

London Boroughs Battle

Contents

Business problem	1
Data sources and preparation	1
Boroughs clustering	5
Conclusions	7

Business problem

London is the capital and largest city of England and the United Kingdom. London is considered to be one of the world's most important global cities and has been called the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment-friendly, and most-popular-for-work city (<https://en.wikipedia.org/wiki/London>)

BestIndSpice, our customer, is a start-up company which main activity is import of **Indian spices**. They intend to open a **brick and mortar shop** in London and asked us to evaluate the London boroughs and **identify the borough with highest potential for this business**. They target British Indian households from London that use to cook at home.

Analytical approach

We considered that one possible approach to **identify the borough with highest potential** for opening an Indian Spice store might be London boroughs clustering using relevant features:

- number of British Indians living in London and their share in total population (higher number/share in a borough - > higher the potential)
- number of Indian restaurants (higher number -> possible lower potential as population has more option to going out)
- number of specialized spices stores (higher number of competitors-> possible lower potential)

Data sources and preparation

We used the following data sources :

- location data about London boroughs; London has 32 borough and we downloaded the list and geographic location from https://en.wikipedia.org/wiki/List_of_London_boroughs; Dataframe df_london_borough was created.

	borough	longitude	latitude
0	Barking and Dagenham	0.1557	51.5607
1	Barnet	-0.1517	51.6252
2	Bexley	0.1505	51.4549
3	Brent	-0.2817	51.5588
4	Bromley	0.0198	51.4039

Figure 1. Top 5 records from df_london_borough

- number of British Indians living in London the share in total population; London population was about 8.2 millions according to latest census from 2011. Around 6.6% of London's population is of Indian origin, largest ethnic minority group (more than 540 thousands in 2011). We used population forecast for 2025 for British Indians and All persons (source: Ethnic group projections (2016-based central trend) from <https://data.london.gov.uk/dataset/ethnic-group-population-projections>)

	borough	Indian_pop_fcast_2025		borough	London_all_pop_fcast_2025
0	City of London	261	0	City of London	7556
1	Barking and Dagenham	15102	1	Barking and Dagenham	237099

Figure 2 London population forecast by borough (2 examples)

- number of Indian restaurants in each borough. (source: Fousquare API)
- number of specialized spices stores (source: Fousquare API)

This information was obtained using Fousquare API. We used a function to get the venues for a specific category.

```

def getNearbyVenuesByCategId(names, latitudes, longitudes, radius, VERSION, LIMIT, categoryId, intent):
    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={}&v={}&categoryId={}&intent={}&radius={}&limit={}'.format(
            CLIENT_ID, CLIENT_SECRET, lat, lng, VERSION, categoryId, intent, radius, LIMIT)
        # print(url)
        # make the GET request
        # print(requests.get(url).json())
        results = requests.get(url).json()['response']['venues']

        # return only relevant information for each venue

        a = [(name,
              lat,
              lng,
              v['id'],
              v['name'],
              v['location']['lat'],
              v['location']['lng'],
              v['location']['distance'],
              checkKey(v['location'], 'city'),
              v['location']['country'],
              checkCategories(v['categories'], 'name'),
              checkCategories(v['categories'], 'shortName'),
              checkCategories(v['categories'], 'id'))
              for v in results]
        # print(a)
        venues_list.append(a)

```

Figure 3 Function used to get venues info by category from Foursquare API

Using <https://developer.foursquare.com/docs/build-with-foursquare/categories/> we found that Indian Restaurant has CategoryId 4bf58dd8d48988d10f941735 and Herbs & Spices Store has CategoryId 52f2ab2ebcbc57f1066b8b2c.

```

df_london_indian_rest = getNearbyVenuesByCategId(names =df_london_borough['borough'],
                                                  latitudes=df_london_borough['latitude'],
                                                  longitudes=df_london_borough['longitude'],
                                                  VERSION='20200331',
                                                  radius =1500,
                                                  LIMIT=100,
                                                  categoryId='4bf58dd8d48988d10f941735', intent='browse')

