London vs Manhattan

A SOCIAL VENUES BASED COMPARATIVE ANALYSIS

Business Scenario

The owner of a successful restaurant chain in London is looking to expand in New York and wants to identify neighbourhoods similar to their current locations in London as they have been very successful in running their business there.

The goal is to identify similar neighbourhoods in Manhattan, New York to launch their next two restuarants. The similarity comparision of the neighbourhoods whould be in terms of the other business/social establishments that exist in the proposed neighbourhoods.

The chain has two categories of restuarants - the first one is in Covent Garden, City of London, London and , it is a fine dining restuarant and offers Indian cuisine, the second one is in Southall, in the Ealing borough , London and offers Indian fast/street food. Both cater to very different clientele and therefore it is important for the business to identify the right kind of neighbourhoods in Manhattan for each restaurant.

Problem Statement

NO.	QUESTION
1.	Is it possible to segment London and Manhattan into neighbourhood clusters and then identify clusters that extend across both cities thereby implying similarity in the neighbourhoods in London and Manhattan?
2.	Do the neighbourhoods of the two restaurants in London fall in different clusters? If yes, for each cluster are we able to identify similar neighbourhoods in Manhattan to expand the restaurant business?

Data Sources and Approach

The data that will be used for this analysis will include four square data to analyze the different neighbourhoods in New York and London and combine them into one single dataset of neighbourhoods , then segment this dataset into different clusters and identify which clusters Southall and Covent Garden fall into respectively and are there neighbourhoods in New York which fall in the same clusters. If there are, then these could potentially be good locations for the launch of the new restuarants.

We will start with identifying the names of all the boroughs of London by webscraping the following page : https://en.wikipedia.org/wiki/List_of_London_boroughs and the neighbourhoods in Manhattan from a pre-existing dataset used earlier in this course. After that we use this list to gather geospatial data like latitude and longitude for each borough/neighbourhood in each city which is used for gathering social venues information from Four Square through an API request.

The main data to be used for this analysis is Four Square data which typically includes the following fields for each location in a given Neighbourhood: Name ,Neighborhood Latitude, Neighborhood Longitude, Venue name, Venue Latitude, Venue Longitude, Venue Category, for top 100 venues within a 500 metre radius.

After identifying the social venues data we will use the K Means algorithm to cluster the data into similar groups and see if we can define the characteristics if each group of neighbourhoods , and then make final recommendations on the locations of the restaurants to be launched in Manhattan..

Data Analysis

London has 32 boroughs and the area of the City of London. Manhattan is itself a borough of New York and has 40 Neighbourhoods. To avoid confusion we will refer to both London boroughs and Manhattan Neighbourhoods as Neighbourhoods going forward.

