Roberto Villalobos February 2021

Capstone Project Final Assignment

Opening a bubble waffle shop in London, United Kingdom

Introduction

The business problem I am going to address in my Capstone Project is a real project of mine which consists in opening a street food stand serving Bubble Waffle (Example image) and hot drinks in the city of London, United Kingdom, where I am currently living.

The city of London is one of the most populated cities in the world and one of the favourite hobbies in Britsh culture is to visit street markets and buy from street stands: from international food to hand-made clothes or other locally produced products.

Desserts and other endulging food have the advantage of generating big margins (a waffle can cost up to £7 in a street market, same as a main dish in the same market) while normally having a more simple process of preparation (flour, sugar, water,.. vs fresh raw materials). This is why I strongly believe there is room in this sector to stablish a Bubble Waffle business.

The main question is where is the best spot in the city to open this stand and maximise both local population living there + daily visitors to have more potential clients.

Data

In order to determine the best location for my street food market, I will look at London' s Boroughs by population and location, and I will complement this data with Foursquare data in order to analyse the most common interest points in these Boroughs and where should I expect more traffic of people that might be interested in buying the product.

The Boroughs' data will be sourced from a table from Wikipedia (link). This table will be converted into a pandas dataframe and then this data will be combined with Foursquare data looking into each Borough. Finally, the place that will be chosen to place our business will be the one with more population concentrated and more points of interest of the type that could complement the purchase of a bubble waffle (Coffe Shops, parks, ..).

Methodology

The methodologies I am going to use for this project will involve some of the techniques and libraries we have learned throughout the course. The following will be the steps of the process:

- Data scrapping: I will get my Borough’s population, area and coordinates from Wikipedia, scrapping one of the tables.
- This table will be transformed into a pandas dataframe in order to be able to handle and analyse the data.

- I will include a measure that calculates the population concentration by dividing the total population of each Borough by the Area, which will give me the number of people by sq mi).
- I will use Foursquare in order to find the main places of interest per Borough and will merge this data with the previous table.
- I will create a ranking of the Top-10 most common places of interest by category in each Borough.
- I will divide my data into 3 clusters using K-means methodology.
- I will display the clusters in a map generated using folium's library.
- I will analyse the different clusters and determine the best Borough in where to open my street food stand.

Results

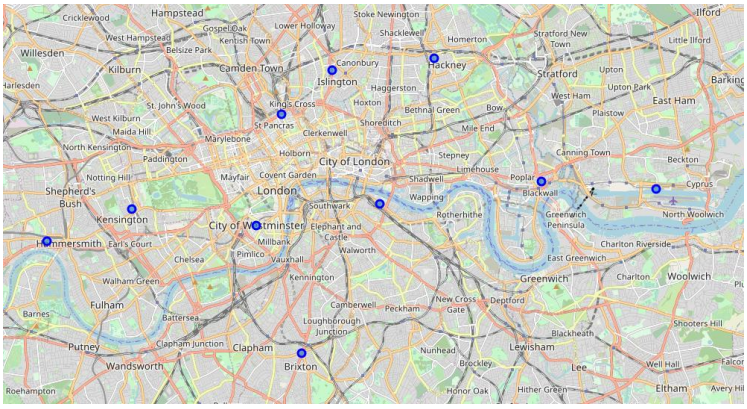
There are 32 Boroughs in the city of London, we get a view of the data available from Wikipedia table in the top 5 rows below:

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2019 est)[1]	Co-ordinates
0	Barking and Dagenham [note 1]	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	212906	.mw-parser-output .geo-default,.mw-parser-outp...
1	Barnet	NaN	NaN	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	395896	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W
2	Bexley	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	248287	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E
3	Brent	NaN	NaN	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	329771	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W
4	Bromley	NaN	NaN	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	332336	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E

Applying the measure to calculate the population concentration by sq mi we find that the Top-10 Boroughs in terms of population concentration are the following:

	index	Borough	Population	Area	Concentration	Lat	Lon
0	28	Tower Hamlets	324745	7.63	42561	51.5099	-0.0059
1	17	Islington	242467	5.74	42241	51.5416	-0.1022
2	10	Hackney	281120	7.36	38195	51.5450	-0.0553
3	18	Kensington and Chelsea	156129	4.68	33360	51.5020	-0.1947
4	5	Camden	270029	8.40	32146	51.5290	-0.1255
5	31	Westminster	261317	8.29	31521	51.4973	-0.1372
6	20	Lambeth	326034	10.36	31470	51.4607	-0.1163
7	11	Hammersmith and Fulham	185143	6.33	29248	51.4927	-0.2339
8	26	Southwark	318830	11.14	28620	51.5035	-0.0804
9	23	Newham	353134	13.98	25259	51.5077	0.0469

We will focus on these 10 Boroughs to run the rest of our project.



