

The Battle of Neighborhoods

Applied Data Science Capstone by IBM

Coursera

Deciding where to open a new venue in London (outside Inner City Limits)

Introduction

London is a great city. It is diverse, multicultural and full of opportunities. But there is also another side to its popularity and appeal. London is an expensive place to live and thrive. Many businesses want to open a venue here and many pay top price to have their window on Piccadilly or Oxford Street. But what about those, who want to start their own business and cannot really afford to open in the City yet? Where is it best to open a new place? Where will it be cheapest and will have enough people living around to be popular? Where the competition is not too overwhelming?

Approach

London, capital of United-Kingdom is a large city; actually much larger than Toronto but smaller than New-York. It shares with them similar characteristics: London is multicultural, modern, full of leisure opportunities and is an attractive business market place. Like in all big cities, it is very very expensive to run a venue in the most popular Boroughs; so venues owners need to have full understanding of business opportunities before launching/investing in their companies. Like the famous UK TV show "Location, location, location!", location is the most important criteria for any property and therefore business investment. So we will use data, maps and data science tools to analysis where is(are) the best place (s) to have a venue in London.

Data Mining

Data Sources

- Wikipedia London page – using web scrapping tool
- Foursquare API – we have personal (free) credential access
- Online data source about London rent
- Folium geo mapping

Data Cleaning

We retrieved data on London Boroughs using web page scrapping on wikipedia London page. After cleaning data and dropping not useful columns, we let with Name, Area, Population, geo coordinates and rent prices.

Because Foursquare can provide venues information around chosen coordinate, we can merge both source and obtain data about . We are able to list venue by types and because Foursquare provides information about venues interest, we can rank them.

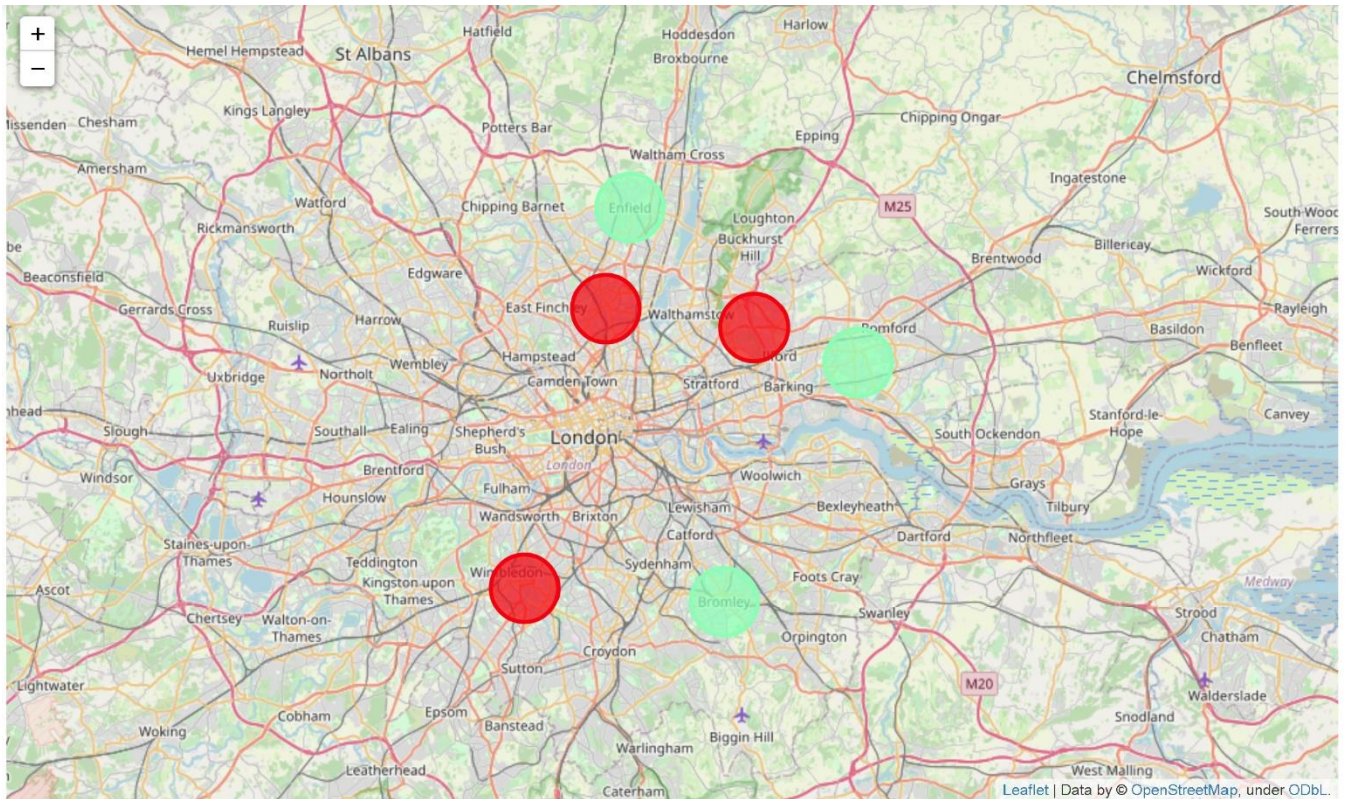
To keep the results easy to understand, we ranked the top 5 venues for 8 Boroughs.

