Finding the best location to open a coffee shop

Marios Djakouris
02/08/2021

Introduction

The London Borough of Bromley is the southeasternmost borough of Greater London with an estimated population of around 332000. This borough can be a great opportunity for any coffee shop owner who seeks to capitalise on London's rich market. Nevertheless, choosing a suitable location may be challenging as Bromley is the largest borough of Greater London by total area. Bromley is made up of 8 Postcode districts, with each one having a Post town which can be considered the centre of activity and commerce for each district. Coffee shops are retailers who need a suitable location with great accessibility, lots of foot traffic and low competition, meaning that neighbouring businesses can significantly impact profitably in a positive or negative way. My main objective is to create maps and charts which will help those who seek to open a coffee shop, bringing more clarity on how location can generate more foot traffic, while also exploring the impact which neighbouring businesses can have on profitability. The ones who will mainly be interested in the best location of coffee shops are the coffee shop owners who wish to enter or expand in the Borough of Bromley market. Furthermore, those who sell similar products or wish to do so may also find this useful. Finally, anyone who wish to live or be located near a location with lots of foot traffic and good accessibility to basic services, including consumers, will probably find interest in this project.

Data

The data sources to create our models are the following:

- I used Wikipedia to get information about each postcode district [1].

- I created and imported a csv file which includes the population of each post code and the coordinates of post towns. Google Map was used to get the Latitude and Longitude for each post town [2] and postcodearea.co.uk provided the population of each postcode district [3].

- Forsquare API I got information about all the venues within the proximity of each Post town [4]

Our database includes the Post Towns, which are the main towns of each Postcode district. I grouped my data by Post town due to being the main focus of our analysis, while joining in a single line post codes with mutual post towns. Also, the csv file I created and imported enabled me to merge on my main dataset each Post town's coordinates and the total population of all Post Code Districts each Post town represents.

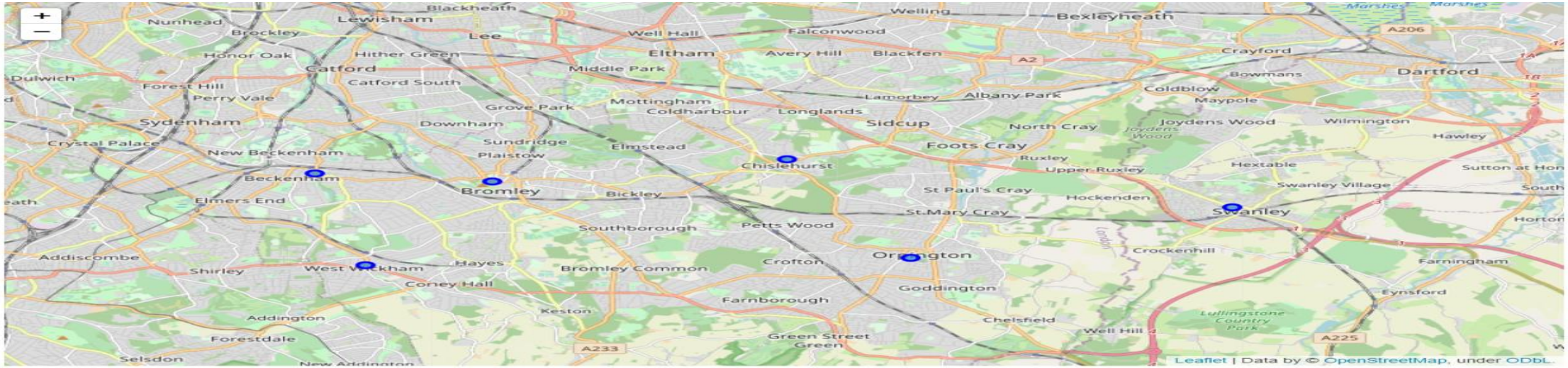
Methodology

My main dataset includes the Postcode district, Post town, Latitude, Longitude and Population.

Out[14]:

	Postcode district	Post town	Latitude	Longitude	Population
0	BR3	BECKENHAM	51.408780	-0.025260	47,411
1	BR1,BR2	BROMLEY	51.406025	0.013156	100,920
2	BR7	CHISLEHURST	51.413785	0.076756	17,322
3	BR5,BR6	ORPINGTON	51.379588	0.103539	92,084
4	BR8	SWANLEY	51.397170	0.173210	22,053
5	BR4	WEST WICKHAM	51.376826	-0.014540	19,367

Using this dataset and python folium, I was able to visualize them and observe the location of post towns, concluding that these post towns are spread out sufficiently throughout the whole Borough, making them suitable for using them as the centre point of our analysis.



