



BATTLE OF NEIGHBORHOOD

Identifying suitable neighbourhoods in New York
and London

Table of Contents

1. Introduction	2
2. Data Acquisition, Cleaning and Evaluation Criteria	2
2.1. Data Source	2
2.2. Data Cleaning	3
2.3. Criteria of Data Evaluation	3
3. Methodology.....	4
3.1. Approach to analysis	4
3.2. Analysis of New York – Manhattan	4
3.3. Explore Query: Analysis of Manhattan Neighborhood	5
3.3.1. Explore: Clustering Data using K-means algorithm.....	5
3.3.2. Explore: Analysis of clusters.....	6
3.4. Search Query: Analysis of Manhattan Neighborhood	10
3.4.1. Search: Clustering Data using K-means algorithm	10
3.4.2. Search: Analysis of Clusters.....	11
3.5. Analysis of London – City of London and Westminster	16
3.5.1. Explore: Clustering Data using K-means algorithm.....	17
3.5.2. Explore: Analysis of Clusters	19
3.6. Search Query: Analysis of City of London and Westminster	24
3.6.1. Search: Clustering Data using K-means algorithm	24
3.6.2. Search: Analysis of Clusters.....	25
4. Results	30
4.1. Recommended Neighborhoods in Manhattan	30
4.2. Recommended Neighborhoods in City of London & City of Westminster	31
5. Observations	32
6. Conclusion.....	32

1. Introduction

New York and London are both very diverse cities and are the financial capitals of their respective countries and well-known financial centers in the world. These cities attract large investment primarily from companies within the Financial Services and FinTech sector. There are also corporate headquarters of large (Fortune 500) corporates and top consulting firms have their offices there. It is not difficult to imagine that some organization will have offices in both cities.

These organizations attract the best and the brightest of global talent and this has also meant that there is significant inward migration. Most people migrating to these countries and jobs are of a certain age group (21+ to 35 years of age) this is primarily due to recruitment methods and immigrations laws in the US and the UK.

People in the 21+ to 35 years of age group, also referred to as Gen Y or Gen Z, have different priorities and ways of living and these often influence their decision. They are looking for a different lifestyle centered around comfort living, they want to stay close to stores, bars and restaurants so they don't have to travel. The older among this age group might be young families and so they look for additional aspects like availability of nursery and possibly avenues to keep partners engaged (where required).

These people are the primary audiences of this report. In this report we will try to analyze the commercial hub within NY and London and recommend which neighborhoods are most suited for these individuals and should be considered should the option of migrating be available.

Online delivery has changed shopping patterns and this has impacted the traditional high street which is disappearing only to be replaced by high-rise office and residential buildings. This presents opportunities for various businesses. Large corporates can expand their workplace, other businesses from the constructions sector can consider the type of properties they want to develop and lastly businesses in the hospitality sector consider investing in these areas.

These businesses are the secondary audience of this report. They will be able to make their investment decisions based on location recommendation within this report.

All this investment will make the city more attractive for individuals from around the world to migrate, benefiting the organization they join, the local economy as well as global make global businesses more competitive.

2. Data Acquisition, Cleaning and Evaluation Criteria

2.1. Data Source

- A. We will obtain information about the neighborhoods via various sources, this will include wikipedia, Google Map and .json files provided during the course.
- B. Data required at this stage is Post Code, Borough, Neighborhood, Latitude and Longitude, information will be taken for various wikipedia, Google Map and .json files and doogal.co.uk has information about Latitude and Longitude for City of London and City of Westminster
- C. Data Wrangling methods will be applied to standardize the data and bring it into the right format for usage
- D. We will use Foursquare API's, to get information about these neighborhoods. We will use the:

- a. **Venue Explore Option:** Returns a list of recommended venues near the current location. This result will let us know what are the popular and what the people living within these locations like doing.
- b. **Venue Search Option:** This will Returns a list of venues near the current location, matching a search category ID that will be provided based on the criteria mentioned above. This will let us conclude if the comfort living that Gen Y and Gen Z are looking for is possible.

2.2. Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There are several challenges that will have to be managed.

- Latitude and Longitude were present in different formats. These were:
 - Decimal degree e.g. 40.446° N 79.982° W
 - Degrees and decimal minutes e.g. 40° 26.767' N 79° 58.933' W

These were all converted into float format i.e. 40.446 and -79.982 so it can be used with Folium and Foursquare.

- Latitude and Longitude data was collated from several websites and cleansed. We worked on postcode districts for London as opposed to individual area postcodes. This was to limit the data calls so they are within Foursquare free service.
- There were several problems in reading data from .json files and while executing data from the Search query the functions used to cleanse data from the Explore query didn't work. This requires multiple conversions to come in the right format.

2.3. Criteria of Data Evaluation

We will explore the neighborhoods within the key commercial areas in these cities, i.e. City of London & Westminster in London and Manhattan in New York. The scope of this report is limited to neighborhoods within these areas.

Findings of this report will be around suitability of places, criteria for preferred neighborhood are as follows:

- Primary Criteria:
 - Gen Y & Gen Z prefer comfort living, we will look at the availability of Residential Accommodation (High Priority) within the neighborhoods as this is to do with moving to these cities
- Secondary Criteria:
 - Presence of Schools
 - Access to Medical & Dental Facilities
 - Gym and Fitness Centers
 - Parks & Playgrounds
 - Childcare Services
 - Laundromat and Laundry Services
 - Presence of Lifestyle Venues
 - Shopping Malls

- Night Clubs and Lounge
- Pubs and Sports Bars
- Movie Theaters and Performing Arts Venues
- Restaurants

We want to pinpoint neighborhoods that provide a consistent quality of facilities and features within a short distance of the neighborhoods.

3. Methodology

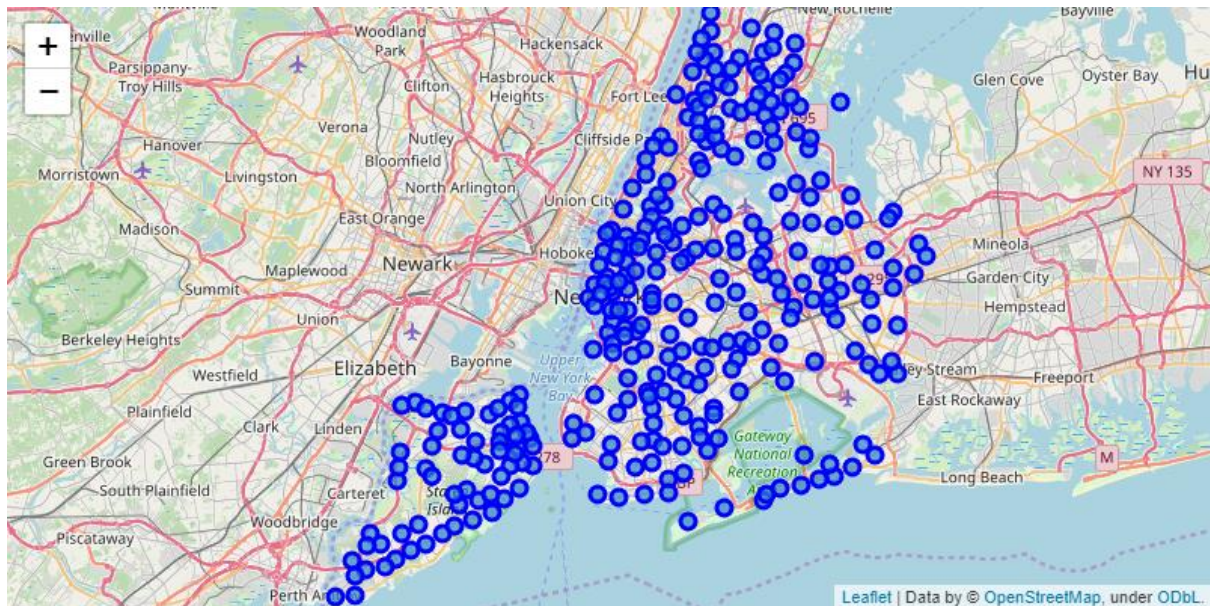
3.1. Approach to analysis

The following steps were followed in collating and analyzing the data:

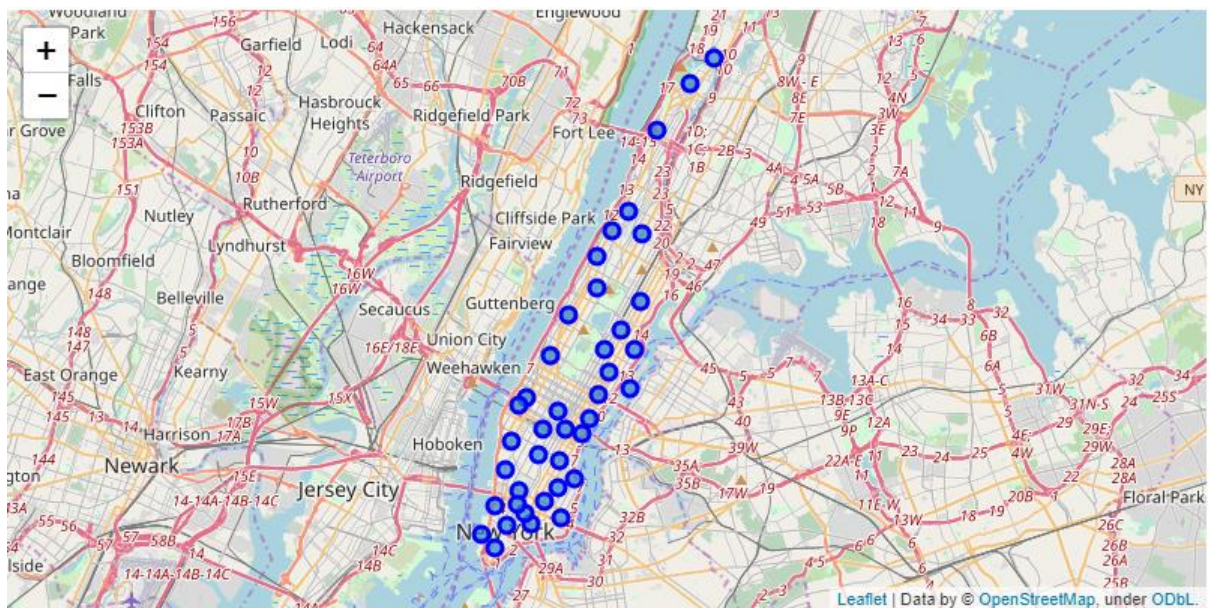
- A. We collated data from various sources and cleansed the data so its in the right format and is consistent
- B. We will explore this data, by visualizing it on Folium Maps to confirm we the right data or the right location
- C. We will use the Foursquare API's to get data about the neighborhood data from Search query and Explore query will be used
- D. We will adopt K-means Clustering as a way of further exploring the data and segmenting it into clusters with similar features. This will be done for all data received from both the Explore and the Search options
- E. The Clusters will be visualized on Folium Maps to ensure that the data is still intact and relevant to the analysis
- F. We will explore that data in each neighborhood further by looking into businesses and services within the neighborhood. Data will be visualized in bar charts to identify if the best clusters which meet the criteria of data evaluation.
- G. Cluster Selection:
 - a. Clusters from the Search query will be selected, based on how strongly they meet the criteria of data evaluation
 - b. Clusters from the Explore query will be selected based on size
- H. The neighborhoods from each selected cluster will be made into Sets
- I. An intersection of the Neighborhoods selected from the Search query cluster and a single Explore query cluster will be the recommended neighborhoods

3.2. Analysis of New York – Manhattan

Data for New York was provided in a .json file which was visualized on a Folium Map, the data included information about all Boroughs and Neighborhoods as well as Latitude and Longitude of the areas. The visualization of the data is below. New York has 5 boroughs and 306 neighborhoods.



For the purpose of the project we focused on the Manhattan borough. To ensure that we had the right data we visualized the data on a Folium Map. Manhattan has 40 neighborhoods.



