How Similar are Neighbourhoods in London, Sydney and New York?

Aaron Armour

Introduction

Deciding to shift to a new city is a big decision to make, even more so when that city is in another country. Imagine that you are in the situation of moving overseas and deciding where to live. An important consideration for you when making this decision might be to find an area which is similar, in terms of the amenities on offer, to where you live presently. In addition to the neighbourhood you choose to live in, you are likely to place some importance on the surrounding neighbourhoods within in the city as well. That is, you would also appreciate a degree of similarity of your new city with your current city.

The audience for this project is people who are planning a shift between two of these cities, and are considering where to live in the new city. This project will also be of interest to people living in one of these three cities and considering which of the other two cities they might like to shift to.

London, Sydney and New York are all the financial capitals of their respective countries. There will be many workers in the banking, insurance and financial services sector (amongst others) who shift between two of these cities. So I expect the findings of this project to be of use to a large group of people.

There are other factors to take into account in deciding where to shift, but this project will provide relevant information for one important aspect of this decision.

I intend to answer the following three specific questions in this project:

1. How similar are neighbourhoods across the different cities; are there similar neighbourhoods in other cities? Or do neighbourhoods tend to be most like other neighbourhoods within that same city?
2. For each neighbourhood under consideration which are the most similar neighbourhoods, one from each of the other two cities?
3. How similar are each pair of cities? (This would allow us to say Sydney is more similar to London than New York, for example.)

Data

In order to limit the number of neighbourhoods under consideration, I have chosen to focus on the central city areas, the inner boroughs of London, the City of Sydney and Manhattan. I will use the following pages on Wikipedia to obtain lists of neighbourhoods in each of the three cities:

- https://en.wikipedia.org/wiki/List_of_areas_of_London
- https://en.wikipedia.org/wiki/City_of_Sydney
- https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods

As an example, the Wikipedia page List of areas of London has the following table:

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
Aldborough Hatch	Redbridge[9]	ILFORD	IG2	020	TQ455895
Aldgate	City[10]	LONDON	EC3	020	TQ334813
Aldwych	Westminster[10]	LONDON	WC2	020	TQ307810
Alperton	Brent[11]	WEMBLEY	HA0	020	TQ185835
Anerley	Bromley[11]	LONDON	SE20	020	TQ345695
Angel	Islington[8]	LONDON	EC1, N1	020	TQ345665
Aperfield	Bromley[11]	WESTERHAM	TN16	01959	TQ425585
Archway	Islington[12]	LONDON	N19	020	TQ285875
Ardleigh Green	Havering[12]	HORNCHURCH	RM11	01708	TQ535895
Arkley	Barnet[12]	BARNET, LONDON	EN5, NW7	020	TQ225955
Arnos Grove	Enfield[12]	LONDON	N11, N14	020	TQ295925
Balham	Wandsworth[13]	LONDON	SW12	020	TQ285735
Bankside	Southwark[14]	LONDON	SE1	020	TQ325795

The Wikipedia page https://en.wikipedia.org/wiki/List_of_places_in_London has a table which classifies each London borough as inner or outer. This information can be used to filter the London neighbourhoods to just those belonging to the inner boroughs.

I used Python's geopy module to obtain the latitude and longitude of each neighbourhood. When this failed I manually obtained the information by using a google search. An example showing how to use geopy is as follows:

In [2]:
```
from geopy.geocoders import Nominatim
locator = Nominatim(user_agent='My_Coursera_Applied_Data_Science_Capstone')

place = locator.geocode('Canary Wharf, London')
place
```

Out[2]:
```
Location(Canary Wharf, Isle of Dogs, London Borough of Tower Hamlets, London, Greater London, England, E14 4HE, United Kingdom, (51.50562, -0.0257
169, 0.0))
```

In [3]:
```
lat = place.latitude
long = place.longitude
print(lat, long)
```

```
51.50562 -0.0257169
```

I used location data from Foursquare, making API calls to their 'venue explore' endpoint to obtain a list of venues from a given neighbourhood. The results obtained from Foursquare are a list of recommended places, below is an example of such a recommendation. The key piece of information I extracted from this data the name of the category that the venue belongs to, which is 'Plaza' in the example below.


```
{'reasons': {'count': 0,
   'items': [{'summary': 'This spot is popular',
     'type': 'general',
     'reasonName': 'globalInteractionReason'}]},
  'venue': {'id': '4c4991b19f2ad13a0fd77653',
   'name': 'Canary Riverside',
   'location': {'address': 'Canary Riverside',
    'lat': 51.50644504299903,
    'lng': -0.028794868546109256,
    'labeledLatLngs': [{'label': 'display',
      'lat': 51.50644504299903,
      'lng': -0.028794868546109256}],
```

```
   'distance': 232,
   'postalCode': 'R M14',
   'cc': 'GB',
   'city': 'Canary Wharf',
   'state': 'Greater London',
   'country': 'United Kingdom',
   'formattedAddress': ['Canary Riverside',
    'Canary Wharf',
    'Greater London',
    'R M14',
    'United Kingdom']},
  'categories': [{'id': '4bf58dd8d48988d164941735',
    'name': 'Plaza',
    'pluralName': 'Plazas',
    'shortName': 'Plaza',
    'icon': {'prefix':
'https://ss3.4sqi.net/img/categories_v2/parks_outdoors/plaza_',
     'suffix': '.png'},
    'primary': True}],
  'photos': {'count': 0, 'groups': []}},
 'referralId': 'e-0-4c4991b19f2ad13a0fd77653-0'}
```

