CHOOSING A HOTEL LOCATION

IBM Data Science Professional Certificate

Capstone Project: Applied Data Science

Antoine Khalife

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.

EXECUTIVE SUMMARY

The senior management of the Library Hotel Collection would like to replicate, at a cultural European city, the niche experience of the Library Hotel in Manhattan, New York.

Three London Wards have been identified as potential candidates using the k-means clustering algorithm. However, it was recommended that other cultural European cities should be explored due to the expected high level of competition from other hotel developers.

The shortlisted London Wards for considerations are:

- Charing Cross, Westminster, London
- Holborn and Convent Garden, Camden, London
- St. James's, Westminster, London

BACKGROUND

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.

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.

DATA REQUIREMENT AND COLLECTION METHOD

The intention is to construct observations with the following feature sets: Venue Categories. The Places API which offers real-time access to Foursquare's global database of rich venue data was used to populate the feature sets.

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 Library Hotel current address was obtained from its official website.

The mapping tools by the Greater London Authority (GLA) were used to download the data for all London Wards.

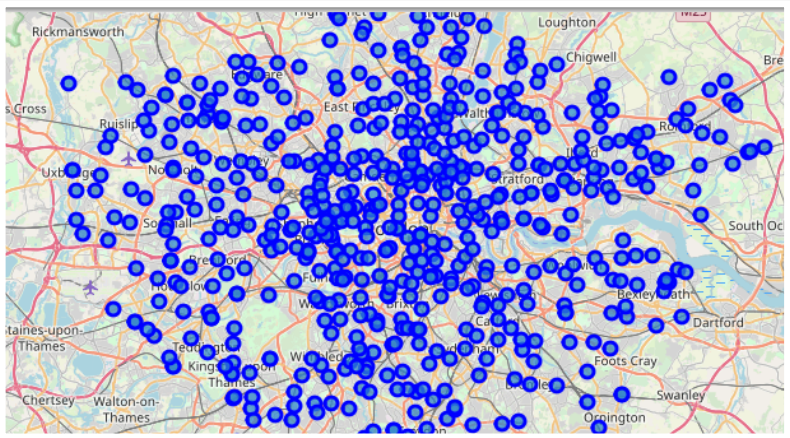
Pyhton Geopy and Geopandas libraries were used together with Nominatim Geocoding service to convert the physical addresses of the London Wards and Library Hotel into latitude and longitude geographic locations.

METHODOLOGY

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:

- Geopy,
- GeoPandas,
- Pandas,
- Folium; and
- Requests



	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

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

PART 3: FOURSQUARE SET-UP

The Foursquare Places API was used to construct a feature set where the top 100 venues in each London Ward were retrieved.

The DataFrame shown on the right is an extract of 5 venues at Abbey, Barking and Dagenham, London. In total, for that specific London Ward there were 24 unique venues.

