IBM Capstone Project NY and London Neighbourhoods

Analysis of Neighborhoods for setting up chains of restaurants and cafes in London and New York

Problem Statement

- New York and London are two of the biggest profile cities on the World

- Setting up restaurant and café establishments in these cities would be costly but potentially very lucrative

- Potential to give high global brand recognition

- Where exactly to open these establishments, beyond the 'centre of town'

Approach to the Problem

- Foursquare (local venue information such as category)
- Python
  - Categorise neighbourhoods into clusters based on establishment type density
  - Map cities and neighbourhoods based on cluster

Methodology

1. Data Sources – Web scraping (Beautiful Soup) public (Wikipedia) data to find names of neighbourhoods in NY and London

2. Find Geolocation Data for Neighbourhoods

3. Use neighbourhood locations with Foursquare API to find venue information of each venue

4. Data Cleaning
   - Change venue types so different types of restaurant are restaurant only

5. Data preparation – create a profile for each neighbourhood with numbers of each type of establishment

Methodology continued…

6. K means Clustering – creates 5 cluster centres form the data
    *   Algorithm sorts the neighbourhoods into these clusters (Euclidean distance)
7. Define cluster centres in terms of number of key establishments
8. Group neighbourhoods for full list of similar neighbourhoods
9. Final report and recommendations

Results – Cluster Centres

Colour	Cluster	Characteristics
Red	Cluster 0	lowest number of Restaurants
Purple	Cluster 1	Highest number of Restaurants
Blue	Cluster 2	medium number of restaurants, no real prevalence of any other type of venue
Light Blue	Cluster 3	Second highest number of restaurants
Light Green	Cluster 4	Second lowest number of restaurants
Orange	Cluster 5	Medium number restaurant -, differentiated from Cluster 2 by having a significant number of cafes, pubs, and coffee shops

Results – Clusters of Interest

Purple - Cluster 1 - Highest number of Restaurants

Light Green - Cluster 4 - Second highest number of restaurants and high concentration of cafes and bars

Orange - Cluster 5 - Medium number restaurants and a good number of cafes and coffee shops

Cluster centre profiles (showing absolute numbers of each venue type):

```
cluster 1                        cluster 4                           cluster 5
                     1                               4                                  5
Restaurant      37.444444        Restaurant      24.000000           Restaurant      19.259259
Coffee Shop      3.277778        Coffee Shop      4.142857           Coffee Shop      5.592593
Bakery           3.000000        Pizza Place      3.142857           Pub              5.518519
Bar              2.944444        Bakery           2.591837           Café             4.703704
Pizza Place      2.666667        Café             2.346939           Bakery           3.037037
Hotel            1.888889        Bar              2.061224           Hotel            2.592593
Sandwich Place   1.777778        Deli / Bodega    1.428571           Burger Joint     1.925926
Cocktail Bar     1.777778        Hotel            1.346939           Clothing Store   1.629630
Dessert Shop     1.666667        Cocktail Bar     1.326531           Grocery Store    1.629630
Café             1.555556        Sandwich Place   1.285714           Pizza Place      1.555556
```

