Best place for opening a lunch restaurant in London

November 11, 2020

Dor Lavy

1. Introduction – Business Problem

Where should I open my restaurant in London

This project is addressed to people who wants to open a restaurant.

If you want to open a restaurant in London it's not enough just to know how to cook. One of the first questions that pops into your mind is "Where would I open it?"

This question is depending on various variables, such as:

1. what food are you cooking?
2. is it an evening restaurant or maybe you aim it for lunch time?
3. will it be an expensive restaurant/fast food/ ethnic food?

In order to succeed you have to successfully address these questions before looking for a place. We are aiming our project for fast food restaurant that aims for lunch time/early afternoon hours.

2. Data Understanding

2.1. We will use web scraping from Wikipedia in order to get all the neighborhoods in London. we will use the following columns from the table: Location,London borough,Post town, Postcode district. Dial code and OS grid ref will be remove during data cleaning since this information has no added value for our goal. The Wikipedia page can be found in https://en.wikipedia.org/wiki/List_of_areas_of_London

2.2. Will add the Latitude and Longitude for each neighborhood using free CSV dataset from www.doogal.co.uk : https://www.doogal.co.uk/london_postcodes.php

2.3. Visualize London map including the neighborhood data frame we described using folium library (same as we did during the course for New York & Toronto)

2.4. Since we want to open our restaurant during lunch time we need to aim for neighborhoods with many schools/workplaces. People are staying there during the day and look for quick lunch place. In order to do so we will use Foursquare API to get the number of workplaces and schools in every neighborhood and summarize it in order to compare between neighborhoods.

2.5. Will recommend the best neighborhood based on the data above

3. Data Cleansing

3.1. We have issues in the 2nd, 4th & 5th columns since they are including '\xa0' instead of space. Therefore we will fix the column name

3.2. Remove "Dial code" & "OS grid ref" columns since we don't need them for our project. Borough & Postcode district information will be sufficient.

3.3. As you can see in the top 4 rows, there is "[x]" sign. This is due to Wikipedia reference note. We don't need it, therefore will remove it.

3.4. check for "Not assigned" / empty cells.

3.5. Will use only 'Post_town' = London for our search

4. Methodology

4.1. London Neighborhoods Data

4.1.1. Will use pandas web scraping to import London neighborhoods information from Wikipedia to pandas dataframe. We received the following dataframe:

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

4.1.2. You can see above that in some Postcode_district we have multiple values. Therefore we must split it, since each Postcode_district refer to a different map coordinates.
Will use chain library from itertools for splitting and duplicate the rows.
Will also use underscore ('_') instead of space ( ) for removing duplicate values.
Result will be as following:

	Location	Post_town	London_borough	Postcode
0	Abbey Wood	LONDON	Bexley, Greenwich	SE2
1	Acton	LONDON	Ealing, Hammersmith and Fulham	W3
1	Acton	LONDON	Ealing, Hammersmith and Fulham	W4
6	Aldgate	LONDON	City	EC3
7	Aldwych	LONDON	Westminster	WC2

4.2. London Neighborhoods Coordinates

4.2.1. Adding Latitude and Longitude for each neighborhood using doogal dataset. The CSV file contains much more specific information than we need. therefore we will remove the sub neighborhoods codes and will keep only the neighborhoods Long/Lat (i.e. AB1 XXX -> AB1)

Will use only the following columns: Latitude, Longitude & Postal code

	Postcode	Latitude	Longitude
0	BR1 1AA	51.401546	0.015415
1	BR1 1AB	51.406333	0.015208
2	BR1 1AD	51.400057	0.016715
3	BR1 1AE	51.404543	0.014195
4	BR1 1AF	51.401392	0.014948

4.2.2. Will remove the sub neighborhoods codes and will keep only the neighborhoods Long/Lat (i.e. AB1 XXX -> AB1)

|	Postcode	Latitude	Longitude
0	BR1	51.401546	0.015415
1	BR2	51.407775	-0.003321
2	BR3	51.407682	-0.032385
3	BR4	51.380973	-0.011615
4	BR5	51.386370	0.064750

4.2.3. Merging neighborhoods data with the coordinates will give the final data frame:

	Location	Post_town	London_borough	Postcode	Latitude	Longitude
0	Abbey Wood	LONDON	Bexley, Greenwich	SE2	51.481603	0.122712
1	Acton	LONDON	Ealing, Hammersmith and Fulham	W3	51.519838	-0.268414
2	Acton	LONDON	Ealing, Hammersmith and Fulham	W4	51.501346	-0.254217
3	Anerley	LONDON	Bromley	SE20	51.409107	-0.059069
4	Angel	LONDON	Islington	N1	51.539442	-0.117873

The data frame above contains the Postcodes (without duplication), Neighborhood names and their Latitude/Longitude

