

Coursera: IBM Data Science Professional Certificate

Course 9: Applied Data Science Capstone

Report of Final Peer-graded Project

**Landing Sites in London:
Recommendation for Emerging
Top Fashion and Jewellery Houses**

Chih-I CHANG

27 Apr 2020

TABLE OF CONTENTS

1. INTRODUCTION: BUSINESS PROBLEM

2. DATA

- 2.1 Demography of London
- 2.2 Geographical Coordinates of London Boroughs' Constituent Areas
- 2.3 Foursquare Location Data

3. METHODOLOGY

- 3.1 Data Selection Criteria
- 3.2 Data Preparation
 - 3.2.1 Demography of London
 - 3.2.2 Geographical coordinates of areas
 - 3.2.3 Foursquare location data
- 3.3 Analysis of Areas
 - 3.3.1 Area overview
 - 3.3.2 Area clustering
 - 3.3.3 Cluster examination

4. RESULTS

5. DISCUSSION

6. CONCLUSION AND OUTLOOK

- 6.1 Conclusion
- 6.2 Limitations and Recommendations

1. INTRODUCTION: BUSINESS PROBLEM

When a brand starts to rise in the industry and wishes to expand its business by reaching new clients, one thing they commonly do is to establish new physical shops in metropolises. In fashion or jewellery industry, the destinations are typically metropolises in Europe or North America, such as London, Paris, Milan and New York.

This project aims to suggest a list of landing sites in London for emerging luxurious fashion and jewellery houses (FJHs) planning to enter the London market as their first European base. These newly-established FJHs share the following characters:

- (1) They define themselves as niche brands in their respective industry.
- (2) Their business model operates on a qualitative production strategy.
- (3) They provide personalised rather than one-fit-all services.
- (4) They target only the first 0.5-1% customers in the market.

Deciding a suitable location to commence the business in a new market is crucially important for these FJHs. Geographically, it has to be easily accessible to their targeted clientele. From a marketing and branding perspective, the choice of such a venue also suggests a brand's image in the market. With the targeted clientele being an extremely demanding, discerning and sophisticated group, the significance of finding a store spot in London that desirably matches the brand's (self-)position cannot be overemphasised for any emerging luxury brand aspiring to build itself into a globally top-tier one.

2. DATA

This project will use the following data:

2.1 Demography of London

These FJHs' clients are mostly successful businesspeople or wives of successful businessmen from 35 to 65 years old. Understanding where in London people from this category live is essential in order for us to narrow down our research scope.

The Greater London area is composed of 32 London boroughs and the City of London. For convenience, all 33 of them will just be referred to as 'boroughs' in this article. A wide range of statistical data of each borough will be looked into, including age, gender, income, education level, etc.

Data in this field will mainly come from the Office for National Statistics (ONS) of the United Kingdom.

2.2 Geographical Coordinates of London Boroughs' Constituent Areas

Each London borough comprises several areas (or districts as a less formal term). Some areas have special historical and cultural connotations, which may have influenced the industrial landscape over London.

Data in this field will mainly be extracted from Wikipedia and using the geopy library.

2.3 Foursquare Location Data

We will use Foursquare location data to analyse each area's features, such as frequent business types in each locality. This will allow us to offer a list of landing sites to recommend to our stakeholders.

3. METHODOLOGY

3.1 Data Selection Criteria

First, we will select the 6 richest boroughs primarily based on residents' weekly income. The corresponding geographical coordinates of these 6 boroughs' component areas will then be collected in order to investigate each individual area's features. In particular, we will determine each area's most common business type(s) and its approximation to other luxury-related sites, such as high-end fashion or jewellery street shops, prestigious department stores, art galleries and concert halls. This step will tremendously help us identify spots that can benefit from economies of agglomeration.

3.2 Data Preparation

3.2.1 Demography of London

We start by collecting the 33 London boroughs' demographical information using the dataset of *Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland* from the ONS for year 2018. This is latest official information available on the Internet. We focus on individual boroughs' population within age group 35-65 years old as this is the age range of the clientele of our stakeholders. The population is divided by gender (female and male).