	Borough	Neighborhood	Latitude	Longitude
6	Manhattan	Marble Hill	40.876551	-73.910660
100	Manhattan	Chinatown	40.715618	-73.994279
101	Manhattan	Washington Heights	40.851903	-73.936900
102	Manhattan	Inwood	40.867684	-73.921210
103	Manhattan	Hamilton Heights	40.823604	-73.949688
104	Manhattan	Manhattanville	40.816934	-73.957385
105	Manhattan	Central Harlem	40.815976	-73.943211
106	Manhattan	East Harlem	40.792249	-73.944182
107	Manhattan	Upper East Side	40.775639	-73.960508
108	Manhattan	Yorkville	40.775930	-73.947118
109	Manhattan	Lenox Hill	40.768113	-73.958860
110	Manhattan	Roosevelt Island	40.762160	-73.949168
111	Manhattan	Upper West Side	40.787658	-73.977059
112	Manhattan	Lincoln Square	40.773529	-73.985338
113	Manhattan	Clinton	40.759101	-73.996119
114	Manhattan	Midtown	40.754691	-73.981669
115	Manhattan	Murray Hill	40.748303	-73.978332
116	Manhattan	Chelsea	40.744035	-74.003116
117	Manhattan	Greenwich Village	40.726933	-73.999914
118	Manhattan	East Village	40.727847	-73.982226
119	Manhattan	Lower East Side	40.717807	-73.980890
120	Manhattan	Tribeca	40.721522	-74.010683
121	Manhattan	Little Italy	40.719324	-73.997305
122	Manhattan	Soho	40.722184	-74.000657
123	Manhattan	West Village	40.734434	-74.006180
124	Manhattan	Manhattan Valley	40.797307	-73.964286
125	Manhattan	Morningside Heights	40.808000	-73.963896
126	Manhattan	Gramercy	40.737210	-73.981376
127	Manhattan	Battery Park City	40.711932	-74.016869
128	Manhattan	Financial District	40.707107	-74.010665
247	Manhattan	Carnegie Hill	40.782683	-73.953256
248	Manhattan	Noho	40.723259	-73.988434
249	Manhattan	Civic Center	40.715229	-74.005415
250	Manhattan	Midtown South	40.748510	-73.988713
271	Manhattan	Sutton Place	40.760280	-73.963556
273	Manhattan	Turtle Bay	40.752042	-73.967708
274	Manhattan	Tudor City	40.746917	-73.971219
275	Manhattan	Stuyvesant Town	40.731000	-73.974052
276	Manhattan	Flatiron	40.739673	-73.990947
301	Manhattan	Hudson Yards	40.756658	-74.000111
0	London	Barking and Dagenham	51.560228	0.171961
1	London	Barnet	51.627300	-0.253760
2	London	Bexley	51.452078	0.069931
3	London	Brent	51.609783	-0.194672
4	London	Bromley	51.601511	-0.066365
5	London	Camden	51.591180	-0.165040
6	London	Croydon	51.593470	-0.083380
7	London	Ealing	51.508383	-0.305200
8	London	Enfield	51.540024	-0.077502
9	London	Greenwich	51.477890	-0.013340
10	London	Hackney	51.531820	-0.061780
11	London	Hammersmith and Fulham	51.482600	-0.212880
12	London	Haringey	51.589270	-0.106405
13	London	Harrow	51.513180	-0.106980
14	London	Havering	51.544610	-0.144260
15	London	Hillingdon	51.484230	-0.096477
16	London	Hounslow	51.467701	-0.361718
17	London	Islington	51.534380	-0.108940
18	London	Kensington and Chelsea	51.522660	-0.207930
19	London	Kingston upon Thames	51.409008	-0.303598
20	London	Lambeth	51.494471	-0.120066
21	London	Lewisham	51.465280	-0.013210
22	London	Merton	51.544520	-0.166860
23	London	Newham	51.519937	0.055882
24	London	Redbridge	51.475773	-0.080698
25	London	Richmond upon Thames	51.480270	-0.237540
26	London	Southwark	51.505734	-0.100002
27	London	Sutton	51.512243	-0.053659
28	London	Tower Hamlets	51.499990	-0.010450
29	London	Waltham Forest	51.581765	-0.276968
30	London	Wandsworth	51.467826	-0.144992
31	London	Westminster	51.628249	0.012986
32	London	City of London	51.508530	-0.125740

London Neighbourhoods

Manhattan Neighbourhoods

Data Analysis

- Even though the average area of Manhattan neighbourhoods is much smaller than London , the concentration of venues in Manhattan neighbourhoods is high, therefore in terms of number of social venues within a neighbourhood, London neighbourhoods and Manhattan neighbourhoods can be considered comparable.

Data Analysis

- Certain venue categories had to be modified to accommodate differences in the way certain establishments are named in UK vs USA. Eg a bar in USA is typically referred to as a pub in UK, a coffee shop in USA is likely to be called a café in UK etc.
- Also venues like beer bar, wine bar, cocktail bar etc were all clubbed into one single category named bar
- There were other similar anomalies in the data eg Gym, Gym/Fitness Centre/Gym Pool/ Gymanstics Gym – but since the frequency of these venues was low, so these were not corrected.

Data Summary

- In all 73 Neighbourhoods were analyzed
- The number of venues identified across these 73 neighbourhoods was 4392
- The features used for K means clustering were the venue categories. A total of 371 unique venue categories were identified
- The neighbourhoods where number of venues identified was less than 10 were excluded from the analysis

K Means Classification: Approach

- Data was prepared for K means Classification algorithm by using the one hot coding technique for all the venue categories and then grouping by neighbourhood and calculating the mean values
- The number of clusters was defined as 5
- For each neighborhood the top 10 most frequently occurring venue categories were identified

K Means Classification

- The algorithm generated these 5 Clusters:

Cluster Name	Details	Dimensions
Cluster 0	These neighbourhoods are dominated by Bars and Coffee Shops , with the top two most frequent venue categories being bars and coffee shops in 10 out of 13 neighbourhoods. The remaining top categories mostly included restaurants. These appear to be vibrant social venues.	13 Neighbourhoods : 2 Manhattan and 11 London
Cluster 1	These neighbourhoods are a mix of restaurants, coffee shops and bars alongwith grocery stores, gyms, parks as the other frequent categories so they appear to be areas that cater to amenities for local residents as well as socialising venues.	36 Neighbourhoods : 29 Manhattan and 7 London

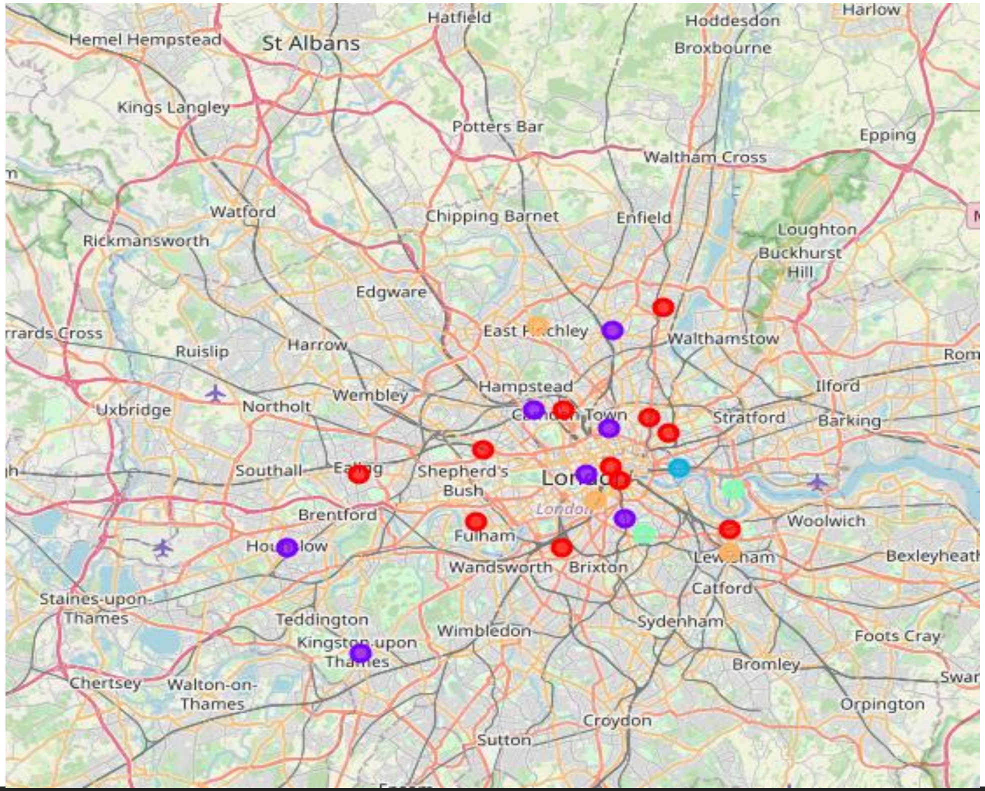
K Means Classification

- The algorithm generated these 5 Clusters :

