of potential areas to select from. The table below indicates that D1, D4, and D24 could be good options – having a strong restaurant market, but not a high concentration of Italian restaurants. We rule out D2 on account of it already have 2 Italian restaurants out of 7.

	District	Total Restaurants	Italian Restaurants
0	D1	13	1
1	D4	13	1
2	D24	11	0
3	D2	7	2
4	D9	4	0
5	D7	2	0
6	D10	1	0
7	D11	1	0
8	D15	1	0
9	D16	1	0
10	D20	1	0
11	D22	1	0
12	D3	1	1

In order to make a more informed choice, we also decide to cluster the districts in the city based on their most common venues. To do this, we first group each venue category up into a "parent" type, namely *Restaurants & Cafes*, *Retail Stores*, *Fitness & Leisure*, and *Other Amenities*. We do this by running keyword searches on the category of each venue, and assigning them to one of the four 'parent groups' accordingly. An example of these allocations is shown below.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Type
0	D1	53.35014	-6.266155	The Black Sheep	53.349708	-6.269066	Gastropub	Restaurants & Cafes
1	D1	53.35014	-6.266155	Dealz	53.350623	-6.263183	Discount Store	Retail Stores
2	D1	53.35014	-6.266155	Blas Cafe	53.351336	-6.267326	Coffee Shop	Restaurants & Cafes
3	D1	53.35014	-6.266155	BoCo	53.351353	-6.269949	Pizza Place	Restaurants & Cafes
4	D1	53.35014	-6.266155	Cineworld	53.350244	-6.267696	Multiplex	Other Amenities

Next, we use one-hot encoding to identify the percentage of venues in each district falling into each of these groups, and use this to rank their most common venue types. This data is then used to build a k-means clustering model, assigning each district to a cluster, shown in the map below.

Finally, we analyse the clusters containing our shortlisted districts – D1, D4, and D24.

D1 and D24 fall into cluster 1. An examination of this group shows that venues in these districts tend to be *Restaurants & Cafes* and *Retail Stores*. D4 falls into cluster 4, and venues in these districts tend to be *Restaurants and Cafes* and *Other Amenities*, with *Retail Stores* less commonly featured.

Results

The analysis has helped us deliver on our key aims. We have identified the most common restaurant types in the city, and our friend has chosen to open a new Italian restaurant based on this information. We have then been able to identify districts that have a high concentration of restaurants, which should provide good footfall, while ruling out those with a high concentration of Italian restaurants already. Finally, through our k-means clustering algorithm, we have been able to group our districts based on their wider venue styles, adding further information which might help our friend to make a choice based on the type of venue that she wants to open. If she wants to capitalise on the lunchtime trade of shoopers and retail staff, she might look to open in an area like D1 or D24, which have higher concentrations of retail stores. If she wishes to open a more relaxed venue away from the hustle and bustle, D4 would be a good option. And if these specific districts are not an option, she can use the clustering output to identify other similar districts that might be more suitable.

Discussion

One point to note in this analysis is that Foursquare doesn't provide a huge amount of data for Dublin – returning 257 venues in total, and only 57 restaurants. It may be that Foursquare is not widely used in the city, and that our analysis might benefit from access to additional datasets, such as Tripadvisor or Google Reviews.

Conclusion

Our analysis concludes that our friend should open an Italian restaurant, and that the best districts in which to do so in Dublin would be D1, D4, or D24, depending on whether she wishes to take advantage of retail customers (where D1 or D24 would be best) or a more relaxed setting in D4.