# Identify A Strategic Location to Open a Brazilian take away restaurant in London, United Kingdom

Marcello Dichiera

Capstone Project

# Introduction

- The success of establishing a new restaurant depends on several factors: demand, brand loyalty, quality of food, competition, and so on.
- In most cases, a restaurant's location plays an essential determinant for its success.
- Whether opening your first full-service restaurant, it’s important to understand what to look out for when choosing a new restaurant location.

# Problem Statement

- Opening a Brazilian take away restaurants in London.
- Objective is to identify the optimal neighborhood location to open.
- Key factors to consider are: spending power of the London population, distribution of Brazilian restaurants, distance to public transport station.
- To identify the ideal London neighborhood clusters group, we use Foursquare API, data scraping, geopy, pycaret to build the clustering model and matplotlib/seaborn to visualize data during our EDA (Exploratory Data Analysis).

# Key Clients

- Brazilian Family of chefs, highly experienced in cooking for small to medium restaurants

# Data Acquisition, Wrangling and Cleaning

- Data scraped (neighbourhoods, postal codes, geographical coordinates, population and income) will be scraped from here: https://www.doogal.co.uk/UKPostcodesCSV.ashx?area=London0 .
- For returning the number of Brazilian restaurants in the vicinity of each neighborhood, we will use Foursquare API, more specifically, its *explore* function.

# Data Cleaning

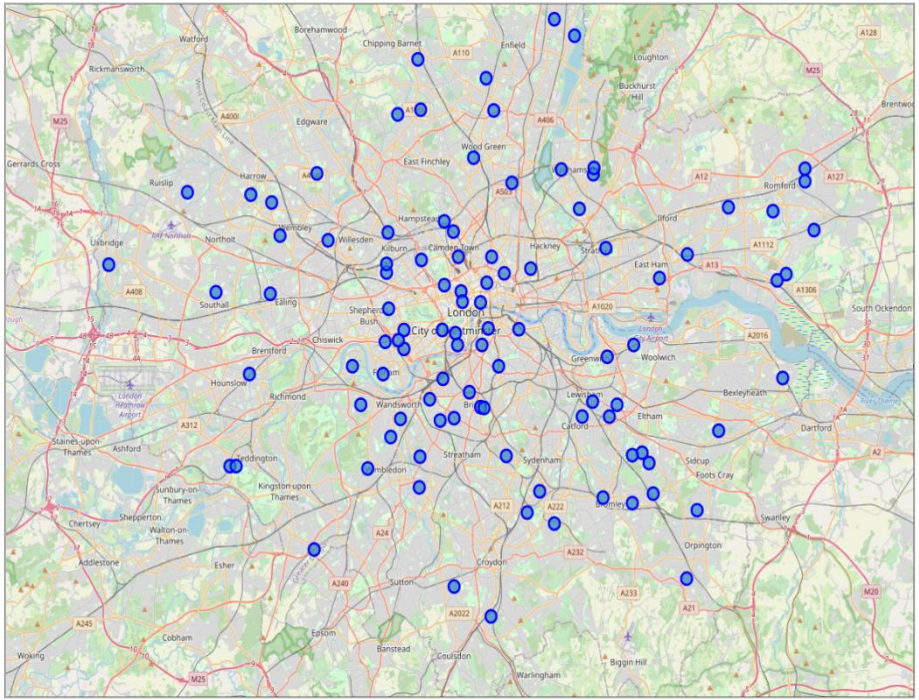
- Drop columns that do not contain useful information for the EDA (Exploratory Data Analysis) and the clustering model.
- Drop of rows with null or non existent values
- Final dataframe has been simplified and sampled to 100 rows because the folium library and the colab platform crashed several times.

**Cleaned dataset**

| | District | Latitude | Longitude | Population | Households | Distance to station | Average Income |
|---|---|---|---|---|---|---|---|
| **0** | Bromley | 51.404543 | 0.014195 | 34.0 | 21.0 | 0.462939 | 63100 |
| **1** | Bromley | 51.408058 | 0.015874 | 38.0 | 37.0 | 0.083058 | 63100 |
| **2** | Bromley | 51.409191 | 0.010068 | 1.0 | 1.0 | 0.489492 | 56100 |
| **3** | Bromley | 51.400462 | 0.016716 | 4.0 | 4.0 | 0.067905 | 63100 |
| **4** | Bromley | 51.401684 | 0.015705 | 14.0 | 6.0 | 0.219358 | 63100 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **140844** | Hillingdon | 51.623983 | -0.495253 | 18.0 | 5.0 | 2.347240 | 54200 |
| **140845** | Hillingdon | 51.626955 | -0.494143 | 22.0 | 15.0 | 2.048740 | 54200 |
| **140846** | Hillingdon | 51.628575 | -0.499204 | 2.0 | 1.0 | 2.191290 | 54200 |
| **140847** | Barnet | 51.643292 | -0.255958 | 11.0 | 6.0 | 1.960300 | 58000 |
| **140848** | Barnet | 51.642309 | -0.256627 | 71.0 | 55.0 | 1.984360 | 58000 |