With the main dataset and my Foursquare account details, I got information about the top 100 venues that are within a radius of 500 meters of each Post town.

Then I used one hot encoding to convert categorical data into binary vectors, creating a dataframe with the first column being the post town and the rest columns being all the venue categories. An integer value of 1 points out the category of a venue, while 0 indicates that a venue doesn't belong to a particular venue category.

Next, a percentage of the mean frequency of occurrence for each category was calculated by grouping the rows of this dataframe by the Post town.

Out[20]:

	Post town	Post town Latitude	Post town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BECKENHAM	51.40878	-0.02526	Kelsey House Bar & Kitchen	51.407673	-0.025917	Cocktail Bar
1	BECKENHAM	51.40878	-0.02526	Ristorante Pizzeria Sapore Vero	51.408288	-0.025251	Italian Restaurant
2	BECKENHAM	51.40878	-0.02526	Caffè Nero	51.407498	-0.027477	Coffee Shop
3	BECKENHAM	51.40878	-0.02526	Breeze Yoga	51.410077	-0.023541	Yoga Studio
4	BECKENHAM	51.40878	-0.02526	Tapeo	51.407164	-0.026438	Tapas Restaurant

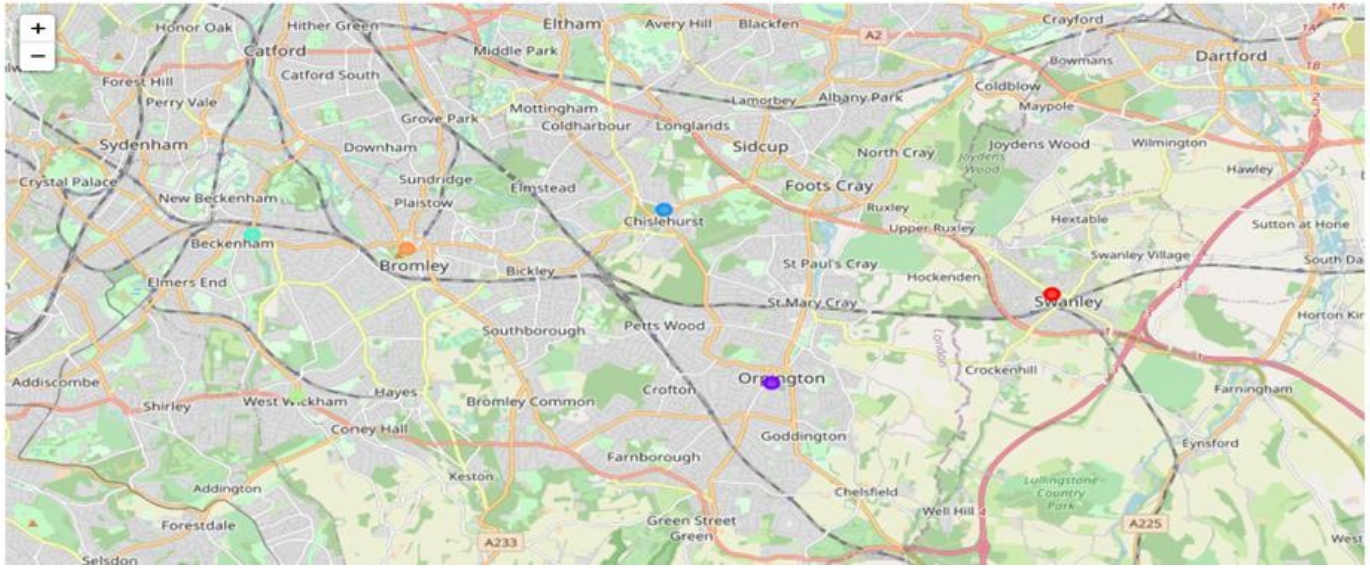
Out[21]:

	Post town	Asian Restaurant	Auto Garage	Bakery	Bar	Bookstore	Bridal Shop	Burger Joint	Café	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	Deli / Bodega	Department Store	Di
0	BECKENHAM	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
1	BECKENHAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	BECKENHAM	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
3	BECKENHAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	BECKENHAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Out[22]:

	Post town	Asian Restaurant	Auto Garage	Bakery	Bar	Bookstore	Bridal Shop	Burger Joint	Café	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	B
0	BECKENHAM	0.000000	0.000000	0.000000	0.044444	0.022222	0.000000	0.000000	0.088889	0.000000	0.022222	0.022222	0.088889	0.0
1	BROMLEY	0.022727	0.022727	0.000000	0.045455	0.022727	0.022727	0.045455	0.045455	0.022727	0.113636	0.000000	0.090909	0.0
2	CHISLEHURST	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.0
3	ORPINGTON	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
4	SWANLEY	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.166667	0.0
5	WEST WICKHAM	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.000000	0.000000	0.058824	0.0

K- means clustering algorithm is a common type of unsupervised learning and in our case, I used it to cluster the Borough of Bromley into 6 cluster groups, one cluster group for each post town. Post towns can be considered the centre of activity within each postcode district making them suitable centroids for clustering.



Next, I created a new dataframe that ranks the top 10 most common venues for each post town. On this dataframe I merged the cluster label, the postcode district, latitude, longitude and population for each Post town. As shown below, this new dataframe states for instance that Beckenham's number one most common venue category is coffee shop, which is a negative factor since too much competition should be avoided.

	Postcode district	Post town	Latitude	Longitude	Population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	
0	BR3	BECKENHAM	51.408780	-0.025260	47,411	3	Coffee Shop	Italian Restaurant	Café	Supermarket	Pharmacy	Platform	Iris
1	BR1,BR2	BROMLEY	51.406025	0.013156	100,920	5	Clothing Store	Coffee Shop	Pub	Bar	Burger Joint	Café	Pa
2	BR7	CHISLEHURST	51.413785	0.076756	17,322	2	Indian Restaurant	Gastropub	Café	Italian Restaurant	Pub	Portuguese Restaurant	Me Re
3	BR5,BR6	ORPINGTON	51.379588	0.103539	92,084	1	Pub	Restaurant	Movie Theater	Hotel	Gym / Fitness Center	Grocery Store	Pa
4	BR8	SWANLEY	51.397170	0.173210	22,053	0	Indian Restaurant	Supermarket	Furniture / Home Store	Coffee Shop	Pizza Place	Fast Food Restaurant	Po Re

Results

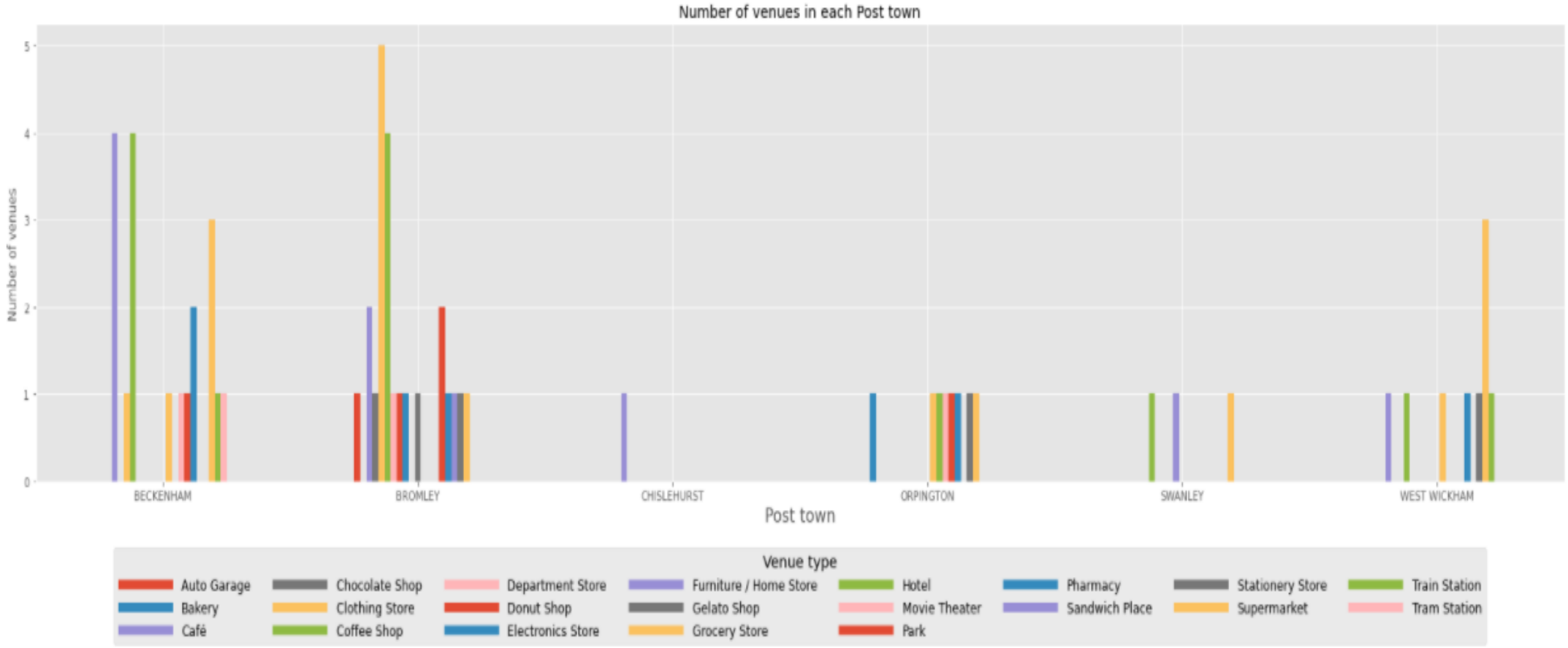
In this section I generated and visualised more targeted results, in an effort to better identify the best location to open a coffee shop. Firstly, I got a list with all the unique venue categories included in the data obtained from Foursquare API, using it to select and use the relevant venue categories. Venue categories like Clothing Stores and Train Stations were included because they increase foot traffic during the day and subsequently the profitability of coffee shops. Restaurant, pubs and bars for instance were removed from my analysis due to being considered irrelevant. Finally, venue categories such as Coffee Shop, Cafés and Donut Shops are competitors and so they were included.