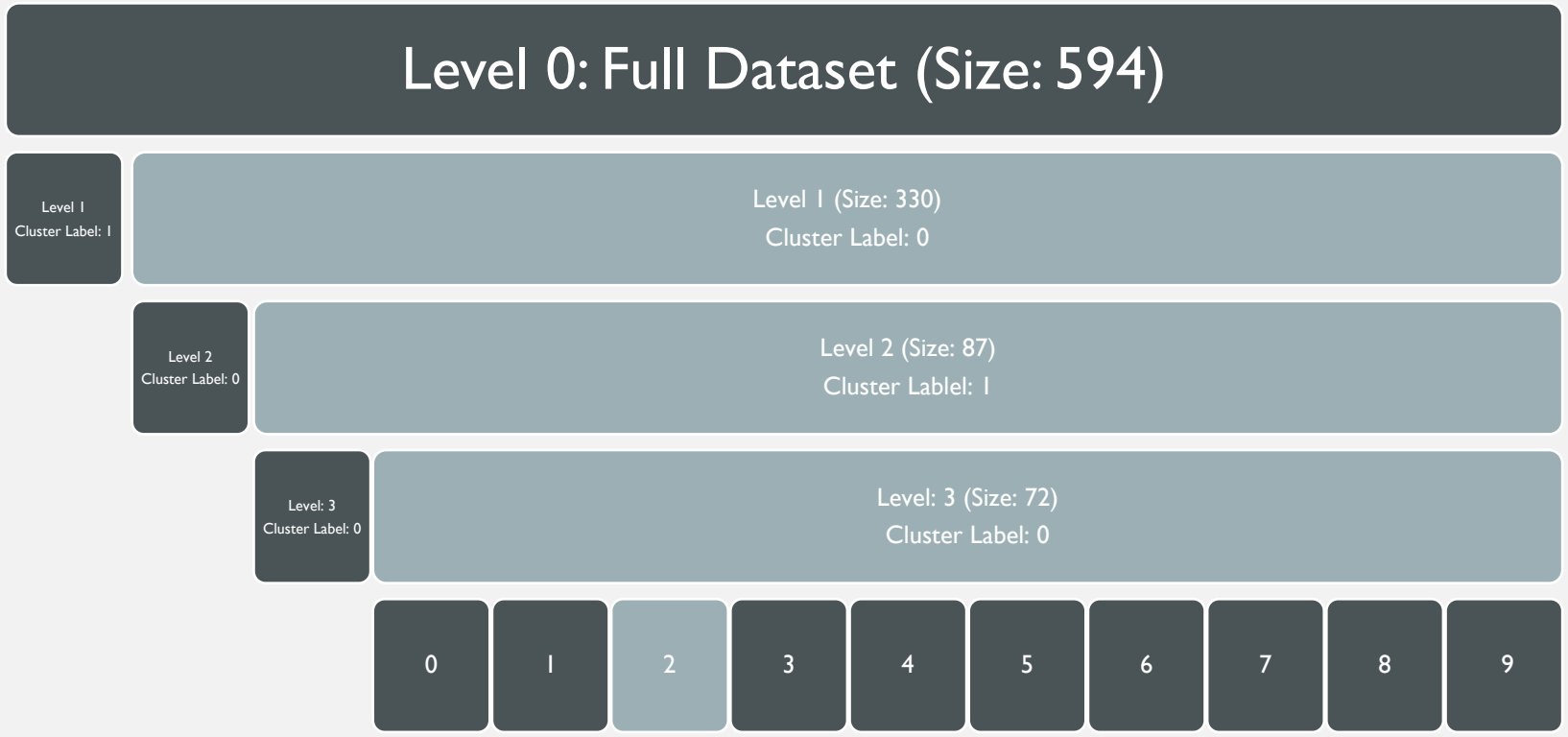
	name	categories	lat	lng
0	Barking Park	Park	51.545217	0.086134
1	Nando's	Portuguese Restaurant	51.539780	0.082297
2	Eastbury Manor House	History Museum	51.532973	0.099741
3	Cristina's	Steakhouse	51.536523	0.076672
4	Mayesbrook Park	Park	51.549842	0.108544

PART 4: ANALYSE EACH WARD IN LONDON

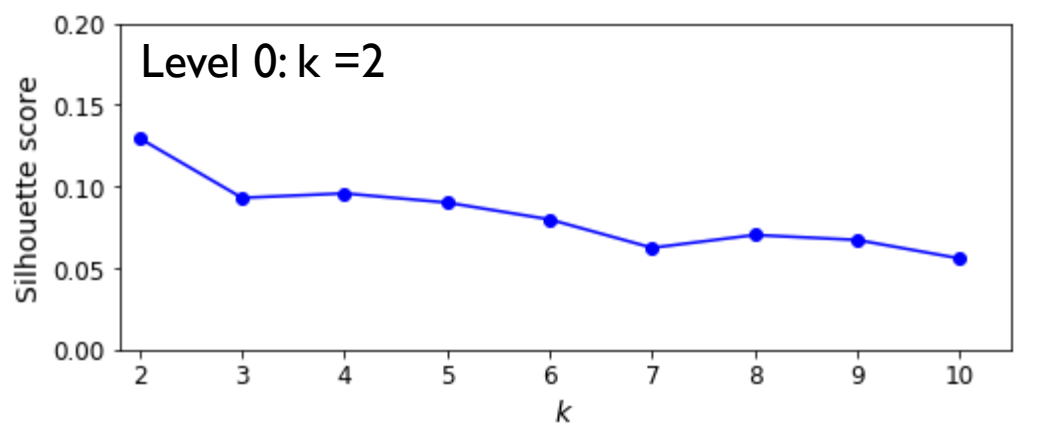
	Address	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
589	Woolwich Riverside,Greenwich,London	Pub	Grocery Store	Park	Coffee Shop	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

