# Choosing a Hotel Location in London

## Preface

The report outlined in this document form part of IBM Data Science Professional Certificate and aims to address the requirements of Course 9: "Applied Data Science Capstone". It shall be noted that the business scenario outlined in this paper while based on a real organisation, it is completely fictitious and has been defined to meet the course requirements.

## 1 Introduction

The Library Hotel by the Library Hotel Collection offers its guests in New York a unique experience centred around books where each of the 10 guestroom floors honours one of the 10 categories of the Dewey Decimal Classification® system and each of the 60 rooms are uniquely adorned with about 50-150 books and artwork exploring a distinctive topic within the category it belongs to (Library Hotel Collection, n.d.).

The portfolio of the Library Hotel Collection is made up of six boutique hotels, of which four are located in New York, one in Budapest and one in Toronto. However, none of the six hotels are centred around the same guest experience in terms of theme. For instance, the Aria Hotel in Budapest, its design and guest experience is inspired by and centred around music (Library Hotel Collection, n.d.).

The senior management are considering an expansion strategy where they would replicate the Library Hotel guest experience in New York at a cultural European city. The city of London has been shortlisted as a potential location of choice.

Spatial location is considered to be one of the most important factors of a new hotel establishment (Yang, et al., 2012), especially in the case of the Library Hotel Collection where the target audience is a niche segment. More importantly, the final hotel location of choice will have a significant impact on the revenue generation of the hotel, both in the short and long run (Johns, et al., 1997).

The objective of this paper is to identify Wards in London which have similar geospatial characteristics to the current address of the Library Hotel in New York. The shortlisted London Wards will then be provided to the senior management of the Library Hotel Collection as potential candidates for a new hotel location.

## 2 Data Requirements and Collection Method

The data requirements were selected to address the business problem of the senior management at the Library Hotel Collection. The intention is to construct observations with the following feature sets: Venue Categories.

The Places API (Foursquare, 2020) which offers real-time access to Foursquare's global database of rich venue data was used to populate the feature sets. The Places API allows access to the full details about a venue including location, tips and categories. For the purposes of this study, I will extract the category type (e.g.: Vietnamese Restaurant) of venues surrounding an address of interest.

Access to venue categories via the Places API is dependent on locational latitude and longitude coordinates as input. Therefore, the data collection exercise first focused on identifying physical addresses of interest, such as:

- Library Hotel in New York; and
- London Wards

The addresses were then converted into latitude and longitude coordinates for later use in the Places API. The following sections describe how the addresses as well as the latitude and longitude information were obtained.

### 2.1.1 Library Hotel: Address

The Library Hotel current address was obtained from its official website at https://libraryhotel.com.

### 2.1.2 London Wards: Addresses

The mapping tools by the Greater London Authority (GLA) were used to download the data for all London Wards (GLA, 2014). According to the dataset, there are 33 Borough Councils and 630 Wards (GLA, 2014).

### 2.1.3 Locational Latitude and Longitude Coordinates

Pyhton Geopy and Geopandas libraries were used together with Nominatim Geocoding service to convert the physical addresses of the London Wards and Library Hotel into latitude and longitude geographic locations.

## 3 Methodology

Figure-1 below illustrates the undertaken data science approach in response to the project problem. In Part 1, all relevant packages and libraries were imported and installed so that Python can get access to code from other modules. As to Part 2, the objective was to address all of the data prerequisite requirements of Part 3, such as: importing, cleansing and formatting addresses of London Wards as well as assigning locational latitude and longitude coordinates for each of those addresses. Part 3 made use of the data in Part 2 to create the feature set for each address by extracting the Venue Categories data from Foursquare's Places API. In Part 4, the outputs of the Foursquare's API were analysed and formatted in preparation of the clustering exercise in Part 5.

In the following sections, each of the individual Parts of the methodology will be further explained and discussed.

