Recommending a South Indian Restaurant in Leeds

COURSERA CAPSTONE PROJECT FOR IBM DATA SCIENCE

OMAR IMRAN – JULY 20

Intro

A relative wants to open a South Indian restaurant in Leeds, England.

Leeds is a diverse city with restaurants of many cuisines, but perhaps a gap in the market for South Indian?

Data

There is an existing **online wikipedia article** which lists different places in Leeds as well as their postcode (https://en.wikipedia.org/wiki/List_of_places_in_Leeds).

Foursquare API will be used to get the data of the venues at each of these places. Notable venue types of interest will be the different types of restaurants, particularly (South) Indian restaurants.

List of places in Leeds   [edit]

Click the symbol in the column heading to sort.
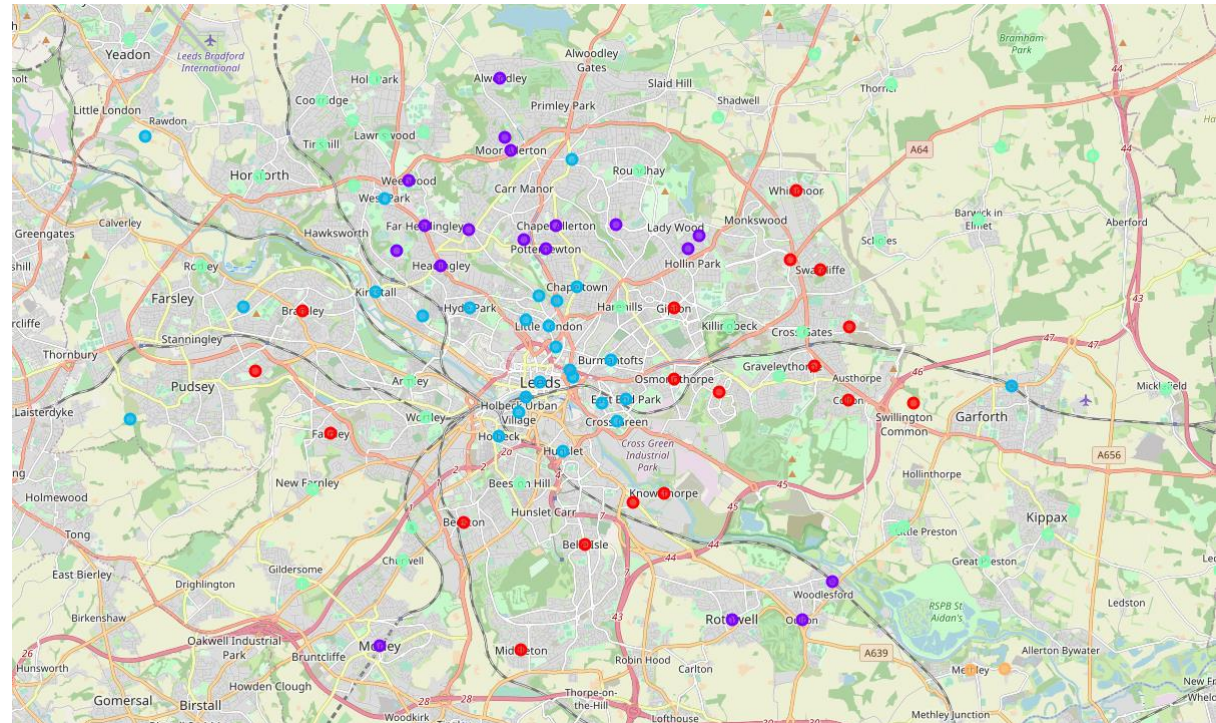
Placename	Leeds City Council Ward	Parliamentary Constituency	Management Area	Post-code	Post Town	Pre-1974 authority	Note
Aberford	Harewood	Elmet & Roth.	NE Outer	LS25	LEEDS	Tadcaster RD	
Adel	Adel & Wharfedale	Leeds NW	NW Outer	LS16	LEEDS	Leeds CB	
Adwalton	Morley North	Morley & Out.	S Outer	BD11	BRADFORD	Morley MB	
Ainsty	Wetherby	Elmet & Roth.	NE Outer	LS22	WETHERBY	Wetherby RD	
Aireborough					LEEDS	Aireborough UD	
Allerton Bywater	Kippax & Methley	Elmet & Roth.	E Outer	WF10	CASTLEFORD	Garforth UD	
Alwoodley	Alwoodley	Leeds NE	NE Outer	LS17	LEEDS	Leeds CB	
Arena Quarter	City & Hunslet	Leeds C	S Inner	LS1	LEEDS	Leeds CB	also LS2
Armley	Armley	Leeds W	W Inner	LS12	LEEDS	Leeds CB	
Arthington	Otley & Yeadon	Leeds NW	NW Outer	LS21	OTLEY	Wharfedale RD	
Austhorpe	Temple Newsam	Leeds E	E Outer	LS15	LEEDS	Tadcaster RD	
Bardsey cum Rigton	Harewood	Elmet & Roth.	NE Outer	LS17	LEEDS	Wetherby RD	