Recommendations for further consideration

From London and New York

New York Neighbourhoods Map
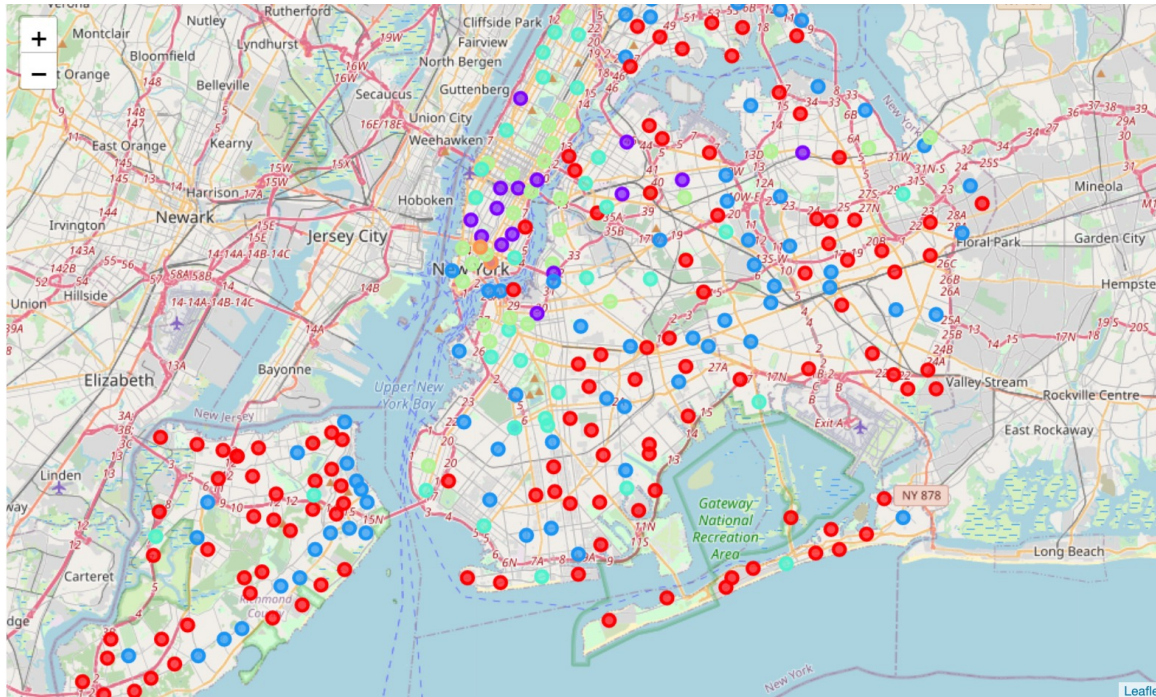
Purple - Cluster 1 - Highest number of Restaurants

Light Green - Cluster 4 - Second highest number of restaurants and high concentration of cafes and bars

Orange - Cluster 5 - Medium number restaurants and a good number of cafes and coffee shops

New York Neighbourhoods of Interest

Neighbourhood	City	Cluster
Astoria	New York	1
Chinatown	New York	1
Chinatown	New York	1
Clinton Hill	New York	1
East Village	New York	1
Flatiron	New York	1
Greenwich Village	New York	1
Jackson Heights	New York	1
Midtown South	New York	1
Murray Hill	New York	1
Murray Hill	New York	1
Noho	New York	1
South Side	New York	1
Sunnyside Gardens	New York	1
Turtle Bay	New York	1
Upper West Side	New York	1
West Village	New York	1
Brompton	London	1

Neighbourhood	City	Cluster
Bay Ridge	New York	4
Bayside	New York	4
Belmont	New York	4
Belmont	New York	4
Belmont	New York	4
Bushwick	New York	4
Carnegie Hill	New York	4
Carroll Gardens	New York	4
Chelsea	New York	4
Chelsea	New York	4
Chelsea	New York	4
Civic Center	New York	4
Clinton	New York	4
Cobble Hill	New York	4
Downtown	New York	4
Elmhurst	New York	4
Financial District	New York	4
Flushing	New York	4
Fort Greene	New York	4
Gramercy	New York	4

Neighbourhood	City	Cluster
Greenpoint	New York	4
Hamilton Heights	New York	4
Hunters Point	New York	4
Inwood	New York	4
Lenox Hill	New York	4
Little Italy	New York	4
Little Neck	New York	4
Midtown	New York	4
North Side	New York	4
Prospect Heights	New York	4
Soho	New York	4
Soho	New York	4
Sutton Place	New York	4
Tribeca	New York	4
Tudor City	New York	4
Upper East Side	New York	4
Washington Heights	New York	4
Woodside	New York	4
Woodside	New York	4
Yorkville	New York	4

Purple - Cluster 1 - Highest number of Restaurants

Light Green - Cluster 4 - Second highest number of restaurants and high concentration of cafes and bars

Orange - Cluster 5 - Medium number restaurants and a good number of cafes and coffee shops