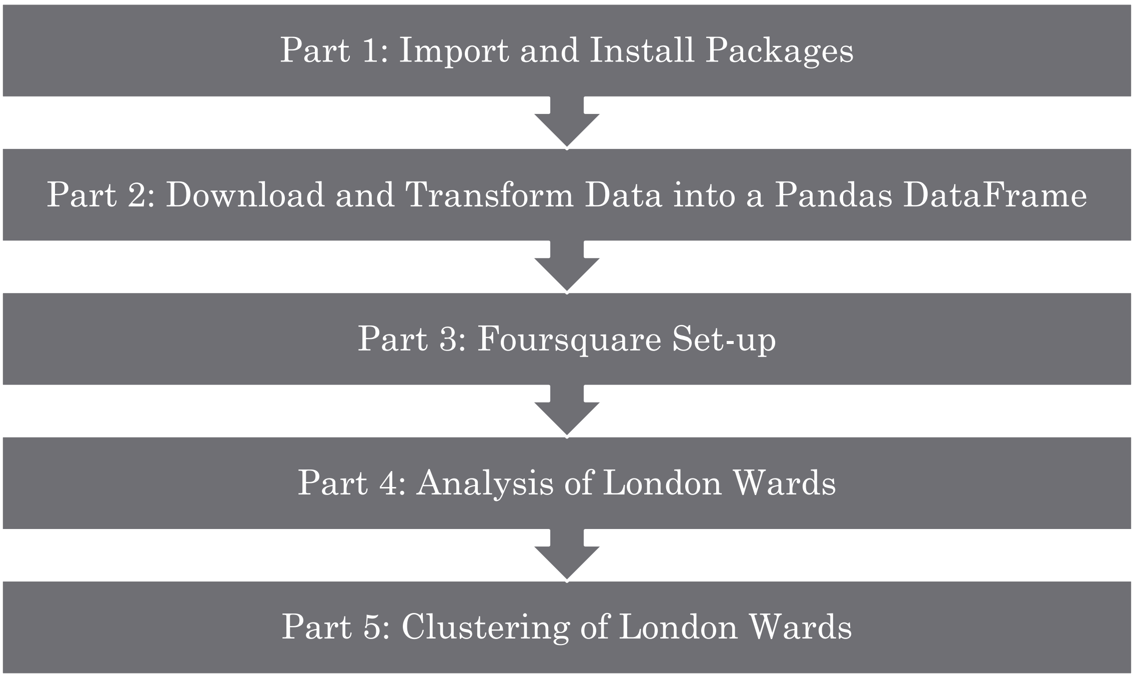


**Figure 1 - Methodology**

## 3.1 Part 1: Import and Install Packages

The import of packages and in libraries allows Python to access codes in other modules that have been pre-written by other developers. For the purposes of this exercise, the following modules were imported:

1. Geopy,
2. GeoPandas,
3. Pandas,
4. Folium; and
5. Requests

The following sub-sections describe the role of each of the modules outlined above. Other modules were also imported in other parts of the project, they will be highlighted and discussed under their own areas of the methodology.

### 3.1.1 Geopy

Geopy is a Python client for several popular geocoding services, it includes geocoder classes for the OpenStreetMap Nominatim, Google Geocoding API, and many other geocoding services (MIT, 2018). In the context of this project, the OpenStreetMap Nominatim geocoding services was used.

### 3.1.2 GeoPandas

GeoPandas is an open source project aimed at simplifying the use of geospatial data in Python, it combines the capabilities of Pandas and negates the need for spatial databases such as PostGIS (GeoPandas, 2019).

#### 3.1.3 Pandas

Pandas is built on top of the Python programming language and its main purpose is to enable real world data analysis using a DataFrame object (Pandas, 2020). In brief, a DataFrame is a data structure which contains labelled axes upon which arithmetic operations can be undertaken (Pandas, 2020).

#### 3.1.4 Folium

Folium is a visualisation package which manipulates data in Python so that it can be visualised in a leaflet map (MIT, n.d.). The package has been used to plot the location of various London Wards over the map of London.

#### 3.1.5 Requests

Requests allows HTTP/1.1 requests to be sent without having the need to manually add query strings to a URL or to form-encode POST data (Apached 2.0, 2020). This package was used to request data from the Foursquare's Places API.

### 3.2 Part 2: Download and Transform Data into a Pandas DataFrame

Part 2 was focused on obtaining the required latitude and longitude data for later use by the Foursquare's Places API. As such, the following steps were followed:

- Step 1: Obtain the address of London Wards
- Step 2: Obtain the address of the Library Hotel
- Step 3: Obtain Latitude and Longitude Data for London Wards and Library Hotel
- Step 4: Create Map of London Wards

#### 3.2.1 Addresses

##### *3.2.1.1 London Wards: Addresses*

The addresses for the London Wards were downloaded from a mapping tool by the GLA (GLA, 2014). The file contained 52 columns of which only two were kept: "Ward name" and "Borough name". Relatedly, the file contained 644 rows, some of which were populated with a NaN or null value. Thus, all rows with a NaN or null value in their columns were deleted. As a result, there were 630 London Wards in contention to become the new hotel location.

##### *3.2.1.2 Library Hotel: Address*

The Library Hotel Address was taken from its official website (Library Hotel Collection, n.d.). The main challenge is that officially, the Library Hotel address falls within the neighbourhood of Murray Hill in Manhattan, New York. However, it is on the North-West boundary of Murray Hill. As such, if the address of Murray Hill is to be given to the geocoding service, then the feature set will be misleading. It is also worth highlighting that the used geocoding service would not recognise the actual physical address of the hotel so that the latitude and longitude data could not be extracted for the purposes of the Places API. As a result, the emphasis was to identify the closest landmark and extract their latitude and longitude data, this resulted in the use of the latitude and longitude data of the Grand Central Terminal Station which is approximately 100 metres away from the Library Hotel physical address.

#### 3.2.1.3 DataFrame: Addresses

Below is an extract of the DataFrame which contained the addresses information:

| | Ward name | Borough name | City |
|---|---|---|---|
| 626 | Vincent Square | Westminster | London |
| 627 | Warwick | Westminster | London |
| 628 | Westbourne | Westminster | London |
| 629 | West End | Westminster | London |
| 630 | Grand Central Terminal | Manhattan | New York |

**Figure 2 - DataFrame Addresses**

### 3.2.2 Latitude and Longitude

So far, the addresses of interest have been captured in a DataFrame Object, the next step was to extract the latitude and longitude for each of those addresses. This was achieved through the use of Geopy and GeoPandas packages together with Nominatim geocoding services.

