Market analysis for a Japanese sushi restaurant company: should they open a restaurant in London or in Madrid?

Introduction

Background

An important Japanese sushi restaurant company decides to start its activities in Europe, since they have plenty of restaurants only in Japan. The owner is considered a visionary, he is fascinated by the idea of **opening a restaurant in an important European capital, but he is not sure which this should be**. In fact, northern and southern European citizens can be considered quite different in their respective ways of living the city. For this reason, **the company asks to compare two very different cities and, more in general, approaches to lifestyle: London and Madrid**. Both cities could be good to expand this business, but the company wants to find the ideal one. To complete this task you have **no specifics requirements besides one**: since the company is very proud of the standard in the quality of food and, more in general, of the "experience" to offer to the consumer, you have to consider that this restaurant is not particularly cheap, so **your analysis should focus only on richest areas of the cities**.

Problem

In order to find the ideal city to open a sushi restaurant, you decide to **focus your research on discovering the most common venues only in the top 5 areas of the cities**, where it is reasonable to think that the life style is more expensive. According to a previous research made by the company, you can start from the considerations reported below.

The **top 5 richest boroughs in London** are:

- Camden;
- Hackney;
- Hammersmith and Fulham;
- Kensington and Chelsea;
- Westminster.

And **top 5 richest boroughs in Madrid** are:

- Centro;
- Chamrtin;
- Chamberi;
- Retiro;
- Salamanca.

Methodology

Coding

Using python to develop the entire model. Different packages were used:

- **bs4**: for web scraping;
- **folium**: to generate maps;
- **geopy**: to convert an address into latitude and longitude values;
- **matplotlib**: to detail maps and eventually plot graphs;
- **numpy**: to exploit some of its mathematical methods;
- **pandas**: to create and manipulate databases;
- **sklearn**: to create the clusters;
- **requests**: to manage http requests.

Data collection

Info about Boroughs in London and Madrid

To find out info about boroughs in London and Madrid, it is sufficient to scrape from these two sources:

- **London**: https://en.wikipedia.org/wiki/List_of_London_boroughs;
- **Madrid**: https://en.wikipedia.org/wiki/Districts_of_Madrid.

After scraping this information, DataFrames were created by pandas, however these had to be "cleaned": GIGO (Garbage in, garbage out) was performed since useless columns and missing/poorly formatted data were reported, as noticeable in Figure 1 and Figure 2.

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2019 est)[1]	Co-ordinates	Nr. in map
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...	25
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	31
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	23
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	12
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	20

Figure 1: DataFrame for London before GIGO

	District Number	Name	District area[n 1] (Ha.)	Population	Population density(Hab./Ha.)	Location	Administrative wards
0	1.0	Centro	522.82	131928	252.34	NaN	Palacio (11)Embajadores (12)Cortes (13)Justici...
1	2.0	Arganzuela	646.22	151965	235.16	NaN	Imperial (21)Acacias (22)Chopera (23)Legazpi (...
2	3.0	Retiro	546.62	118516	216.82	NaN	Pacífico (31)Adelfas (32)Estrella (33)Ibiza (3...
3	4.0	Salamanca	539.24	143800	266.67	NaN	Recoletos (41)Goya (42)Fuente del Berro (43)Gu...
4	5.0	Chamartín	917.55	143424	156.31	NaN	El Viso (51)Prosperidad (52)Ciudad Jardín (53)...

Figure 2: DataFrame for Madrid before GIGO

Hence, the CIGO operation consisted in eliminating *NaNs* and dropping all the columns not useful to the purpose of this analysis, such as "Local Authority" for London and "Administrative wards" for Madrid. After the cleaning, here is how the two Dataframes looked like (see Figure 3):

	Borough	Area (sq mi)	Population (2019)
0	Barking and Dagenham	13.93	212906
1	Barnet	33.49	395896
2	Bexley	23.38	248287
3	Brent	16.70	329771
4	Bromley	57.97	332336
5	Camden	8.40	270029
6	Croydon	33.41	386710
7	Ealing	21.44	341806
8	Enfield	31.74	333794
9	Greenwich	18.28	287942
10	Hackney	7.36	281120
11	Hammersmith and Fulham	6.33	185143
12	Haringey	11.42	268647
13	Harrow	19.49	251160
14	Havering	43.35	259552
15	Hillingdon	44.67	306870
16	Hounslow	21.61	271523
17	Islington	5.74	242467
18	Kensington and Chelsea	4.68	156129
19	Kingston upon Thames	14.38	177507
20	Lambeth	10.36	326034
21	Lewisham	13.57	305842
22	Merton	14.52	206548
23	Newham	13.98	353134
24	Redbridge	21.78	305222
25	Richmond upon Thames	22.17	198019
26	Southwark	11.14	318830
27	Sutton	16.93	206349
28	Tower Hamlets	7.63	324745
29	Waltham Forest	14.99	276983
30	Wandsworth	13.23	329677
31	Westminster	8.29	261317