These results were used to populate a DataFrame where each row contains a neighbourhood and the category of a venue in that neighbourhood returned by the Foursquare query. I applied one-hot encoding to convert the categories into dummy variables. I then grouped by neighbourhood and averaged to obtain a vector of observed proportions of venues in each category for a given neighbourhood. In this way, I obtained a feature vector for each neighbourhood. (This technique was taught in the week 3 lab of the Applied Data Science Capstone on Coursera.org.)

I aimed to have data on 50-70 neighbourhoods in each of the three cities. I dropped neighbourhoods where there are too few venues returned by Foursquare.

Using the information from the Wikipedia pages I obtained 187 neighbourhoods from the Inner Boroughs of London, 48 neighbourhoods from the City of Sydney and 85 Neighbourhoods from Manhattan, New York. Balancing the need for a meaningful number of data points against the quality of those data points, I decided to keep only Sydney neighbourhoods with 30 or more venues obtained from Foursquare. For London and New York I chose to keep only neighbourhoods where Foursquare returned 50 or more venues. This meant I ended up with 61 neighbourhoods from London, 30 from Sydney and 62 from New York. In total, across these 153 neighbourhoods the data obtained from Foursquare had venues from 404 unique venue categories.

Methodology

To address question 1, I chose to apply the method of clustering. I used the K-Means algorithm with the number of clusters set to three, being the number of cities we are examining. If the neighbourhoods in each city are most similar to other neighbourhoods in that city then we would expect to see a separate cluster for each city. This provided the motivation for choosing the number of clusters to be three.

As mentioned earlier, the way we are measuring similarity of neighbourhoods is by the types of amenities they offer. Each neighbourhood has an associated feature vector which is the proportions of the different venues available in that neighbourhood. Now, to be more precise, the similarity of two neighbourhoods is measured by the (Euclidean) distance between the features vectors corresponding to the two neighbourhoods. For brevity, I shall

simply refer to distances between neighbourhoods, however it is important to remember this means *distance between their corresponding feature vectors, and not any sort of geographic distance*.

Prior to attempting to answer question 2, I calculated the average distances between all neighbourhoods in each of the three cities. This gives a yardstick which shows how similar the neighbourhoods within a given city are.

For each pair of cities, I calculated the matrix of distances between neighbourhoods in the first city from neighbourhoods in the second city. Then for a given neighbourhood, I found the most similar neighbourhood in a different city by finding which entry was the smallest in a row or column of the relevant distance matrix. For instance to find which neighbourhood in Sydney was most similar to Highbury in London, I would look at the row for Highbury in the London-Sydney matrix of distances, find the smallest element in this row and the column which this element lies in would tell me which neighbourhood in Sydney is most similar to Highbury.

Finally to answer question 3, I used the distance matrices to compute two similarity measures at the city level. The first measure was simply an average of the distances of the neighbourhoods in one city from the neighbourhoods in the other.

The second measure of similarity, which I have termed “matching distance”, is a little more complex. For a pair of cities, say London and Sydney, consider pairing neighbourhoods (one from London, one from Sydney) in such a way that each neighbourhood belongs to at most one pair (i.e. once a neighbourhood has been used it cannot be repeated) and in such a way that the sum of distances between neighbourhoods in a pair is minimised. This minimal sum of distances is then averaged by dividing by the number of neighbourhoods in the city which has the lesser neighbourhoods (since this will be the number of neighbourhood pairings between the two cities as neighbourhoods cannot be repeated). This matching distance measures the similarity of the two cities as a whole, considering their neighbourhoods in aggregate and simultaneously.

For both of these measures, a lower value is interpreted as meaning “more similar”. So if either measure was 0.2 for City1 and City2, but the measure for City1 and City3 was 0.5 then we would consider City1 more similar to City2 than to City3.

Results

Question 1: similarity of neighbourhoods across different cities

The table below shows the results of the clustering of the neighbourhoods:

Cluster label	London	Sydney	New York
0	4	28	0
1	39	0	0
2	18	2	62

Table 1: number of neighbours in each cluster by city

London appears the most diverse with neighbourhoods belonging to each cluster. On the other hand, New York neighbourhoods all belong to a single cluster.