Methodology

1. Importing libraries

2. Scraping Leeds neighbourhoods from Wikipedia Geocoding the places in Leeds – the places are geocoded one by one and corresponding latitude and longitude are stored in the dataframe. This dataframe is saved for reference.

3. Superimpose places on map

4. Explore venues with Foursquare

5. **Clustering analysis**

6. **Clustering analysis repeated for custom created feature**

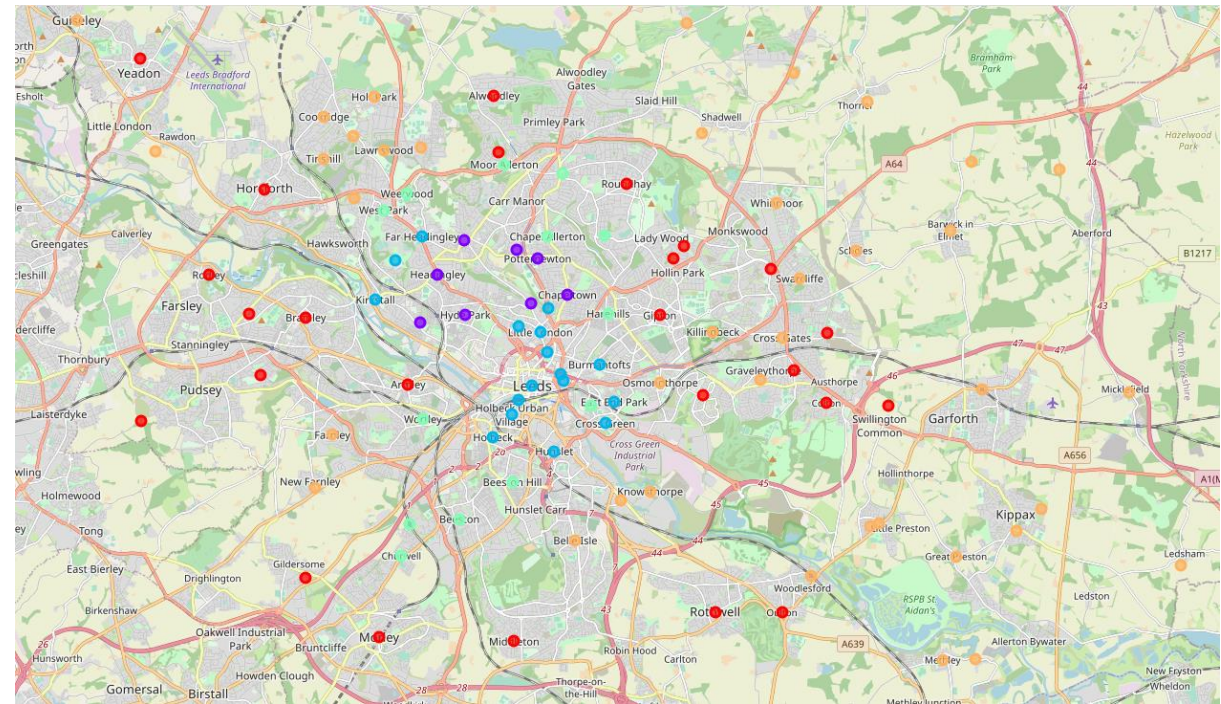7. **Clustering analysis repeated by normalising all features**