	Name	Population
0	Centro	131928
1	Arganzuela	151965
2	Retiro	118516
3	Salamanca	143800
4	Chamartín	143424
5	Tetuán	153789
6	Chamberí	137401
7	Fuencarral-El Pardo	238756
8	Moncloa-Aravaca	116903
9	Latina	233808
10	Carabanchel	243998
11	Usera	134791
12	Puente de Vallecas	227595
13	Moratalaz	94197
14	Ciudad Lineal	212529
15	Hortaleza	180462
16	Villaverde	142608
17	Villa de Vallecas	104421
18	Vicálvaro	70051
19	San Blas-Canillejas	154357
20	Barajas	46876

Figure 3: DataFrames of London (left) and Madrid (right) after GIGO

Despite both Dataframes at this point had a more “logical” aspect, there was still space for improvements: since the aim of this analysis is to retrieve information about places and venues, we needed to gather data about latitude and longitude of these areas. A couple of things must be pointed out now: regarding both London and Madrid, we already had data about coordinates, however it was decided to exploit **Foursquare** (see next paragraph) for a more precise analysis. Regarding the “morphology” of the DataFrames, as it was shown in Figure 3, there was a different denomination of one column, namely “Borough” in London and “Name” in Madrid: both columns reports the same information – the name of the area – so it was appropriate to uniform this field.

Info about Venues

To find out about venues and places, **Foursquare** was used. This made possible to retrieve information about places in the cities – London and Madrid in this case but, more in general, from all over then world - and then to incorporate them in the code: this was crucial since the business model will be based on this real-world location data. Here's the info gathered:

- **Name** of the Borough;
- **Latitude** of the Borough;
- **Longitude** of the Borough;
- **Venue**: name of the Venue;
- **Venue Latitude**: latitude of Venue;
- **Venue Longitude**: longitude of Venue;
- **Venue Category**: category of Venue.

The usage of Foursquare in combination with geopy made possible to proceed with the analysis, in fact once obtained coordinates for both cities, here is how the new DataFrames appeared (*Figure 4*):

	Borough	Area (sq mi)	Population (2019)	Latitude	Longitude
0	Barking and Dagenham	13.93	212906	51.554117	0.150504
1	Barnet	33.49	395896	51.653090	-0.200226
2	Bexley	23.38	248287	51.441679	0.150488
3	Brent	16.70	329771	51.563826	-0.275760
4	Bromley	57.97	332336	51.402805	0.014814
5	Camden	8.40	270029	51.542305	-0.139560
6	Croydon	33.41	386710	51.371305	-0.101957
7	Ealing	21.44	341806	51.512655	-0.305195
8	Enfield	31.74	333794	51.652085	-0.081018
9	Greenwich	18.28	287942	51.482084	-0.004542
10	Hackney	7.36	281120	51.543240	-0.049362
11	Hammersmith and Fulham	6.33	185143	51.492038	-0.223640
12	Haringey	11.42	268647	51.601474	-0.111782
13	Harrow	19.49	251160	51.596827	-0.337316
14	Havering	43.35	259552	51.544385	-0.144307
15	Hillingdon	44.67	306870	51.542519	-0.448335
16	Hounslow	21.61	271523	51.468613	-0.361347
17	Islington	5.74	242467	51.538429	-0.099905
18	Kensington and Chelsea	4.68	156129	51.498480	-0.199043
19	Kingston upon Thames	14.38	177507	51.409627	-0.306262
20	Lambeth	10.36	326034	51.501301	-0.117287
21	Lewisham	13.57	305842	51.462432	-0.010133
22	Merton	14.52	206548	51.410870	-0.188097
23	Newham	13.98	353134	51.530000	0.029318
24	Redbridge	21.78	305222	51.576320	0.045410
25	Richmond upon Thames	22.17	198019	51.440553	-0.307639
26	Southwark	11.14	318830	51.502922	-0.103458
27	Sutton	16.93	206349	51.357464	-0.173627
28	Tower Hamlets	7.63	324745	51.520300	0.029300
29	Waltham Forest	14.99	276983	51.598169	-0.017837
30	Wandsworth	13.23	329677	51.457027	-0.193261
31	Westminster	8.29	261317	51.500444	-0.126540