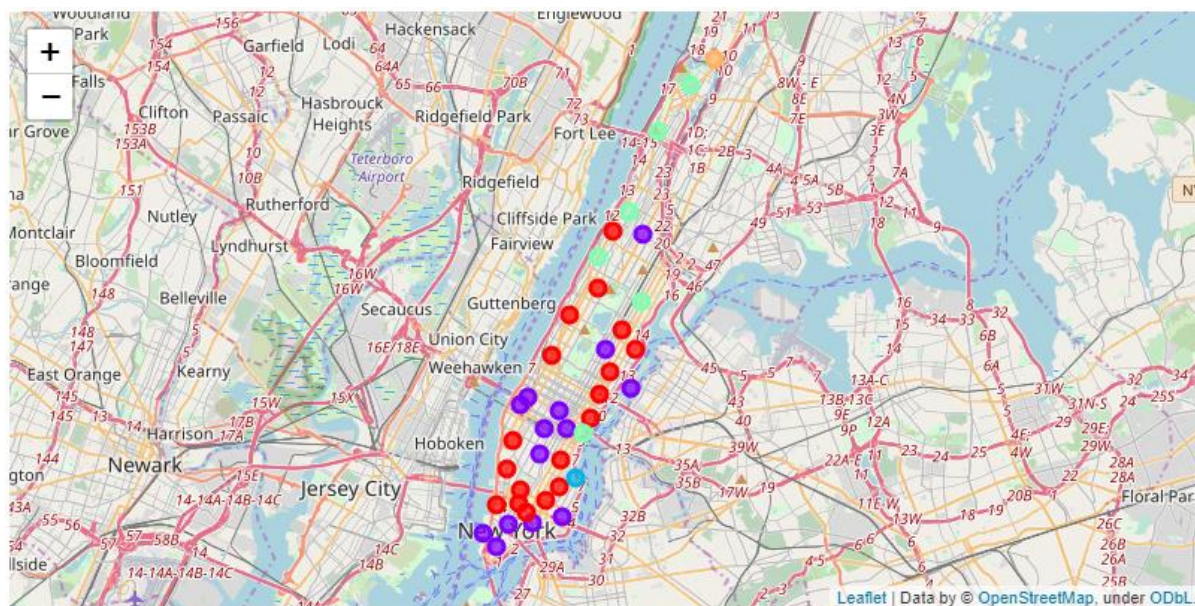
3.3. Explore Query: Analysis of Manhattan Neighborhood






Foursquare API Explore was used to get recommended venues within 500 meters of the neighborhood latitude and longitude information with us.

As a result of this we identified 3279 recommended venues in 330 unique categories.

3.3.1. Explore: Clustering Data using K-means algorithm

We used k-means clustering to develop 5 clusters from the neighbourhood data we have. The result was then visualized on Folium maps.



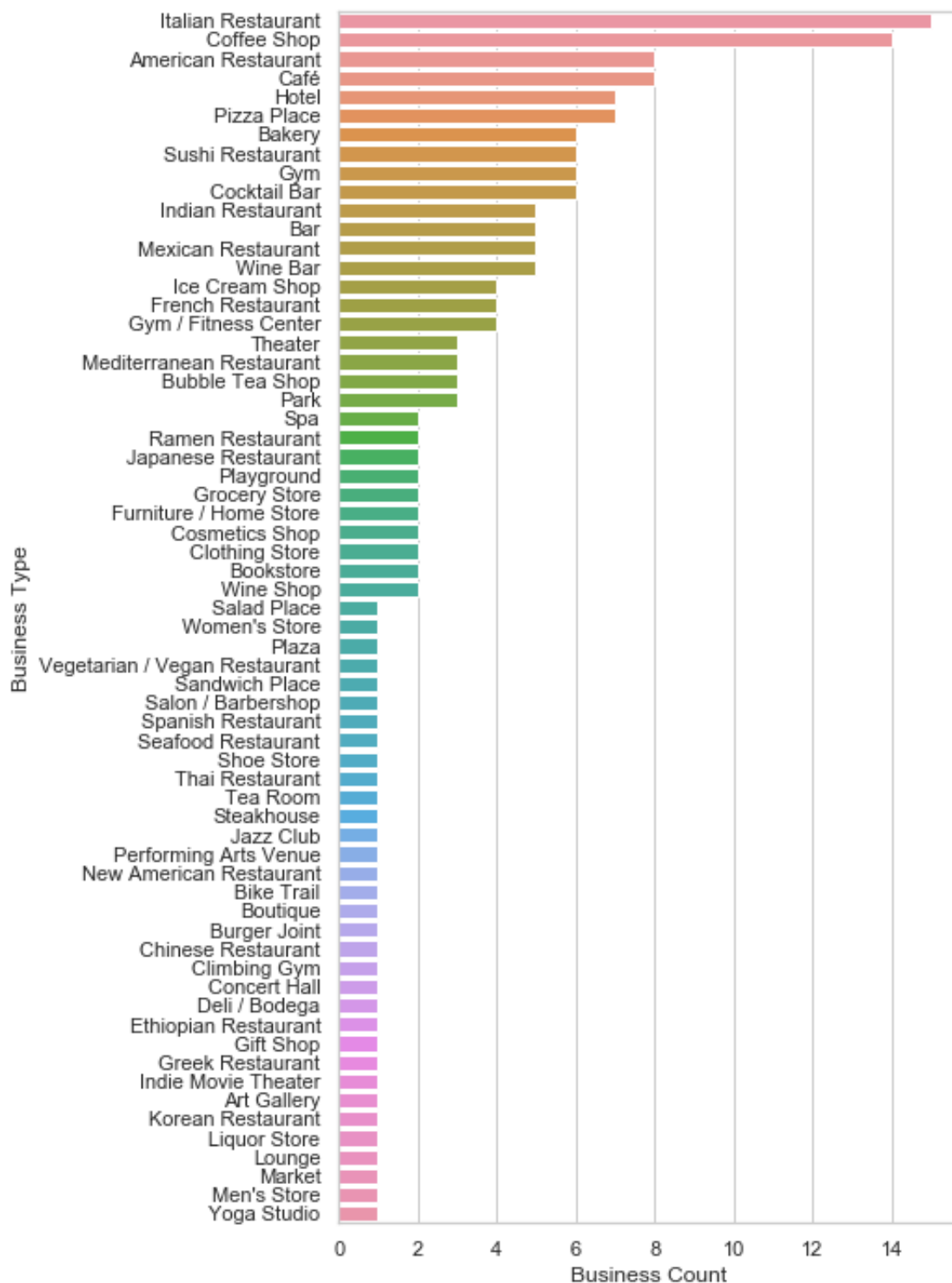
Label	Neighbourhoods in Cluster	Count
	Cluster 1: Lincoln Square, Little Italy, Manhattan Valley, Upper West Side, Sutton Place, Tribeca, Greenwich Village, Lenox Hill, Manhattanville, East Village, Chelsea, Soho, Carnegie Hill, Noho, West Village, Gramercy, Turtle Bay, Yorkville	18
	Cluster 2: Financial District, Roosevelt Island, Clinton, Flatiron, Hudson Yards, Lower East Side, Battery Park City, Upper East Side, Chinatown, Murray Hill, Civic Center, Midtown South, Central Harlem, Midtown	14
	Cluster 3: Stuyvesant Town	1
	Cluster 4: Morningside Heights, Tudor City, East Harlem, Washington Heights, Hamilton Heights, Inwood	6
	Cluster 5: Marble Hill	1

3.3.2. Explore: Analysis of clusters

3.3.2.1. Cluster 1

Cluster 1 includes 18 neighborhoods of Manhattan and is the largest cluster. It includes Lincoln Square, Little Italy, Manhattan Valley, Upper West Side, Sutton Place, Tribeca, Greenwich Village, Lenox Hill, Manhattanville, East Village, Chelsea, Soho, Carnegie Hill, Noho, West Village, Gramercy, Turtle Bay, Yorkville.

Recommended venues within the cluster are.

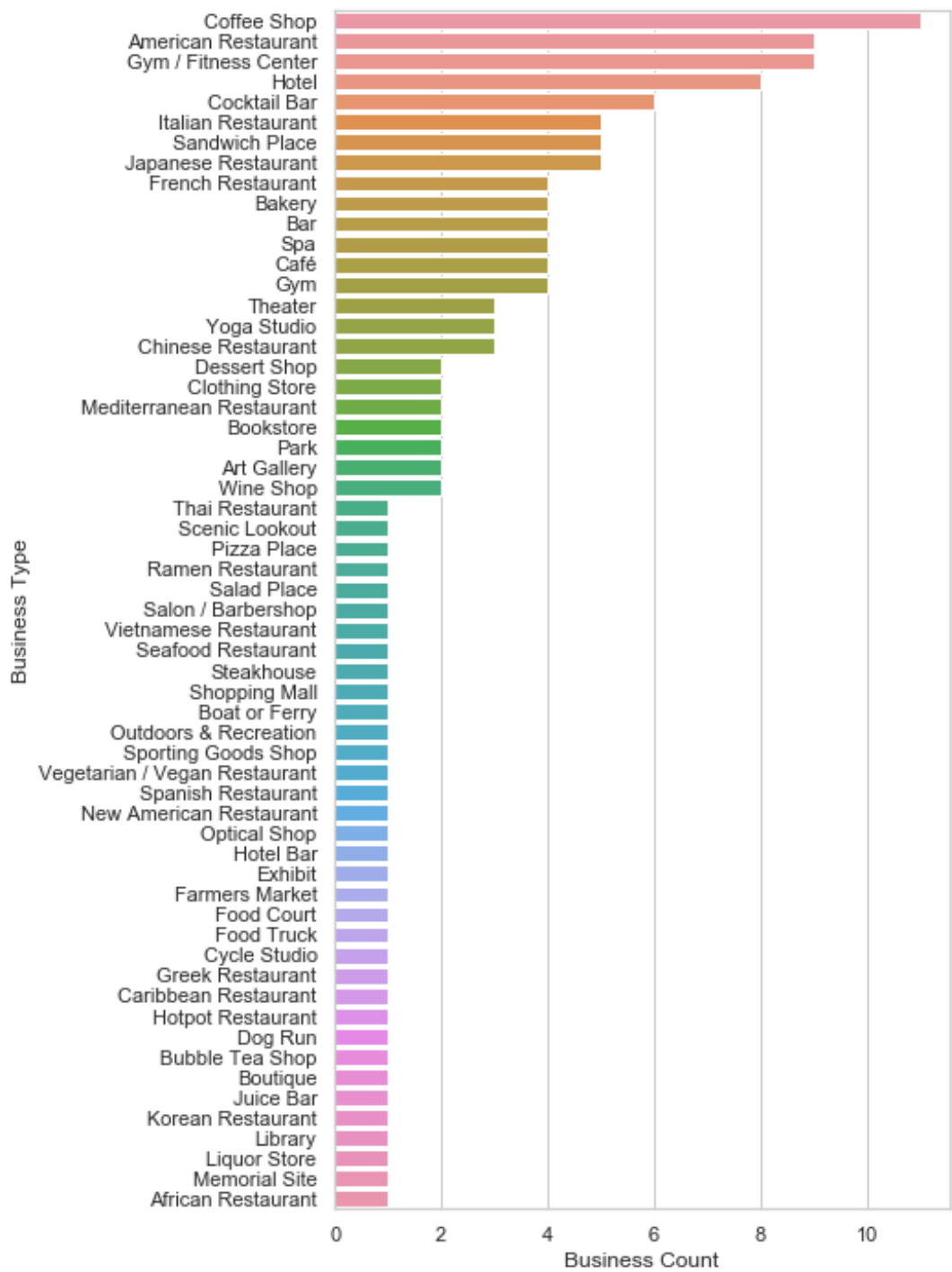


As the largest cluster, in line with the methodology we would use this cluster as the first cluster for intersection with the cluster from the search clusters.

3.3.2.2. Cluster 2

Cluster 2 includes 14 neighborhoods and is the second biggest cluster output from the k-means algorithm. It includes Financial District, Roosevelt Island, Clinton, Flatiron, Hudson Yards, Lower East

Side, Battery Park City, Upper East Side, Chinatown, Murray Hill, Civic Center, Midtown South, Central Harlem, Midtown. The recommended venues within the cluster are in the barchart below.