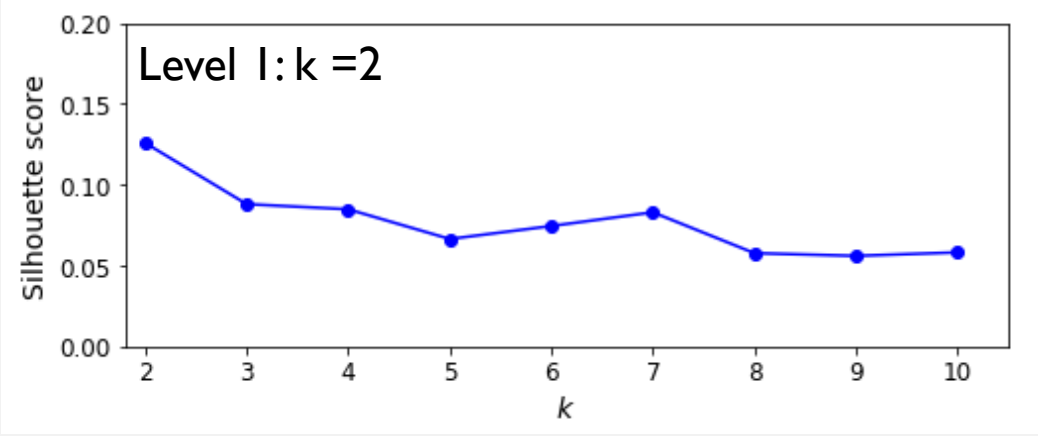
	Address	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
218	Grand Central Terminal,Manhattan,New York	Theater	Hotel	Plaza	American Restaurant	Korean Restaurant	Gym	Gym / Fitness Center	Japanese Restaurant	Boxing Gym	Food Truck