```

	borough	Borough Latitude	Borough Longitude	Venue Id	Venue Name	Venue Latitude	Venue Longitude	Venue Distance	Venue City	Venue Country	Venue Category
0	Barking and Dagenham	51.5607	0.1557	4f3561afe4b0debe1f187692	Saffron Restaurant	51.565345	0.171533	1211	-NA-	United Kingdom	Indian Restaurant
1	Barking and Dagenham	51.5607	0.1557	52cdc68311d23a96916b2892	Tarka	51.552033	0.162370	1069	Dagenham	United Kingdom	Indian Restaurant
2	Barking and Dagenham	51.5607	0.1557	4bfd89eff7c82d7f4b558e04	Shirins Tandoori	51.572158	0.166567	1480	London	United Kingdom	Indian Restaurant
3	Barking and Dagenham	51.5607	0.1557	5638ff6060b2753ece75b57a	Clay Oven Indian Takeaway	51.563209	0.146845	673	-NA-	United Kingdom	Indian Restaurant
4	Barnet	51.6252	-0.1517	4b4cebf1f964a5204bc426e3	Boulevard Club	51.614627	-0.149623	1185	London	United Kingdom	Indian Restaurant

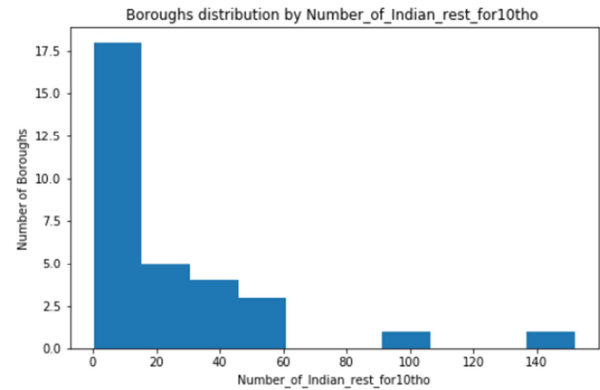
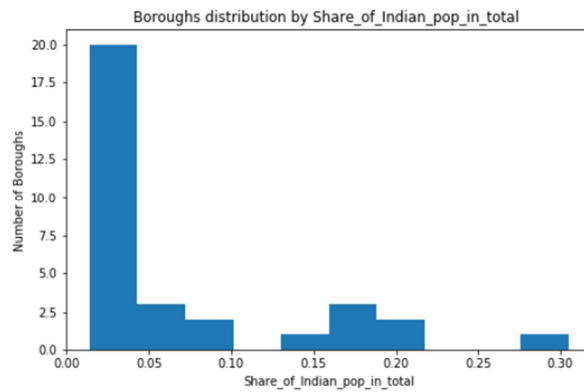
Figure 4 Indian restaurants sample returned by Foursquare API

Then we aggregated the data at borough level:

	borough	Number_of_Indian_restaurants
0	Barking and Dagenham	4
1	Barnet	7
2	Bexley	8
3	Brent	28
4	Bromley	15

All sources were centralized in df_London_Cum_Info . Missing values for number_of_spices_stores were replaced by 0.

New feature was created: Number_of_Indian_rest_for10tho (number of Indians restaurants on 10 thousand British Indians).



Boroughs clustering

There are many models for clustering. For current analysis we chose **k-means algorithm** which despite its simplicity is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data (Coursera notebook ML0101EN-Clus-K-Means-Customer-Seg-py-v1)

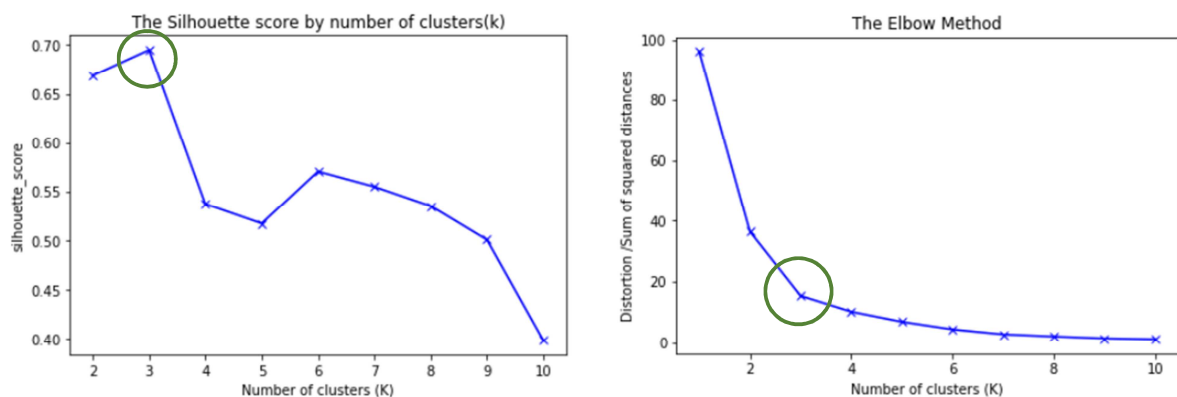
For simplicity we used for clustering only three features:

- Indian_pop_fcast_2025
- Share_of_Indian_pop_in_total
- Number_of_Indian_rest_for10tho

The features were standardized using StandardScaler()

A key parameter for k-means algorithm is number of clusters. Two methods were considered to determine optimal k: Silhouette score and Elbow method. “The optimal number of clusters is somehow subjective and depends on the method used for measuring similarities and the parameters used for partitioning “

(<https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/>)






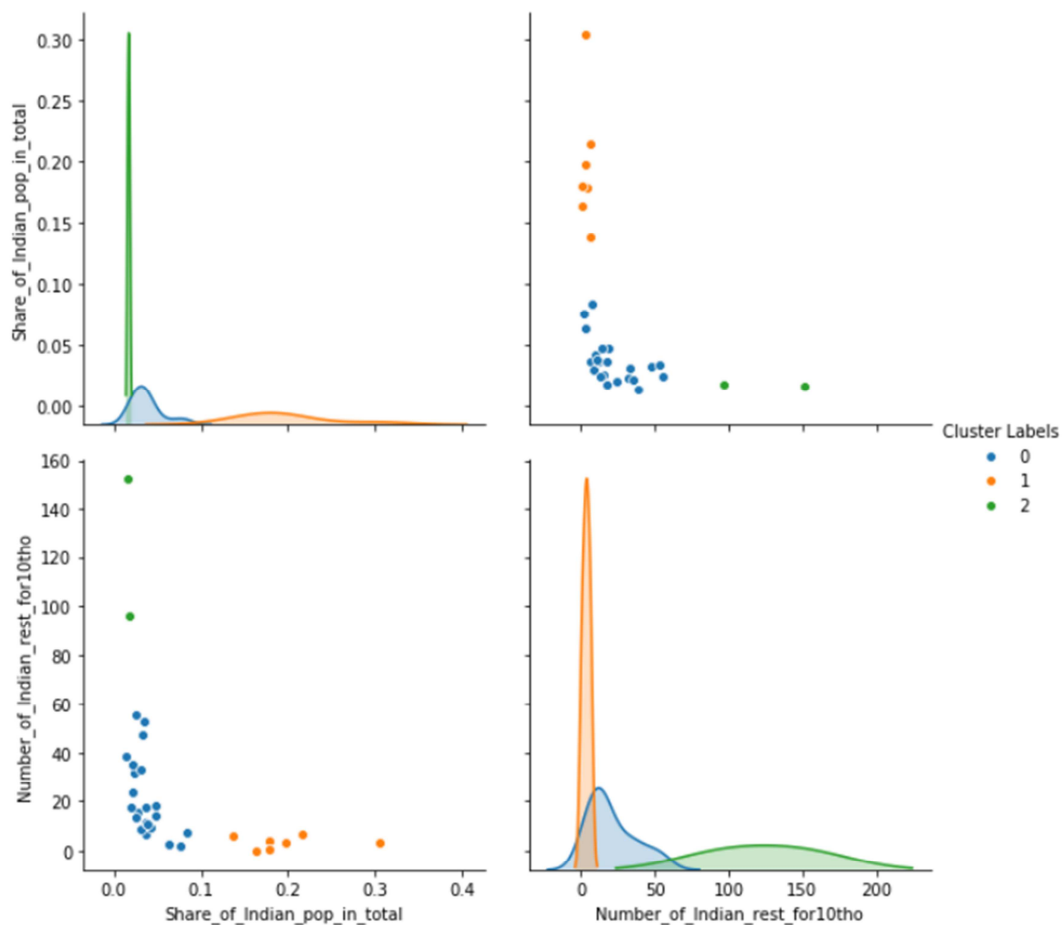
Both methods indicated k=3 as “optimal” so the final solution was run with 3

```
df_London_Cum_Info['Cluster Labels'].value_counts()
```

```
0    23
1     7
2     2
Name: Cluster Labels, dtype: int64
```

Figure 5 Number of Borough in each cluster

-  **Cluster0** with 23 boroughs with **medium share** of British Indian population and # of Indian restaurants/10thou pop
-  **Cluster1** with 7 boroughs with **high share** of British Indian population and **low** # of Indian restaurants /10thou pop
-  **Cluster2** with only 2 boroughs quite central, very **low share of British Indian** population and **high number of Indian restaurant** /10thou pop



Cluster1 seems to be the one with the highest potential for our customer. The main two competitors (The Spice Shop and Spice Mountain) are located in a more central area.