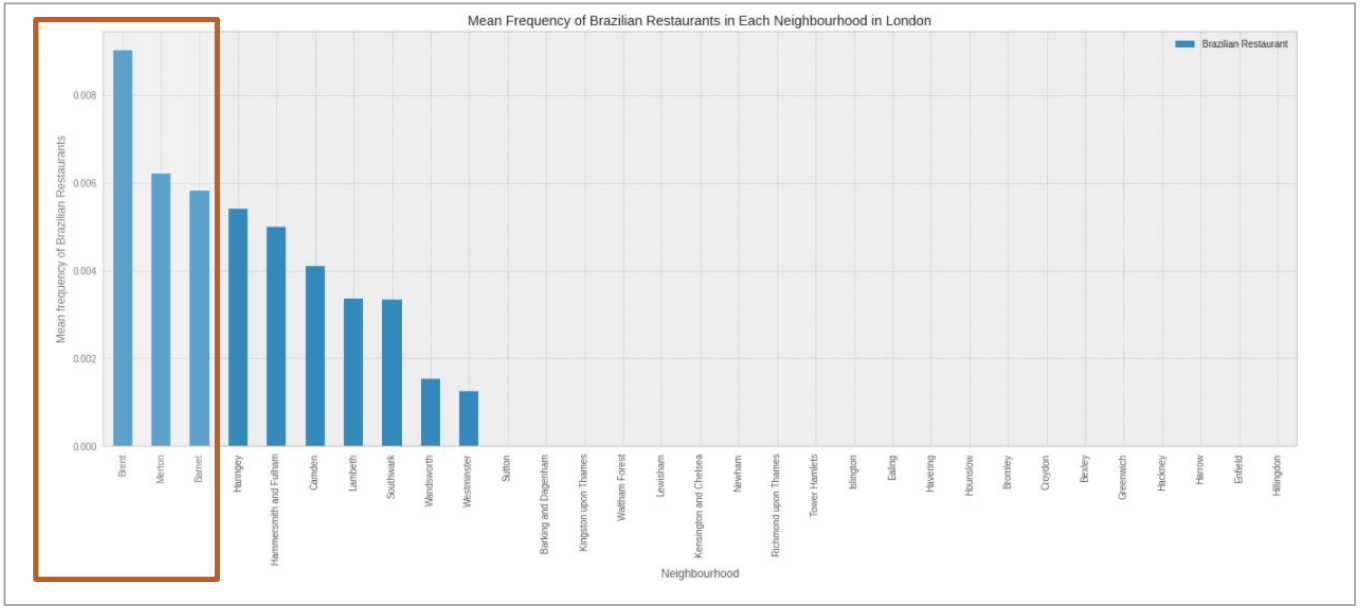
# Exploratory Data Analysis

- Folium mapping of city of London and neighborhoods



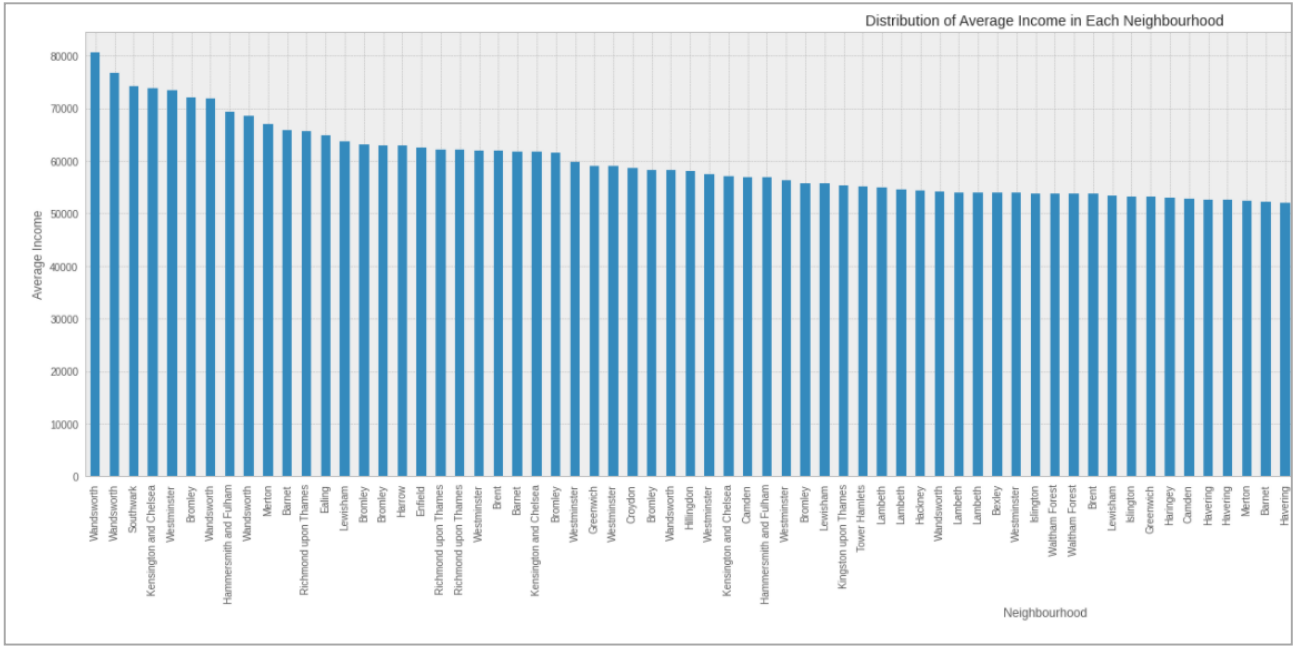
# Frequency Distribution of Brazilian Restaurants

