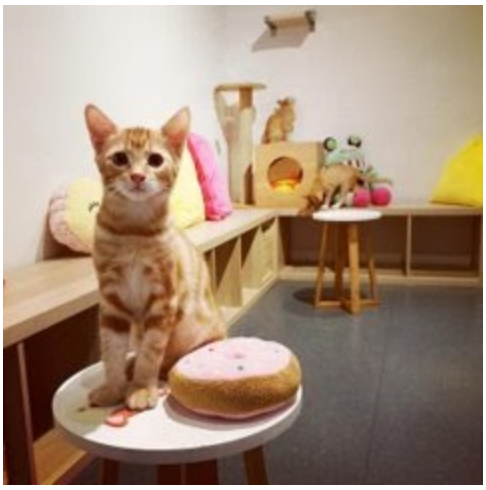
Report

Pet Cafe in London



Submission for Week 5 Capstone Project

Battle of the Neighbourhoods

Business Problem

- The business problem is to identify the **best location to situate a new pet cafe in London, in order to balance accessibility, business prospects and costs**.
- This is a business opportunity because pet cafes are a relatively untapped market in London, with only 7 existing outlets compared to about 10,000 conventional coffee shops.
- In addition, in the post-lockdown London, avenues for greater social interaction and relaxation would be increasingly popular after Londoners have endured months of isolation during COVID.

Data

In order to assess the suitability of a location, we will consider the following factors and obtain the necessary data:

1. **Popularity**: Assume that high density of F&B/retail outlets suggests a popular shopping and entertainment district where footfall would be high. We will use Foursqaure data to assess this.
2. **Competition**: Avoid situating near other pet cafes. Also consider business prospects based on existing businesses in the area. We will use Foursquare data to look at location of existing competitors, and business data from the Greater London Authority (data.london.gov.uk).
3. **Accessibility to residents**: Persistence of COVID-19 pandemic means business from locals outweighs that of tourists. We will use population density data from GLA to assess this.

4. **Affordability**: Assume rental prices increase closer downtown. We will use Wikipedia data to assess proximity to downtown.

Methodology

In Step 1: Data acquisition and preparation, we will conduct web-scraping from Foursquare and Wikipedia, as well as importing relevant data from the GLA database.

In Step 2: Exploratory data analysis and inferential analysis, we will establish locations of competitors (existing pet cafes), conduct Pearson correlation analysis to establish features that affect business outcomes the most, as well as study and cross-verify data on proximity to city center by looking at GLA and Wikipedia data.

In Step 3: K-Means clustering, we will use clustering to study borough profiles and establish which are popular F&B, retail and entertainment areas.

Thereafter, we will provide conclusions and recommendations.

Step 1: Data acquisition and preparation

In order to generate a dataframe with the London Borough names and coordinates of the Boroughs, we will use BeautifulSoup to scrape a Wikipedia page on the Boroughs, and Geocoder to generate the accompanying coordinates.

The Wikipedia page also contains a numerical listing of the Boroughs by their proximity to the city center of London, which we will also scrape from the page as "Nr. in map".

This data will be required for later usage when using geolocation to study the Boroughs. The dataframe containing this data is below:

Borough	Nr. in map	Latitude	Longitude
Barking and Dagenham	25	51.627300	-0.253760
Barnet	31	51.452078	0.069931
Bexley	23	51.609783	-0.194672
Brent	12	51.601511	-0.066365
Bromley	20	51.591180	-0.165040
Camden	11	51.593470	-0.083380
Croydon	19	51.508383	-0.305200
Ealing	13	51.540024	-0.077502
Enfield	30	51.477890	-0.013340
Greenwich	22	51.531820	-0.061780
Hackney	9	51.482600	-0.212880
Hammersmith and Fulham	4	51.589270	-0.106405
Haringey	29	51.513180	-0.106980
Harrow	32	51.544610	-0.144260
Havering	24	51.484230	-0.096477
Hillingdon	33	51.471393	-0.351374
Hounslow	14	51.534380	-0.108940
Islington	10	51.522660	-0.207930
Kensington and Chelsea	3	51.410881	-0.291933
Kingston upon Thames	16	51.494471	-0.120066
Lambeth	6	51.465280	-0.013210
Lewisham	21	51.544520	-0.166860
Merton	17	51.519937	0.055882
Newham	27	51.475773	-0.080698
Redbridge	26	51.480270	-0.237540
Richmond upon Thames	15	51.505734	-0.100002
Southwark	7	51.512243	-0.053659
Sutton	18	51.499990	-0.010450
Tower Hamlets	8	51.581765	-0.276968