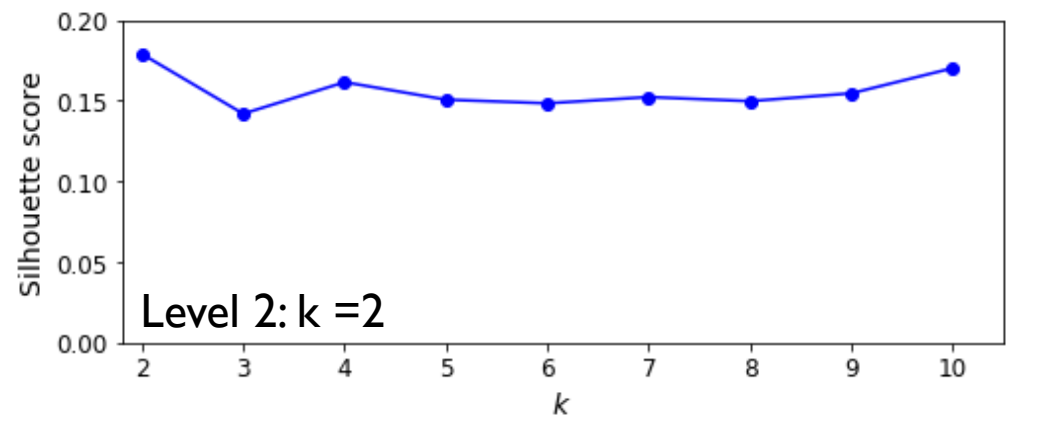
PART 5: CLUSTER LONDON WARDS

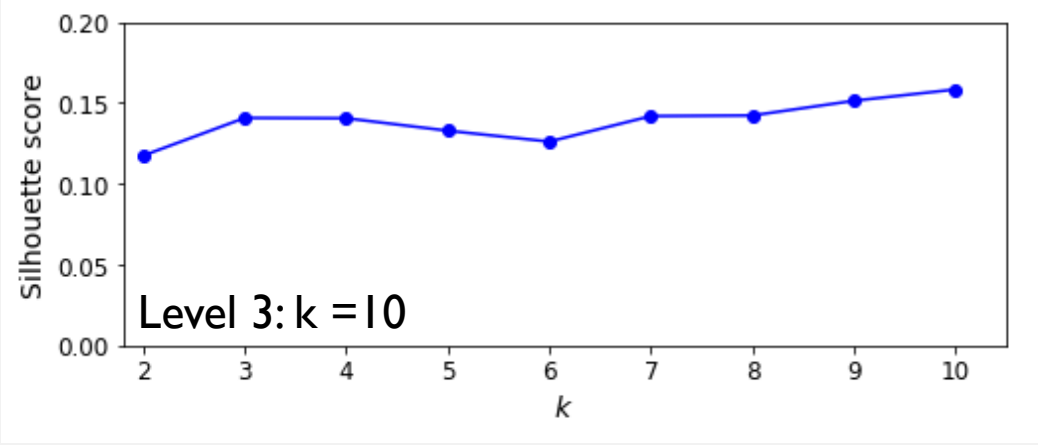


PART 5: CLUSTER LONDON WARDS





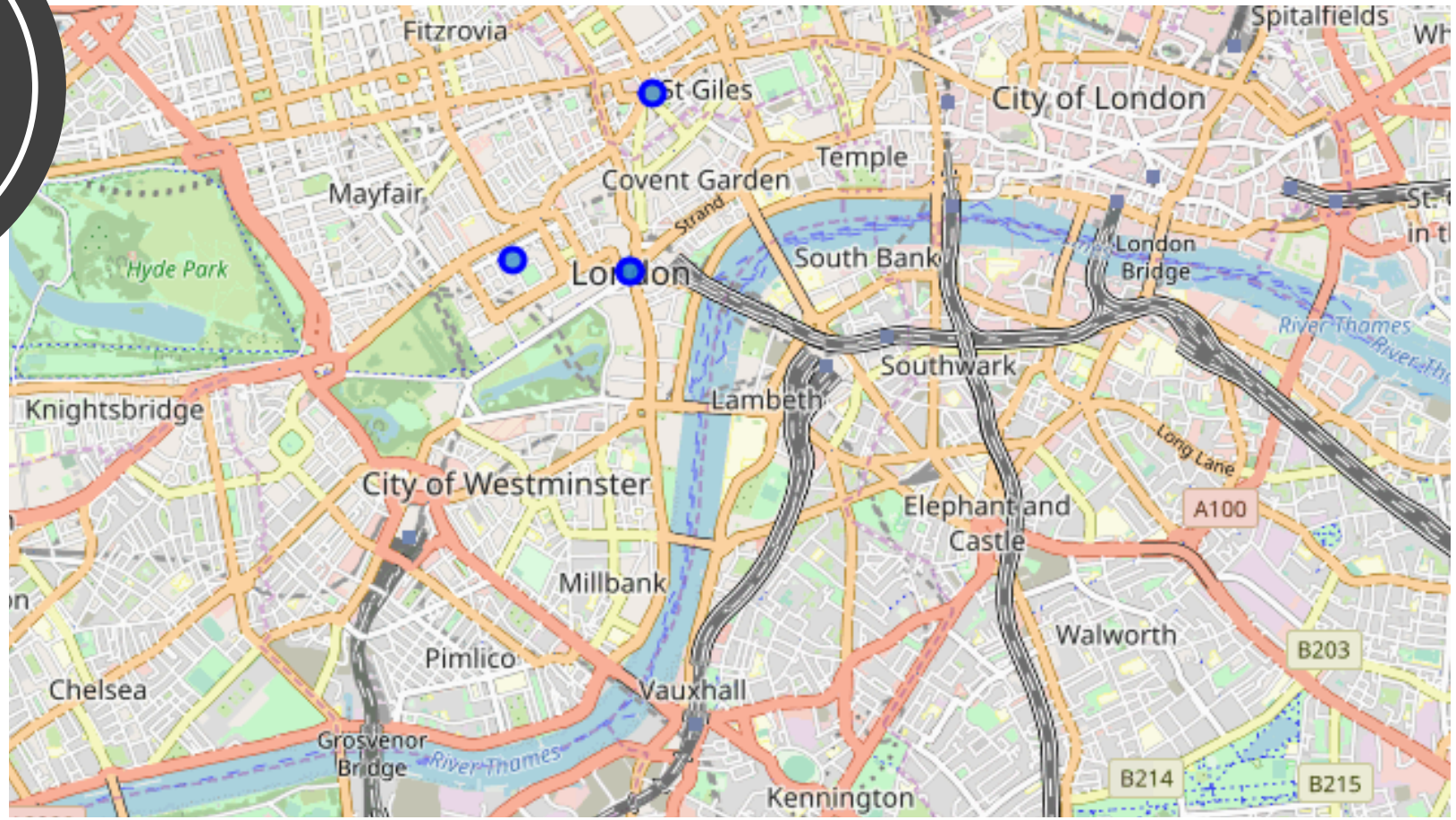




RESULTS: EXAMINE LONDON LOCATIONS FOR LIBRARY HOTEL

	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

RESULTS: EXAMINE LONDON LOCATIONS FOR LIBRARY HOTEL



DISCUSSION

Recommendation to the Library Hotel:

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.

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, then a commercial agreement for API geocoding services will be required.

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.

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

THANK YOU