- Higher number of Brazilian restaurants in a neighborhood, the stronger the competition.



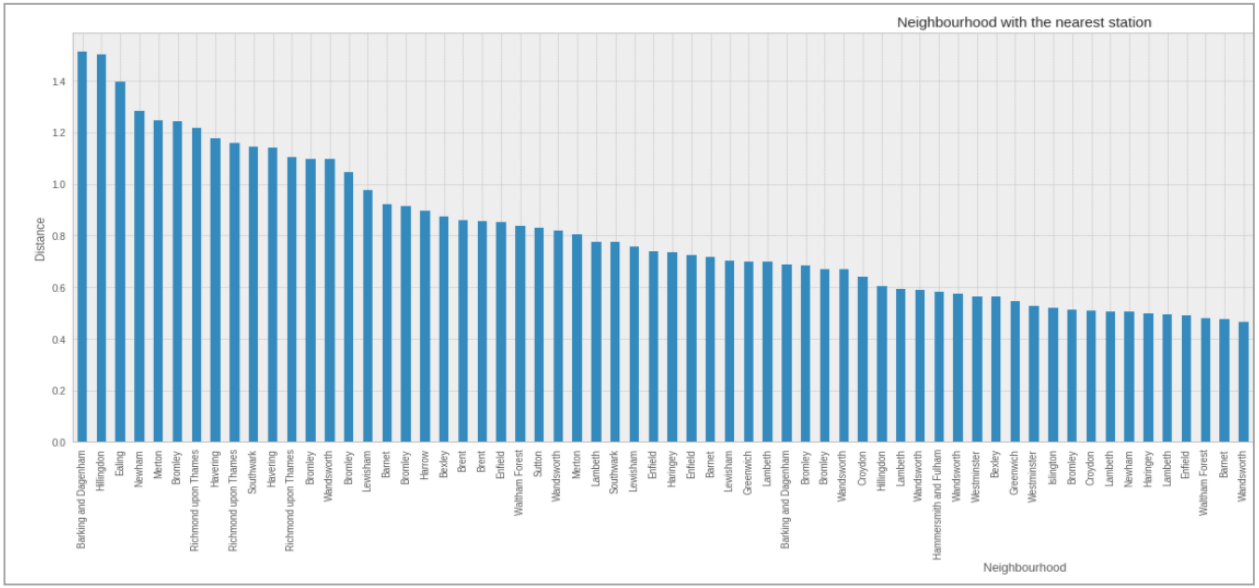
# Distribution of Median Household Income

- Neighborhoods distributed towards around the mean can readily afford and indulge themselves



# Distance to Station by neighborhoods

- Most people live in London without having a car and the public transport is quite important to reach the areas and venues of preference