Geographically the cluster has two main areas: NW with 5 borough and E with 2 boroughs.

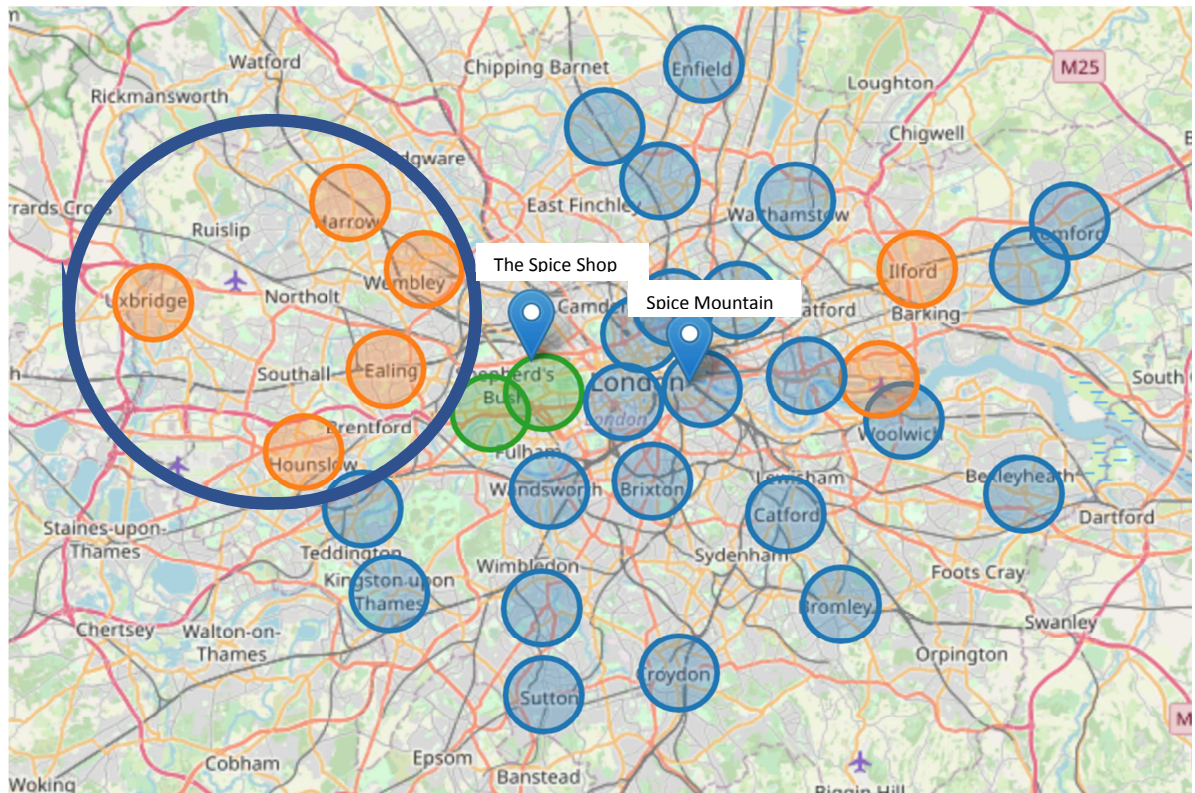


Figure 6 London Borough grouped by their potential for Indian Spice store

Conclusions

Our customer, BestIndSpice, asked us to evaluate the London boroughs and **identify the boroughs with highest potential for an Indian Spice store.**

As they target British Indian households from London that use to cook at home, we considered the best boroughs being that with high number and share of British Indians and low number of Indian restaurants.

We used k-means to explore and cluster London boroughs and identified a cluster with 7 boroughs with the highest potential. This cluster is spread on to geographical areas: NW with 5 borough and E with 2 boroughs

We recommend focusing on NW area where 5 boroughs with high potential are concentrated.