The images below display these clusters on a map of each city. Red markers represent

cluster 0, blue markers represent cluster 1, and green markers represent cluster 2.

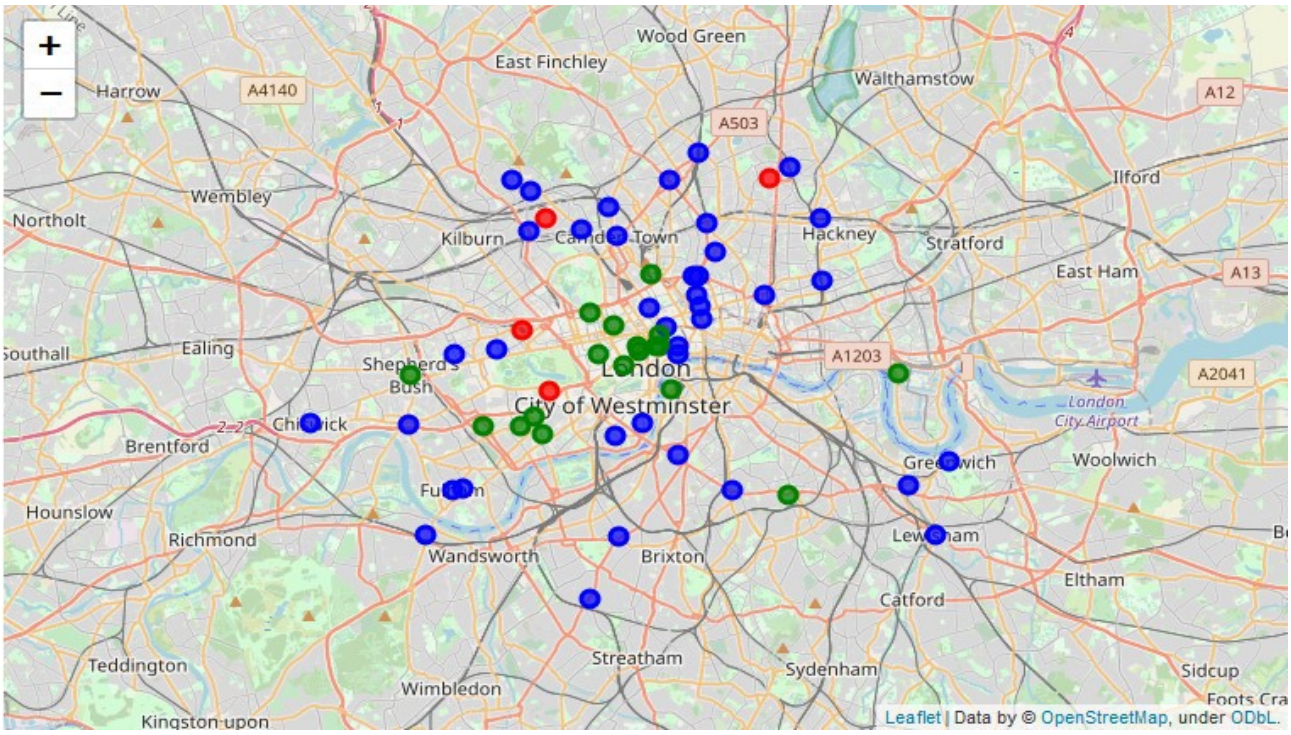

Image 1: Clusters within London

It is interesting to note that some of the London neighbourhoods in cluster 2 appear to be geographically close to each other.