We also investigated the education level of residents in each borough. Although we do not have access to data regarding education level of population from age group 35-65, we have that of qualifications of working age (16-64) population for our reference, provided by the ONS's study *Qualifications by Economic Activity Status, Borough* in 2018. Percentage of working age population having NVQ4+ qualification (college/university degree level or above) of each borough is the criteria.

Finally, we examine income of residents in each borough using the ONS's 2019 dataset *Earnings by Place of Residence, Borough*. The figures provides are median earnings (gross weekly rate in GBP) of full time workers.

Since we need to find locations near rich people to open shops, income should be the primary factor in our consideration. We therefore select 6 richest boroughs for further exploration, which are City of London, Kensington and Chelsea, Richmond upon Thames, Wandsworth, Islington, and Westminster (in descending order).

Table 1: The 6 richest boroughs in London

Borough	Population	Male Population	Female Population	Percentage of Degree Holders	Weekly Income Median in GBP
City of London	3,561	2,119	1,442	100	1013.4
Kensington and Chelsea	67,374	33,861	33,513	72.2	912.9
Richmond upon Thames	88,474	43,300	45,174	72.9	820.2
Wandsworth	124,617	61,838	62,779	76	815.8
Islington	83,224	43,456	39,768	74	800.8
Westminster	104,205	55,141	49,064	78	799.9

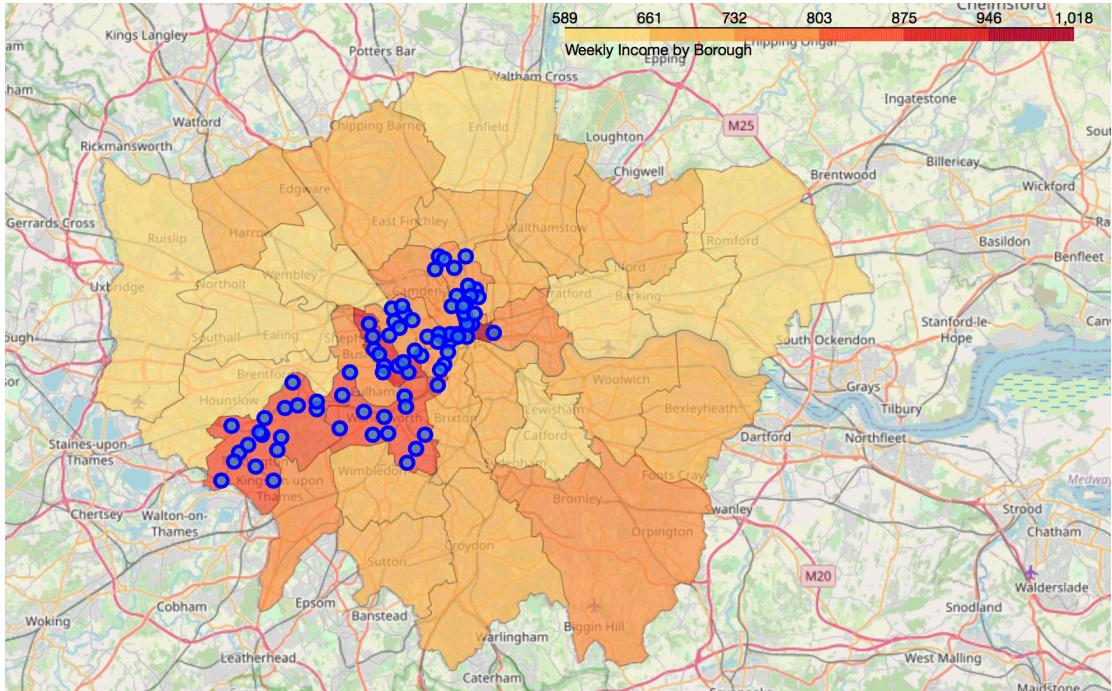
3.2.2 Geographical coordinates of areas

We then use *List of areas of London* from Wikipedia to seize names of the constituent areas under the 6 boroughs in concern. We read the table and clean the data before retrieving the geographical coordinates of a total of 80 areas using the geopy library. The result is compiled in another dataframe showing areas, their corresponding borough names and coordinates, ready for later investigation.