# Clustering Modeling

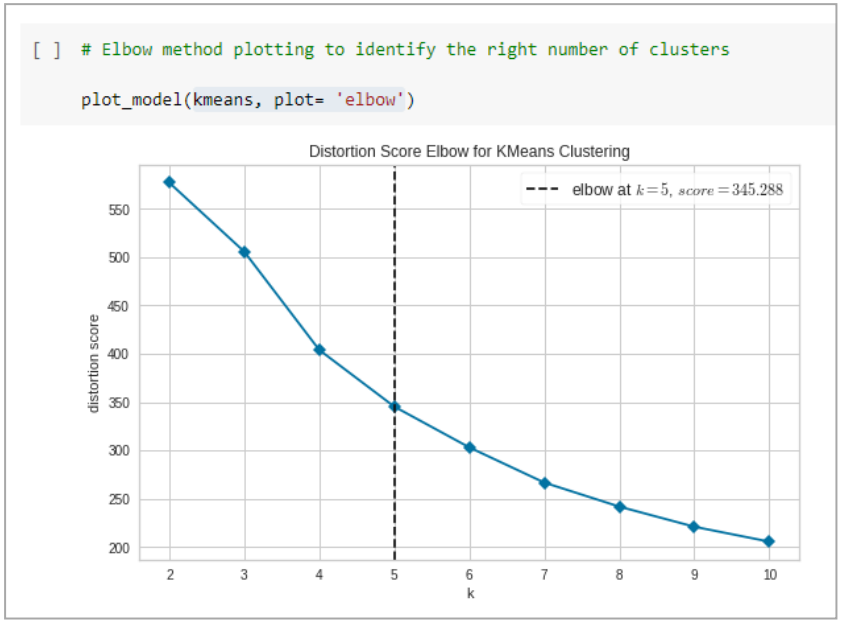
- **Data Pre-processing of cluster model (Pycaret):**
- PyCaret is an open source low-code machine learning library in Python that aims to reduce the hypothesis to insights cycle time in a ML experiment. It enables data scientists to perform end-to-end experiments quickly and efficiently (source: https://pycaret.org ).
- The library allows to implement a simple preprocessing just with few lines of codes.

```
[ ]  #we import the clustering model of pycaret

     from pycaret.clustering import *

     exp_clu101 = setup(df_final, normalize = True,
                        ignore_features = ['Neighbourhood'],
                        session_id = 123)
```

# K-Means Clustering model

- To implement a k-means model usually is important to assign a number of clusters.
- Pycaret highlights the suitable number of clusters calculated with the squared error as a performance metric.
- The suitable number of clusters is 5:

```
[ ]  # Elbow method plotting to identify the right number of clusters

     plot_model(kmeans, plot= 'elbow')
```