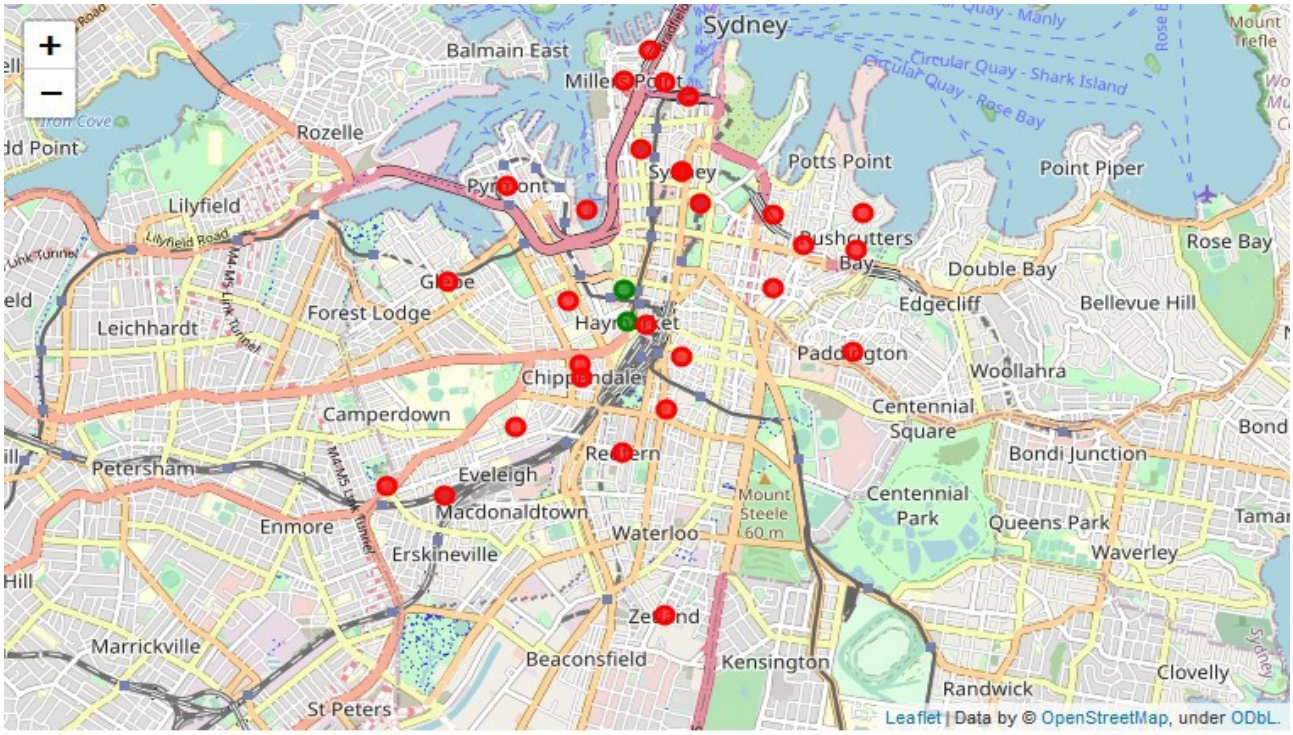

Image 2: Clusters within Sydney

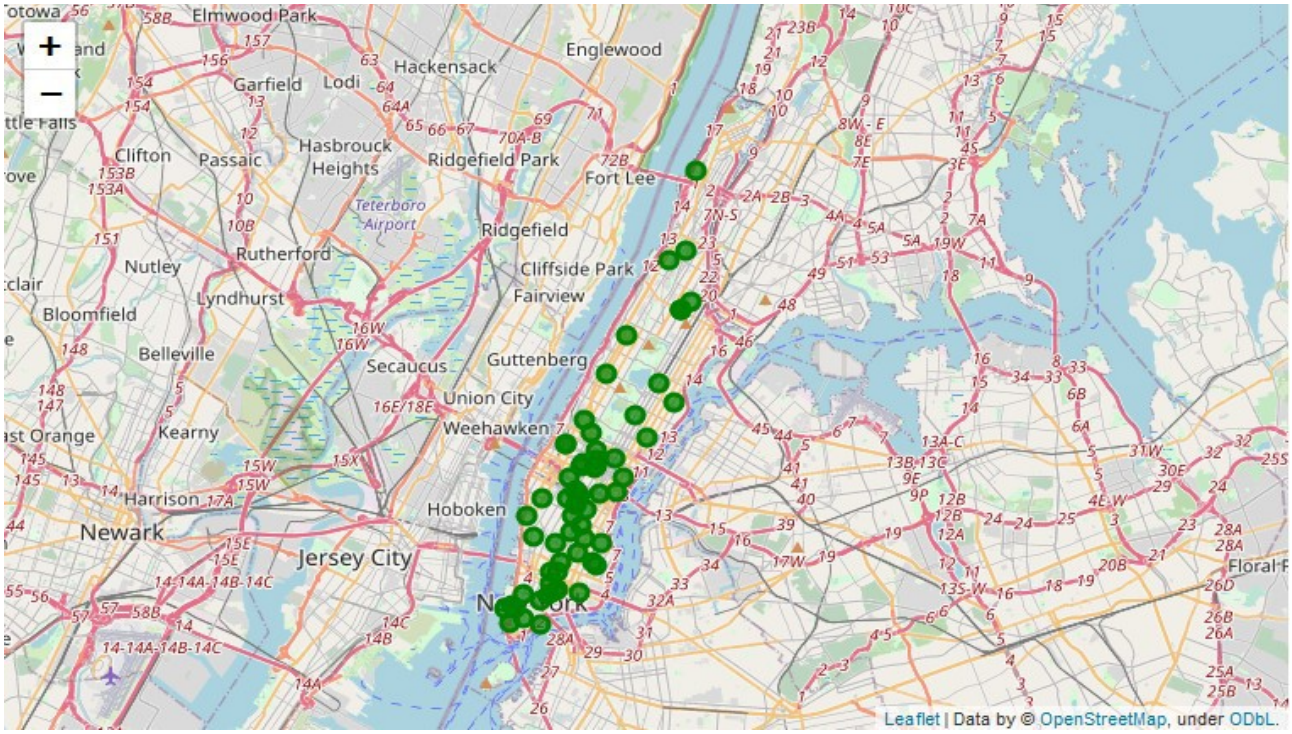

Image 3: Cluster within New York

Question 2: most similar neighbourhoods from other cities

The table below gives the average distance between (feature vectors of) neighbourhoods in each of the three cities. This also gives further evidence to support the claim that of these three cities London has the most variety.

	Intra city average neighbourhood feature distance
London	0.2116
Sydney	0.2011
New York	0.2000

Table 2: Average feature distance between neighbours in each city

The tables below give a summary of the analysis to find for a given neighbourhood, the most similar neighbourhoods in the other two cities. The full results are presented in the appendix, which list each neighbourhood and the neighbourhood from each of the two other cities which are most similar to the given neighbourhood.

The table below tells us, for example, that Woolloomooloo in Sydney is the most similar neighbourhood to 24 neighbourhoods from London. We observe that, in general, the majority of neighbourhoods within a given city are most similar to a very small number of neighbourhoods from the other two cities. As an example of this, just two neighbourhoods - The Rocks and Woolloomooloo are the most similar neighbourhood for 41, or roughly two thirds, of London's neighbourhoods. If we also add East Sydney and Wynyard then these four neighbourhood account for 54, or roughly 90%, of London's neighbourhoods.

Sydney neighbourhood	Number of London neighbourhoods
Chinatown (Sydney)	1
Darling Harbour	1
Darlinghurst	3
East Sydney	7
Glebe	2
The Rocks	17
Woolloomooloo	24
Wynyard	6

London neighbourhood	Number of Sydney neighbourhoods
Bayswater	2
Brompton	1
Chalk Farm	14
Hackney Central	2
Hammersmith	2
Highbury	1
Knightsbridge	1
Shoreditch	3
Stoke Newington	2
West Hackney	2

Table 3: Most similar neighbourhoods between London and Sydney

New York neighbourhood	Number of London neighbourhoods
Central Harlem	4
Columbus Circle	1
Financial District	6
Flower District	3
Garment District	2
Little Australia	3
Lower East Side	6
Meatpacking District	1
Midtown	1
Midtown East	1
Murray Hill	1
NoHo	1
SoHo	4
Times Square	5
Tribeca	3
Tudor City	12
Turtle Bay	1
Upper East Side	1
Yorkville	5