In practice, the Nominatim geocoding service was unstable and frequently timed out, only one geocoding attempt was successful, and the results of that attempt were used throughout the project. This made it harder to assess the quality of the geocoding service, as a result, 10 addresses were picked at random and had their latitude and longitude checked manually. Even though the results were fairly accurate, quality issues emerged throughout the project and were handled appropriately so that the final results were not affected.

In addition to the above, there were 34 London Wards which had null values and were dropped from the dataset. This brought down the total number of London Wards in contention for a new hotel location to 596 potential locations. Below is an extract of the DataFrame showing the Latitude and Longitude for 5 London Ward addresses.

| | Ward name | Borough name | City | Address | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 0 | Abbey | Barking and Dagenham | London | Abbey,Barking and Dagenham,London | 51.546127 | 0.091834 |
| 1 | Alibon | Barking and Dagenham | London | Alibon,Barking and Dagenham,London | 51.548280 | 0.153244 |
| 2 | Becontree | Barking and Dagenham | London | Becontree,Barking and Dagenham,London | 51.540311 | 0.126524 |
| 3 | Chadwell Heath | Barking and Dagenham | London | Chadwell Heath,Barking and Dagenham,London | 51.567986 | 0.127994 |
| 4 | Eastbrook | Barking and Dagenham | London | Eastbrook,Barking and Dagenham,London | 51.551565 | 0.161543 |

**Figure 3 - DataFrame Showing Latitude and Longitude**

### 3.2.3 Create Map of London Wards

As part of a final visual quality check, all of the 596 London Wards were plotted over the map where it became evident that for three addresses the geocoding was unsuccessful as one address was located in the South-East of England, a second address was plotted in the North-East outside of London, and lastly one address was plotted in Canada. As a consequence, these three addresses were removed which meant the dataset was reduced to 593 London Wards in total.

## 3.3 Part 3: Foursquare Set-up

Initially, the Places API was used to get the Venue Categories of the first address in the DataFrame. The main reason for doing that was to meet the minimum criteria of acceptance testing prior to scaling it to the 593 London Wards and Library Hotel address. The behaviour of the API met expectations where the requirements were to retrieve the top 100 venues over a 1-mile radius for the geospatial coordinates of Abbey, Barking and Dagenham, London. In total, for that specific address there were 24 unique venues, below is an extract of the first five venues in the dataset.

| | name | categories | lat | lng |
|---|---|---|---|---|
| 0 | Barking Park | Park | 51.545217 | 0.086134 |
| 1 | Nando's | Portuguese Restaurant | 51.539780 | 0.082297 |
| 2 | Eastbury Manor House | History Museum | 51.532973 | 0.099741 |
| 3 | Cristina's | Steakhouse | 51.536523 | 0.076672 |
| 4 | Mayesbrook Park | Park | 51.549842 | 0.108544 |

**Figure 4- Venues: Abbey, Barking and Dagenham**

The same exercise was then repeated across all of the 593 London Wards and Library Hotel address in New York. This has resulted in the shortlisting of 39,271 venues and 436 unique categories.

## 3.4 Part 4: Analyse Each Ward in London

To this point, we have identified London Wards of interest, extracted their latitude and longitude coordinates which were then used by the Places API to retrieve 39,271 venues. The next step was to group the venues per London Ward. One-hot-encoding was applied to convert the categorical variables into binary data, see Figure-5 below.

| | Address | Accessories Store | Afghan Restaurant | African Restaurant | Airport | Airport Lounge | Airport Service | Airport Terminal | American Restaurant | Animal Shelter | ... | Wine Bar | Wine Shop | Winery | Wings Joint | Women's Store |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Abbey,Barking and Dagenham,London | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 |
| 1 | Abbey,Barking and Dagenham,London | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 |
| 2 | Abbey,Barking and Dagenham,London | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 |
| 3 | Abbey,Barking and Dagenham,London | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 |
| 4 | Abbey,Barking and Dagenham,London | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 |

**Figure 5 - One Hot Encoding**

The one-hot-encoding was followed by the grouping of venues per London Ward by taking the mean of the frequency of occurrence of each category. As an example, see column "Wine Bar" or "Wine Shop" for Abbey Road, Westminster, London in the DataFrame shown in Figure-6.

| | Address | Accessories Store | Afghan Restaurant | African Restaurant | Airport | Airport Lounge | Airport Service | Airport Terminal | American Restaurant | Animal Shelter | ... | Wine Bar | Wine Shop | Winery | Wings Joint |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Abbey Road,Westminster,London | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.02 | 0.01 | 0.0 | 0.0 |
| 1 | Abbey Wood,Greenwich,London | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.00 | 0.00 | 0.0 | 0.0 |
| 2 | Abbey,Barking and Dagenham,London | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.00 | 0.00 | 0.0 | 0.0 |
| 3 | Abbey,Merton,London | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.00 | 0.00 | 0.0 | 0.0 |
| 4 | Abingdon,Kensington and Chelsea,London | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.01 | 0.00 | 0.0 | 0.0 |

**Figure 6 - Venue Grouping per London Ward**

Lastly, the dataset was re-arranged to show the top 10 most common venues per London Ward and around the Library Hotel, see Figure-7 and 8.