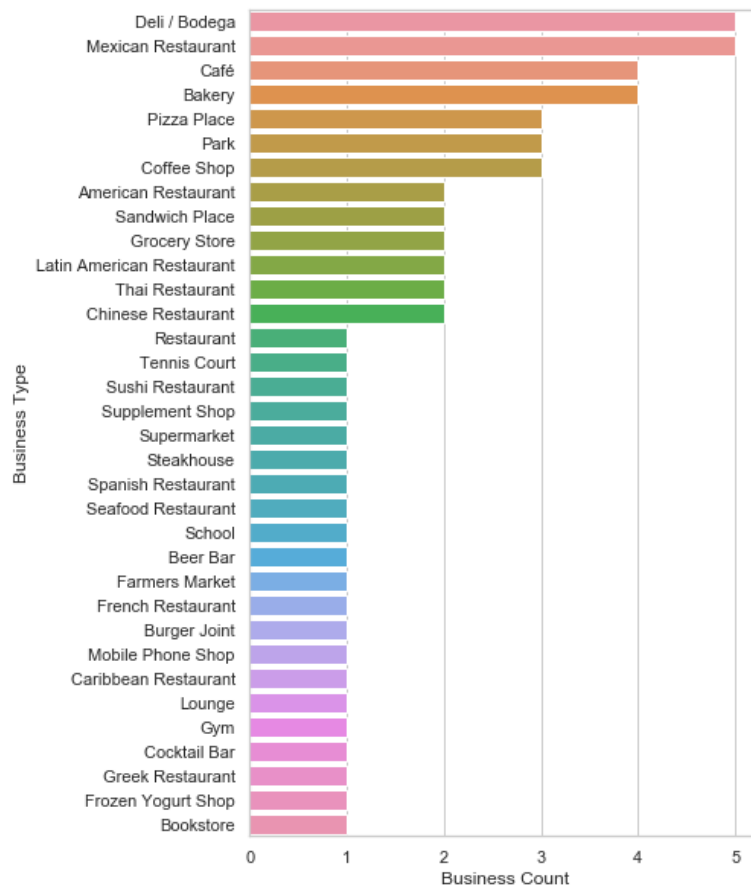
3.3.2.3. Cluster 3

Cluster 3 includes just one neighbourhood Stuyvesant Town. The recommended venues within the cluster are in the barchart below.



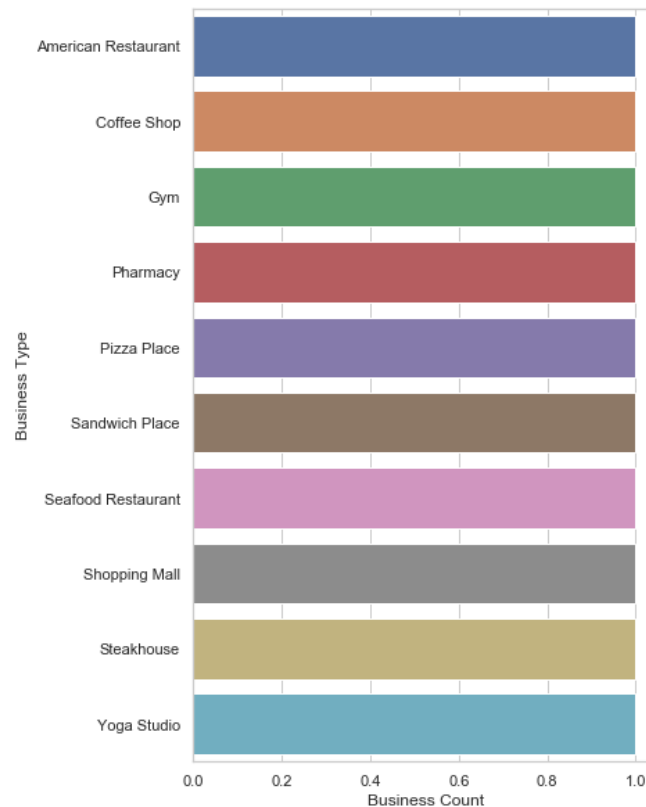
3.3.2.4. Cluster 4

Cluster 4 includes 6 neighborhoods in Manhattan. It includes Morningside Heights, Tudor City, East Harlem, Washington Heights, Hamilton Heights, Inwood. The recommended venues within the cluster are in the barchart below.



3.3.2.5. Cluster 5

Cluster 5 includes 1 neighborhood on Marble Hill. The recommended venues within the cluster are in the barchart below.



3.4. Search Query: Analysis of Manhattan Neighborhood

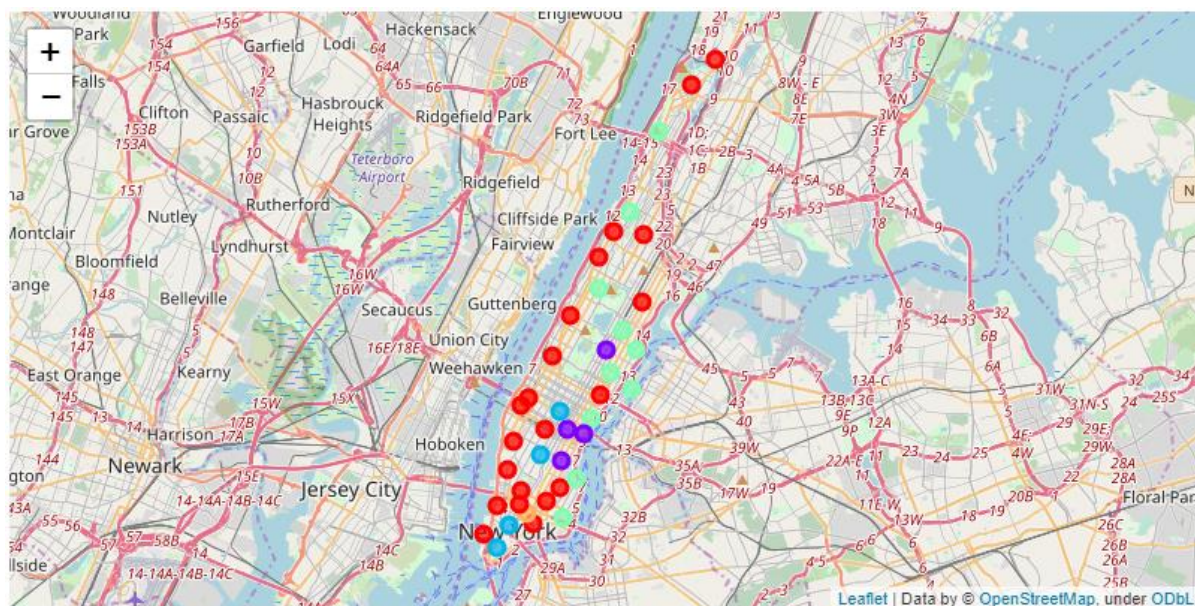
Foursquare API Search query was used to search for specific category of venues within 500 meters of the neighborhood latitude and longitude information with us.






As a result of this we identified 3434 recommended venues in 379 unique categories.

While sending the search query a category list with the category ID's for specific categories was sent.

3.4.1. Search: Clustering Data using K-means algorithm

We used k-means clustering to develop 5 clusters from the neighbourhood data we have. The result was then visualized on Folium maps.



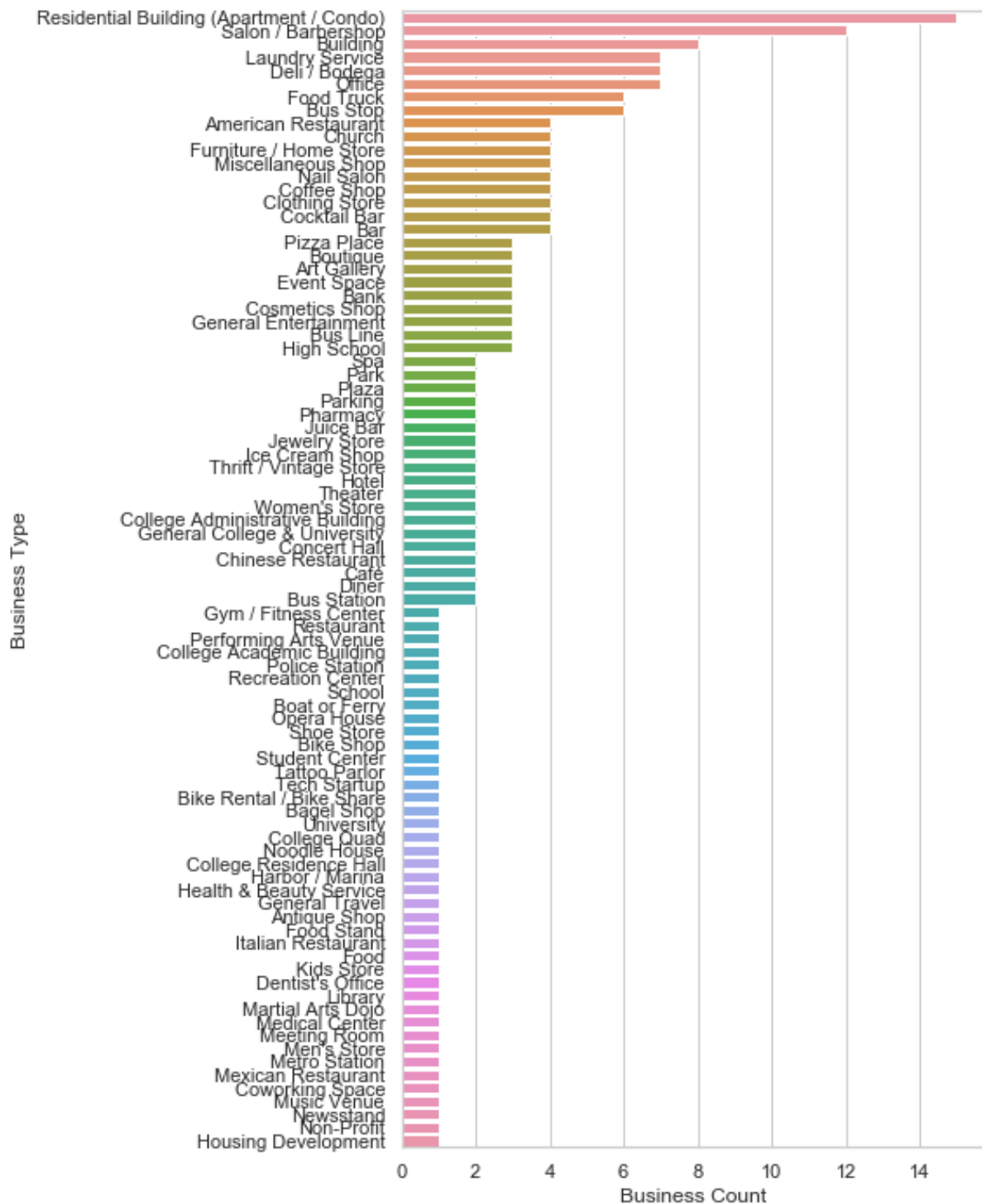
Label	Neighbourhoods in Cluster	Count
	Cluster 1: Battery Park City, Noho, Lincoln Square, Chinatown, Greenwich Village, West Village, East Village, Manhattanville, Upper West Side, Inwood, Morningside Heights, Tribeca, Central Harlem, Clinton, Hudson Yards, East Harlem, Midtown South, Sutton Place, Soho, Chelsea, Marble Hill	21
	Cluster 2: Murray Hill, Gramercy, Tudor City, Upper East Side	4
	Cluster 3: Midtown, Flatiron, Financial District, Civic Center	4
	Cluster 4: Yorkville, Manhattan Valley, Turtle Bay, Lenox Hill, Stuyvesant Town, Washington Heights, Carnegie Hill, Roosevelt Island, Hamilton Heights, Lower East Side	10
	Cluster 5: Little Italy	1

3.4.2. Search: Analysis of Clusters

3.4.2.1. Cluster 1

Cluster 1 is the largest cluster that has 21 neighborhoods, it is the largest cluster identified. The neighborhoods include Battery Park City, Noho, Lincoln Square, Chinatown, Greenwich Village, West Village, East Village, Manhattanville, Upper West Side, Inwood, Morningside Heights, Tribeca, Central Harlem, Clinton, Hudson Yards, East Harlem, Midtown South, Sutton Place, Soho, Chelsea, Marble Hill.

