

Applied Data Science Capstone

Juan Andres Ron,

26 June, 2020

Contents

Introduction: Business Problem	2
Data:	2
Methodology	3
Exploratory Data Analysis	4
Population:	4
Income:	5
Tourists:	6
Population, income and tourism:	8
Cluster Analysis and Modelling	9
The first winner: Westminster	11
Foursquare - Westminster Ice Cream Shops	11
Discussion & Conclusion	14

Introduction: Business Problem

The main objective of this project is to analyze London city, to understand/explore the best neighborhood to set up an Italian ice cream shop. This state-of-the-art shop targets the upper-level income population with a premium portfolio of innovative ice cream flavors.

In this analysis, we are introducing boroughs public information: average individual income, tourists and population density.

Hotel rooms per borough and population per square kilometer are valuable measures to predict the number of future clients. Furthermore, as we want to reach a specific target of wealthy individuals, we need the measure "income".

Data:

- Population per borough ("Population per square kilometer")¹
- Average income per borough.²
- Hotels rooms per borough.³
- Number of existing ice cream shops and frozen yoghurt shops in the neighborhoods of the selected borough, according to Foursquare.

¹ <https://data.london.gov.uk/dataset/office-national-statistics-ons-population-estimates-borough>.

² <https://data.london.gov.uk/dataset/earnings-place-residence-borough>.

³ https://www.london.gov.uk/sites/default/files/visitor_accommodation_-_working_paper_88.pdf

Methodology

London is a complex urban agglomerate, with a large number of areas and different forms to classify them. It can get partitioned by political boroughs, electoral wards, neighborhoods, commuting zones or even by their local underground station influence. Most of the public information is available on a borough level and, that is why our first approach will explore them. As we said in the introduction, we are going to see which are the ones with the highest population density, tourists and income. We are looking for a place with wealthy people and high traffic (tourists and permanent residents).

First, we are going to describe each area and select some potential boroughs for our shop. Then, we are doing a K-Means clustering of London boroughs according to this information (population, income and tourists). Afterwards, we are using Foursquare data to understand in the selected borough which neighborhoods have the most quantity of ice cream shops, and indirect competition shops (like Frozen Yoghurt), to finally decide the winner location. The idea is to find an opportunity, good location but not so many competitors.

Public information comes in different formats: csv, excel files, pdf, among others. Before submitting this information to Jupyter Notebook, I had to do some editing with Microsoft Excel and all files were saved as csv. These edit work included separating merged cells, and eliminating formats that were chosen by the file's author. I had also to extract information from a PDF table that wasn't available in other type of formats. I understand there might be some way to do this using Python technology but the use of this techniques exceeds the objective of this projects.

Exploratory Data Analysis

Population:

Population per square information is available in a csv file called “housing-density-borough” that includes real information and estimation of the London population. We were only interested in two columns: “Population_per_square_kilometre” and “Borough”, for the year 2020.

	Population_per_square_kilometre
count	33.00
mean	7767.07
std	4181.09
min	2240.60
25%	4693.40
50%	6178.80
75%	11443.80
max	16583.50

As we can see the average “Population_per_square_kilometre” in London boroughs is: 7776.07 residents and 25% of them have 11443.80 or more individuals (top quartile).

These are the boroughs with higher density:

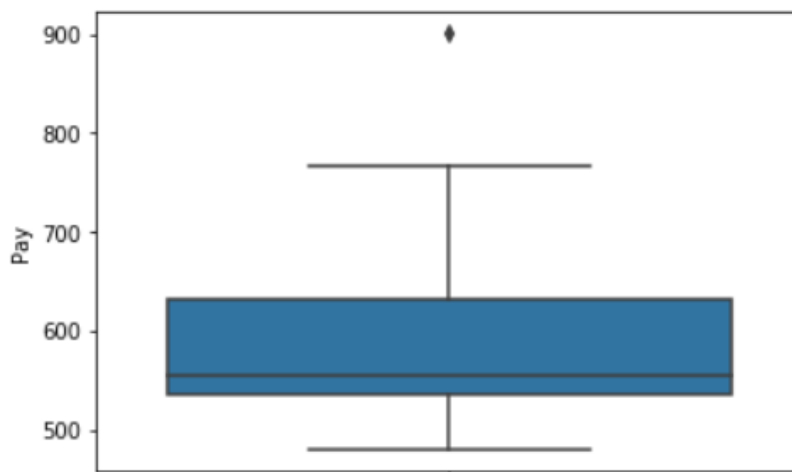
Borough	Population_per_square_kilometre
Tower Hamlets	16583.5
Islington	16344.7
Hackney	15197.0
Kensington and Chelsea	13262.4
Lambeth	12693.9
Westminster	12099.5
Camden	11812.4
Hammersmith and Fulham	11631.1