| | Address | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 589 | Woolwich Riverside,Greenwich,London | Pub | Grocery Store | Park | Coffee Shop | Plaza | Supermarket | Fast Food Restaurant | Clothing Store | Bakery | Bus Stop |
| 590 | Worcester Park,Sutton,London | Pub | Grocery Store | Bus Stop | Pharmacy | Coffee Shop | Park | Supermarket | Train Station | Japanese Restaurant | Steakhouse |
| 591 | Wormholt and White City,Hammersmith and Fulham... | Pub | Bakery | Café | Indian Restaurant | Gym / Fitness Center | Thai Restaurant | Chinese Restaurant | Coffee Shop | Burger Joint | Pizza Place |
| 592 | Yeading,Hillingdon,London | Fast Food Restaurant | Grocery Store | Clothing Store | Indian Restaurant | Electronics Store | Coffee Shop | Supermarket | Hotel | Pizza Place | Video Game Store |
| 593 | Yiewsley,Hillingdon,London | Supermarket | Pub | Grocery Store | Lake | Park | Bed & Breakfast | Harbor / Marina | Bus Station | Bar | Electronics Store |

**Figure 7 - Most Common Venues in London Wards**

| | Address | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 218 | Grand Central Terminal,Manhattan,New York | Theater | Hotel | Plaza | American Restaurant | Korean Restaurant | Gym | Gym / Fitness Center | Japanese Restaurant | Boxing Gym | Food Truck |

**Figure 8 - Most Common Venues Around Library Hotel**

## 3.5 Part 5: Cluster London Wards

K-means clustering was selected as the machine algorithm of choice. The aim of the K-means algorithm is to group similar data points together with the purpose of identifying patterns without having to label the outcome (Garbade, 2018). This unsupervised learning approach aligns with the interests of this project where the objective is to identify London Wards with similar geospatial patterns to the neighbourhood of the Library Hotel in Manhattan, New York.

As shown in Figure-9 below, a divide and conquer clustering approach was taken into account, this is aimed at overcoming the limitation of clustering where the main concern of the clustering algorithm is to group similar objects, ignoring the fact that while they are similar, they may be of different levels (Khalilian, Boroujeni, Mustapha, & Sulaiman, 2009).

Referring back to Figure-9, at Level-0, the full dataset, 593 London Wards plus the Library Hotel address, was subjected to the clustering analysis where the Library Hotel was identified to have a Cluster Label of 0. This reduced the dataset of interest to the ones of Level-1, Cluster Label 0, which was made-up of 329 London Wards plus the Library Hotel address. This dataset was further divided into two clusters, where the

Library Hotel was identified to have a Cluster Label of 1. As such, Level 2, Cluster Label 1, was now subjected to a cluster analysis where the 86 London Wards plus the Library Hotel address were further clustered into two groups. The results indicated that the Library Hotel under Level 3 had a Cluster Label of 0. Consequently, Level 3, Cluster Label 0, was now grouped into 10 clusters where the Library Hotel had a Cluster Label of 2. Under Level 4, Cluster Label 2, there were 3 London Wards shortlisted as candidate locations for a new Library Hotel.

In the next section of the report, I will examine and discuss in detail the shortlisted London Ward locations. However, it is worth noting at this point that by taking a data reduction technique via a divide and conquer approach the quality of results was increasingly higher as this helped to amplify the identity of each of the 10 clusters at Level 4, this will be more evident in the following section where cluster 2 will be explored.