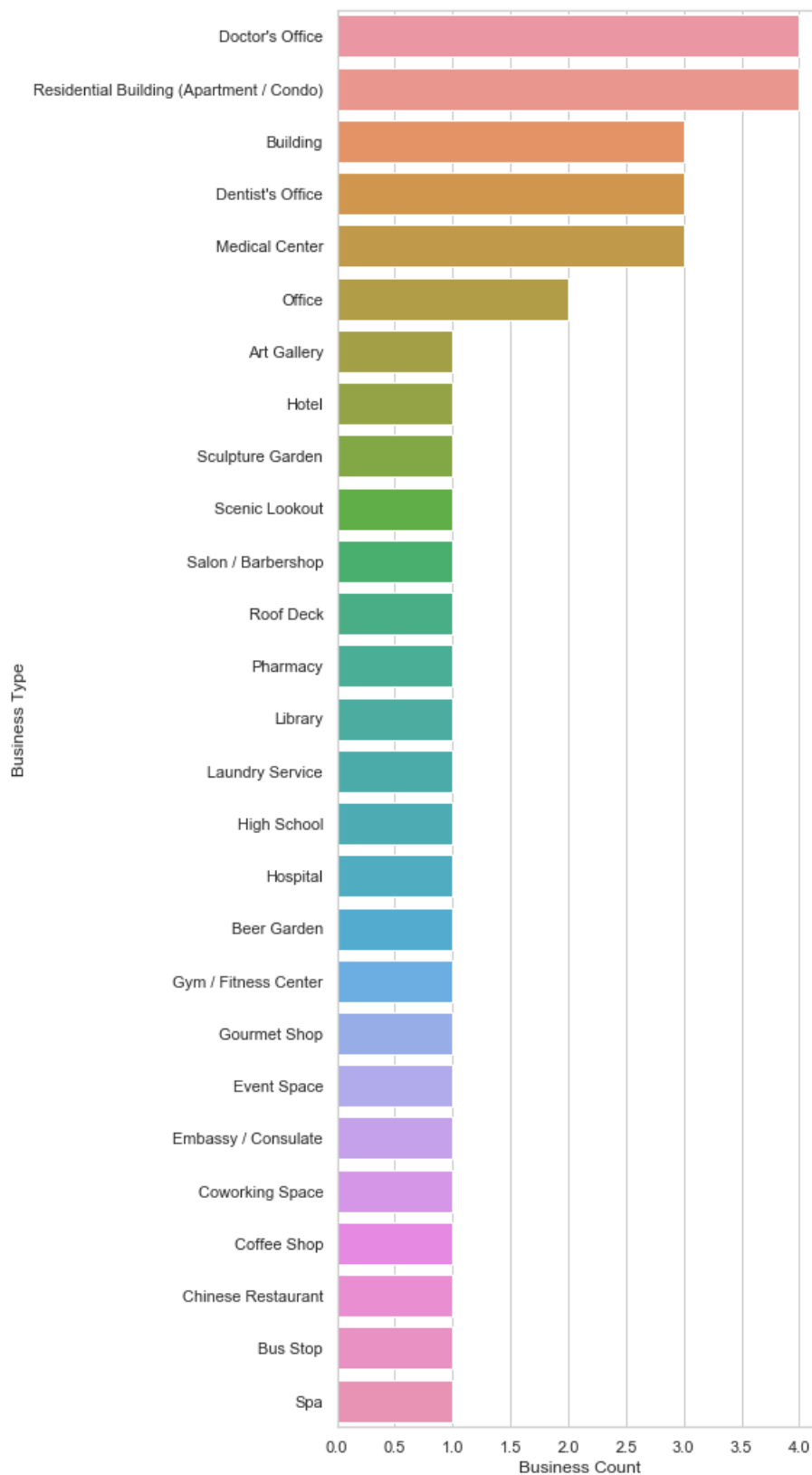
Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.



The primary criteria for our target audience is the availability of residential Accommodations. This criteria is best met in this cluster. Furthermore, all secondary criteria are also met. We will use this cluster in further analysis to identify neighborhoods suitable for the target audience.

3.4.2.2. Cluster 2

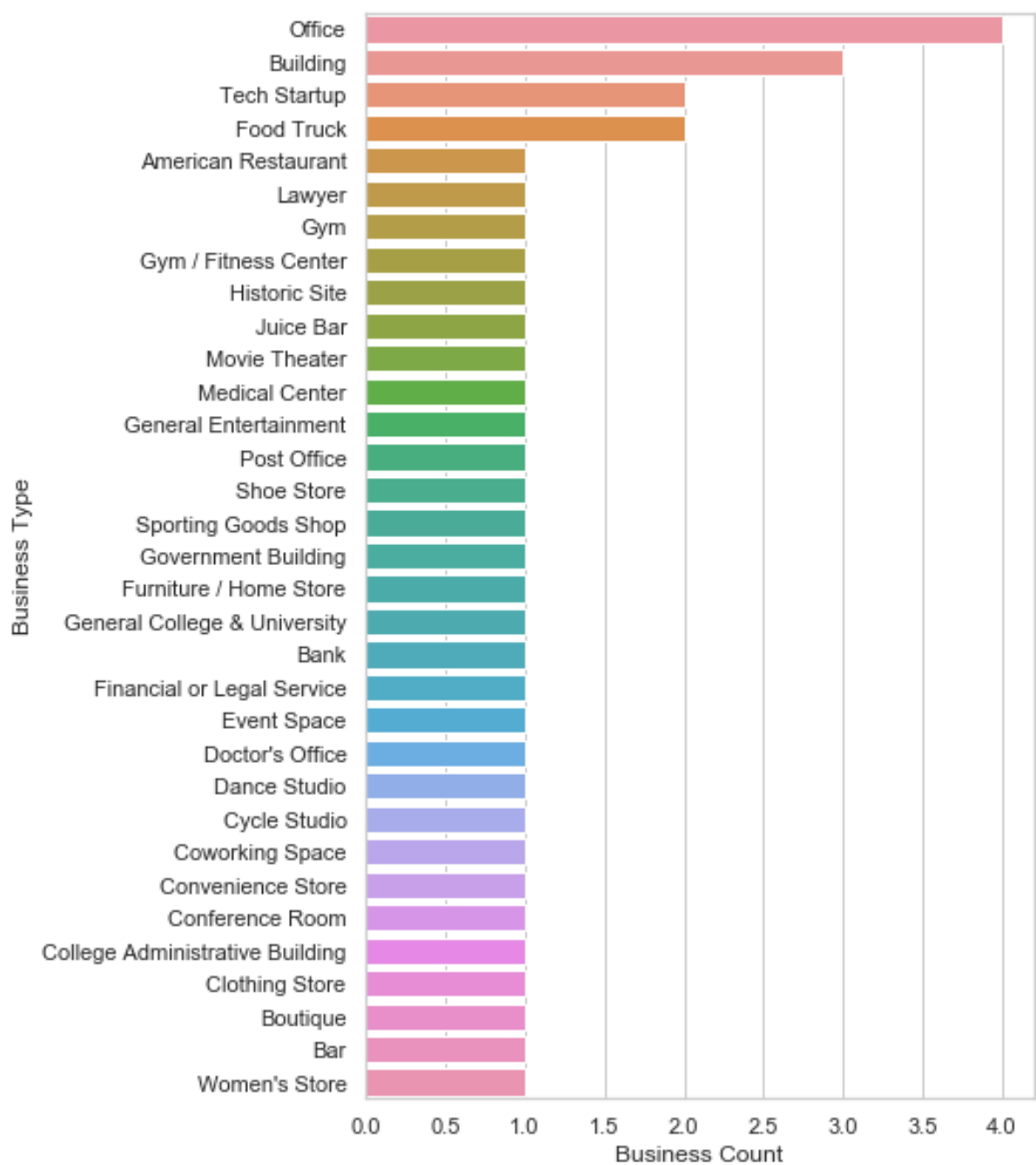
Cluster 2 has 4 neighborhoods Murray Hill, Gramercy, Tudor City, Upper East Side. Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.



3.4.2.3. Cluster 3

Cluster 3 has 4 neighborhoods, which include Midtown, Flatiron, Financial District, Civic Center.

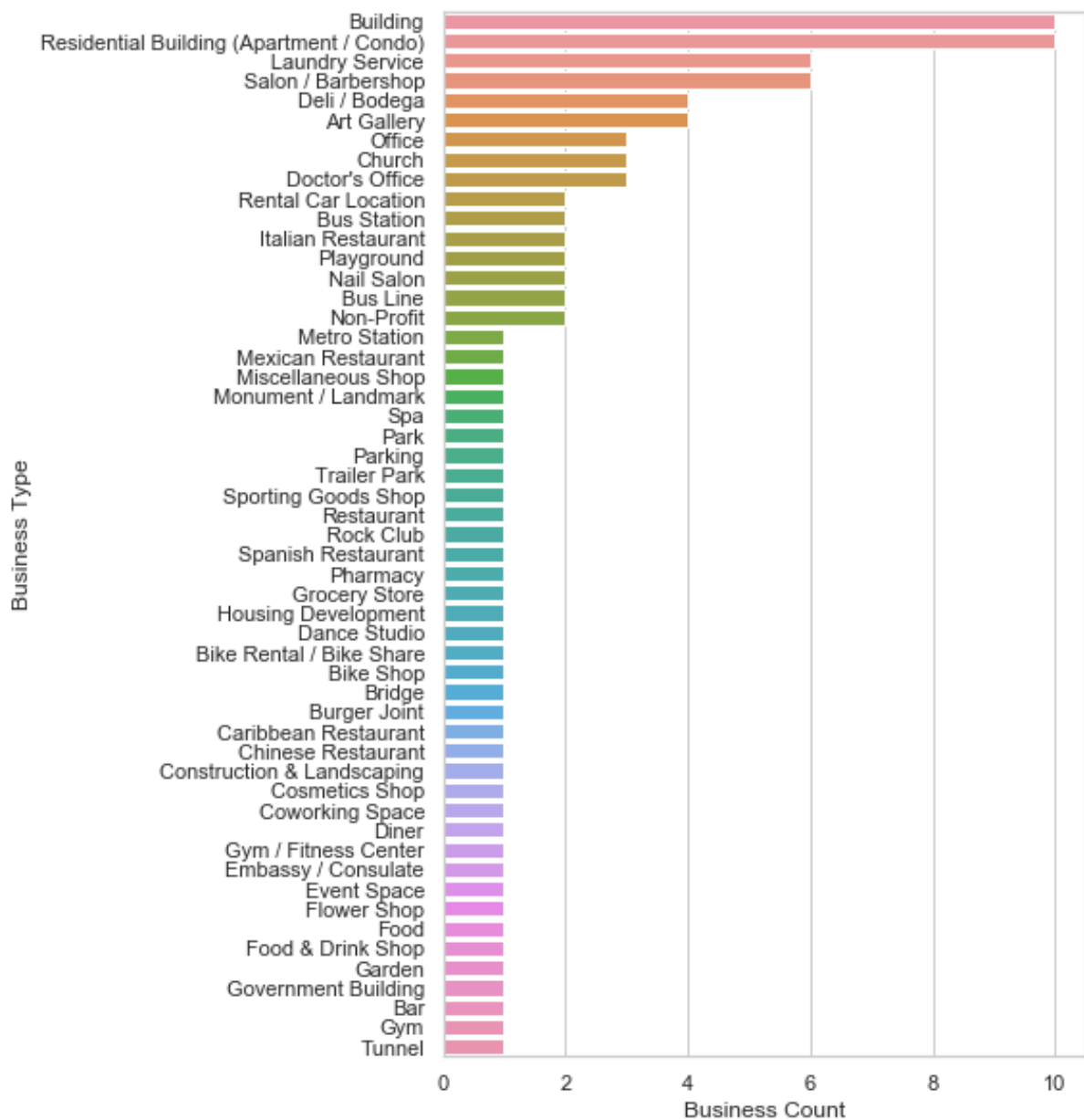
Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.



3.4.2.4. Cluster 4

Cluster 4 is the second largest cluster that has 10 neighborhoods. The neighborhoods included in this cluster are : Yorkville, Manhattan Valley, Turtle Bay, Lenox Hill, Stuyvesant Town, Washington Heights, Carnegie Hill, Roosevelt Island, Hamilton Heights, Lower East Side.

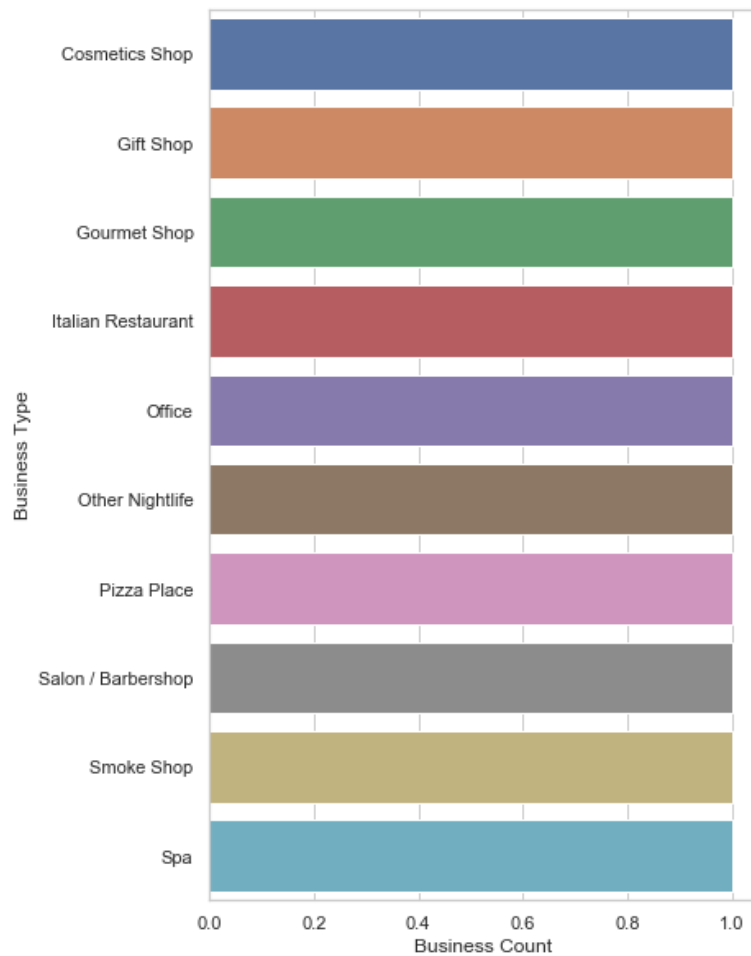
Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.