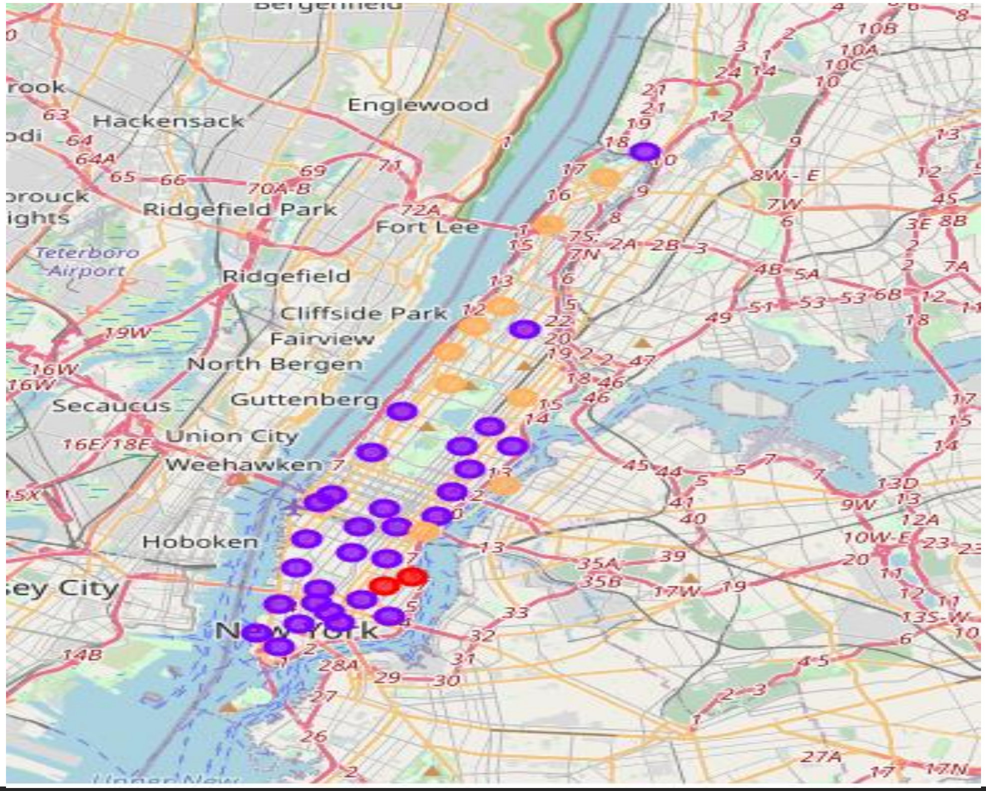
Cluster Name	Details	Dimensions
Cluster 2	Just a single neighbourhood – so we will ignore this cluster	1 Neighbourhood : 0 Manhattan and 1 London
Cluster 3	Just two neighbourhoods, with the most frequent category being a bus stop – so we will ignore this cluster	2 Neighbourhoods : 0 Manhattan and 2 London
Cluster 4	In this cluster 10 out of 12 neighbourhoods had Coffee Shops as the most frequently occurring social venue	12 Neighbourhoods : 9 Manhattan and 3 London

K Means Classification

LONDON CLUSTERS



MANHATTAN CLUSTERS



K Means Classification: Results

Revisiting our original problem statement :

1. ***Is it possible to segment London and Manhattan into neighbourhood clusters and then identify clusters that extend across both cities thereby implying similarity in the neighbourhoods in London and Manhattan?*** Yes : The algorithm was able to classify the combined dataset of London and Manhattan neighbourhoods into 5 Clusters. 2 of the 5 had very few datapoints so these clusters were disregarded. 3 Clusters extend across both London and Manhattan with a good mix of neighbourhoods from both cities.

2. ***Do the neighbourhoods of the two restaurants in London fall in different clusters? If yes, for each cluster are we able to identify similar neighbourhoods in Manhattan to expand the restaurant business?*** Yes : Ealing and City of London were the two neighbourhoods that we were analyzing. Ealing falls in Cluster 0 and City of London falls in Cluster 1 and we are able to identify similar neighbourhoods in Manhattan.

K Means Classification: Results

- ✓ Neighbourhoods similar to Ealing where we could consider launching the Indian street food restaurant are : Stuyvesant Town and East Village
- ✓ Neighbourhoods similar to City of London where we could consider launching the Indian fine dining restaurant are : Battery Park City, Financial District, Carnegie Hill, Noho, Civic Center , Midtown South, Sutton Place, Turtle Bay, Flatiron ,Hudson Yards

Other Considerations

1. One of the limitations of this approach due to time constraints and API request limits has been that we have only requested data for the top 100 venues in a 500 metre radius. Since London is less concentrated than Manhattan therefore this study needs to be redone with a larger radius for London neighbourhoods to get more accurate results

2. Another point to bear in mind is that we have taken the latitude and longitude coordinates of the neighbourhoods as the starting point to define the area around it. These coordinates may not always be the best points to identify the social venues around it as venues may not be distributed evenly around these coordinates .

3. Further analysis can be carried out by exploring the venues list differently based on the above considerations to fine tune this study.

4. Additionally multiple other factors like real estate cost, population density , presence of other Indian restaurants and their ratings can be used to further drill down into the identified neighbourhoods and provide a more precise list of possible options to consider.