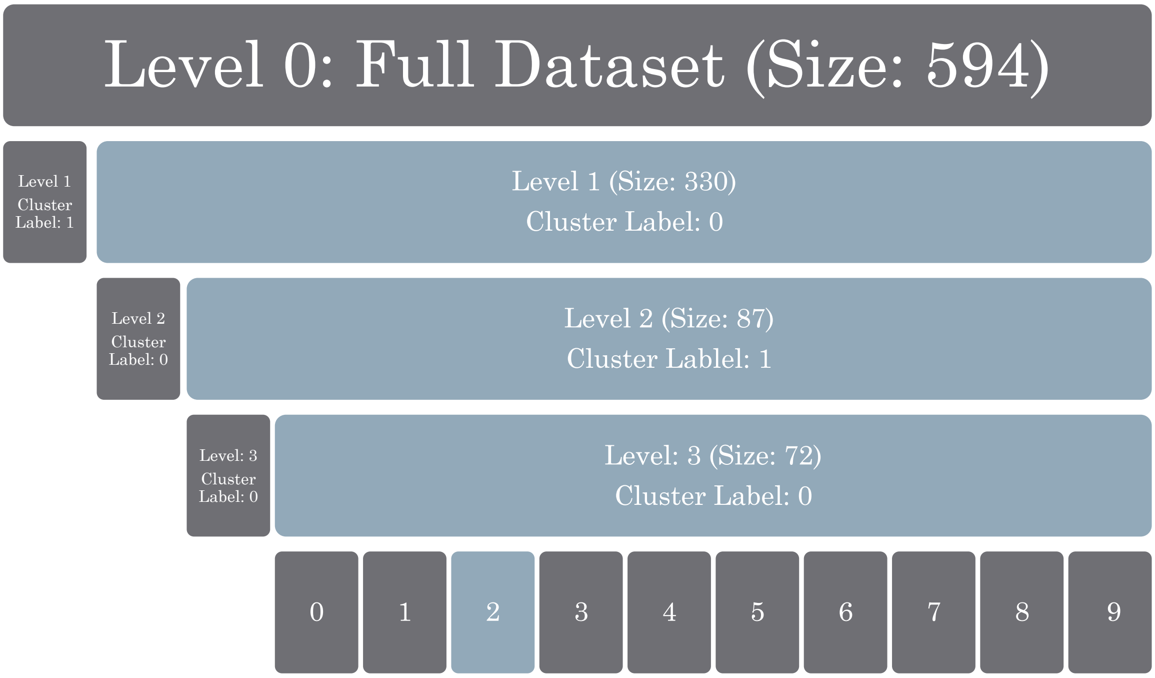


**Figure 9 - Clustering Hierarchy**

At this point, it is worth addressing the two limitations of the K-means algorithm and the handling techniques deployed to address them. The first limitation of the K-means algorithm is the selection of centroids. Often, and as was the case in this project, the location of centroids is unknown to us, this could lead to a suboptimal solution due to ineffective centroid initializations. A solution to address the shortcomings of centroid initializations is by running the algorithm multiple times with different random initializations and by keeping the best solution. The number of random initializations is controlled by the n_init hyperparameter. By default, the hyperparameter is set to 10, this was amended to 150 which means that the whole algorithm described earlier in Figure-9 ran 150 times at each clustering Level where the best solution was maintained based on the best inertia score. The inertia scored was determined by taking the mean squared distance between each instance and its closest centroid.

The second limitation of K-means is the selection of the optimal number of clusters. Similarly, to centroid initializations, a poor selection of number of clusters could lead to a suboptimal solution. This was addressed by calculating the silhouette score. In brief, the silhouette score is a measure of how similar an object is to its own cluster compared to other clusters. At each Level (e.g.: Level 0, Level 1, etc.) the silhouette score was calculated for k clusters ranging from 2 to 10. Figure 10 to 13 show the silhouette score and the optimal k cluster for each of Level-0, Level-1, Level-2 and Level-3.