Income:

Income information is available in an csv file called “earnings-residence-borough” that includes information from the Annual Survey of Hours and Earnings (ASHE), ONS. According to the Office for National Statistics *“This data set provides information about earnings of employees who are living in an area, who are on adult rates and whose pay for the survey pay-period was not affected by absence”*⁴. I choose for this analysis the weekly pay information, for the year 2018.

	Pay
count	33.00
mean	589.35
std	90.94
min	479.10
25%	535.90
50%	554.70
75%	632.50
max	902.00

As we can see the average “Pay” in London boroughs is: £589.35 and 25% of residents earn £632.50 or more (top quartile). That is £2530 a month. I found interesting for this variable the Boxplot graphic, as it shows some extreme pay values, that correspond to the City of London borough.



The following boroughs are the ones with the highest paid residents:

⁴ <https://data.london.gov.uk/dataset/earnings-place-residence-borough>

Borough	Pay
City of London	902.0
Westminster	766.6
Wandsworth	689.9
Islington	687.6
Hammersmith and Fulham	681.3
Richmond upon Thames	678.2
Kensington and Chelsea	669.3
Camden	634.7

As an early finding, we can observe that Westminster (highlighted in yellow) has a concentrated well-paid population. Hammersmith and Fulham and Camden have some likeness.

Tourists:

There is a very complete working paper regarding tourism, where I could extract the information of hotel rooms per boroughs⁵.

	Number_of_bedrooms
count	33.00
mean	4416.27
std	7365.71
min	235.00
25%	614.00
50%	1929.00
75%	4590.00
max	38480.00

The average number of bedrooms per London boroughs is: 7365.71 and the top 25% them have at least 4590 (top quartile). Hotel rooms tend to be concentrated in a few boroughs.

These boroughs are: Westminster, Camden, Kensington and Chelsea, Hillingdon, Tower Hamlets, City of London, Southwark and Newham. Once again Westminster appears in our analysis.

⁵ https://www.london.gov.uk/sites/default/files/visitor_accommodation_-_working_paper_88.pdf

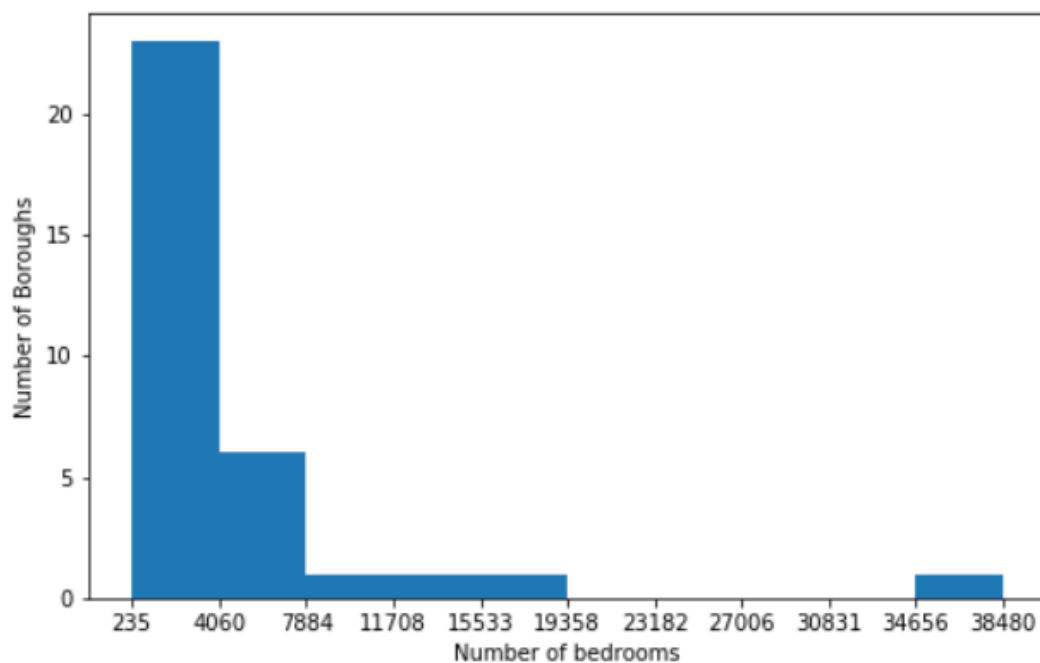
Hillingdon and Newham are new appearances. London Heathrow is located in Hillingdon, and that probably explains its large amounts of hotels rooms. Newham has the London Airport and some mayor transports hubs like Stratford.