Table 2: Area coordinates (80 rows x 4 columns)

	Area	Borough	Latitude	Longitude
0	Aldgate	City of London	51.5142477	-0.0757186
1	Aldwych	Westminster	51.513103	-0.1149205
2	Angel	Islington	51.5319458	-0.1061056
3	Archway	Islington	51.5654371	-0.1349977
4	Balham	Wandsworth	51.4456449	-0.1503643
...

Before exploring these areas using Foursquare, let's use the folium library to create a map of London that helps us visualise how these areas are spread in the metropolis. The map also gives a glance on the wealth distribution across different boroughs. The information of London's borough border is downloaded from the Internet.



Graph 1: Income-scaled map of London locating the 6 richest boroughs' 80 areas

3.2.3 Foursquare location data

We make Foursquare API requests in order to obtain a list of 100 nearby venues in each of the 80 areas under the 6 boroughs. The searching radius is set to 500m around the areas' geographical coordinates. We collect 3,734 pieces of venue information, comprising venue names, venue coordinates and venue categories. There are 283 unique venue categories in total.

3.3 Analysis of Areas

3.3.1 Area overview

We use one hot coding to compare the times of each venue category appearing in each area and take the mean of the occurrence frequency of each category. The resulting dataframe has 80 rows as there are 80 distinct areas.

Table 3: Venue category frequency in each area (80 rows x 284 columns)

	Area	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	...
0	Aldgate	0.000000	0.0	0.01	0.000000	...
1	Aldwych	0.000000	0.0	0.00	0.015385	...
2	Angel	0.016667	0.0	0.00	0.000000	...
3	Archway	0.000000	0.0	0.00	0.000000	...
4	Balham	0.000000	0.0	0.00	0.000000	...
...

We then explore each area's 10 most common venues based on venue category frequencies, and create a dataframe.

Table 4: The 10 most common venues in each area (80 rows x 11 columns)

	Area	1 st Most Common Venue	2 nd Most Common Venue	3 rd Most Common Venue	4 th Most Common Venue	...
0	Aldgate	Coffee Shop	Hotel	Cocktail Bar	Gym/Fitness Center	...
1	Aldwych	Pub	Hotel	Theater	Coffee Shop	...
2	Angel	Pub	Coffee Shop	Café	Burrito Place	...
3	Archway	Coffee Shop	Grocery Store	Pub	Pizza Place	...
4	Balham	Coffee Shop	Pub	Pizza Place	Indian Restaurant	...
...

We are now ready to segment these areas into distinct clusters for categorisation purpose.

3.3.2 Area clustering

We run k-means to cluster the areas into 15 clusters, and create a master table listing areas, boroughs, geographical coordinates, cluster labels and most common venues. The master table can sequentially be cast into a map using the folium library.

Table 5: The 10 most common venues in each area (80 rows x 15 columns)

	Area	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	...
0	Aldgate	City of London	51.514248	-0.075719	1	Coffee Shop	...
1	Aldwych	Westminster	51.513103	-0.114920	9	Pub	...
2	Angel	Islington	51.531944	-0.106106	1	Pub	...
3	Archway	Islington	51.565437	-0.134998	6	Coffee Shop	...
4	Balham	Wandsworth	51.445644	-0.150364	6	Coffee Shop	...
...