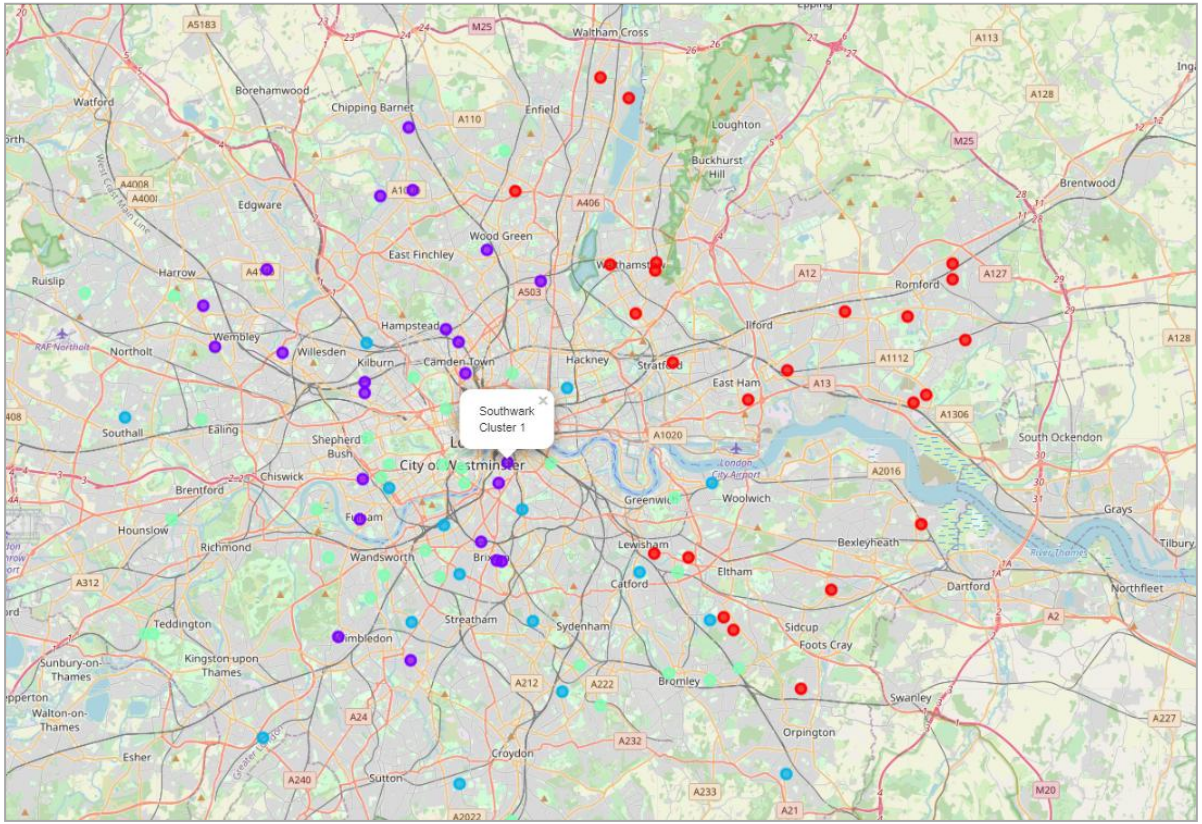


# Cluster Labels

- Here below we have the final data frame with the neighborhoods and the clusters (from 0 to 4 as in Python the counting starts from zero so 5 clusters: 0,1,2,3,4).

| | Neighbourhood | Latitude | Longitude | Population | Households | Distance to station | Average Income | Brazilian Restaurant | Cluster |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Westminster | 51.528918 | -0.206048 | 71.0 | 20.0 | 0.565596 | 44500 | 0.00125 | Cluster 1 |
| 1 | Westminster | 51.496019 | -0.137685 | 46.0 | 27.0 | 0.433721 | 57500 | 0.00125 | Cluster 3 |
| 2 | Westminster | 51.522106 | -0.149017 | 3.0 | 1.0 | 0.316775 | 73500 | 0.00125 | Cluster 3 |
| 3 | Westminster | 51.513050 | -0.131427 | 5.0 | 3.0 | 0.294475 | 54000 | 0.00125 | Cluster 3 |
| 4 | Westminster | 51.535685 | -0.171829 | 6.0 | 2.0 | 0.180020 | 59900 | 0.00125 | Cluster 3 |

# Clusters on the London Map



# Cluster Description

- Cluster 0:
- Low spending power
- No presence of competition
- mid/high distance to the station

| | Neighbourhood | Latitude | Longitude | Population | Households | Distance to station | Average Income | Brazilian Restaurant | Cluster |
|---|---|---|---|---|---|---|---|---|---|
| 8 | Bromley | 51.425513 | 0.052632 | 34.0 | 17.0 | 1.098220 | 45800 | 0.0 | 0 |
| 9 | Bromley | 51.399587 | 0.100309 | 71.0 | 28.0 | 0.683057 | 42300 | 0.0 | 0 |
| 10 | Bromley | 51.430848 | 0.045822 | 35.0 | 11.0 | 1.045150 | 45800 | 0.0 | 0 |
| 32 | Waltham Forest | 51.584746 | -0.033925 | 30.0 | 11.0 | 0.409922 | 50700 | 0.0 | 0 |
| 33 | Waltham Forest | 51.582277 | -0.002204 | 89.0 | 31.0 | 0.478209 | 53700 | 0.0 | 0 |
| 34 | Waltham Forest | 51.563496 | -0.015653 | 6.0 | 6.0 | 0.837416 | 46700 | 0.0 | 0 |
| 35 | Waltham Forest | 51.585813 | -0.001676 | 58.0 | 29.0 | 0.097137 | 53700 | 0.0 | 0 |
| 40 | Enfield | 51.657057 | -0.020592 | 1.0 | 1.0 | 0.723715 | 38700 | 0.0 | 0 |
| 42 | Enfield | 51.666158 | -0.040539 | 76.0 | 30.0 | 0.853476 | 38300 | 0.0 | 0 |
| 43 | Enfield | 51.616572 | -0.100123 | 16.0 | 5.0 | 0.737417 | 46200 | 0.0 | 0 |
| 47 | Newham | 51.542189 | 0.010360 | 36.0 | 15.0 | 0.503985 | 50200 | 0.0 | 0 |
| 48 | Newham | 51.525888 | 0.063026 | 95.0 | 33.0 | 1.283330 | 40400 | 0.0 | 0 |
| 66 | Lewisham | 51.458596 | -0.002975 | 40.0 | 17.0 | 0.702157 | 53500 | 0.0 | 0 |
| 70 | Havering | 51.585085 | 0.206607 | 69.0 | 31.0 | 0.356473 | 52600 | 0.0 | 0 |
| 71 | Havering | 51.527812 | 0.188181 | 83.0 | 30.0 | 1.176440 | 44000 | 0.0 | 0 |
| 72 | Havering | 51.551881 | 0.215761 | 100.0 | 47.0 | 0.351363 | 52100 | 0.0 | 0 |
| 73 | Havering | 51.578145 | 0.206487 | 56.0 | 23.0 | 0.414984 | 52600 | 0.0 | 0 |
| 74 | Havering | 51.524680 | 0.179122 | 171.0 | 57.0 | 1.142380 | 44000 | 0.0 | 0 |
| 86 | Greenwich | 51.457183 | 0.021233 | 60.0 | 30.0 | 0.699302 | 53200 | 0.0 | 0 |
| 92 | Bexley | 51.443029 | 0.121412 | 88.0 | 39.0 | 0.874050 | 54000 | 0.0 | 0 |
| 93 | Bexley | 51.471565 | 0.185028 | 69.0 | 21.0 | 0.565264 | 44200 | 0.0 | 0 |
| 94 | Barking and Dagenham | 51.561947 | 0.175161 | 26.0 | 11.0 | 1.512370 | 43700 | 0.0 | 0 |
| 95 | Barking and Dagenham | 51.538752 | 0.090788 | 116.0 | 35.0 | 0.688178 | 49900 | 0.0 | 0 |
| 96 | Barking and Dagenham | 51.564448 | 0.131013 | 61.0 | 24.0 | 0.423293 | 40800 | 0.0 | 0 |