	Borough	Population	Latitude	Longitude
0	Centro	131928	40.417653	-3.707914
1	Arganzuela	151965	40.396954	-3.697289
2	Retiro	118516	40.411150	-3.676057
3	Salamanca	143800	40.427045	-3.680602
4	Chamartín	143424	40.458987	-3.676129
5	Tetuán	153789	40.460578	-3.698281
6	Chamberí	137401	40.436247	-3.703830
7	Fuencarral-El Pardo	238756	40.556346	-3.778591
8	Moncloa-Aravaca	116903	40.439495	-3.744204
9	Latina	233808	40.403532	-3.736152
10	Carabanchel	243998	40.374211	-3.744676
11	Usera	134791	40.383894	-3.706446
12	Puente de Vallecas	227595	40.383553	-3.654535
13	Moratalaz	94197	40.405933	-3.644874
14	Ciudad Lineal	212529	40.448431	-3.650495
15	Hortaleza	180462	40.472549	-3.642552
16	Villaverde	142608	40.345610	-3.695956
17	Villa de Vallecas	104421	40.373958	-3.612163
18	Vicálvaro	70051	40.396584	-3.576622
19	San Blas-Canillejas	154357	40.428919	-3.604002
20	Barajas	46876	40.473318	-3.579845

Figure 4: Final Dataframes obtained for London (left) and Madrid (right)

At this point it is evident how the operation of retrieving coordinates made possible to locate the boroughs in these cities. Another thing to point out is that London's Dataframes had a column, namely "Area (sq mi)", not available in Madrid's one (not available from the source): instead of dropping this information, it was decided to preserve the column to give further information to the client. A similar problem was identified for the column "Population", in fact it was clearly stated that the numbers for London were referring to year 2019 while Madrid did not provide any information about the year but only the amount of people living in a particular borough.

Once identified the location of each borough in the cities, the venues in the top 5 richest areas in the cities were retrieved (*Figure 5*):

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Camden	Pub	Coffee Shop	Café	Burger Joint	Italian Restaurant	Ice Cream Shop	Beer Bar	Vegetarian / Vegan Restaurant	Caribbean Restaurant	Vietnamese Restaurant
10	Hackney	Coffee Shop	Pub	Café	Supermarket	Brewery	Flea Market	Beer Store	Sporting Goods Shop	Boutique	Yoga Studio
11	Hammersmith and Fulham	Café	Pub	Coffee Shop	Hotel	Gym / Fitness Center	Grocery Store	Sandwich Place	Thai Restaurant	Breakfast Spot	Portuguese Restaurant
18	Kensington and Chelsea	Café	Pub	Italian Restaurant	Persian Restaurant	Burger Joint	Clothing Store	Supermarket	Breakfast Spot	Mediterranean Restaurant	Filipino Restaurant
31	Westminster	Coffee Shop	Pub	Sandwich Place	Historic Site	Outdoor Sculpture	Plaza	Café	Monument / Landmark	Hotel	Garden

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Centro	Plaza	Spanish Restaurant	Hotel	Gourmet Shop	Bookstore	Hostel	Tapas Restaurant	Restaurant	Department Store	Mexican Restaurant
4	Chamartín	Restaurant	Spanish Restaurant	Mediterranean Restaurant	Grocery Store	Gym	Tapas Restaurant	Plaza	Supermarket	Cocktail Bar	Bar
5	Chamberí	Spanish Restaurant	Tapas Restaurant	Bar	Café	Restaurant	Theater	Bakery	Plaza	Mediterranean Restaurant	Beer Bar
12	Retiro	Spanish Restaurant	Plaza	Garden	Supermarket	Dog Run	Diner	Jazz Club	Dessert Shop	Board Shop	Pizza Place
13	Salamanca	Restaurant	Spanish Restaurant	Tapas Restaurant	Furniture / Home Store	Italian Restaurant	Burger Joint	Mediterranean Restaurant	Bakery	Ice Cream Shop	Café

Figure 5: Venues in the top 5 richest boroughs in London (up) and Madrid (down)