

Report – Capstone Project

Deciding where to open a new venue in London

Introduction

1) Background -

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. With such a competitive market and high revenues it is challenging for newcomers to find a great place to make their first steps and competition is fiercer than ever. In this situation, there is no such thing as too much data to help make that choice.

2) Problem Statement -

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?

If we consider all of these questions, it might be a good idea to turn to data on the outer London boroughs, to look into the numbers with a bit more scrutiny.

3) Interest -

The analysis will be of interest to the following groups:

- a. First time entrepreneurs, who want to start their first business. Below dataset will give a comprehensive insight into where best to open a new venue, to maximize the value for money.
- b. People who already run a business and want to branch out. Given the extra information, it may provide valuable information before decision making.

Data acquisition and cleaning

1) Data sources -

As data sources the below were chosen:

1. Wikipedia list of London Boroughs with coordinates.
2. Foresquare data on the most popular venues in the respective boroughs
3. Online based data on rent in London boroughs

2) Data cleaning -

Once the dataset of London boroughs has been downloaded, we must edit the dataset provided to only have information, necessary for our problem. Wikipedia provided information on political situation, headquarters of the borough council etc. that will not be required. After cleaning, we will only be left with Name, Area, Population, Coordinates and rent prices for each borough. Since the dataset will only include outer boroughs, all the inner ones will be omitted as well. Foursquare provides a dataset of venues around the specific coordinates or venues, if we use the “Explore” function in the Developer tab. Once requested, we get a full breakdown of all recorder venues around the boroughs of interest.

3) Feature selection -

For convenience, we transform the dataset only to show top 5 places to work with. After that, we merge the data frames together for a comprehensive set of values, worth analyzing.

Out[27]:

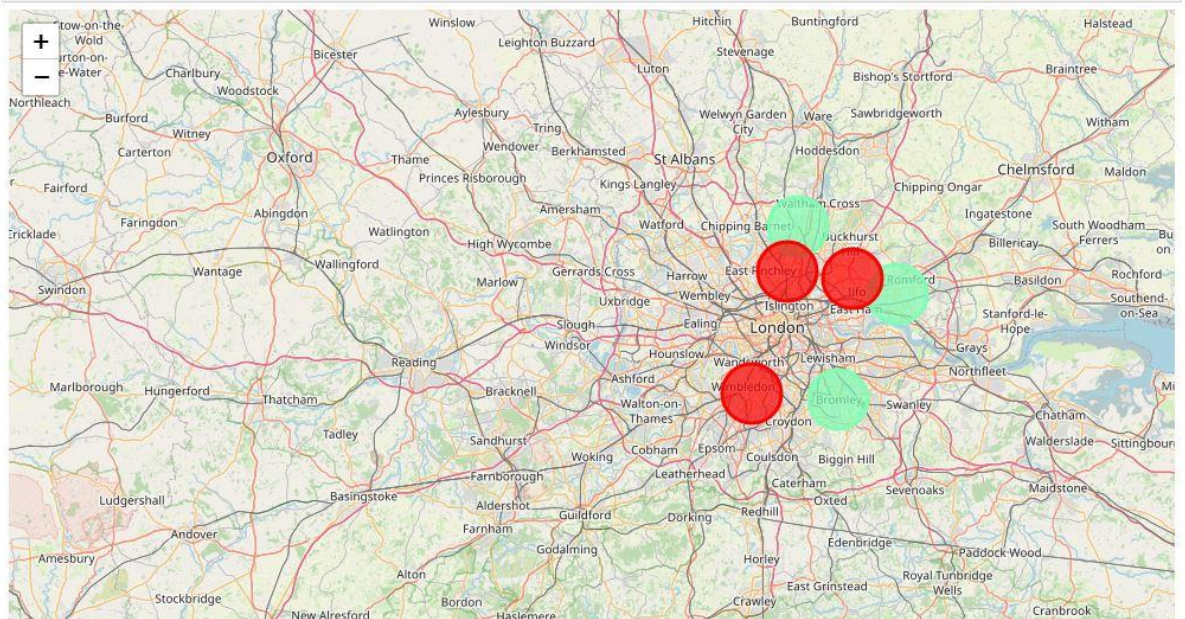
	Borough	Area	Population	Latitude	Max_Rent	Longitude	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barking and Dagenham	13.93	194352	51.554117	102.25	0.150504	2	Supermarket	Park	Grocery Store	Coffee Shop	Pub
1	Bexley	23.38	236687	51.402805	97.00	-82.936864	2	Pizza Place	Coffee Shop	Ice Cream Shop	Discount Store	Sandwich Place
2	Bromley	57.97	317899	51.402805	118.50	0.014814	2	Coffee Shop	Pub	Grocery Store	Park	Gym / Fitness Center
3	Enfield	31.74	320524	51.652085	102.25	-0.081018	2	Coffee Shop	Pub	Turkish Restaurant	Café	Park
4	Haringey	11.42	263386	51.587930	107.75	-0.105410	0	Café	Pub	Turkish Restaurant	Park	Coffee Shop
5	Havering	43.35	242080	51.004361	86.00	-2.337475	1	Airfield	Hotel	Electronics Store	Food & Drink Shop	Women's Store
6	Merton	14.52	203223	51.410803	123.75	-0.188099	0	Pub	Park	Coffee Shop	Café	Bar
7	Redbridge	21.78	288272	51.576320	118.50	0.045410	0	Pub	Park	Café	Coffee Shop	Restaurant