3.3.3 Cluster examination

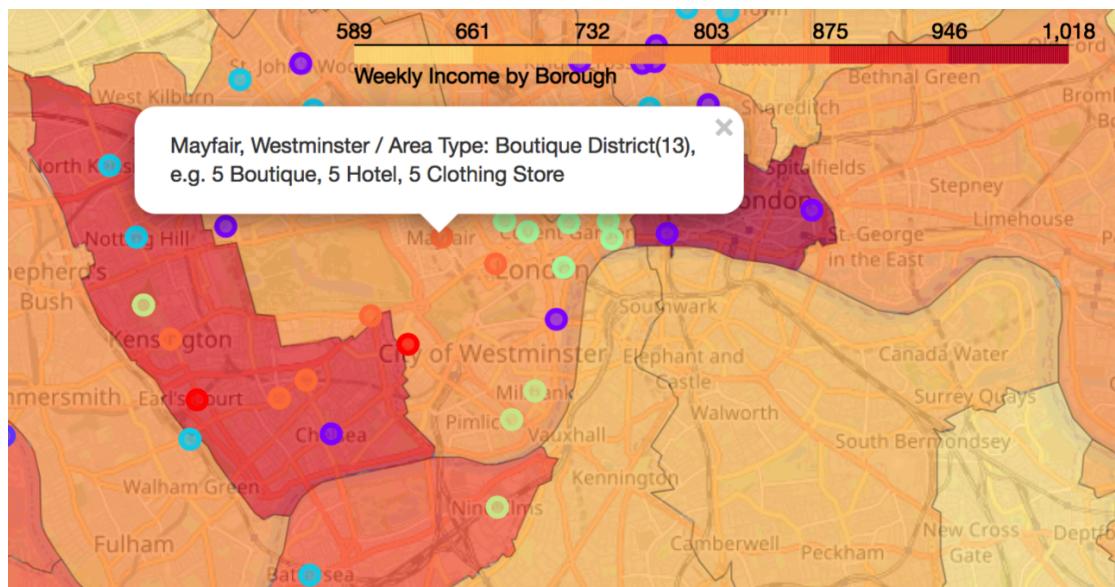
We study the number of the top 3 most frequent venue categories of each area in order to define the cluster names. The 15 cluster labels we come up with are listed in the table below. They are integrated into the master table and the map of London.

Table 6: 15 cluster names

0: Hotel District	1: Foreign Community'	2: French Community
3: Market Area	4: Stadium	5: Social Life-Pub
6: Social Life-Coffee Shop	7: Local Community	8: Home Store
9: Theatre District	10: Park & Sport	11: Grocery Store
12: Park	13: Boutique District	14: Wine Shop

Table 7: The 10 most common venues in each area (80 rows x 17 columns)

	Area	...	Cluster Labels	1st Most Common Venue	2 nd Most Common Venue	...	Description	Label
0	Aldgate	...	1	Coffee Shop	Hotel	...	10 Coffee Shop, 10 Hotel, 6 Cocktail Bar	Foreign Community
1	Aldwych	...	9	Pub	Hotel	...	6 Pub, 6 Hotel, 5 Theater	Theatre District
2	Angel	...	1	Pub	Coffee Shop	...	5 Pub, 4 Coffee Shop, 3 Café	Foreign Community
3	Archway	...	6	Coffee Shop	Grocery Store	...	6 Coffee Shop, 4 Grocery Store, 3 Pub	Social Life-Coffee Shop
4	Balham	...	6	Coffee Shop	Pub	...	7 Coffee Shop, 6 Pub, 4 Pizza Place	Social Life-Coffee Shop
...



Graph 2: 15 venue clusters marked on the income-scaled map of London

4. RESULTS

The demographic investigation of London's 33 boroughs suggests that west and southwest London tends to be richer than elsewhere. A unique exception is City of London in the core of the city, which is the richest borough of all.

The 6 boroughs selected to be studied under scrutiny are City of London, Kensington and Chelsea, Richmond upon Thames, Wandsworth, Islington and Westminster, ordered by weekly income median. These boroughs consist of a total of 80 constituent areas, which can be grouped into 15 clusters using k-means that runs on each area's 10 most frequent venues.