3.4.2.5. Cluster 5

Cluster 5 has one neighborhood on Little Italy.

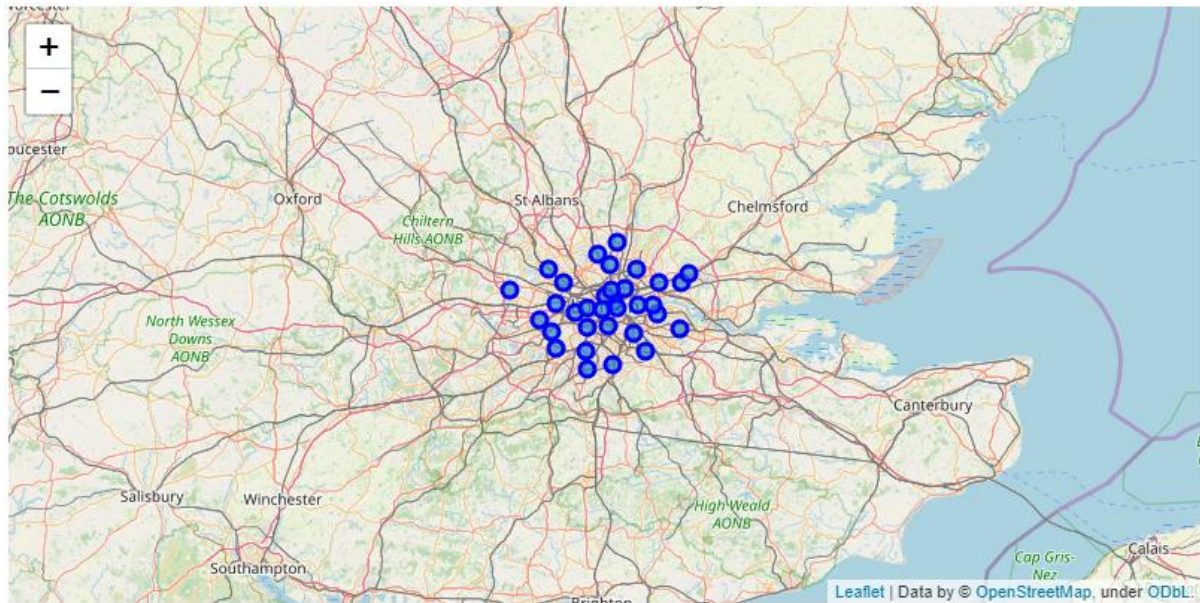
Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.



3.5. Analysis of London – City of London and Westminster

Data for London wasn't readily available, while we had some information about postcodes it wasn't matched with Latitudes and Longitudes. Various websites were scraped and data was downloaded and merged to develop a coherent version. Visualization of this data was critical to understand the quality and accuracy of information we had.

London has 33 boroughs. We didn't identify the neighborhoods within London, this was because it wasn't relevant to the analysis.



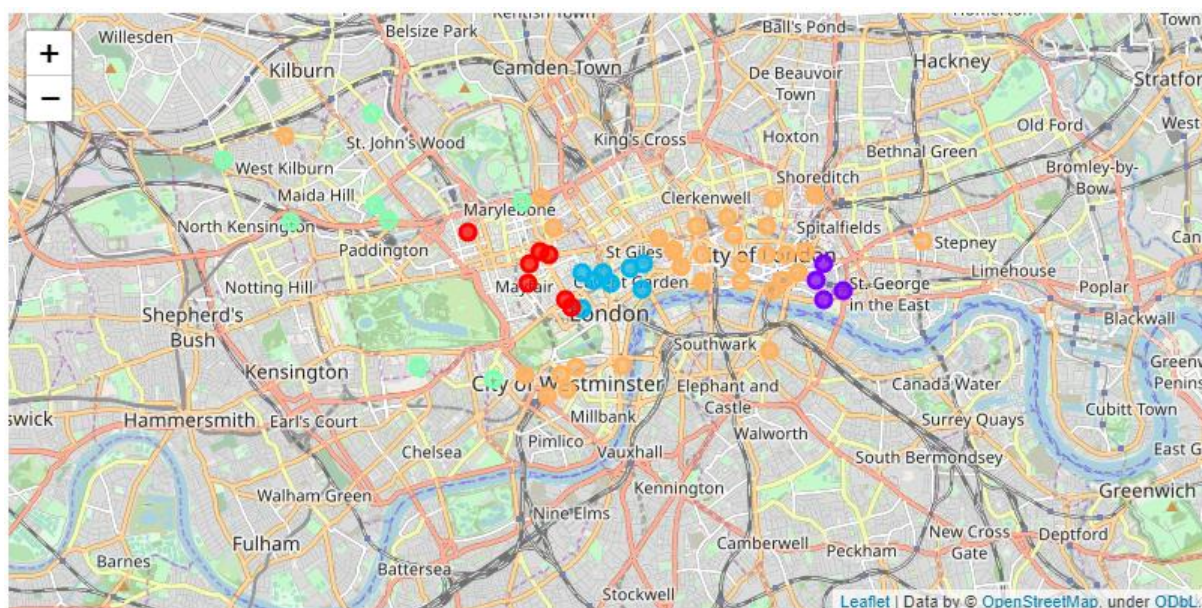
Further analysis was done regarding City of London and City of Westminster and information of the neighborhoods was scraped through websites.








A total of 59 neighborhoods were identified within the City of London and City of Westminster.

3.5.1. Explore: Clustering Data using K-means algorithm

We used k-means clustering to develop 5 clusters from the neighbourhood data we have. The result was then visualized on Folium maps.



Label	Neighbourhoods in Cluster	Count
	Cluster 1: 'Mayfair (south), Piccadilly', Harley Street, 'Mayfair (north), Grosvenor Square', Marylebone(in 2 parts), 'Mayfair (east), Hanover Square, Savile Row, Royal Academy', Oxford Street (west)	7
	Cluster 2: "Lloyd's of London, Fenchurch Street", 'Monument, Billingsgate', 'St Mary Axe, Aldgate', 'Tower Hill, Tower of London'	4
	Cluster 3: 'Soho (north west)', 'Soho (south east); Chinatown, Soho Square', 'Covent Garden', 'Leicester Square, St. Giles', "All but the west of St James's", 'Charing Cross', 'Drury Lane, Kingsway, Aldwych', 'Somerset House, Temple (west)'	8
	Cluster 4: "St John's Wood, Primrose Hill (south), Marylebone (north), Lisson Grove (north)", 'Maida Hill district: Maida Hill, Maida Vale, Little Venice (part)', 'South Kensington district: South Kensington, Knightsbridge (part)', "North Kensington, Kensal Town, Ladbroke Grove (north), Queen's Park (part)", 'Notting Hill, Ladbroke Grove (south), Holland Park (part)', 'Rest of Belgravia (north of Eaton Square), Knightsbridge (eastern half) and Chelsea (a corner)', 'Paddington, Bayswater, Hyde Park, Westbourne Green, Little Venice (part), Notting Hill (part)', 'Portland Place, Regent Street'	8
	Cluster 5: 'Guildhall', 'Shoreditch', 'Cornhill, Gracechurch Street, Lombard Street', 'Great Portland Street, Fitzrovia', 'Barbican', 'Parliament Square, Nearby Westminster School to Westminster Cathedral to CCAL, University of the Arts, London', 'Cannon Street', 'Bankside, South Bank, Southwark', 'Fetter Lane', "Whitehall, Buckingham Palace, the Houses of Parliament and west of St James's", 'Whitechapel, Spitalfields, Shadwell, Limehouse, Stepney, Portsoken', "St Paul's", 'South or lower Belgravia and Chelsea, from Sloane Square to Victoria Station to Grosvenor Waterside basin', "Kilburn, Brondesbury, West Hampstead, Queen's Park", 'Mansion House', 'Clerkenwell, Farringdon', "Northern 'Victoria' around Buckingham Gate ", 'Temple', "Lincoln's Inn Fields, Royal Courts of Justice, Chancery Lane", "St Bartholomew's Hospital", 'Hatton Garden', 'Old Broad Street, Tower 42', "Broadway, Queen Anne's Gate and Old Queen Street ", 'High Holborn', "Euston, Regent's Park, Baker Street, Camden	32

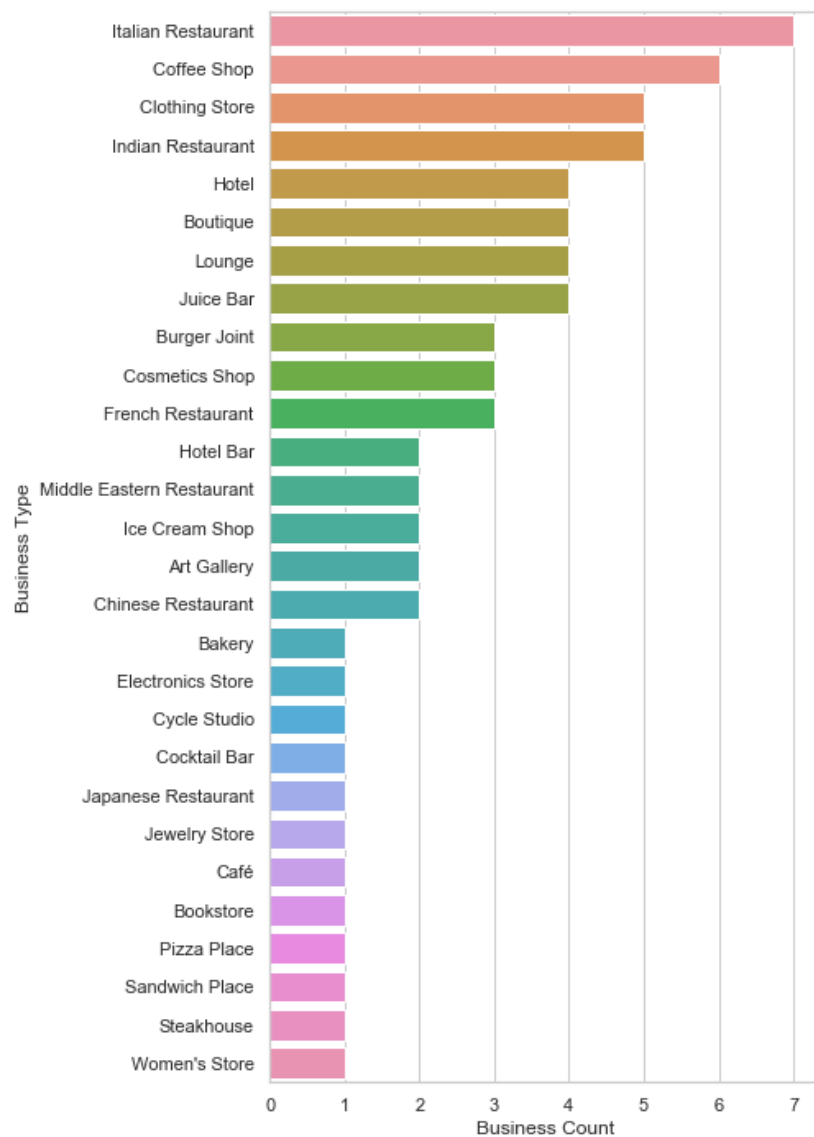
	Town, Somers Town", 'Bank of England', 'Between Vauxhall Bridge, Grosvenor Bridge and Victoria Station, includes Pimlico', 'Blackfriars', "St Luke's, Bunhill Fields"	
--	---	--

3.5.2. Explore: Analysis of Clusters

3.5.2.1. Cluster 1

Cluster 1 has 7 neighborhoods, these are 'Mayfair (south), Piccadilly', Harley Street, 'Mayfair (north), Grosvenor Square', Marylebone(in 2 parts), 'Mayfair (east), Hanover Square, Savile Row, Royal Academy', Oxford Street (west).