Data Analysis

1) Clustering -

Once the boroughs are selected, we will place them onto the map and cluster the boroughs to analyze the similar ones against each other and within the clusters themselves. Each of the clusters will be compared against the popularity of the venues, worth considering for a new business venture, as well as the hospitability of the venture climate. Depending on the area, population and rent, each of the clusters offer different advantages and disadvantages in terms of venue choices.

Out[28]:



Results

Reviewing each cluster as per method above, we learn:

1. First cluster is heavy favoring pubs – predominantly as centers of socializing of the local communities. First cluster boroughs have majority of people living further from the central London, keeping to the old ways of pub socializing.
2. Second cluster, and the most controversial. Enfield is one of the bigger outer London boroughs, and the data that we get can differ massively within the same borough: those living closer to M25 in the suburban houses tend to stay indoors and spend more time with their families and cook at home. On the other hand, those living closer to the inner borough limits may be more outgoing, and because of this contribute the most to our dataset. Regardless, with not too high rent, big and a very diverse population, Enfield is worth considering
3. Biggest and the most diverse cluster. Most of the boroughs are not heavy drinkers, but neither are they very outgoing. Because I would not call visit to the grocery store as a social route! However, Haringey and Bromley are good fits to our target boroughs: both have entertainment venues as most common, both rather big and quite populated. In this situation perhaps Haringey has a slight financial edge - cheaper rent a more incline to pubs and coffee shops in the area.
4. Within top 5 places of interest in every borough is an ethnic restaurant. Because of the different ethnicities in the boroughs, some choices will be more favorable amongst the specific group in the area.
5. Rent price is not so much a factor for going out - the demand is not affected by difference in costs. There is a spike in rent price going into London, but further away the cost is not too much of an issue.

Findings

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 considered, and cannot be obtained from Foursquare Developer:

1. Higher ethnic presence in each borough can and will influence the popularity of a given cuisine.
2. Closer proximity to Inner boroughs and better transport links allows people to travel to the neighboring borough and impact the measurements
3. Many small venues are not registered in Foursquare and are marketed via word-of-mouth, and are not considered

Regardless, the analysis provided an insight into what people like and opt for, when it comes to going out in their own neighborhoods.

Conclusion

Finally to conclude this project, I have had a good trial run at solving a real-life problem, using available data to find a business solution - choosing to open a venue in London. I have made use of some frequently used python libraries to manipulate data, use Foursquare API to explore the information on the Boroughs I looked into and managed to make a map of results that allowed me to illustrate my point graphically and quite clearly to anyone not familiar with data manipulation. And answer the questions that needs deep evaluation of data.