Map 1. London Boroughs located in the map.

After using Foursquare data and running the process to extract the venue category by Borough, we find out the Top-5 most common venue categories per borough are the ones following:

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Camden	Coffee Shop	Café	Hotel	Bookstore	Park
1	Hackney	Coffee Shop	Pub	Café	Bakery	Brewery
2	Hammersmith and Fulham	Pub	Café	Coffee Shop	Indian Restaurant	Park
3	Islington	Pub	Coffee Shop	Café	Park	Bakery
4	Kensington and Chelsea	Garden	Italian Restaurant	Café	Juice Bar	Restaurant
5	Lambeth	Coffee Shop	Caribbean Restaurant	Pub	Pizza Place	Cocktail Bar
6	Newham	Hotel	Airport Service	Coffee Shop	Sandwich Place	Bus Station
7	Southwark	Coffee Shop	Hotel	Park	Bakery	Garden
8	Tower Hamlets	Coffee Shop	Hotel	Park	Italian Restaurant	Sandwich Place
9	Westminster	Hotel	Coffee Shop	Sandwich Place	Café	Theater

I decide to create 3 clusters and, after applying K-means and merging our results with the initial table, we see the boroughs are distributed into our clusters as shown below:

Borough	Population	Area	Concentration	Lat	Lon	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Tower Hamlets	324745	7.63	42561	51.5099	-0.0059	1	Coffee Shop	Hotel	Park	Italian Restaurant	Sandwich Place
Islington	242467	5.74	42241	51.5416	-0.1022	0	Pub	Coffee Shop	Café	Park	Bakery
Hackney	281120	7.36	38195	51.5450	-0.0553	0	Coffee Shop	Pub	Café	Bakery	Brewery
Kensington and Chelsea	156129	4.68	33360	51.5020	-0.1947	1	Garden	Italian Restaurant	Café	Juice Bar	Restaurant
Camden	270029	8.40	32146	51.5290	-0.1255	1	Coffee Shop	Café	Hotel	Bookstore	Park
Westminster	261317	8.29	31521	51.4973	-0.1372	1	Hotel	Coffee Shop	Sandwich Place	Café	Theater
Lambeth	326034	10.36	31470	51.4607	-0.1163	0	Coffee Shop	Caribbean Restaurant	Pub	Pizza Place	Cocktail Bar
Hammersmith and Fulham	185143	6.33	29248	51.4927	-0.2339	0	Pub	Café	Coffee Shop	Indian Restaurant	Park
Southwark	318830	11.14	28620	51.5035	-0.0804	1	Coffee Shop	Hotel	Park	Bakery	Garden
Newham	353134	13.98	25259	51.5077	0.0469	2	Hotel	Airport Service	Coffee Shop	Sandwich Place	Bus Station