# Cluster description

- Cluster 1:
- Mid spending power
- Med/high presence of competition
- Medium distance to the station

| | Neighbourhood | Latitude | Longitude | Population | Households | Distance to station | Average Income | Brazilian Restaurant | Cluster |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Westminster | 51.528918 | -0.206048 | 71.0 | 20.0 | 0.565596 | 44500 | 0.001250 | 1 |
| 18 | Lambeth | 51.455669 | -0.113461 | 9.0 | 5.0 | 0.777299 | 54000 | 0.003367 | 1 |
| 21 | Lambeth | 51.455202 | -0.109609 | 63.0 | 27.0 | 0.506102 | 54000 | 0.003367 | 1 |
| 22 | Lambeth | 51.489647 | -0.112059 | 2.0 | 1.0 | 0.495755 | 54600 | 0.003367 | 1 |
| 23 | Lambeth | 51.463865 | -0.124353 | 22.0 | 9.0 | 0.402040 | 54900 | 0.003367 | 1 |
| 24 | Brent | 51.582646 | -0.274947 | 57.0 | 19.0 | 0.353669 | 50900 | 0.009009 | 1 |
| 25 | Brent | 51.533333 | -0.206048 | 92.0 | 46.0 | 0.101633 | 53700 | 0.009009 | 1 |
| 26 | Brent | 51.566621 | -0.319504 | 11.0 | 5.0 | 0.860764 | 61900 | 0.009009 | 1 |
| 27 | Brent | 51.549053 | -0.310892 | 69.0 | 24.0 | 0.457962 | 47800 | 0.009009 | 1 |
| 28 | Brent | 51.546484 | -0.263667 | 62.0 | 19.0 | 0.857033 | 40900 | 0.009009 | 1 |
| 44 | Barnet | 51.614525 | -0.194904 | 70.0 | 23.0 | 0.715983 | 65800 | 0.005814 | 1 |
| 45 | Barnet | 51.644459 | -0.175013 | 44.0 | 21.0 | 0.474490 | 61700 | 0.005814 | 1 |
| 46 | Barnet | 51.617341 | -0.172074 | 37.0 | 16.0 | 0.923050 | 52200 | 0.005814 | 1 |
| 57 | Southwark | 51.498460 | -0.105932 | 26.0 | 14.0 | 0.410530 | 48300 | 0.003333 | 1 |
| 64 | Haringey | 51.591314 | -0.120335 | 43.0 | 16.0 | 0.734368 | 53100 | 0.005405 | 1 |
| 65 | Haringey | 51.577388 | -0.082158 | 77.0 | 28.0 | 0.498445 | 44400 | 0.005405 | 1 |
| 75 | Merton | 51.422477 | -0.224531 | 19.0 | 8.0 | 1.247450 | 67000 | 0.006211 | 1 |
| 76 | Merton | 51.411943 | -0.173447 | 85.0 | 32.0 | 0.805762 | 52500 | 0.006211 | 1 |
| 79 | Camden | 51.518421 | -0.132706 | 1.0 | 1.0 | 0.269255 | 46800 | 0.004098 | 1 |
| 81 | Camden | 51.537440 | -0.135533 | 24.0 | 11.0 | 0.407740 | 48500 | 0.004098 | 1 |
| 82 | Camden | 51.551023 | -0.139781 | 23.0 | 10.0 | 0.071352 | 57000 | 0.004098 | 1 |
| 83 | Camden | 51.556426 | -0.148735 | 34.0 | 15.0 | 0.185730 | 52900 | 0.004098 | 1 |
| 98 | Hammersmith and Fulham | 51.491253 | -0.207484 | 6.0 | 5.0 | 0.154938 | 56900 | 0.005000 | 1 |
| 99 | Hammersmith and Fulham | 51.473862 | -0.209332 | 39.0 | 11.0 | 0.581015 | 69400 | 0.005000 | 1 |