Borough	Number_of_bedrooms
Westminster	38480
Camden	18038
Kensington and Chelsea	15151
Hillingdon	9885
Tower Hamlets	6757
City of London	6064
Southwark	5802
Newham	4810

Another way to observe the concentration of hotels rooms in a few boroughs with a histogram.

X = Number of bedrooms

Y = Quantity of boroughs



Population, income and tourism:

Merging all this different tables we can identify which boroughs has the maximum level population density, income, and hotel rooms.

Borough	Population_per_square_kilometre	Pay	Number_of_bedrooms
Camden	11812.4	634.7	18038
Kensington and Chelsea	13262.4	669.3	15151
Westminster	12099.5	766.6	38480

These are the boroughs that performed in the top quartile of each attribute: Camden, Kensington and Chelsea, and Westminster.

Cluster Analysis and Modelling.

The objective of this analysis is to understand which boroughs are alike considering the attributes we have been evaluating (pay, bedrooms, population per square). To do that we proceed with a K-mean clustering, using Sklearn package. The number of clusters defined is = 5.

These are the 5 clusters with their average's values:

Labels	Population_per_square_kilometre	Pay	Number_of_bedrooms
0	4395.89	562.27	1878.40
1	8117.87	512.84	1636.43
2	2770.70	902.00	6064.00
3	12099.50	766.60	38480.00
4	13186.62	639.57	6840.22

We are interested in cluster 3 due to its high values in all variables. As we can see it is represented by only one borough: Westminster.

Borough	Population_per_square_kilometre	Pay	Number_of_bedrooms	Labels
Westminster	12099.5	766.6	38480	3

High Income – High population density – High number of hotel bedrooms

Cluster 2 is only represented by The City of London. It is interesting but not many people live there and it hasn't got many hotels.

Borough	Population_per_square_kilometre	Pay	Number_of_bedrooms	Labels
City of London	2770.7	902.0	6064	2

Maximum Income – Low population density – Above average number of hotel bedrooms

Cluster 4 is also a valid option. There are some very touristic boroughs with a large number of wealthy residents.

Borough	Population_per_square_kilometre	Pay	Number_of_bedrooms	Labels
Kensington and Chelsea	13262.4	669.3	15151	4
Islington	16344.7	687.6	3633	4

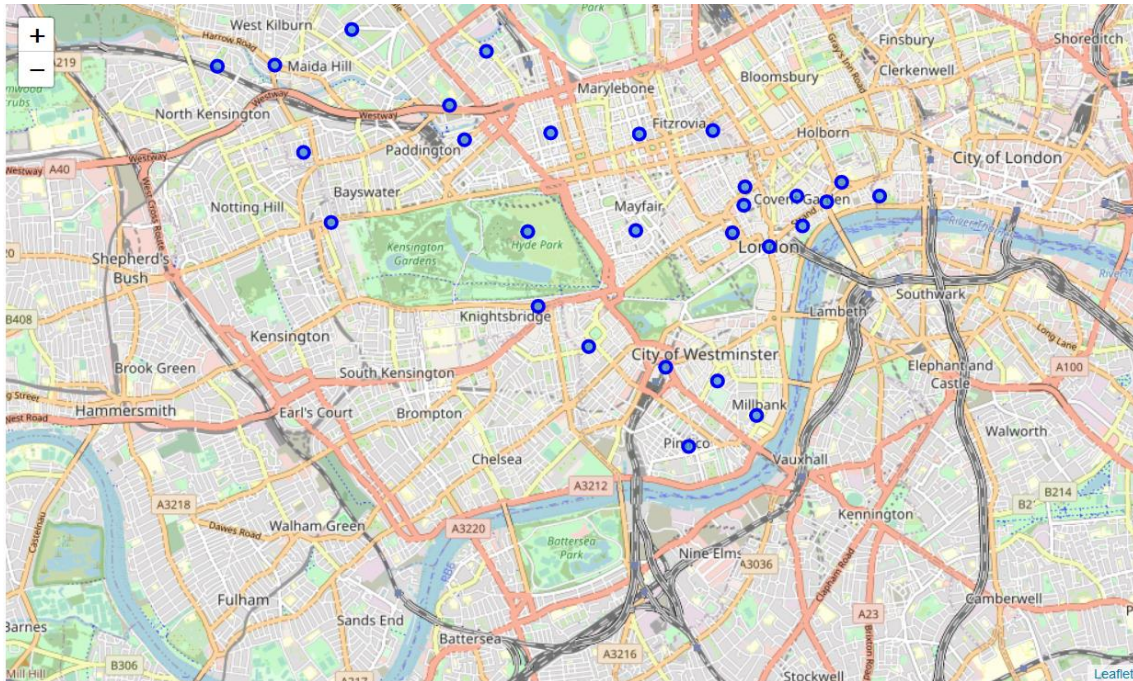
Borough	Population_per_square_kilometre	Pay	Number_of_bedrooms	Labels
Wandsworth	9710.8	689.9	1393	4
Hammersmith and Fulham	11631.1	681.3	4590	4
Southwark	11443.8	589.4	5802	4
Hackney	15197.0	555.6	1929	4
Tower Hamlets	16583.5	627.9	6757	4
Camden	11812.4	634.7	18038	4
Lambeth	12693.9	620.4	4269	4