London Neighbourhoods Map
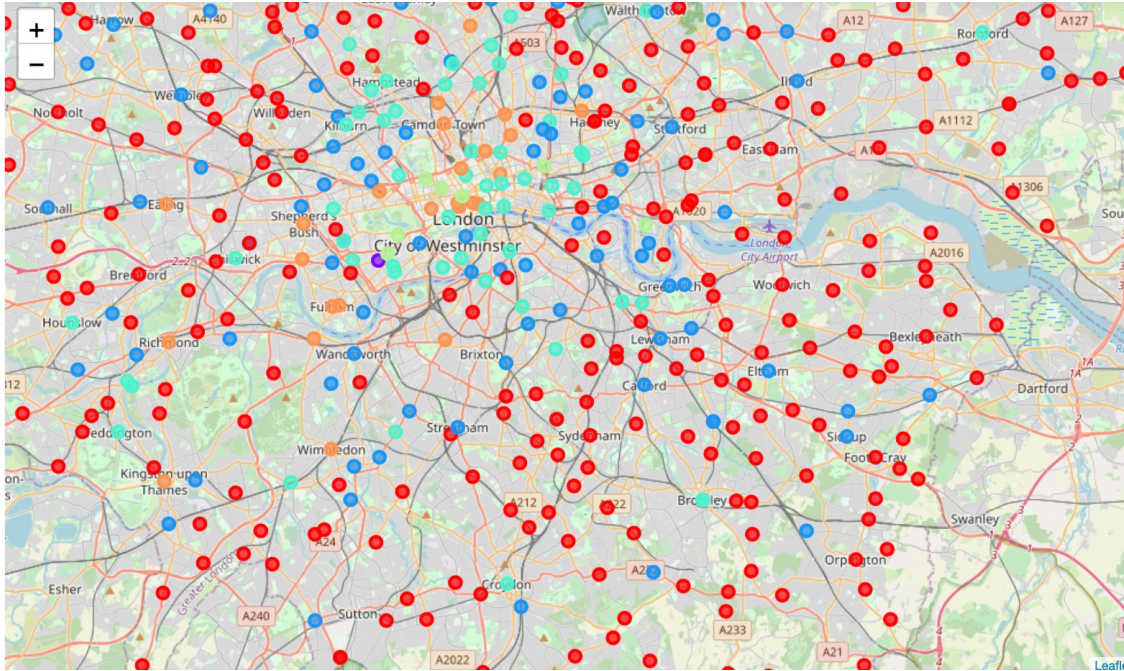
Purple - Cluster 1 - Highest number of Restaurants

Light Green - Cluster 4 - Second highest number of restaurants and high concentration of cafes and bars

Orange - Cluster 5 - Medium number restaurants and a good number of cafes and coffee shops

London Neighbourhoods of Interest

Neighbourhood	City	Cluster
Brompton	London	1

Canary Wharf	London	4
Church End	London	4
Church End	London	4
Fitzrovia	London	4
Knightsbridge	London	4
Marylebone	London	4

Chinatown	London	5
Chinatown	London	5
Soho	London	5
Soho	London	5
Aldwych	London	5
Camden Town	London	5
Clapham	London	5
Clerkenwell	London	5
Covent Garden	London	5
Crouch End	London	5
Ealing	London	5
Fulham	London	5
Hackney Central	London	5
Hammersmith	London	5
Highbury	London	5

Islington	London	5
Kentish Town	London	5
Kingston upon Thames	London	5
Mayfair	London	5
Paddington	London	5
Parsons Green	London	5
Pentonville	London	5
Putney	London	5
Richmond	London	5
Shepherd's Bush	London	5
St Pancras	London	5
Wimbledon	London	5

Purple - Cluster 1 - Highest number of Restaurants

Light Green - Cluster 4 - Second highest number of restaurants and high concentration of cafes and bars

Orange - Cluster 5 - Medium number restaurants and a good number of cafes and coffee shops

Conclusions

- London has very few neighborhoods in cluster 1

- New York has zero neighbourhoods in cluster 5

- New York might be better suited for opening restaurant chain whilst London may be better for Café chain

- Potential difficulty in analysis due to category types which might need better standardization
  - Pubs and bars are the same thing but due to cultural differences, are classed differently