While many clusters share with one another common venues such as pubs, bars and restaurants, there are 2 clusters showing distinct features. They are cluster 0 (Hotel District) and 13 (Boutique District).

Table 8: Cluster 0 (Hotel District) (3 rows x 14 columns)

	Area	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	...	Description	Label
10	Belgravia	Westminster	Hotel	Boutique	Restaurant	...	7 Hotel, 3 Boutique, 2 Restaurant	Hotel District
22	Earls Court	Kensington and Chelsea	Hotel	Café	Garden	...	20 Hotel, 4 Café, 3 Garden	Hotel District
54	Paddington	Westminster	Hotel	Café	Coffee Shop	...	22 Hotel, 13 Café, 5 Coffee Shop	Hotel District

Table 9: Cluster 13 (Boutique District) (7 rows x 14 columns)

	Area	Borough	1st Most Common Venue	2 nd Most Common Venue	3rd Most Common Venue	...	Description	Label
12	Brompton	Kensington and Chelsea	Hotel	Bakery	Burger Joint	...	10 Hotel, 6 Bakery, 4 Burger Joint	Boutique District
38	Kensington	Kensington and Chelsea	Café	Clothing Store	History Museum	...	5 Café, 3 Clothing Store, 2 History Museum	Boutique District
39	Kew	Richmond upon Thames	Garden	Café	Botanical Garden	...	5 Garden, 5 Café, 5 Botanical Garden	Boutique District
41	Knightsbridge	Westminster	Café	Boutique	Hotel	...	7 Café, 5 Boutique, 5 Hotel	Boutique District
46	Mayfair	Westminster	Boutique	Hotel	Clothing Store	...	5 Boutique, 5 Hotel, 5 Clothing Store	Boutique District
62	South Kensington	Kensington and Chelsea	Café	Exhibit	Hotel	...	8 Café, 6 Exhibit, 5 Hotel	Boutique District
64	St James's	Westminster	Boutique	Hotel	Clothing Store	...	6 Boutique, 4 Hotel, 4 Clothing Store	Boutique District

The great majority of these 2 clusters (9 out of 10 areas) spread around Hyde Park in the slightly western part of London city centre, as shown in Graph 2. They fall under the Borough of Westminster and Borough of Kensington and Chelsea, and are within or adjacent to the richest regions in London. We will make further discussion related to our business problem in the next chapter.

5. DISCUSSION

This study is intended to provide a list of landing sites in London for emerging luxury brands seeking to establish their first shop in Europe. The stakeholders have unconventional business strategies and they only target a specific group of customers on the very top of the pyramid. The locations we are looking for need to be 1) geographically close to rich neighbourhoods and 2) within areas dominated by luxury industry, such as boutiques and high-end department stores.

It is noticed that although City of London in the city heart is the richest borough of all, the composing areas' most common business types are coffee shops and foreign restaurant. There seems to be no boutique according to Foursquare location data. We therefore remove City of London from our consideration list.

Kensington and Chelsea is the second richest borough in west central London. We note that cluster 0 and 13 appear more relevant to our research. However, in areas of Earls Court (cluster 0), Brompton, Kensington and South Kensington (cluster 13), Foursquare does not show any significant luxury site. Interestingly, to the east of Kensington and Chelsea is situated Westminster. A total of 4 areas in this borough seem to match with our research goal. They are:

- (1) Belgravia: cluster 0, with 7 hotel, 3 boutiques and 2 restaurants,
- (2) St James's: cluster 13, with 6 boutiques, 4 hotels and 4 clothing stores,
- (3) Mayfair: cluster 13, with 5 boutiques, 5 hotels and 5 clothing stores, and
- (4) Knightsbridge: cluster 13, with 7 cafés, 5 boutiques and 5 hotels.