	Borough	Area	Population	Latitude	Max_Rent	Longitude	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barking and Dagenham	13.93	194352	51.554117	102.25	0.150504	2	Supermarket	Park	Grocery Store	Coffee Shop	Pub
1	Bexley	23.38	236687	39.969238	97.00	-82.936864	2	Pizza Place	Coffee Shop	Ice Cream Shop	Discount Store	Chinese Restaurant
2	Bromley	57.97	317899	51.402805	118.50	0.014814	2	Pub	Grocery Store	Coffee Shop	Park	Pizza Place
3	Enfield	31.74	320524	51.652085	102.25	-0.081018	2	Coffee Shop	Turkish Restaurant	Park	Pub	Supermarket
4	Haringey	11.42	263386	51.587930	107.75	-0.105410	0	Café	Pub	Turkish Restaurant	Coffee Shop	Park
5	Havering	43.35	242080	51.004361	86.00	-2.337475	1	Airfield	IT Services	Electronics Store	Food & Drink Shop	Women's Store
6	Merton	14.52	203223	51.410870	123.75	-0.188097	0	Pub	Park	Café	Coffee Shop	Bar
7	Redbridge	21.78	288272	51.576320	118.50	0.045410	0	Pub	Park	Coffee Shop	Restaurant	Italian Restaurant

Data Cleaning

Clustering

We compute data to obtain clusters on place them on the London map. Clusters group boroughs by most similar characteristics. This helps us to reduce noise of information (reducing the amount of data) and give us tools to target our analysis



Results

The most popular social venues, outside of Inner London boroughs are Pubs and Coffee shops Northern boroughs are more prone to visiting pubs, whereas southern boroughs are most likely coffee shop and restaurants. Within top 5 places of interest in every borough is an ethnic restaurant Rent price is not so much a factor for going out - the demand is not affected by difference in costs

Discussion:

Looking at the data, Havering, Bromley and Enfield are the best places outside of Central London where a new venue is worth opening. However, a lot of information is not taken into account, and cannot be obtained from Foursquare Developer:

Higher ethnic presence in a given borough can and will influence the popularity of a given cuisine. Closer proximity to Inner boroughs and better transport links allows people to travel to the neighbouring borough and impact the measurements. Many small venues are not registered in Foursquare and are marketed via word-of-mouth, and are not taken into account. Regardless, the analysis provided an insight into what people like and opt for, when it comes to going out in their own neighborhoods.

Conclusion:

This has been a long but very interesting training. We used real life data to answer business questions and it is incredibly powerful. If I was running this project now, I will certainly propose to my project owner/client to use another API than Foursquare, because Foursquare data has not been actualized much the last few years. We learn that quality of data is a pre-requirement to good data science analysis.