Clustering analysis #1

Cluster 0 seems fast food dominant, Cluster 1 seems Italian restaurant dominant, Cluster 2 seems Thai/Indian dominant, Cluster 3 seems to be Vietnamese dominant which in this case due to ordering convention means there is a lack of restaurants in this area, Cluster 4 have only one unspecified cuisine restaurant.
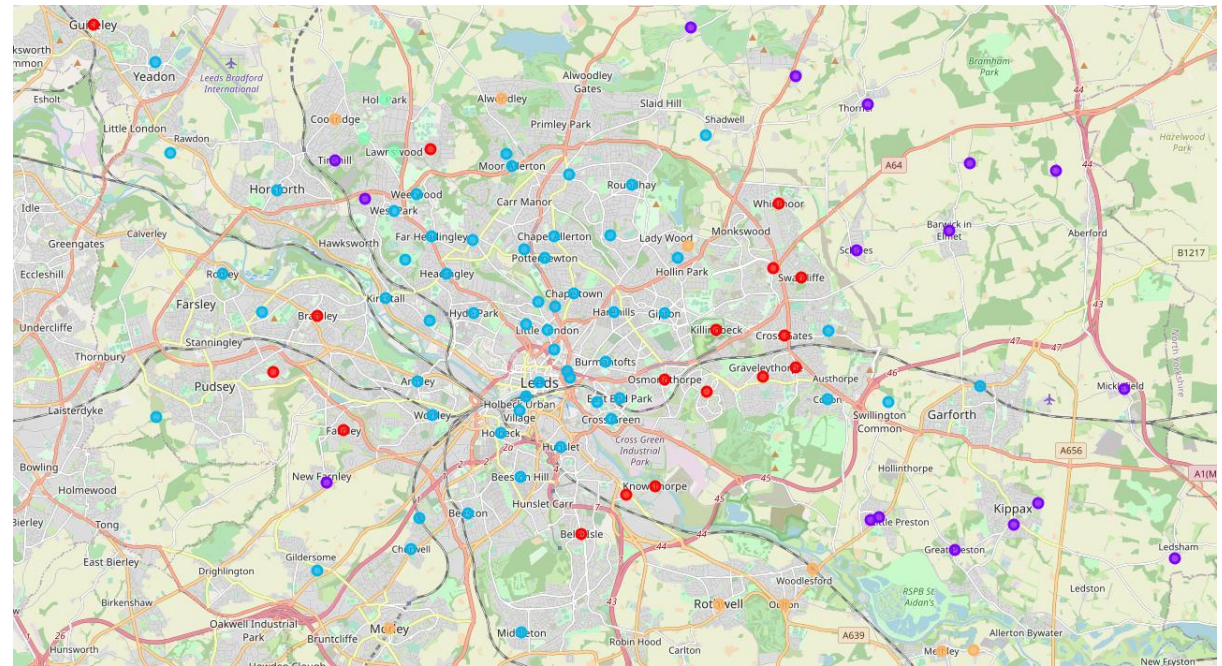
Clustering analysis #2

Cluster 0 has 3-7 total restaurants in each place of varying cuisine, cluster 1 has 18-21 restaurants in each place of mainly Thai/Italian cuisine, cluster 2 has 12-18 restaurants in each place of Thai/Indian cuisine, cluster 3 has 8-13 restaurants of varying cuisine and cluster 4 has mainly Jewish restaurants which by the numbering convention actually means it has 0-1 total restaurants per place.

Clustering analysis #3

Cluster 0 has 1-3 restaurants per place, primarily fast food. Cluster 1 has 0 restaurants per place. Cluster 2 has variety of types of cuisines and is hard to define as a cluster. Cluster 3 has only one Italian restaurant per place. Cluster 4 is dominated by unspecified 'restaurant' type.

Discussion

➢Firstly, there is not enough data on restaurants. Even when increasing the venue reach to 2km (quite a large allowance for a town centre)

➢From the first analysis - Cluster 2 seems of most interest for an Indian restaurant owner, and cluster 2 is based centrally on the map. Thus, there may be a lot of competition in central Leeds for an Indian restaurant and perhaps more of a market niche in other areas (particularly cluster 3 as there are no restaurants here. On the other hand, this could mean that there is only a taste for Indian food in central Leeds.

➢The further 2 analyses did not deliver any surprising insights. Thus, the new feature didn't seem to contribute much to useful analysis. It was rather expected that there would be more restaurant in the centre of Leeds. The new feature didn't seem to highlight how Indian restaurants varied in high population areas, which was the initial aim.

Conclusion

➢If more restaurants existed within a 2km radius of the centres of the different neighbourhoods in Leeds, then clustering could be more insightful. What is clear is that there is a lack of restaurants in many areas in outer Leeds or perhaps just a low number of restaurants registering their location close to the town centre.

➢So, there is indeed **a market gap in outer-Leeds town centres**. Also, the dominant restaurant cuisine in central Leeds seems to be Thai food, followed by Indian. Thus, there is **a taste for India in central Leeds which could be profited upon**.

➢In summary, I would advise to open a restaurant just outside the centre of Leeds, to capture the Leeds market interest in Indian food, but to stay further away from dense market competition.