# Cluster description

- Cluster 2:
- Mid spending power
- Low presence of competition
- Medium/high distance to the station

| | Neighbourhood | Latitude | Longitude | Population | Households | Distance to station | Average Income | Brazilian Restaurant | Cluster |
|---|---|---|---|---|---|---|---|---|---|
| 14 | Bromley | 51.362141 | 0.090192 | 73.0 | 38.0 | 1.241690 | 61600 | 0.000000 | 2 |
| 19 | Lambeth | 51.449614 | -0.139185 | 83.0 | 29.0 | 0.697894 | 49700 | 0.003367 | 2 |
| 20 | Lambeth | 51.429065 | -0.088232 | 102.0 | 36.0 | 0.594626 | 50100 | 0.003367 | 2 |
| 38 | Kensington and Chelsea | 51.487289 | -0.188915 | 109.0 | 81.0 | 0.458343 | 57100 | 0.000000 | 2 |
| 50 | Wandsworth | 51.428941 | -0.172968 | 148.0 | 53.0 | 0.363703 | 51300 | 0.001538 | 2 |
| 52 | Wandsworth | 51.471192 | -0.150287 | 77.0 | 36.0 | 0.465168 | 54100 | 0.001538 | 2 |
| 56 | Southwark | 51.478168 | -0.095108 | 127.0 | 45.0 | 1.143710 | 44000 | 0.003333 | 2 |
| 61 | Sutton | 51.357886 | -0.139529 | 117.0 | 38.0 | 0.832058 | 44600 | 0.000000 | 2 |
| 67 | Lewisham | 51.429587 | 0.036631 | 144.0 | 63.0 | 0.976703 | 44200 | 0.000000 | 2 |
| 68 | Lewisham | 51.450695 | -0.012703 | 173.0 | 66.0 | 0.756707 | 55700 | 0.000000 | 2 |
| 78 | Ealing | 51.518183 | -0.373996 | 108.0 | 27.0 | 1.395690 | 47700 | 0.000000 | 2 |
| 80 | Camden | 51.550415 | -0.204611 | 186.0 | 105.0 | 0.385689 | 51400 | 0.004098 | 2 |
| 88 | Greenwich | 51.489310 | 0.038067 | 146.0 | 56.0 | 0.546876 | 52000 | 0.000000 | 2 |
| 90 | Croydon | 51.398141 | -0.067555 | 188.0 | 85.0 | 0.509361 | 45200 | 0.000000 | 2 |
| 91 | Kingston upon Thames | 51.378398 | -0.277722 | 173.0 | 70.0 | 0.209291 | 55400 | 0.000000 | 2 |
| 97 | Tower Hamlets | 51.530888 | -0.063582 | 118.0 | 50.0 | 0.452014 | 55100 | 0.000000 | 2 |

# Cluster description

- Cluster 3:
- Mid/high spending power
- Low presence of competition
- Medium distance to the station