*Above average income – Maximum population density – High number of hotel
bedrooms*

Cluster 0 and 1 are not good options for our shop. They might be considered for a niche alternative or a different type of ice-cream shop.

The first winner: Westminster

We chose this borough to put our shop, but where? The borough is quite extended and has many districts. According to Wikipedia⁶ there are more than 30 different districts. Let's analyze where they are located.



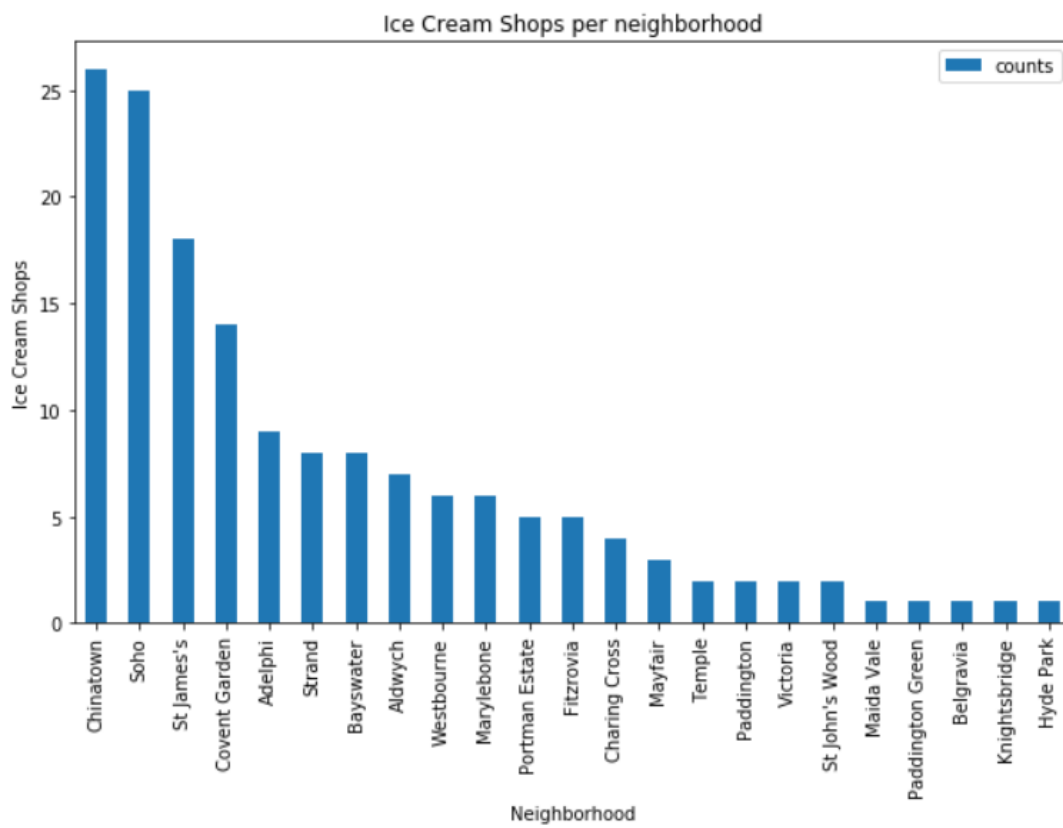
Foursquare - Westminster Ice Cream Shops

Now we are going to analyze the quantity of ice cream shops in each neighborhood of Westminster.