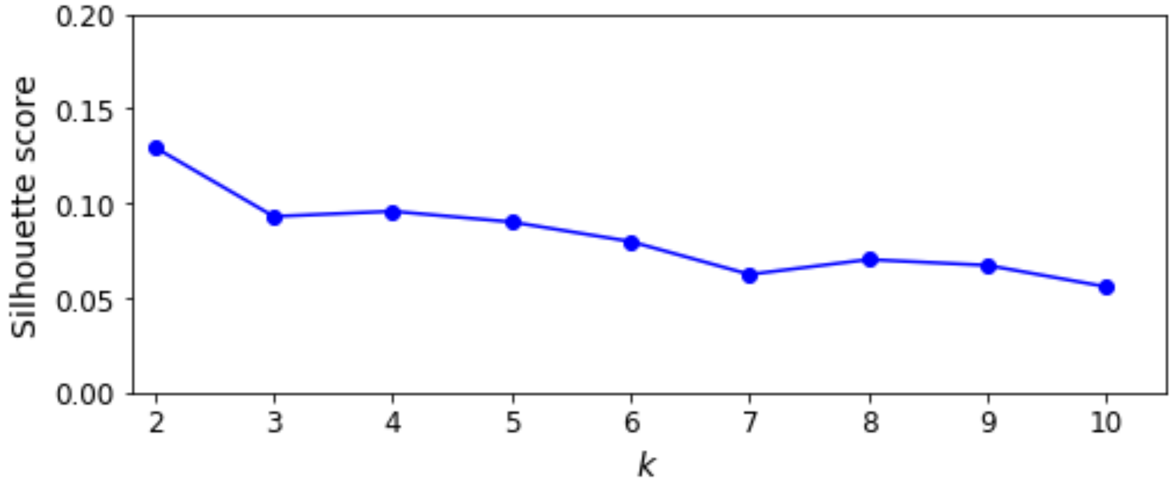


**Figure 10 - Level 0: Selecting Optimal Number of Clusters (k=2)**

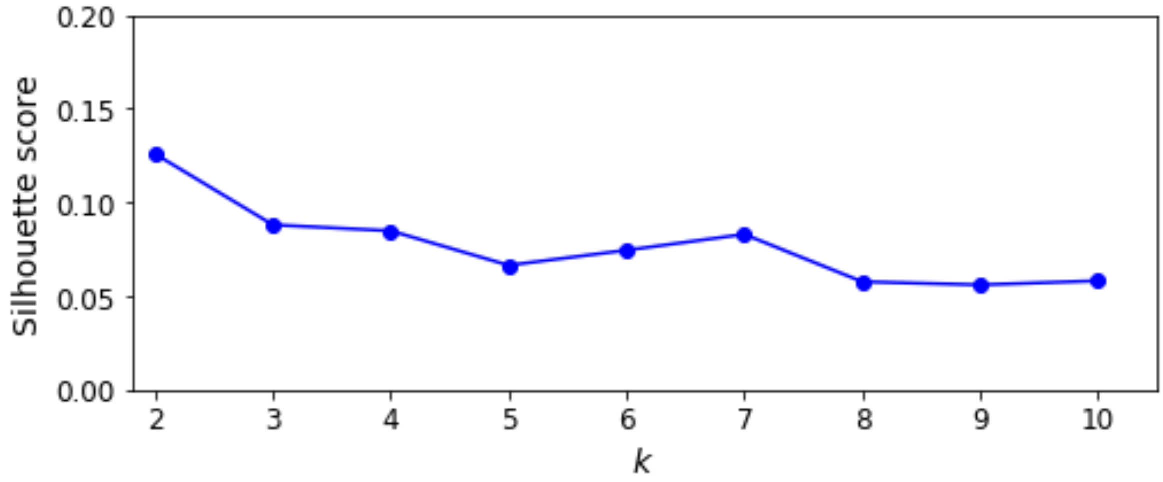


**Figure 11 - Level 1: Selecting Optimal Number of Clusters (k=2)**

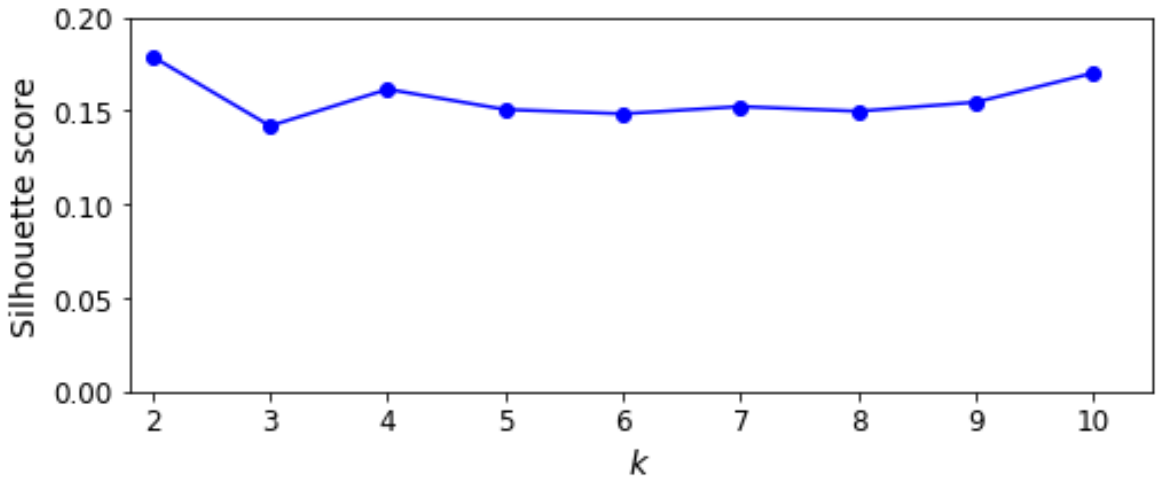


**Figure 12 - Level 2: Selecting Optimal Number of Clusters (k=3)**

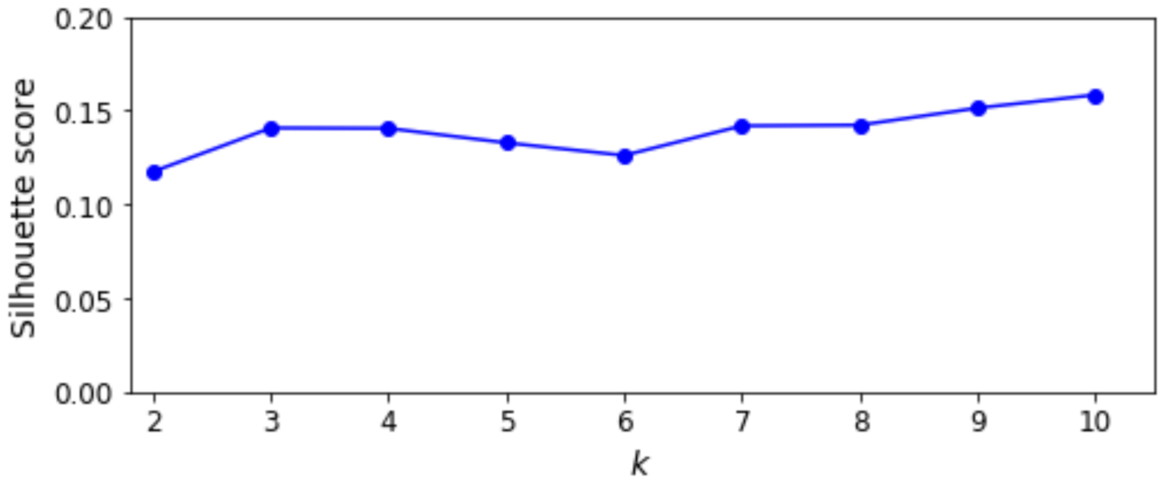


**Figure 13 - Level 3: Selection Optimal Number of Cluster (k=10)**

# 4 Results: Examine London Locations for Library Hotel

The initial results of Level 4, Cluster Label 2, indicated four potential London Ward locations. However, upon closer inspection, two of the four London Ward addresses had the same latitude and longitude coordinates and at the same time the “string type” (str) object address did not match the latitude and longitude shown in the DataFrame. This is reflective of the challenges previously raised in relation to the geocoding services, this had no implications on the final result as the latitude and longitude still resulted in the correct feature set via the Places API; nevertheless, a cleansing exercise was required to the non-feature elements of the DataFrame object so that the final visual representation of the DataFrame reflects the computations of the K-means algorithm.