London neighbourhood	Number of New York neighbourhoods
Brompton	2
Canary Wharf	4
Chalk Farm	2
Chinatown (London)	3
Hammersmith	1
Marylebone	5
Nag's Head	2
Putney	4
Shepherd's Bush	10
Shoreditch	20
South Kensington	2
St Giles	7

Table 4: Most similar neighbourhoods between London and New York

New York neighbourhood	Number of Sydney neighbourhoods
Bowery	2
Carnegie Hill	2
Financial District	2
Flower District	3
Hudson Heights	10
Koreatown	1
Little Australia	1
Lower East Side	4
Tudor City	5

Sydney neighbourhood	Number of New York neighbourhoods
Chinatown (Sydney)	13
East Sydney	30
Newtown	2
The Rocks	13
Woolloomooloo	3
Wynyard	1

Table 5: Most similar neighbourhoods between Sydney and New York

Question 3: similarity between pairs of cities

The two tables below provide the results for my two measures of similarity at the city level. Both measure lead to the same conclusion, that London and New York are the most similar cities of the three. Sydney is more similar to London than to New York.

	London	Sydney	New York
London	0.2116	0.2426	0.2246
Sydney		0.2011	0.2577
New York			0.2000

Table 6: Average neighbourhood feature distances

	London	Sydney	New York
London		0.1959	0.1919
Sydney			0.2236
New York			

Table 7: Average matching distance between cities

Discussion

While these results are useful in addressing the questions we have posed, it is important not to read too much into this analysis. We have only examined the similarity of neighbourhoods by the amenities on offer, there are other factors to consider.

I do have a couple of concerns about the approach of reducing the data down to a feature vector as this removes information. For instance two of the dummy variables may refer to French restaurants and Italian restaurant and these are more similar than say a hospital and a tourist attraction, yet in both cases the corresponding dummy variables are equally distinct from each other. Also, the granularity of the venue categories could affect the

results. In Foursquare some restaurants are categorised by cuisine at the country level, whereas others are broken into further categories, e.g. Japanese restaurants may be categorised as Ramen, or Sushi restaurants. The dummy variables corresponding to Ramen and Sushi restaurants are distinct, yet both are types of Japanese restaurant.