| | Neighbourhood | Latitude | Longitude | Population | Households | Distance to station | Average Income | Brazilian Restaurant | Cluster |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Westminster | 51.496019 | -0.137685 | 46.0 | 27.0 | 0.433721 | 57500 | 0.001250 | 3 |
| 2 | Westminster | 51.522106 | -0.149017 | 3.0 | 1.0 | 0.316775 | 73500 | 0.001250 | 3 |
| 3 | Westminster | 51.513050 | -0.131427 | 5.0 | 3.0 | 0.294475 | 54000 | 0.001250 | 3 |
| 4 | Westminster | 51.535685 | -0.171829 | 6.0 | 2.0 | 0.180020 | 59900 | 0.001250 | 3 |
| 5 | Westminster | 51.512436 | -0.113124 | 4.0 | 1.0 | 0.160563 | 59000 | 0.001250 | 3 |
| 6 | Westminster | 51.497664 | -0.151261 | 32.0 | 17.0 | 0.526160 | 56300 | 0.001250 | 3 |
| 7 | Westminster | 51.489636 | -0.136115 | 9.0 | 6.0 | 0.208241 | 61900 | 0.001250 | 3 |
| 11 | Bromley | 51.406526 | 0.007654 | 36.0 | 17.0 | 0.353919 | 63100 | 0.000000 | 3 |
| 12 | Bromley | 51.403588 | 0.036137 | 55.0 | 21.0 | 0.669057 | 62900 | 0.000000 | 3 |
| 13 | Bromley | 51.408445 | 0.056699 | 26.0 | 11.0 | 0.321041 | 72000 | 0.000000 | 3 |
| 15 | Bromley | 51.392147 | -0.040713 | 7.0 | 6.0 | 0.913357 | 58300 | 0.000000 | 3 |
| 16 | Bromley | 51.409708 | -0.055161 | 52.0 | 20.0 | 0.511303 | 55700 | 0.000000 | 3 |
| 17 | Hackney | 51.528557 | -0.089805 | 11.0 | 10.0 | 0.315468 | 54300 | 0.000000 | 3 |
| 29 | Richmond upon Thames | 51.423487 | -0.360588 | 65.0 | 25.0 | 1.158880 | 62200 | 0.000000 | 3 |
| 30 | Richmond upon Thames | 51.423439 | -0.354513 | 35.0 | 16.0 | 1.217460 | 62200 | 0.000000 | 3 |
| 31 | Richmond upon Thames | 51.477996 | -0.239360 | 20.0 | 8.0 | 1.104870 | 65600 | 0.000000 | 3 |
| 36 | Kensington and Chelsea | 51.497809 | -0.188800 | 52.0 | 18.0 | 0.376699 | 73900 | 0.000000 | 3 |
| 37 | Kensington and Chelsea | 51.508964 | -0.204267 | 2.0 | 1.0 | 0.256720 | 61700 | 0.000000 | 3 |
| 39 | Kensington and Chelsea | 51.491967 | -0.194232 | 13.0 | 8.0 | 0.068090 | 49800 | 0.000000 | 3 |
| 41 | Enfield | 51.634244 | -0.107997 | 38.0 | 16.0 | 0.489925 | 62600 | 0.000000 | 3 |
| 49 | Wandsworth | 51.456981 | -0.231109 | 80.0 | 25.0 | 1.096980 | 68500 | 0.001538 | 3 |
| 51 | Wandsworth | 51.459925 | -0.163482 | 71.0 | 25.0 | 0.668297 | 80600 | 0.001538 | 3 |
| 53 | Wandsworth | 51.448371 | -0.153267 | 29.0 | 16.0 | 0.574968 | 71800 | 0.001538 | 3 |
| 54 | Wandsworth | 51.449184 | -0.192121 | 32.0 | 12.0 | 0.819897 | 58200 | 0.001538 | 3 |
| 55 | Wandsworth | 51.439621 | -0.201794 | 69.0 | 31.0 | 0.588494 | 76700 | 0.001538 | 3 |
| 58 | Southwark | 51.498154 | -0.075531 | 38.0 | 16.0 | 0.775141 | 74200 | 0.003333 | 3 |
| 59 | Harrow | 51.571271 | -0.339808 | 35.0 | 13.0 | 0.897249 | 62900 | 0.000000 | 3 |
| 60 | Hounslow | 51.473786 | -0.341075 | 97.0 | 35.0 | 0.294194 | 51600 | 0.000000 | 3 |
| 62 | Hillingdon | 51.572291 | -0.402240 | 61.0 | 20.0 | 0.603354 | 58100 | 0.000000 | 3 |
| 69 | Lewisham | 51.450751 | 0.014012 | 74.0 | 35.0 | 0.116572 | 63700 | 0.000000 | 3 |
| 77 | Ealing | 51.517266 | -0.320989 | 35.0 | 11.0 | 0.413129 | 64800 | 0.000000 | 3 |
| 84 | Islington | 51.523194 | -0.107313 | 8.0 | 5.0 | 0.366736 | 53800 | 0.000000 | 3 |
| 85 | Islington | 51.537215 | -0.102420 | 22.0 | 15.0 | 0.520079 | 53300 | 0.000000 | 3 |

# Cluster description

- Cluster 4:
- Low spending power
- No presence of competition
- Highest distance to the station

| | Neighbourhood | Latitude | Longitude | Population | Households | Distance to station | Average Income | Brazilian Restaurant | Cluster |
|---|---|---|---|---|---|---|---|---|---|
| 63 | Hillingdon | 51.533017 | -0.479713 | 1129.0 | 0.0 | 1.50396 | 45800 | 0.0 | 4 |

# CONCLUSION

- The most promising group of neighborhoods for opening an Brazilian Restaurant, with a niche in Brazilian cuisine, appears to be 'Cluster Label 3'.
- The medium high spending power of the neighborhoods allows them to readily afford the slightly upscaled prices of the client's Brazilian restaurant menu.
- The number of competitors is pretty low, that would definitely help build a brand in the area for people curious to try Brazilian cuisine.
- Our client could more specifically consider Richmond and some areas of Kensington and Chelsea characterized by a medium high spending average.
- However cluster 3 shows a medium distance to the station and that probably could be considered as an insight to activate a solid promotion to the locals to make sure there is a significant and remunerative client base to limit any potential issue of the less easy to reach of the restaurant from people of other areas of London.
- In conclusion, the extensive analysis above would greatly increase the likelihood of the restaurant's success.