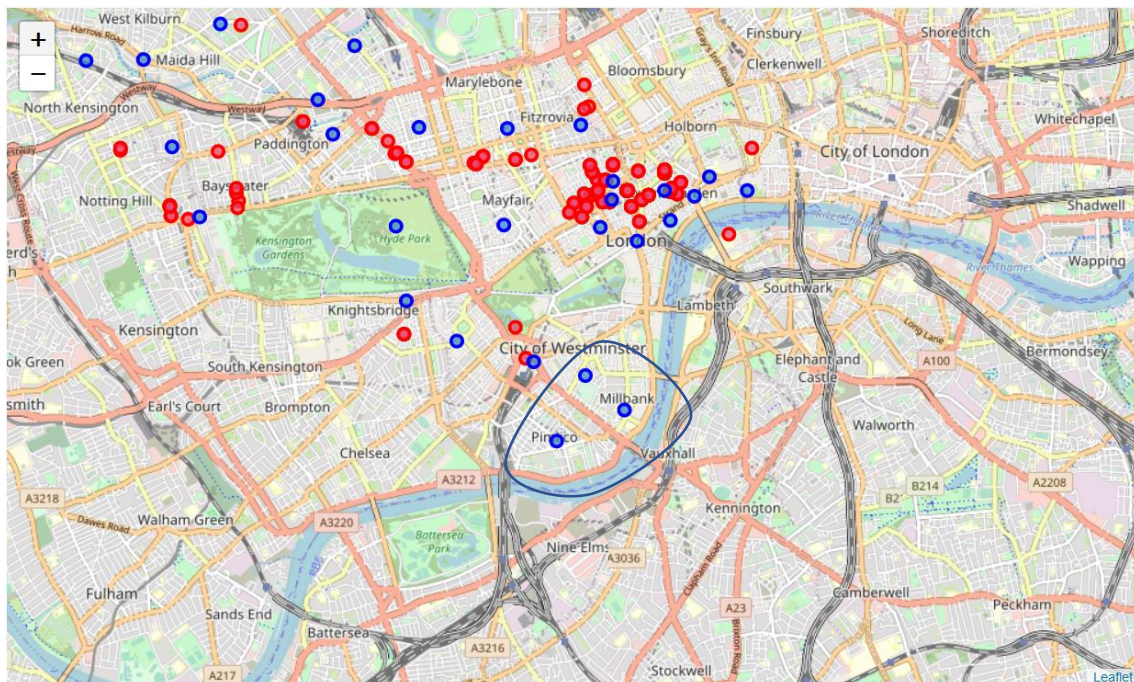
As we can observe in the following bar charts most of the ice-cream shops are located near Chinatown, Soho, St James and Covent Garden. It makes sense as these are very touristic places.

In the map that follows it is easy to observe the neighborhoods with low quantity of ice cream shops: Millbank, Pimlico, Queens Park, Lisson Grove, Kensal Town and Westbourne Green. One thing to consider, that we are not analyzing in this project is the density of commercial zones. There are some places that might not have space to rent a venue.

⁶ https://en.wikipedia.org/wiki/List_of_districts_in_the_City_of_Westminster



In blue neighborhoods center. In red, ice-cream shops.



How about yoghurt shops? As we can see in the following table there are not many shops of this kind in the borough, and most of them are concentrated in St James, Chinatown, Adelphi and Soho.

unique_values	counts
St James's	2
Chinatown	2
Adelphi	2
Soho	2
Westbourne	1
Covent Garden	1
Charing Cross	1

It seems that frozen yoghurt is not as popular as ice cream or it is not well served in the area. In any case they do not represent a competitive danger.

Discussion & Conclusion

In this project we have analyzed London boroughs to find a suitable place for a deluxe ice cream shop considering income, population density and quantity of tourists. We discovered that Westminster might be a suitable option, since it has a wide wealthy population and a lot of tourism. But there are other options like Camden, Kensington and Chelsea.

We found out with Foursquare information that in Westminster there are some specific zones that haven't got many ice cream shops, that might be suitable options. These are: Millbank, Pimlico, Queens Park, Lisson Grove, Kensal Town and Westbourne Green.

We couldn't find information regarding zoning and shop rents per borough that are important aspects to consider to proceed with this business.