Conclusion

There are neighbourhoods which are more similar to neighbourhoods in other cities than to other neighbourhoods in the same city, as shown by the cluster analysis. London has more variety in its neighbourhoods according to my measure of similarity. Perhaps this is not surprising as it is likely to be the most multicultural city of the three given its proximity to a wide variety of other countries. My analysis has shown that for a given neighbourhood in a particular city, we can usually find a neighbourhood in each of the other two cities which is relatively close to this given neighbourhood (i.e. they have a feature distance similar to, or less than the average feature distance in the city the neighbourhood belongs to). I have also discovered that London and New York are the most similar cities of the three, according to my meaures.

Appendix

The tables below list the complete results for each neighbourhood; the cluster label, which neighbourhoods from the other two cities it is most similar to, and the feature distance to these most similar neighbourhoods.

London neighbourhood	Cluster labels	Sydney		New York	
		Neighbourhood	Feature distance	Neighbourhood	Feature distance
Aldwych	1	The Rocks	0.1951	Midtown	0.1939
Angel	1	Woolloomooloo	0.1866	Tudor City	0.1761
Balham	1	Wynyard	0.2026	Yorkville	0.1989
Bayswater	1	Woolloomooloo	0.1638	Tudor City	0.1700
Belsize Park	0	Woolloomooloo	0.1670	Tudor City	0.1782
Bethnal Green	1	Wynyard	0.1833	Financial District	0.1859
Bloomsbury	1	Wynyard	0.1760	Financial District	0.2000
Brompton	2	The Rocks	0.1681	Murray Hill	0.1573
Camberwell	1	Woolloomooloo	0.2029	Tudor City	0.1874
Camden Town	1	Woolloomooloo	0.1872	Lower East Side	0.1940
Canary Wharf	2	East Sydney	0.1802	Tribeca	0.1518
Chalk Farm	1	East Sydney	0.1425	Lower East Side	0.1615
Chinatown (London)	2	The Rocks	0.1949	Times Square	0.1442
Chiswick	1	Woolloomooloo	0.1942	Tudor City	0.1836
Clapham	1	Woolloomooloo	0.1756	Yorkville	0.1699
Clerkenwell	1	Woolloomooloo	0.1685	Financial District	0.1715
Covent Garden	2	The Rocks	0.2095	Times Square	0.1447
Deptford	1	The Rocks	0.1975	Flower District	0.1849
Earls Court	2	Darling Harbour	0.2404	Garment District	0.2344
Farringdon	1	Wynyard	0.2034	Turtle Bay	0.1847
Finsbury	1	Woolloomooloo	0.1912	Tudor City	0.2078
Finsbury Park	1	Woolloomooloo	0.1687	Little Australia	0.1929
Fitzrovia	2	Wynyard	0.1796	Financial District	0.1504
Frognal	1	Darlinghurst	0.1799	SoHo	0.1879
Fulham	1	Woolloomooloo	0.1958	SoHo	0.1850
Greenwich	1	The Rocks	0.2224	Central Harlem	0.1983
Hackney Central	1	The Rocks	0.1614	Lower East Side	0.1647
Hammersmith	1	Woolloomooloo	0.1584	Tudor City	0.1566
Hampstead	1	Darlinghurst	0.1848	SoHo	0.1935
Highbury	1	Glebe	0.1853	Tudor City	0.1902
Holborn	1	Woolloomooloo	0.1849	Tudor City	0.1916
Holloway	1	Woolloomooloo	0.2207	Tribeca	0.2197
Islington	1	Woolloomooloo	0.2014	Central Harlem	0.1973