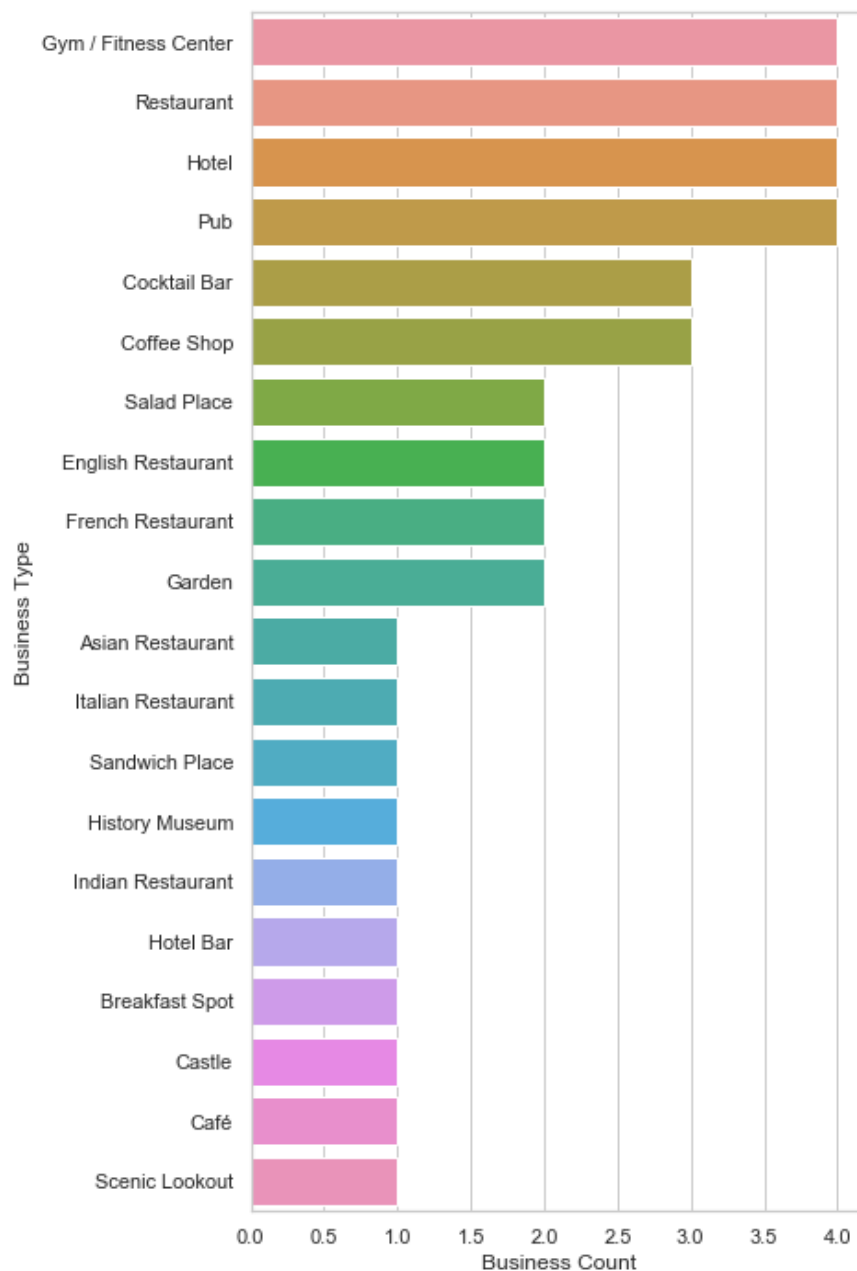
Category of the recommended venues gave several results which are in the bar chart below.



3.5.2.2. Cluster 2

Cluster 2 has 4 neighborhoods, these are “Lloyd's of London, Fenchurch Street”, 'Monument, Billingsgate', 'St Mary Axe, Aldgate', 'Tower Hill, Tower of London'.

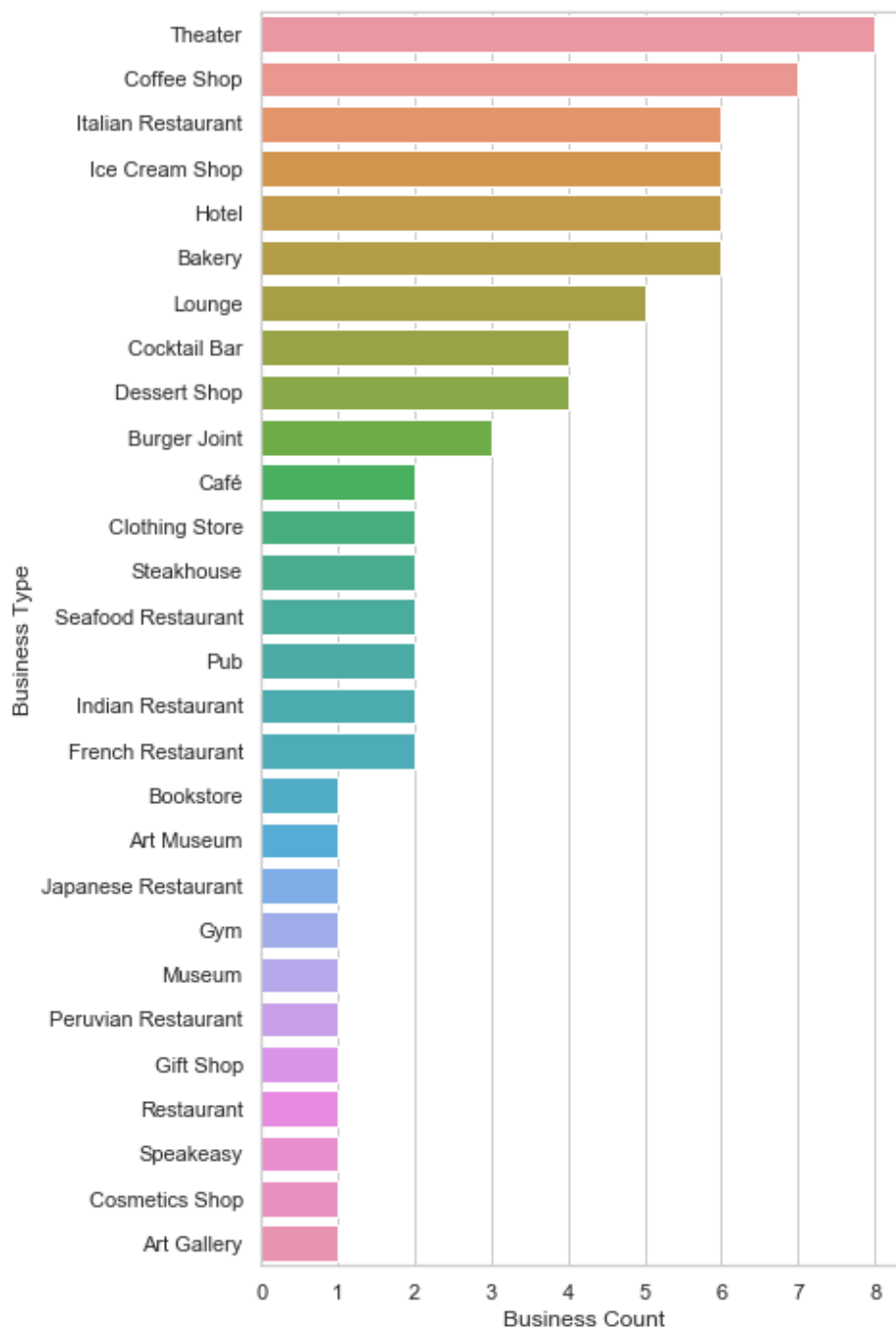
Category of the recommended venues within this neighbourhood is in the bar chart below.



3.5.2.3. Cluster 3

Cluster 3 has 8 neighborhoods, these are : 'Soho (north west)', 'Soho (south east); Chinatown, Soho Square', 'Covent Garden', 'Leicester Square, St. Giles', "All but the west of St James's", 'Charing Cross', 'Drury Lane, Kingsway, Aldwych', 'Somerset House, Temple (west)'.

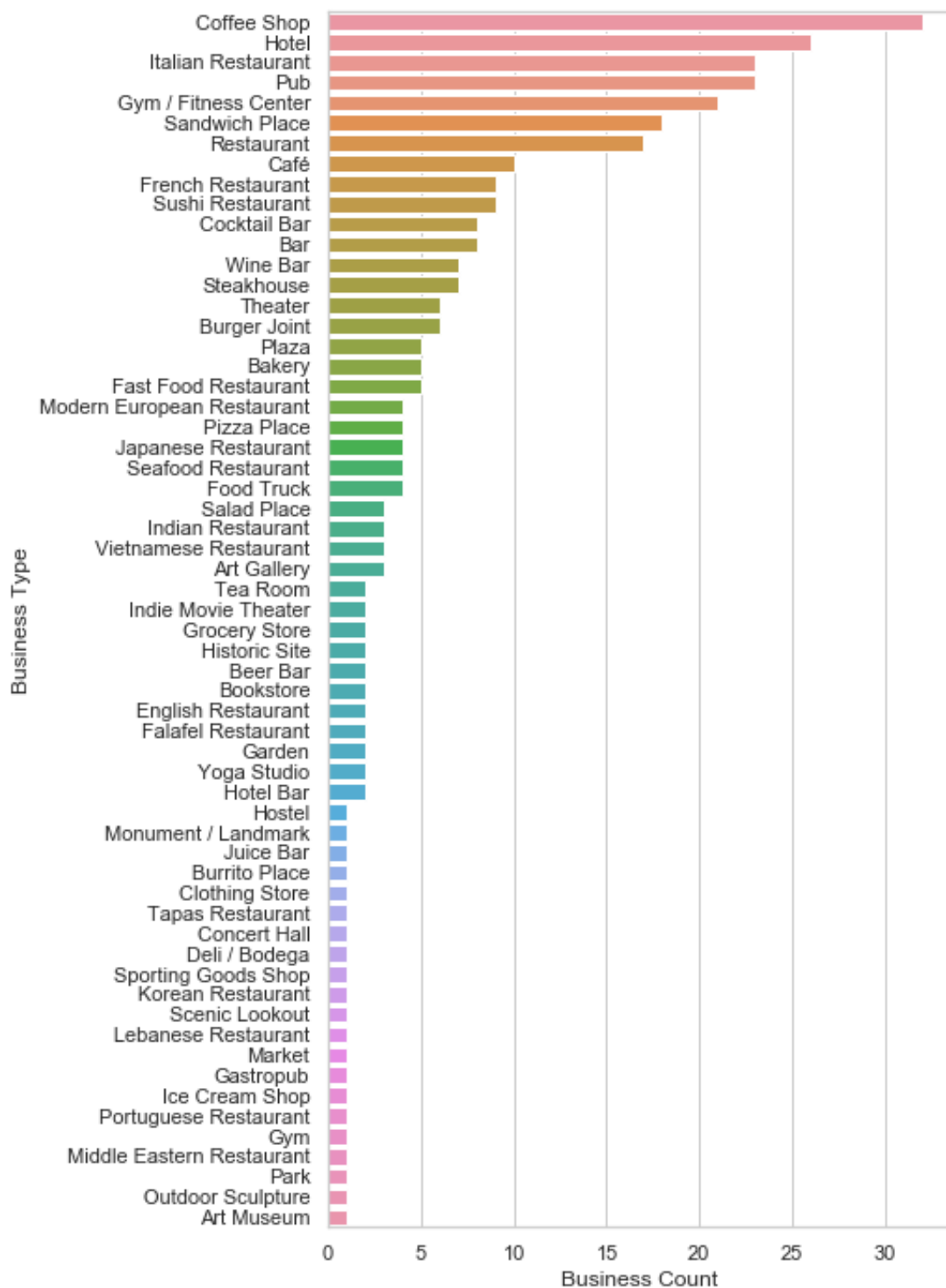
Category of the recommended venues within this neighbourhood is in the bar chart below.



3.5.2.4. Cluster 4

Cluster 4 has 8 neighborhoods, these are "St John's Wood, Primrose Hill (south), Marylebone (north), Lisson Grove (north)", 'Maida Hill district: Maida Hill, Maida Vale, Little Venice (part)', 'South Kensington district: South Kensington, Knightsbridge (part)', "North Kensington, Kensal Town, Ladbroke Grove (north), Queen's Park (part)", 'Notting Hill, Ladbroke Grove (south), Holland Park (part)', 'Rest of Belgravia (north of Eaton Square), Knightsbridge (eastern half) and Chelsea (a corner)', 'Paddington, Bayswater, Hyde Park, Westbourne Green, Little Venice (part), Notting Hill (part)', 'Portland Place, Regent Street'

Category of the recommended venues within this neighbourhood is in the bar chart below.