We retrieve a series of venue categories most relevant to our study: 'Boutique', 'Jewelry Store', 'Department Store', and 'Art Gallery'. We find out that out of 26 boutiques in London, 6 are located in St James's, 5 in Mayfair and Knightsbridge respectively, 3 in Belgravia. The only 5 jewelry stores in London are located majorly in Knightsbridge (3), with the rest in Mayfair and St James's. Department stores share a seemingly similar distribution: 2 out of 7 in Knightsbridge, 1 in Mayfair and 1 in St James's. In addition, art galleries are also available in Mayfair (4 out of 29) and St James's (1). The information above can be summarised into a dataframe. The percentage in the table refers to the portion of a given store type out of all the stores of the same kind in London.

Table 10: List of 4 prospective landing sites and their typical business types

	Boutique	Jewelry Store	Department Store	Art Gallery
St James's	6 (23%)	1 (20%)	1 (14%)	1 (7%)
Mayfair	5 (20%)	1 (20%)	1 (14%)	4 (14%)
Knightsbridge	5 (20%)	3 (60%)	2 (29%)	na
Belgravia	3 (12%)	na	na	na
Total in London	26	5	7	29

Given the industrial landscape demonstrated by the table above, we may provide the following suggestions for our FJH stakeholders:

- (1) The most boutique-densed areas are St James's, Mayfair and Knightsbridge in Westminster. Adjacent to some of the richest London neighbourhoods, these areas may be potential and profitable landing sites to enter the British and European markets.
- (2) There seems to be a more vivid hub of jewellery industry in Knightsbridge, which may be a preferred area for jewellery houses.
- (3) FJHs leaning towards an artistic side for their business may give more consideration to Mayfair as the area is the most artistic of all.
- (4) Despite having fewer venues related to our research subject, Belgravia is just on the border between Westminster and Kensington and Chelsea. The area may also be taken into account for FJHs really hunting for a peculiar, undiscovered spot to commence their business.

6. CONCLUSION AND OUTLOOK

6.1 Conclusion

The objective of this project is to offer a list of locations in London for a particular group of newly-established FJHs planning to launch their first physical shops in the UK. Through research on London's demography, specifically residents' income level in each borough, and k-means clustering analysis based on each area's typical business activities, we propose St James's, Mayfair, Knightsbridge and Belgravia as prospective starting points for our stakeholders to expand their business. All of these four areas are near London's richest neighbourhoods and high density of luxury boutiques, yet each in a slightly different scale and manner. Stakeholders are best recommended to further perform evaluation of each individual area against their own business type and strategies.

6.2 Limitations and Recommendations

Some improvements may still be made to this project. First, on the demographic level, more detailed and comprehensive income information on each borough's constituent areas may help locate the very areas where our stakeholder's potential clients live. Second, the statistic analysis on areas was implemented quantitatively on limited categorial data. It did not take into account the fact that some areas have special historical and cultural connotations that may have led to London's industrial landscape today. From a marketing and branding perspective, qualitative investigation is required to provide a truly thorough understanding of the London market that would eventually benefit stakeholders' business decision.

REFERENCES

Earnings by Place of Residence, Borough. Office for National Statistics. Nov 2019.
(Available online at
<https://data.london.gov.uk/dataset/earnings-place-residence-borough> <25 Apr 2020>)

Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland. Office for National Statistics. 26 Jun 2019. (Available online at
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland><25 Apr 2020>)

Foursquare API. Foursquare.
(Available online at <https://developer.foursquare.com> <25 Apr 2020>)

List of areas of London. Wikipedia. (Available online at
https://en.wikipedia.org/wiki/List_of_areas_of_London <25 Apr 2020>)

London Borough Border Data: london_boroughs_proper. Carto. 1 Jul 2015.
(Available online at
https://joshuaboyd1.carto.com/tables/london_boroughs_proper/public <25 Apr 2020>)

Qualifications by Economic Activity Status, Borough. Office for National Statistics. Jul 2019. (Available online at
<https://data.london.gov.uk/dataset/qualifications-economically-active-population-gender-borough> <25 Apr 2020>)