London neighbourhood	Cluster labels	Sydney		New York	
		Neighbourhood	Feature distance	Neighbourhood	Feature distance
Kentish Town	1	Woolloomooloo	0.1969	Yorkville	0.1797
Knightsbridge	0	Woolloomooloo	0.1813	Little Australia	0.1898
Lambeth	2	East Sydney	0.1673	Financial District	0.1685
Lewisham	1	East Sydney	0.2250	Central Harlem	0.2051
Marylebone	2	Woolloomooloo	0.1705	Lower East Side	0.1441
Mayfair	2	The Rocks	0.2373	Meatpacking District	0.1903
Millbank	1	East Sydney	0.1726	Tudor City	0.1701
Nag's Head	2	The Rocks	0.2052	Times Square	0.1360
Notting Hill	1	Woolloomooloo	0.2036	Yorkville	0.1797
Oval	1	Woolloomooloo	0.1980	Tudor City	0.2236
Paddington (London)	0	Woolloomooloo	0.1927	Garment District	0.2395
Parsons Green	1	Woolloomooloo	0.1860	SoHo	0.1771
Peckham	2	The Rocks	0.2121	NoHo	0.1855
Pentonville	1	Woolloomooloo	0.1679	Tudor City	0.1592
Pimlico	1	Woolloomooloo	0.2247	Midtown East	0.2411
Putney	1	East Sydney	0.1752	Yorkville	0.1487
Shepherd's Bush	2	Chinatown (Sydney)	0.1990	Central Harlem	0.1523
Shoreditch	1	Wynyard	0.1517	Financial District	0.1304
Soho	2	The Rocks	0.2150	Times Square	0.1559
South Kensington	2	The Rocks	0.1723	Little Australia	0.1497
St Giles	2	The Rocks	0.1724	Columbus Circle	0.1549
St James's	2	The Rocks	0.2276	Upper East Side	0.1771
St Luke's	2	East Sydney	0.2168	Tribeca	0.1661
St Pancras	2	The Rocks	0.1738	Flower District	0.1630
Stoke Newington	0	Darlinghurst	0.1848	Lower East Side	0.2105
Swiss Cottage	1	The Rocks	0.1707	Flower District	0.1615
Temple	1	The Rocks	0.2191	Times Square	0.1754
West Hackney	1	Glebe	0.1690	Lower East Side	0.1932

Table 8: Full results for London neighbourhoods

Sydney neighbourhood	Cluster labels	London		New York	
		Neighbourhood	Feature distance	Neighbourhood	Feature distance
Central Park	0	Chalk Farm	0.1544	Lower East Side	0.1959
Chinatown (Sydney)	2	Brompton	0.1881	Koreatown	0.1718
Chippendale	0	Chalk Farm	0.1505	Lower East Side	0.1992
Circular Quay	0	Hackney Central	0.1682	Lower East Side	0.1908
Darling Harbour	0	Shoreditch	0.1753	Tudor City	0.1864
Darlinghurst	0	Chalk Farm	0.1652	Tudor City	0.1988
Darlington	0	Highbury	0.2189	Tudor City	0.2890
Dawes Point	0	Hackney Central	0.1711	Flower District	0.2048
East Sydney	0	Chalk Farm	0.1425	Flower District	0.1679
Elizabeth Bay	0	Chalk Farm	0.2015	Hudson Heights	0.2268
Glebe	0	West Hackney	0.1690	Hudson Heights	0.2120
Haymarket	2	Bayswater	0.2093	Bowery	0.1922
Kings Cross	0	Chalk Farm	0.1835	Hudson Heights	0.2209
Macdonaldtown	0	West Hackney	0.2089	Hudson Heights	0.2434
Millers Point	0	Hammersmith	0.1942	Tudor City	0.2477
Newtown	0	Chalk Farm	0.1588	Lower East Side	0.1772
Paddington (Sydney)	0	Chalk Farm	0.1724	Hudson Heights	0.2234
Pyrmont	0	Chalk Farm	0.1746	Carnegie Hill	0.2390
Railway Square	0	Bayswater	0.1704	Bowery	0.1808
Redfern	0	Chalk Farm	0.2287	Hudson Heights	0.2838
Rushcutters Bay	0	Knightsbridge	0.2028	Carnegie Hill	0.2499
St James	0	Chalk Farm	0.1693	Tudor City	0.1825
Strawberry Hills	0	Stoke Newington	0.2183	Hudson Heights	0.2591
Surry Hills	0	Stoke Newington	0.2141	Hudson Heights	0.2584
Sydney CBD	0	Chalk Farm	0.1425	Flower District	0.1679
The Rocks	0	Shoreditch	0.1602	Financial District	0.1770
Ultimo	0	Chalk Farm	0.2124	Hudson Heights	0.2381
Woolloomooloo	0	Hammersmith	0.1584	Little Australia	0.1842
Wynyard	0	Shoreditch	0.1517	Financial District	0.1679
Zetland	0	Chalk Farm	0.2158	Hudson Heights	0.2278