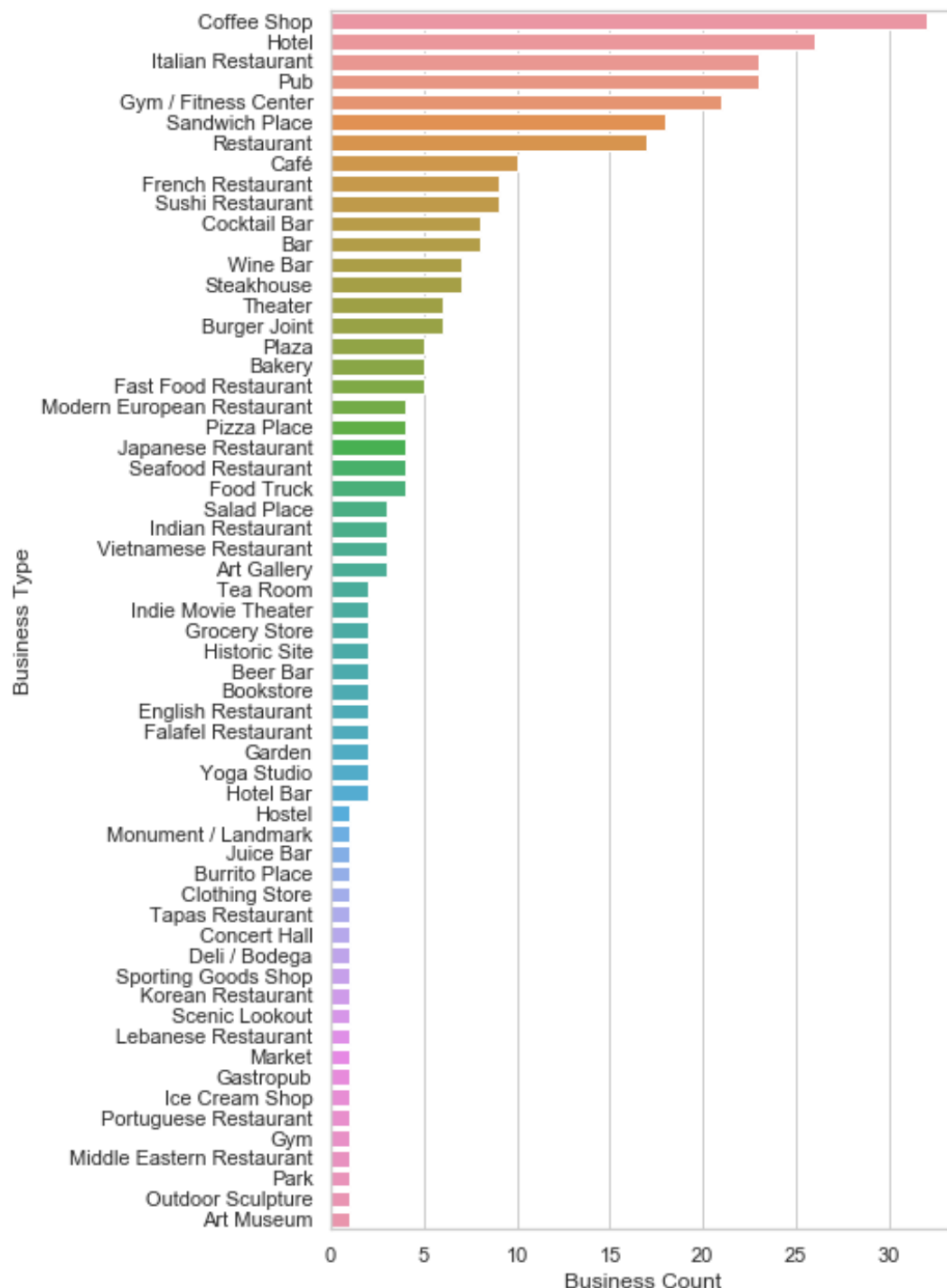


3.5.2.5. Cluster 5

Cluster 5 is the largest cluster that has 32 neighborhoods in it. These are 'Guildhall', 'Shoreditch', 'Cornhill, Gracechurch Street, Lombard Street', 'Great Portland Street, Fitzrovia', 'Barbican', 'Parliament Square, Nearby Westminster School to Westminster Cathedral to CCAL, University of the Arts, London', 'Cannon Street', 'Bankside, South Bank, Southwark', 'Fetter Lane', "Whitehall, Buckingham Palace, the Houses of Parliament and west of St James's", 'Whitechapel, Spitalfields, Shadwell, Limehouse, Stepney, Portsoken', "St Paul's", 'South or lower Belgravia and Chelsea, from

Sloane Square to Victoria Station to Grosvenor Waterside basin', 'Kilburn, Brondesbury, West Hampstead, Queen's Park', 'Mansion House', 'Clerkenwell, Farringdon', "Northern 'Victoria' around Buckingham Gate ", 'Temple', "Lincoln's Inn Fields, Royal Courts of Justice, Chancery Lane", "St Bartholomew's Hospital", 'Hatton Garden', 'Old Broad Street, Tower 42', "Broadway, Queen Anne's Gate and Old Queen Street ", 'High Holborn', "Euston, Regent's Park, Baker Street, Camden Town, Somers Town", 'Bank of England', 'Between Vauxhall Bridge, Grosvenor Bridge and Victoria Station, includes Pimlico', 'Blackfriars', "St Luke's, Bunhill Fields"

Category of the recommended venues within this neighbourhood is in the bar chart below.

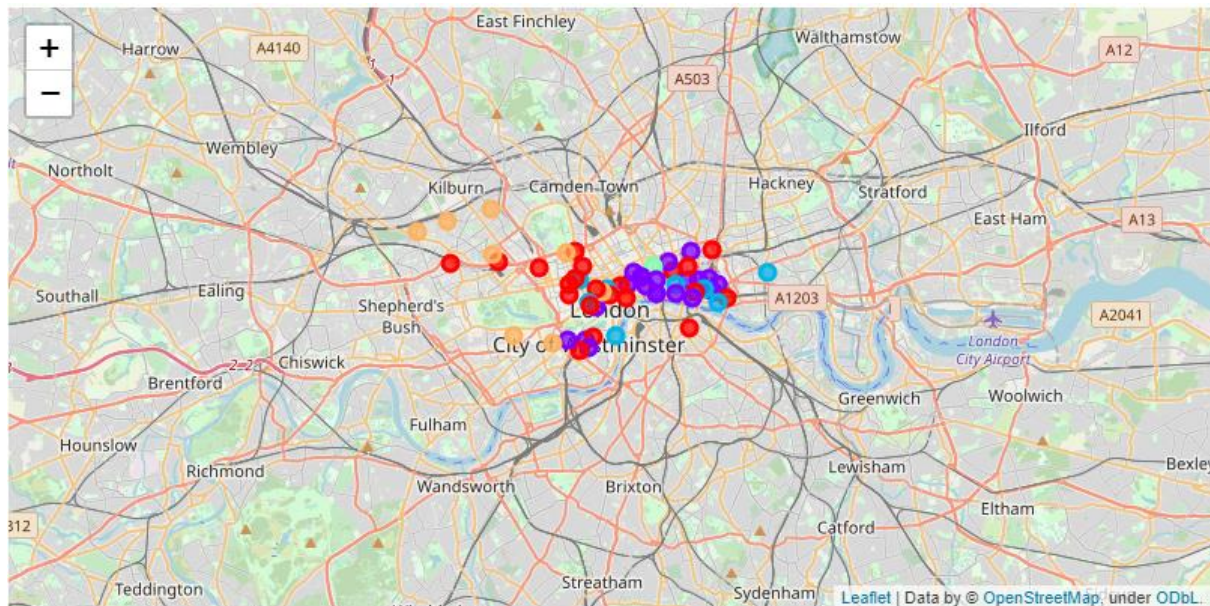




Cluster 5 will be the cluster used in further analysis to identify a suitable intersection with the Clusters from the search query.




3.6. Search Query: Analysis of City of London and Westminster

3.6.1. Search: Clustering Data using K-means algorithm

We used k-means clustering to develop 5 clusters from the neighbourhood data we have. The result was then visualized on Folium maps.



Label	Neighbourhoods in Cluster	Count
	Cluster 1: 'Mayfair (east), Hanover Square, Savile Row, Royal Academy', 'Tower Hill, Tower of London', "Broadway, Queen Anne's Gate and Old Queen Street ", 'Mansion House', 'Mayfair (north), Grosvenor Square', 'St Bartholomew's Hospital', 'Shoreditch', 'Paddington, Bayswater, Hyde Park, Westbourne Green, Little Venice (part), Notting Hill (part)', 'Between Vauxhall Bridge, Grosvenor Bridge and Victoria Station, includes Pimlico', 'Marylebone (2 parts)', 'Charing Cross', 'Notting Hill, Ladbroke Grove (south), Holland Park (part)', 'Soho (north west)', 'Bankside, South Bank, Southwark', 'Barbican', 'Somerset House, Temple (west)', "Euston, Regent's Park, Baker Street, Camden Town, Somers Town", 'Great Portland Street, Fitzrovia', 'Harley Street', 'Covent Garden'	21
	Cluster 2: "St Luke's, Bunhill Fields", "Northern 'Victoria' around Buckingham Gate ", 'Drury Lane, Kingsway, Aldwych', "All but the west of St James's", 'Blackfriars', 'Temple', "Lincoln's Inn Fields, Royal Courts of Justice, Chancery Lane", 'Fetter Lane', 'South or lower Belgravia and Chelsea, from Sloane Square to Victoria Station to Grosvenor Waterside basin', 'Clerkenwell, Farringdon', 'High Holborn', 'Guildhall', 'Old Broad Street, Tower 42', 'St Mary Axe, Aldgate', "Lloyd's of London, Fenchurch Street", 'Parliament Square, Nearby Westminster School to Westminster Cathedral to CCAL, University of the Arts, London', 'Bank of England', 'Cannon Street'	21

	Cluster 3: 'Oxford Street (west)', 'Mayfair (south), Piccadilly', 'Cornhill, Gracechurch Street, Lombard Street', 'Leicester Square, St. Giles', 'St Paul's', 'Monument, Billingsgate', 'Whitehall, Buckingham Palace, the Houses of Parliament and west of St James's', 'Whitechapel, Spitalfields, Shadwell, Limehouse, Stepney, Portsoken'	8
	Cluster 4: Hatton Garden	1
	Cluster 5: 'Rest of Belgravia (north of Eaton Square), Knightsbridge (eastern half) and Chelsea (a corner)', 'Soho (south east); Chinatown, Soho Square', 'South Kensington district: South Kensington, Knightsbridge (part)', 'Maida Hill district: Maida Hill, Maida Vale, Little Venice (part)', 'Portland Place, Regent Street', 'St John's Wood, Primrose Hill (south), Marylebone (north), Lisson Grove (north)', 'Kilburn, Brondesbury, West Hampstead, Queen's Park', 'North Kensington, Kensal Town, Ladbroke Grove (north), Queen's Park (part)'	8