4.4 London Neighborhoods Visualization - Folium

Using Folium library we will visualize all the London neighborhoods including their Postcodes on map
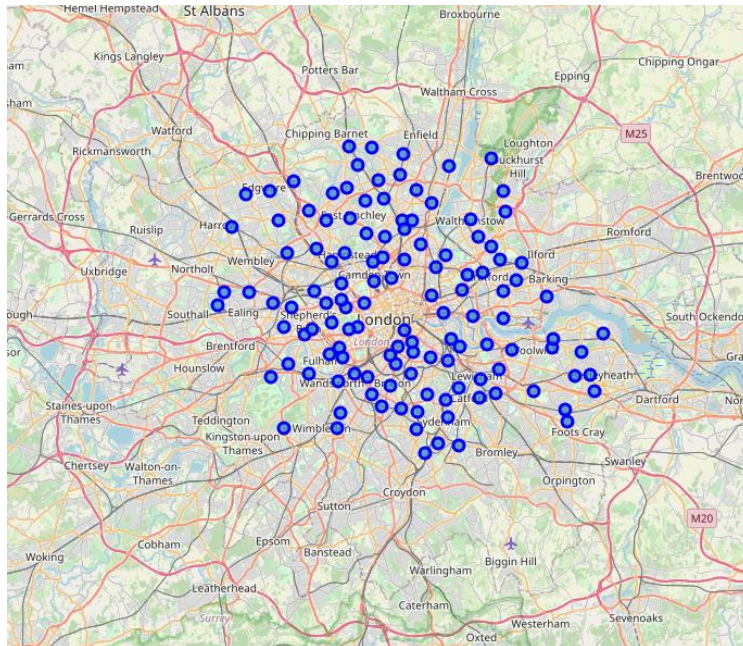
4.5 London Restaurants Query – Foursquare API

Will get the restaurant in each neighborhood using foursquare API with the following information:

Category ID: 4bf58dd8d48988d121941735 (restaurants)

Limit: 70 (runtime and API limitations)

Radius: 1000 (in meters)

We will put the information we get in a data frame as follow:

	Postcode	London_borough_Latitude	London_borough_Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	W3	51.519838	-0.268414	Bamboo Lounge	51.528255	-0.271092	Hookah Bar
1	W4	51.501346	-0.254217	High Road House	51.492901	-0.254711	Lounge
2	N1	51.539442	-0.117873	The Lexington	51.531669	-0.111359	Rock Club
3	N1	51.539442	-0.117873	Simmons Bar	51.531897	-0.120764	Cocktail Bar
4	N1	51.539442	-0.117873	The Parcel Yard	51.532374	-0.123933	Pub
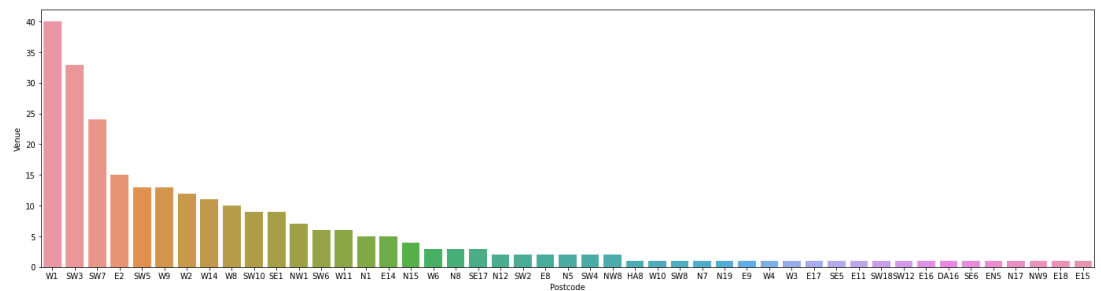
4.6 London Potential Clients Query – Foursquare API

We need to check how many workplaces we have for each postcode.

Will do the same process we did for the restaurants but this time to locate potential clients.

After examining the Foursquare API - i decided that the most suitable clients for our lunch time restaurant would be college students and people going for lunch break during work.

Will get the following information using:

Category ID:

1. 4d4b7105d754a06372d81259 (College and University)
2. 4d4b7105d754a06375d81259 (Professional and other places (Government places, office etc..)

Limit: 70 (runtime and API limitations)

Radius: 1000 (in meters)

We will put the information we get in a data frame as follow:

|	Postcode	London_borough_Latitude	London_borough_Longitude	Client	Client_Latitude	Client_Longitude	Client_Category
0	SE2	51.481603	0.122712	J & D Pallet Services Ltd	51.477697	0.123973	Office
1	SE2	51.481603	0.122712	The Belvedere Clinic	51.483525	0.121878	Doctor's Office
2	SE2	51.481603	0.122712	Passing You Driving School	51.479510	0.127379	Driving School
3	SE2	51.481603	0.122712	Home Security Surveys UK	51.476251	0.117953	Police Station
4	W3	51.519838	-0.268414	Dixons Carphone HQ	51.521592	-0.260953	Office
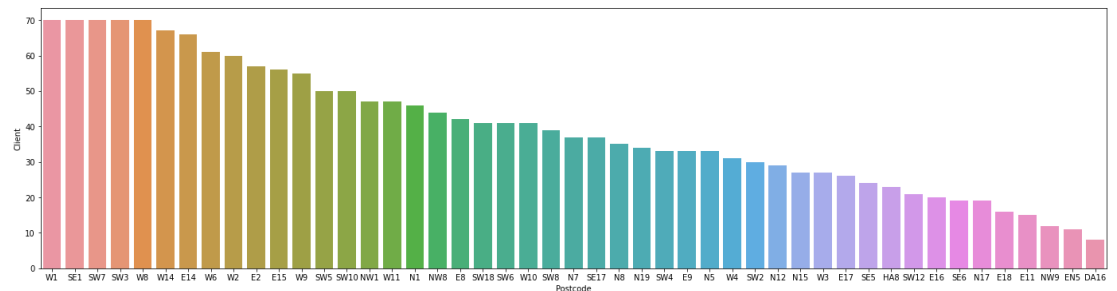
## 4.7 Calculating the Clients/Restaurants ratio and Visualizing