The shortlisted London Wards for consideration by the senior management of the Library Hotel Collection are presented in the below DataFrame, Figure-14. From the shortlisted addresses below, it is evident that there is an emerging theme around “Art, Theatre and Culture”. More importantly, by looking at the map in Figure-15, the shortlisted London

Wards are adjacent to each other; while this creates a centralised destination for the sought after niche target segment, it is likely though that properties will come at a significant premium not only due to the wide range of Art, Culture and Theatre activities covered by the shortlisted London Wards but it also appears that the three London Wards are of short-distance to touristic landmarks such as Big Ben and Buckingham Palace. Thus, the senior management of the Library Hotel Collection is likely not only to be facing competition for locations from other hotel developers with a similar proposition but they will also be competing with other hotel developers with a different proposition who are trying to target various market segments, such as tourists seeking to visit landmarks.

| Address | Latitude | Longitude | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Charing Corss,Westminster,London | 51.507322 | -0.127647 | Theater | Ice Cream Shop | Hotel | Bookstore | Dessert Shop | Bakery | Plaza | Gourmet Shop | Garden | Cocktail Bar |
| Grand Central Terminal,Manhattan,New York | 40.752806 | -73.977179 | Theater | Plaza | Hotel | Korean Restaurant | American Restaurant | Japanese Restaurant | Gym | Gym / Fitness Center | Sushi Restaurant | Burger Joint |
| Holborn and Covent Garden,Camden,London | 51.515825 | -0.125985 | Theater | Ice Cream Shop | Hotel | Coffee Shop | Steakhouse | Cocktail Bar | Bakery | Liquor Store | Arts & Crafts Store | Beer Bar |
| St. ımes's,Westminster,London | 51.507908 | -0.136573 | Hotel | Clothing Store | Art Gallery | Indian Restaurant | Boutique | Theater | Ice Cream Shop | Lounge | Art Museum | Cocktail Bar |