Step 2: Exploratory data analysis

After importing and cleaning business and population data as well as data on the profiles of each Borough, we applied Pearson correlation analysis to find out what features affected business survival in each Borough the most.

```
In [21]: #Correlation with output variable
         cor_target = abs(cor["Two-year_business_survival_rates_(started_in_2013)"])
         #Selecting highly correlated features
         relevant_features = cor_target[cor_target>0.5]
         relevant_features

Out[21]: Population_density_(per_hectare)_2017              0.522354
         Proportion_of_population_of_working-age,_2015     0.536294
         Two-year_business_survival_rates_(started_in_2013)    1.000000
         Average_Public_Transport_Accessibility_score,_2014    0.501066
         Name: Two-year_business_survival_rates_(started_in_2013), dtype: float64
```

Based on the above Pearson Correlation Analysis, the features most correlated with business survival, are:

1. Population density
2. Proportion of population of working age
3. Public transport accessibility

After comparing these features with the proximity to the city center, we see that business survival rates are higher in boroughs that are further from London city center. This may be due to the lower rental rates and competition from other businesses in boroughs further from the city center.

In addition, after considering the three most important features, we see that **Hackney** scores the best as the Borough with good business outcomes, good in population density, good proportion of population of working age, and good public transport accessibility.

	Borough	Nr. in map	year_business_survival_rates_(started_in_2013)	Population_density_(per_hectare)_2017	Proportion_of_population_of_working-age_2015
25	Richmond upon Thames	15	78.8	34.4	64.5
4	Bromley	20	78.6	21.8	62.6
22	Merton	17	78.4	55.3	67.2
10	Hackney	9	76.8	144.0	72.1
19	Kingston upon Thames	16	76.8	47.1	67.2
13	Harrow	32	76.5	50.0	64.5
16	Hounslow	14	76.2	48.0	67.8
27	Sutton	18	76.0	46.2	64.3
7	Ealing	13	75.8	63.3	66.8
30	Wandsworth	5	75.8	93.7	72.8
14	Havering	24	75.3	22.6	62.3
6	Croydon	19	75.3	44.7	64.9
15	Hillingdon	33	75.0	26.0	65.6
24	Redbridge	26	74.7	53.9	65.0
18	Kensington and Chelsea	3	74.5	131.1	69.3

Our Foursquare API call collected 2843 venues across 271 categories, which we grouped by Borough location and popularity.

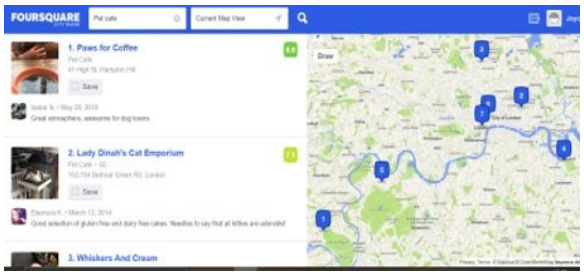
Top 5 Boroughs where Londoners go for their coffee fix are:

1. Bromley
2. Islington
3. Hackney
4. Bexley
5. Richmond upon Thames

This result from the Foursquare call can be seen in the below table. We could consider locating the pet cafe in the vicinity of these Boroughs, in order to leverage the popularity of these Boroughs as cafe areas:

In addition, Foursquare also provides the information that the existing competitor pet cafes are located in the following Boroughs, as shown in the screengrab further below:

1. Richmond upon Thames
2. Tower Hamlets
3. Islington
4. Greenwich
5. Camden
6. Bromley

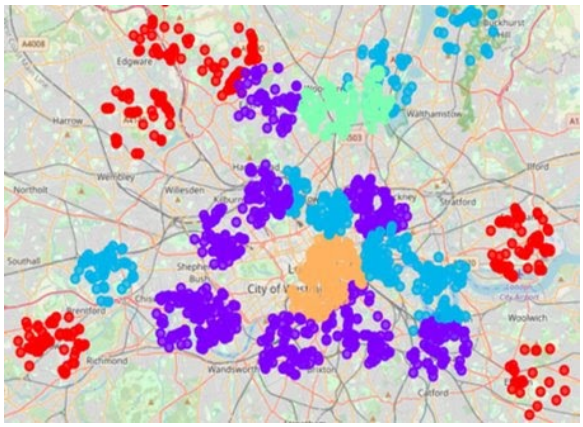


We should consider avoiding these Boroughs for our pet cafe, so that we do not have to compete with them.

Step 3: K-Means clustering

Finally, we used K-Means clustering to study borough profiles and establish which are popular F&B, retail and entertainment areas.

The resultant clusters, as visualised on the Folium map of London, is shown below:



As expected, the extreme clusters (cluster 0 and cluster 4, shown in red and orange) are located in central and outer London respectively. These locations may be too remote or too competitive for our pet cafe. In order to strike a good balance between popularity of Boroughs and rental/business costs, we should choose boroughs where clusters 1, 2, or 3 are (shown in blue, purple and green), and avoid the extreme clusters.

Conclusions

In this concluding section, we will bring together the results from the various above analyses.

From steps 1 and 2 and our analysis of Wikipedia and GLA data, we have obtained the following insights:

1. Most important for business survival are: (a) population density, (b) proportion of population of working age, and (c) public transport accessibility.
2. Among the boroughs, the boroughs that are top 5 for business survival are: Richmond upon Thames, Bromley, Merton, Hackney, and Kingston upon Thames. This implies that it is not critical to be close to city center for achieving good business outcomes, as there are benefits from locating further, such as lower rental rates and competition from other businesses.
3. Among these top 5, when we cross-reference with Wikipedia data on proximity, only Hackney is located in Inner London and has a moderate proximity of number 9 out of 32. The other 4 are in Outer London and substantially further from city center.
4. Furthermore, Hackney scores better than the other 4 boroughs when we consider GLA data for (a) population density, (b) proportion of population of working age, and (c) public transport accessibility.

From geolocation data and our analysis of Foursquare data, we have obtained the following insights:

1. Boroughs that cafes/coffee shops are most popular: Bexley, Bromley, Hackney, Islington, Richmond upon Thames.
2. Among these top 5, when we cross-reference with Wikipedia data on proximity, only Hackney and Islington are located in Inner London and has a moderate proximity of number 9 and 10 respectively out of 32. The other 3 are in Outer London and substantially further from city center.
3. Boroughs were existing competitor pet cafes are located: Richmond upon Thames, Tower Hamlets, Islington, Greenwich, Camden, Bromley. We should avoid these boroughs as such

competition could affect our business survival.

From step 3 and our analysis of results from K-Means clustering, we have obtained the following insights:

1. As expected, the extreme clusters (cluster 0 and cluster 4) are located in central and outer London respectively. In order to strike a good balance between popularity of borough and rental/business costs, we should choose boroughs where clusters 1, 2, or 3 are, and avoid the extreme clusters.

Taking all insights together, we can conclude that Hackney is an ideal borough for opening our pet cafe, for the following reasons:

1. Hackney scores well in business survival, affordability and accessibility to residents.
2. Hackney does not have an existing pet cafe as a competitor.
3. Hackney is moderately close to the city center and also still considered Inner London, thus helping to strike a balance between accessibility/popularity and business costs.
4. Hackney is not located in the extreme clusters where popularity is most or least intense. This would also help strike a balance between accessibility and competition.

Evaluation

This analysis is based on publicly-available population, geolocation and business data, some of which were updated two years ago. This analysis can be improved by using newer data once available. This is especially pertinent as the impact of COVID over the past year may have brought about significant changes in trends.

In addition, there was difficulty obtaining suitable data on rental costs of retail spaces matched according to boroughs. Availability of this datapoint could also improve the analysis.