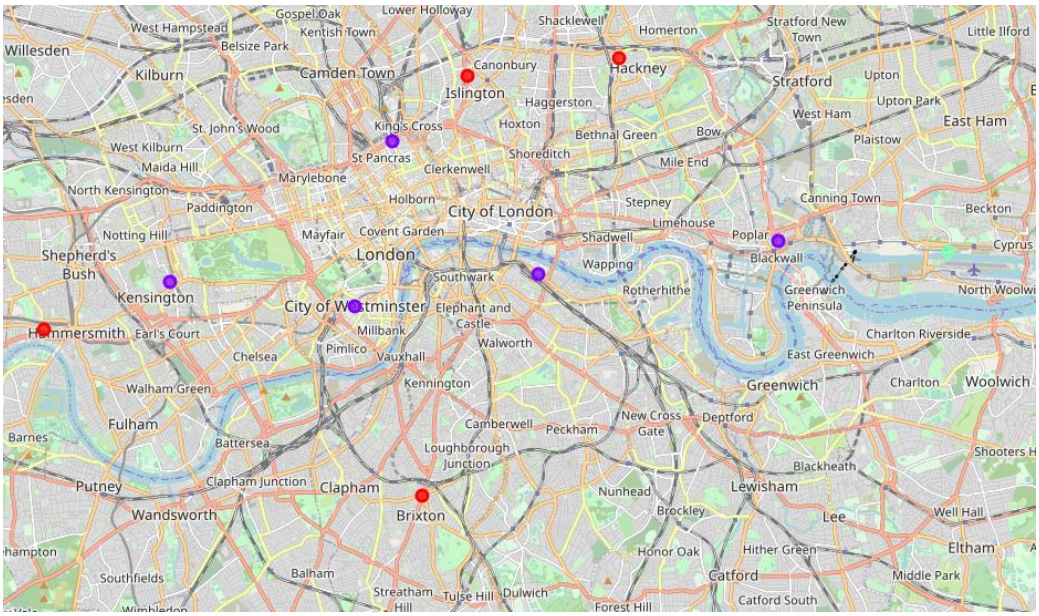


Figure 2. London boroughs with the clusters applied (0= Light green, 1= Red, 2= Purple).

Cluster 1

Cluster 1 has 4 boroughs: Islington, Hackney, Lambeth, and Hammersmith and Fulham.

	Borough	Concentration	Lat	Lon	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Islington	42241	51.5416	-0.1022	0	Pub	Coffee Shop	Café	Park	Bakery
2	Hackney	38195	51.5450	-0.0553	0	Coffee Shop	Pub	Café	Bakery	Brewery
6	Lambeth	31470	51.4607	-0.1163	0	Coffee Shop	Caribbean Restaurant	Pub	Pizza Place	Cocktail Bar
7	Hammersmith and Fulham	29248	51.4927	-0.2339	0	Pub	Café	Coffee Shop	Indian Restaurant	Park

This cluster main characteristic is all boroughs have pubs in top-3 positions, indicating these activity is focused on going out and having some drinks and food (pubs, restaurants, cafes).

Cluster 2

Cluster 2 has 5 boroughs: Tower Hamlets, Kensington and Chelsea, Camden, Westminster, and Southwark.

	Borough	Concentration	Lat	Lon	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Tower Hamlets	42561	51.5099	-0.0059	1	Coffee Shop	Hotel	Park	Italian Restaurant	Sandwich Place
3	Kensington and Chelsea	33360	51.5020	-0.1947	1	Garden	Italian Restaurant	Café	Juice Bar	Restaurant
4	Camden	32146	51.5290	-0.1255	1	Coffee Shop	Café	Hotel	Bookstore	Park
5	Westminster	31521	51.4973	-0.1372	1	Hotel	Coffee Shop	Sandwich Place	Café	Theater
8	Southwark	28620	51.5035	-0.0804	1	Coffee Shop	Hotel	Park	Bakery	Garden

The main difference in this cluster is the presence of Hotel in high positions (hint for tourism) and Park/Garden in most of the boroughs. These kinds of venues indicate tourist activity and other leisure activities such as walking and having take-aways (coffee, sandwich, Juice).

Cluster 3

Cluster 3 only has 1 borough: Newham.

	Borough	Concentration	Lat	Lon	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Newham	25259	51.5077	0.0469	2	Hotel	Airport Service	Coffee Shop	Sandwich Place	Bus Station

We can clearly see Newham Borough is focused on the Airport and its services (Hotel, Airport, Bus Station), so we could say this area is not a destination for its points of interest in terms of tourism but rather an area to be close to the airport, with no major points of interest that indicate leisure activities that could drive bubble waffle sales.

Recommendations

After taking into consideration our results and findings in the previous section of our project, I strongly recommend focusing on the boroughs in Cluster 2.

Within this cluster and considering the population concentration in the area in order to maximise both the population living in the area + visitors, **I would recommend stablishing our Bubble Waffle stand in Tower Hamlets,** where we can definitely benefit from must-sees in London such as Tower of London and Tower Bridge that attract millions of tourists every year.

Conclusions

To sum up, this capstone project has allowed us to find the optimum borough to place our Bubble Waffle business in London: Tower Hamlets Borough.

We got to this conclusion by applying different techniques and methodologies studied across the courses offered by Coursera. Just to mention some of them, we have:

- Used pandas, numpy, matplotlib, and folium libraries
- Used Foursquare data in order to explore the venues in each borough
- Scrapped data from the internet
- Used K-means to split our data into clusters to help us in our decision

London is definitely one of the most visited places in Europe for its history and importance in Europe, attracting millions of tourists and visitors every year and making it a perfect city to stablish new businesses and launching new projects.