**Figure 14 - Recommendation: Level 4, Cluster Label 7**

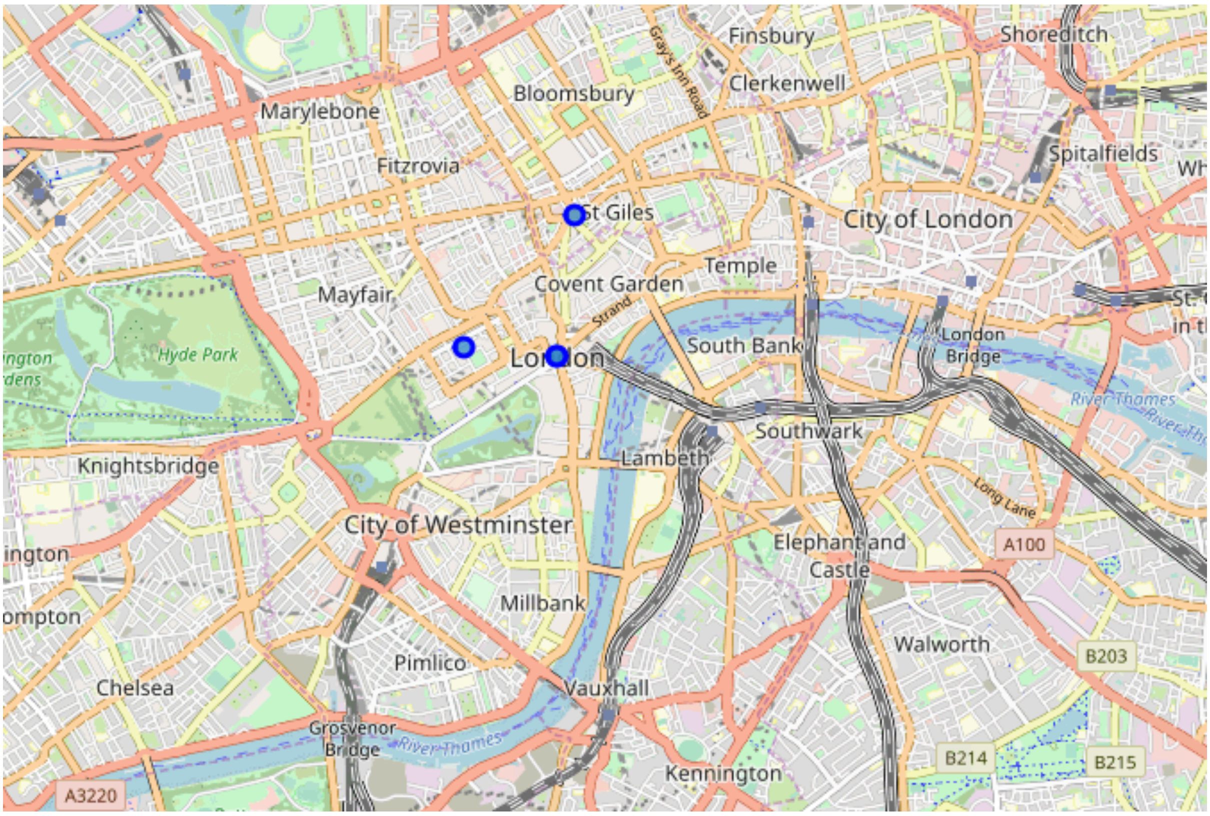


**Figure 15 - Map of Recommended Locations**

# 5 Discussion

## 5.1 Recommendation to Library Hotel Collection

It is recommended that the senior management of the Library Hotel Collection considers the shortlisted addresses in London. However, considering the choice is limited and the cost is likely to be prohibitive. It is recommended that other cultural European cities such as Berlin, Vienna and so forth are explored. This is likely to increase the variety of location and then based on localised market and customer research studies alternative cities may prove to be a superior alternative to London from a: market, customer and economic perspective.

## 5.2 Recommendation of Further Algorithmic Works

The study has highlighted the importance of data integrity as the project initially started with 630 London Wards and ended up with 593 London Wards suitable for clustering analysis. This is equivalent to around 5.9% loss in data via cleansing. If this algorithm is to be scaled to other cities and commercialised, then a data procurement strategy will need to be set in place.

Another area of consideration is the geocoding services, this was one of the least successful part of the coding exercise; however, this is mostly due to the choice of free services to extract geospatial coordinates. If this algorithm is to be scaled, then a commercial agreement for API geocoding services will be required.

Regarding feature sets, even though using the venue categories from the Places API led to a plausible explanation to the recommended solution. A more thorough literature review as well as engagement with hoteliers and guests is required to understand "what is a relevant feature" when choosing a hotel location. This project did not explore any features in terms of importance and rather relied on unsupervised learning to identify patterns for rationalisation.

Lastly, the coding of the K-means "divide and conquer approach", as written, requires human intervention to instruct the algorithm on how to proceed from one level to another. If there is a desire to scale the algorithm to other cities and/or scenarios, then the code would need to be re-written to cater for automation.

# 6 Conclusion

This thesis was set out to identify London Wards of interest to the senior management of the Library Hotel Collection. Initially, the emphasis was on obtaining the latitude and longitude data of London Wards and Library Hotel. The access to geospatial coordinates was an important milestone as it enabled the construction of features sets using Foursquare's Places API. Then, a k-means clustering algorithm was used to identify three locations with similar geospatial patterns to the neighbouring area of the Library Hotel in New York. Even though a shortlist of locations was presented to the senior management of the Library Hotel Collection for consideration, it was recommended that other cultural European cities should be explored due to the likelihood of high competition from other hotel developers in the identified London Wards. Finally, the shortcomings of the algorithm were pointed out and recommendations were made regarding further works to improve it.

## Works Cited


- Apached 2.0, 2020. *Requests: HTTP for Humans.* [Online]
  Available at: https://requests.readthedocs.io/en/master/
  [Accessed February 2020].

- Foursquare, 2020. *Places.* [Online]
  Available at: https://enterprise.foursquare.com/products/places
  [Accessed February 2020].

- Garbade, M., 2018. *Understanding K-means Clustering in Machine Learning.* [Online]
  Available at: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

- GeoPandas, 2019. *GeoPandas 0.7.0.* [Online]
  Available at: https://geopandas.org/index.html
  [Accessed February 2020].

- GLA, 2014. *Excel Mapping Template for London Boroughs, and Wards.* [Online]
  Available at: https://data.london.gov.uk/dataset/excel-mapping-template-for-london-boroughs-and-wards
  [Accessed February 2020].

- Johns, N., Howcrfot, B. & Drake, L., 1997. The use of data envelopment analysis to monitor hotel productivity. *Progress in Tourism and Hospitality Research,* 3(2), pp. 119-127.

- Khalilian, M., Boroujeni, F. Z., Mustapha, N. & Sulaiman, M. N., 2009. *K-Means Divide and Conquer Clustering.* Bangkok, s.n., pp. 306-309.

- Library Hotel Collection, n.d. *The Experience: The Library Concept.* [Online]
  Available at: https://libraryhotel.com
  [Accessed February 2020].

- MIT, 2018. *Geocoding library for Python.* [Online]
  Available at: https://github.com/geopy/geopy
  [Accessed February 2020].

- MIT, n.d. *Folium.* [Online]
  Available at: https://github.com/python-visualization/folium
  [Accessed February 2020].

- New York University, 2014. *Spatial Data Repository.* [Online]
  Available at: https://geo.nyu.edu/catalog/nyu_2451_34572
  [Accessed February 2020].

- NYC Department of Planning, 2014. *City of Neighborhoods.* [Online]
  Available at: https://www1.nyc.gov/site/planning/data-maps/city-neighborhoods.page#nycmap
  [Accessed February 2020].

- Pandas, 2020. *Pandas.* [Online]
  Available at: https://pandas.pydata.org/docs/index.html#
  [Accessed February 2020].

- Yang, Y., Wong, K. & Wang, T., 2012. How do hotels choose their location? Evidence from hotels in Beijing. *International Journal of Hospitality Management,* 31(3), pp. 675-685.