4.7.1 Will use the 2 dataframes (Restaurants & Clients) to create one dataframe the calculates the ratio between clients to restaurants for every postcode (neighborhood).

	Postcode	Venue	Client	ratio	London_borough	Latitude	Longitude
0	E15	1	56	56.0	Waltham Forest	51.544029	-0.002063
1	SW18	1	41	41.0	Wandsworth	51.458681	-0.185747
2	W10	1	41	41.0	Kensington and Chelsea	51.529068	-0.215875
3	SW8	1	39	39.0	Wandsworth	51.478732	-0.118861
4	N7	1	37	37.0	Islington	51.555353	-0.128727

4.7.2 Visualize on bar chart the ratio per postcode
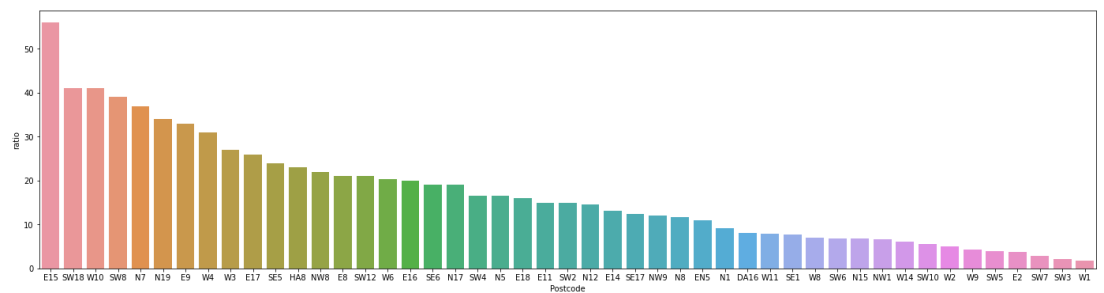
4.7.3 In order to understand the area behind the postcode – Let's mark the postcodes with top 5 ratio on London map. These neighborhoods are the most suitable for new restaurant during lunch time.
We choose top 5 in order to give the customer alternative in case the top neighborhood is not possible for some reason.
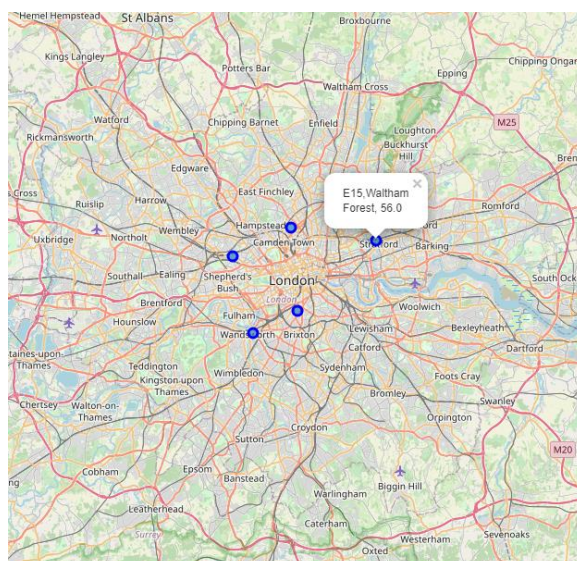
5. Results

As we described in sections 4.7.2 & 4.7.3 we found the most suitable neighborhoods for opening a restaurant by diving the number of clients by the number of competitive restaurants.

6. Discussion

We found that the best neighborhood will be E15, Walhalm. If the customer will look for alternatives he can check the top 5 other neighborhoods we visualize on the London map. This might happen since there are more variables we didn't take into account in this project. As discussed in the introduction, restaurant owner might reject neighborhood for other reasons such as crime rates, real estate price etc. Therefore we are not only recommending but giving the full picture for the client to choose based on our research.

7. Conclusion

During this project we were able to use datasets from Wikipedia & doogal.co.uk with the Foursquare API in order to address our business problem – where should we open a restaurant for lunch time in London. During the process we defined which datasets we need, cleaned it and visualize it. At the end, we were able to recommend a specific restaurant for our client.

8. Reference:
    8.1. Wikipedia (London neighborhoods list):
        https://en.wikipedia.org/wiki/List_of_areas_of_London
    8.2. Doogal.co.uk (London neighborhoods coordinates):
        https://www.doogal.co.uk/london_postcodes.php
    8.3. Foursquare API – Retrieving the Restaurants and potential clients around the neighborhoods.