I calculated the total number of venue categories in each post towns, doing so by grouping again by Post town the generated one hot encoding dataframe. This time though, I used the sum function instead of the mean, while also I included only the relevant venue categories I mentioned earlier.

```
array(['Cocktail Bar', 'Italian Restaurant', 'Coffee Shop', 'Yoga Studio',
       'Tapas Restaurant', 'Fish & Chips Shop', 'Grocery Store',
       'Supermarket', 'Café', 'Mediterranean Restaurant', 'Diner',
       'Indian Restaurant', 'Pub', 'Park', 'Kebab Restaurant',
       'Clothing Store', 'Deli / Bodega', 'Portuguese Restaurant',
       'Steakhouse', 'Pharmacy', 'Nightclub', 'Bar', 'Bookstore',
       'Tram Station', 'Pizza Place', 'Fast Food Restaurant',
       'Movie Theater', 'Irish Pub', 'Platform', 'Train Station',
       'Turkish Restaurant', 'Department Store', 'Electronics Store',
       'Gelato Shop', 'Donut Shop', 'Burger Joint', 'Asian Restaurant',
       'Sandwich Place', 'Gym / Fitness Center', 'Stationery Store',
       'Noodle House', 'Chocolate Shop', 'Sushi Restaurant',
       'Shopping Mall', 'Auto Garage', 'Bridal Shop', 'Gastropub', 'Food',
       'Bakery', 'Mexican Restaurant', 'Restaurant', 'Hotel',
       'Furniture / Home Store'], dtype=object)
```

	Post town	Auto Garage	Bakery	Café	Chocolate Shop	Clothing Store	Coffee Shop	Department Store	Donut Shop	Electronics Store	Furniture / Home Store	Gelato Shop	Grocery Store	Hotel	Movie Theater
0	BECKENHAM	0	0	4	0	1	4	0	0	0	0	0	1	0	1
1	BROMLEY	1	0	2	1	5	4	1	1	1	0	1	0	0	0
2	CHISLEHURST	0	0	1	0	0	0	0	0	0	0	0	0	0	0
3	ORPINGTON	0	1	0	0	0	0	0	0	0	0	0	1	1	1
4	SWANLEY	0	0	0	0	0	1	0	0	0	1	0	0	0	0
5	WEST WICKHAM	0	0	1	0	0	1	0	0	0	0	0	1	0	0

Next, using the resulting dataframe I decided that the best method of visualisation in this case was a bar chart, clearly showing the number of venues with a positive or negative influence within each Post town. Bromley for instance has the largest number of clothing stores which is positive, however it's also having the largest number of coffee shops which is negative since they are direct competitors.