Table 9: Full results for Sydney neighbourhoods

New York neighbourhood	Cluster Labels	London		Sydney	
		Neighbourhood	Feature distance	Neighbourhood	Feature distance
Alphabet City and Loisaida	2	Shoreditch	0.1470	East Sydney	0.1954
Astor Row	2	Shepherd's Bush	0.1831	East Sydney	0.2077
Battery Park City	2	St Giles	0.1746	East Sydney	0.1996
Bowery	2	Shepherd's Bush	0.1691	Chinatown (Sydney)	0.1737
Carnegie Hill	2	Chalk Farm	0.1698	East Sydney	0.1813
Central Harlem	2	Shepherd's Bush	0.1523	East Sydney	0.1970
Chelsea (New York)	2	Marylebone	0.4005	Woolloomooloo	0.4359
Chinatown (New York)	2	Shepherd's Bush	0.1860	Chinatown (Sydney)	0.1959
Civic Center	2	Shoreditch	0.1958	Chinatown (Sydney)	0.2005
Columbus Circle	2	St Giles	0.1549	East Sydney	0.1913
Diamond District	2	Chinatown (London)	0.1715	The Rocks	0.2274
Downtown Manhattan	2	Canary Wharf	0.1765	East Sydney	0.2052
East Village	2	Putney	0.1665	Chinatown (Sydney)	0.1801
Financial District	2	Shoreditch	0.1304	Wynyard	0.1679
Five Points	2	Shepherd's Bush	0.1832	Chinatown (Sydney)	0.2011
Flatiron District	2	Shoreditch	0.1515	East Sydney	0.1929
Flower District	2	Shoreditch	0.1435	East Sydney	0.1679
Garment District	2	Brompton	0.1707	The Rocks	0.2035
Gramercy Park	2	Shoreditch	0.1611	East Sydney	0.2022
Greenwich Village	2	Shoreditch	0.1640	East Sydney	0.1965
Hamilton Heights	2	Marylebone	0.1785	East Sydney	0.2053
Harlem	2	Shepherd's Bush	0.1523	East Sydney	0.1970
Hell's Kitchen	2	St Giles	0.1575	The Rocks	0.2011
Herald Square	2	Shoreditch	0.2199	Chinatown (Sydney)	0.1857
Hudson Heights	2	Chalk Farm	0.1662	East Sydney	0.1790
Koreatown	2	Shoreditch	0.1876	Chinatown (Sydney)	0.1718
Lenox Hill	2	Marylebone	0.1509	East Sydney	0.1892
Lincoln Square	2	St Giles	0.1714	East Sydney	0.1895
Little Australia	2	South Kensington	0.1497	Woolloomooloo	0.1842
Little Brazil	2	Chinatown (London)	0.1530	The Rocks	0.2162
Little Germany	2	Shoreditch	0.1497	Newtown	0.1960
Little Italy	2	Shepherd's Bush	0.1661	Chinatown (Sydney)	0.1721
Little Syria	2	Shoreditch	0.2098	The Rocks	0.2224
Lower East Side	2	Marylebone	0.1441	East Sydney	0.1697

New York neighbourhood	Cluster Labels	London		Sydney	
		Neighbourhood	Feature distance	Neighbourhood	Feature distance
Manhattan Valley	2	Shepherd's Bush	0.1725	East Sydney	0.2040
Meatpacking District	2	Shepherd's Bush	0.1594	East Sydney	0.1881
Midtown	2	St Giles	0.1562	The Rocks	0.2014
Midtown East	2	Shoreditch	0.1577	The Rocks	0.1952
Midtown West	2	St Giles	0.1575	The Rocks	0.2011
Murray Hill	2	Brompton	0.1573	Chinatown (Sydney)	0.1868
NoHo	2	Shoreditch	0.1649	Woolloomooloo	0.2136
NoMad	2	Shoreditch	0.2059	Chinatown (Sydney)	0.1937
Radio Row	2	Canary Wharf	0.1577	East Sydney	0.1930
Rockefeller Center	2	Chinatown (London)	0.1745	The Rocks	0.2297
Rose Hill	2	Canary Wharf	0.1790	Chinatown (Sydney)	0.2016
SoHo	2	South Kensington	0.1603	East Sydney	0.1838
South Street Seaport	2	Shoreditch	0.1416	East Sydney	0.1724
Stuyvesant Square	2	Putney	0.1671	Chinatown (Sydney)	0.1953
Stuyvesant Town	2	Shoreditch	0.1913	Newtown	0.2171
Sugar Hill	2	Shoreditch	0.2048	East Sydney	0.2117
Tenderloin	2	Shoreditch	0.1993	Chinatown (Sydney)	0.1753
Theater District	2	Nag's Head	0.1523	The Rocks	0.2460
Times Square	2	Nag's Head	0.1360	The Rocks	0.2243
Tribeca	2	Canary Wharf	0.1518	East Sydney	0.1791
Tudor City	2	Hammersmith	0.1566	East Sydney	0.1809
Turtle Bay	2	Marylebone	0.1568	The Rocks	0.1860
Union Square	2	Shoreditch	0.1619	East Sydney	0.1944
Upper East Side	2	Shoreditch	0.1658	The Rocks	0.2087
Upper West Side	2	Putney	0.1726	East Sydney	0.2014
West Harlem	2	Shepherd's Bush	0.1523	East Sydney	0.1970
West Village	2	St Giles	0.1783	East Sydney	0.2112
Yorkville	2	Putney	0.1487	East Sydney	0.1849

Table 10: Full results for New York neighbourhoods