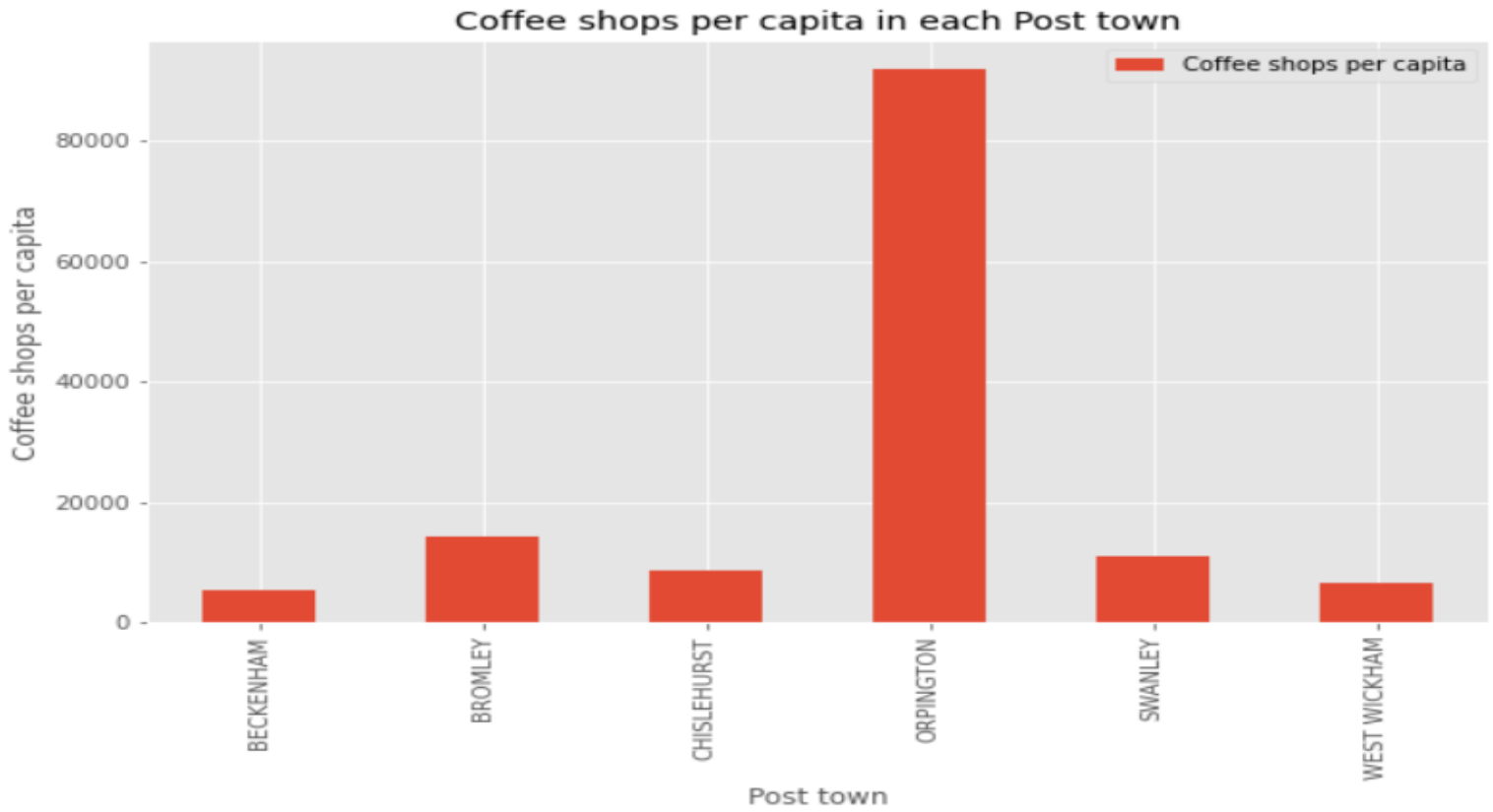
Cafes and coffee shops are direct competitors who can greatly influence profitability negatively. Therefore, I created a model which measures the total number of people who will be served by each direct competitor and our new coffee shop. This model assumes that all coffee shops and cafés within the same post town will serve an equal number of people, serving only people living in their respective postcode district. Firstly, I filtered the data I extracted from Foursquare API, creating a new dataframe which includes only the venues whose Venue category is Coffee shop or Café.

	Post town	Post town Latitude	Post town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	BECKENHAM	51.408780	-0.025260	Caffè Nero	51.407498	-0.027477	Coffee Shop
8	BECKENHAM	51.408780	-0.025260	Rendez-Vous Cafe	51.408005	-0.025303	Café
11	BECKENHAM	51.408780	-0.025260	Em and Lou's Kitchen	51.409528	-0.025227	Café
15	BECKENHAM	51.408780	-0.025260	Fee & Brown	51.409188	-0.025156	Coffee Shop
16	BECKENHAM	51.408780	-0.025260	Starbucks	51.406618	-0.028772	Coffee Shop
25	BECKENHAM	51.408780	-0.025260	Love 'A' Slice	51.406668	-0.030324	Café
34	BECKENHAM	51.408780	-0.025260	Costa Coffee	51.406864	-0.028106	Coffee Shop
36	BECKENHAM	51.408780	-0.025260	Village Bistro Cafe	51.407497	-0.027607	Café
50	BROMLEY	51.406025	0.013156	Costa Coffee	51.405653	0.015279	Coffee Shop
65	BROMLEY	51.406025	0.013156	Costa Coffee	51.404458	0.016645	Coffee Shop
68	BROMLEY	51.406025	0.013156	Patisserie Valerie	51.405164	0.016470	Café
69	BROMLEY	51.406025	0.013156	Caffè Nero	51.402653	0.015616	Coffee Shop
71	BROMLEY	51.406025	0.013156	Starbucks	51.405300	0.015538	Coffee Shop

Next, I merged on my main dataframe a “Competitors” column which includes the total number of cafes and coffee shops within each post town, adding also a column which calculates the total coffee shops per capita with the following calculation: “Population” column / (“Competitors” column+ 1). The +1 represents our new coffee shop. As shown below, the Coffee shops per capita column shows for instance that we can assume 5267.888889 people in Beckenham are served by each coffee shop and café.

	Post town	Postcode district	Latitude	Longitude	Population	Competitors	Coffee shops per capita
0	BECKENHAM	BR3	51.408780	-0.025260	47411	8	5267.888889
1	BROMLEY	BR1,BR2	51.406025	0.013156	100920	6	14417.142857
2	CHISLEHURST	BR7	51.413785	0.076756	17322	1	8661.000000
3	ORPINGTON	BR5,BR6	51.379588	0.103539	92084	0	92084.000000
4	SWANLEY	BR8	51.397170	0.173210	22053	1	11026.500000
5	WEST WICKHAM	BR4	51.376826	-0.014540	19367	2	6455.666667

Finally, a bar chart was used to display the number of coffee shops per capita within each post town. As show, Orpington has a large customer potential when taking only into account post code district population and number of direct competitors within the same Post town.



Discussion

The models discussed point out the locations which have the higher number of neighbouring venues with positive influence, the locations with the highest direct competition and also bring into the equation the total population of each post code district. All post code districts have a Post town which is the centre of activity. Also, these Post towns are sufficiently spread, therefore they were suitable centroids for the K- means clustering algorithm which was used to explore each post code district. Nevertheless, this project has several limitations needed to be addressed. Firstly, I was using the free Foursquare API account version and got a limited number of venues in my dataset, thus an updated account is needed to incorporate a larger number of venues and generate more reliable results. Moreover, our models don't take into consideration the cost, for instance a location closer to a post town will probably be more expensive. Thus, our models may not be suitable for those without substantial initial capital, in this case a K- means clustering algorithm which includes a larger number of areas within the Borough of Bromley may be more suitable. Also, a more detailed analysis on positive impact on profitably could prove to be useful, for example assessing how accessible is each post town to the whole population of Greater London. Finally, constant feedback and continuous improvements are needed when these models are applied in the real world.

Conclusion

Investors who seek to make profit in the coffee market need to utilise dynamic up to date models which are able to navigate this fluid, fast-paced environment, especially when they need to choose a specific location. The objective of this project was to create models which can aid in the selection of a suitable location to open a coffee shop in The London Borough of Bromley, attempting to close this information gap.

References

- [1] https://en.wikipedia.org/wiki/BR_postcode_area
- [2] https://www.google.com/maps
- [3] https://www.postcodearea.co.uk/postaltowns/bromley/
- [4] https://developer.foursquare.com/