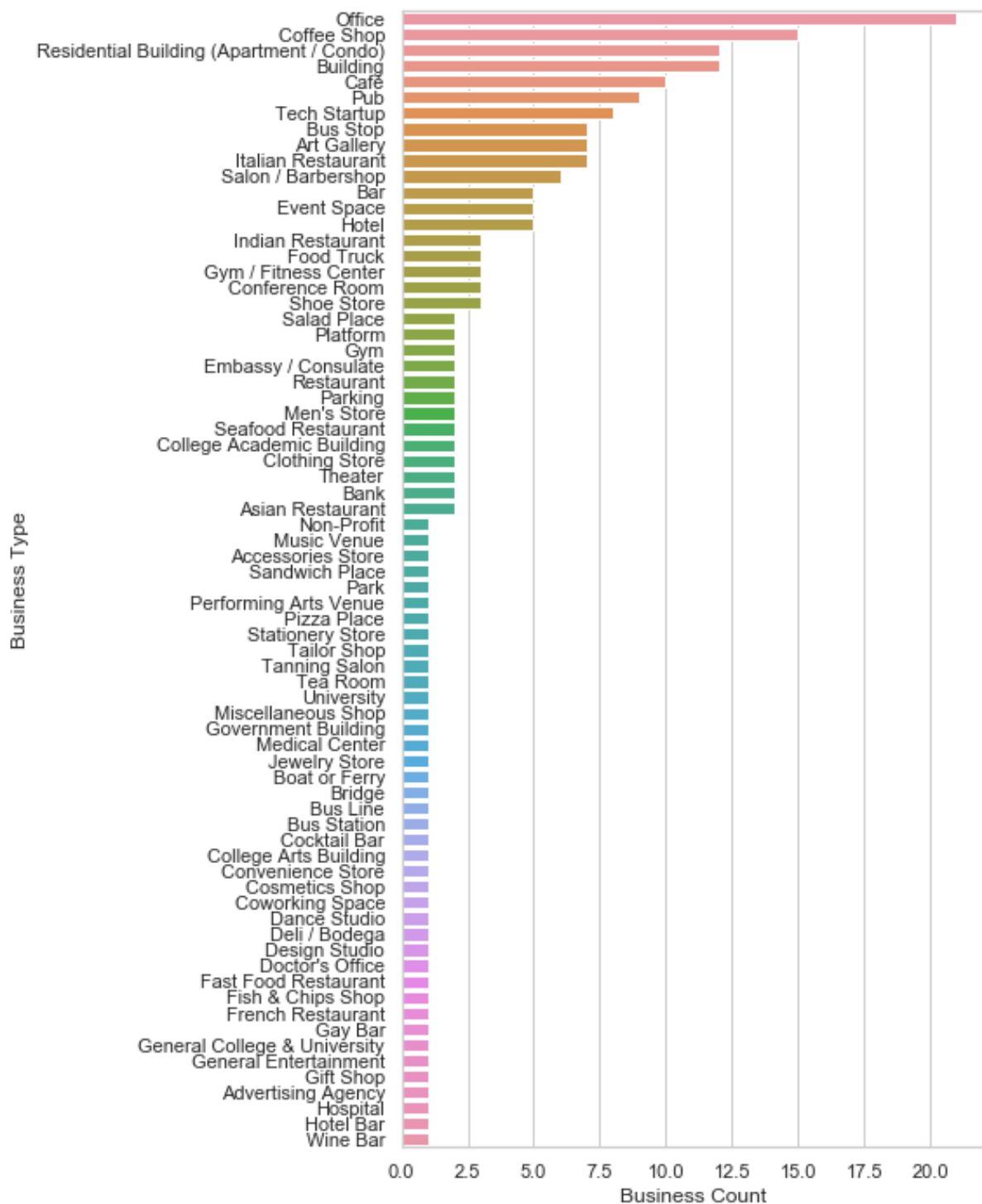
3.6.2. Search: Analysis of Clusters

3.6.2.1. Cluster 1

Cluster 1 (similar to Cluster 2) has 21 neighborhoods as part of it, these are 'Mayfair (east), Hanover Square, Savile Row, Royal Academy', 'Tower Hill, Tower of London', 'Broadway, Queen Anne's Gate and Old Queen Street ', 'Mansion House', 'Mayfair (north), Grosvenor Square', 'St Bartholomew's Hospital', 'Shoreditch', 'Paddington, Bayswater, Hyde Park, Westbourne Green, Little Venice (part), Notting Hill (part)', 'Between Vauxhall Bridge, Grosvenor Bridge and Victoria Station, includes Pimlico', 'Marylebone (2 parts)', 'Charing Cross', 'Notting Hill, Ladbroke Grove (south), Holland Park (part)', 'Soho (north west)', 'Bankside, South Bank, Southwark', 'Barbican', 'Somerset House, Temple (west)', 'Euston, Regent's Park, Baker Street, Camden Town, Somers Town', 'Great Portland Street, Fitzrovia', 'Harley Street', 'Covent Garden'.

Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.

In comparison to Cluster 2, its clear that Cluster 1 has significantly greater Residential accommodation and will therefore be considered for analysis.

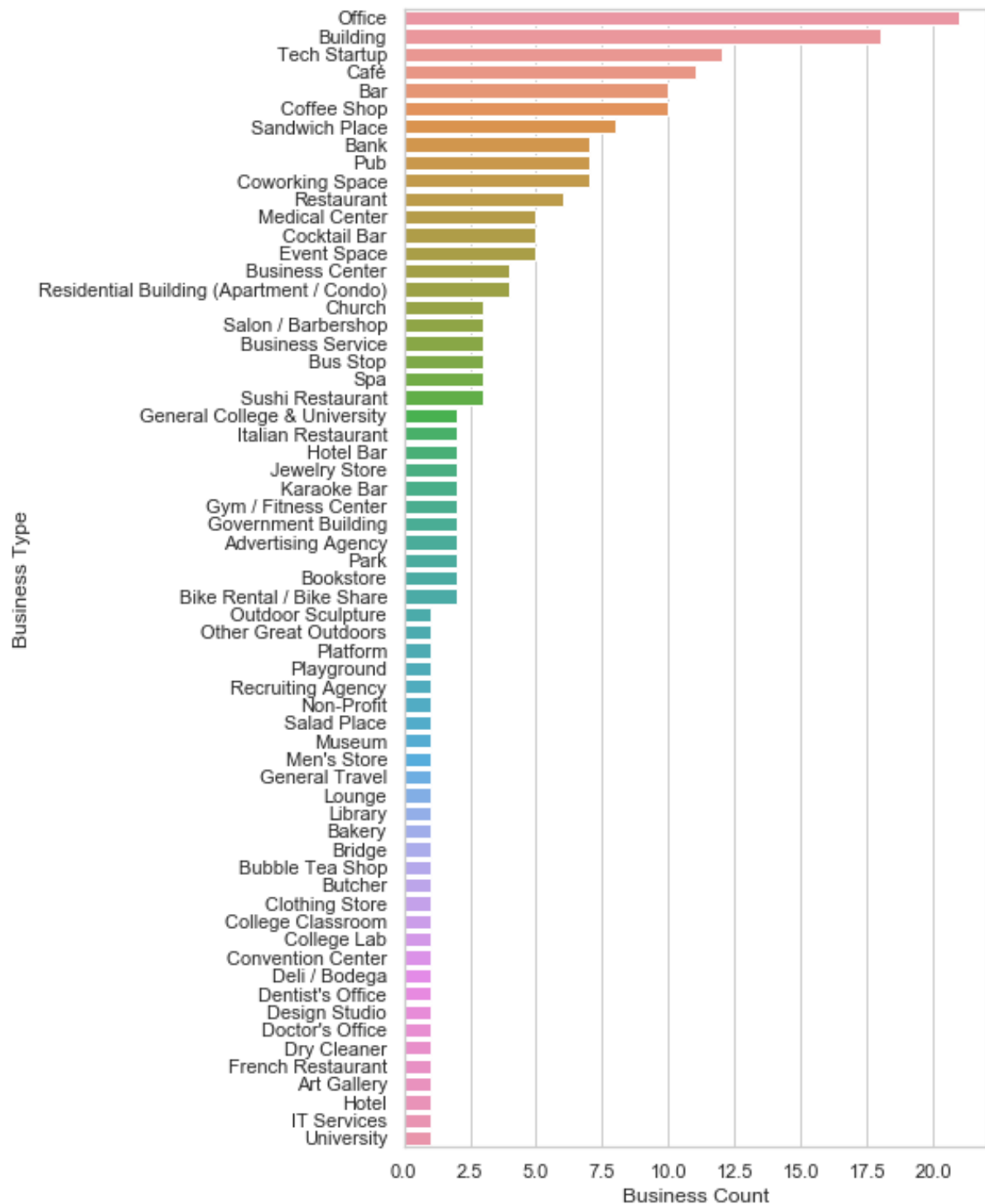


3.6.2.2. Cluster 2

Cluster 2 (similar to Cluster 1) has 21 neighborhoods in it, these are "St Luke's, Bunhill Fields", "Northern 'Victoria' around Buckingham Gate ", 'Drury Lane, Kingsway, Aldwych', "All but the west of St James's", 'Blackfriars', 'Temple', "Lincoln's Inn Fields, Royal Courts of Justice, Chancery Lane", 'Fetter Lane', 'South or lower Belgravia and Chelsea, from Sloane Square to Victoria Station to Grosvenor Waterside basin', 'Clerkenwell, Farringdon', 'High Holborn', 'Guildhall', 'Old Broad Street, Tower 42', 'St Mary Axe, Aldgate', "Lloyd's of London, Fenchurch Street", 'Parliament Square, Nearby

Westminster School to Westminster Cathedral to CCAL, University of the Arts, London', 'Bank of England', 'Cannon Street'

Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.

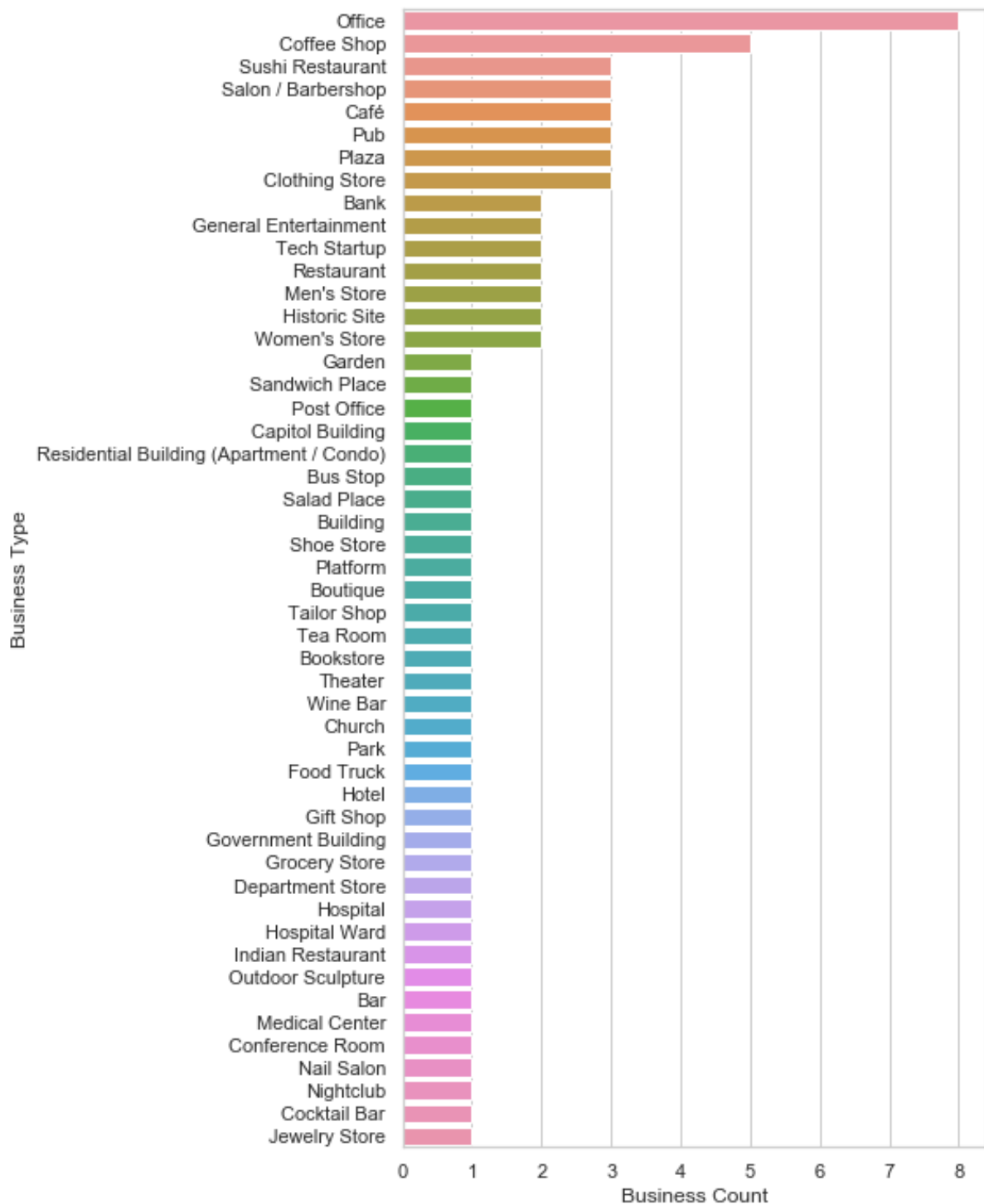


The number of Residential Accommodation is significantly lower in Cluster 2 when compared to Cluster 1

3.6.2.3. Cluster 3

Cluster 3 has 8 neighborhoods within it, these are 'Oxford Street (west)', 'Mayfair (south), Piccadilly', 'Cornhill, Gracechurch Street, Lombard Street', 'Leicester Square, St. Giles', 'St Paul's', 'Monument, Billingsgate', 'Whitehall, Buckingham Palace, the Houses of Parliament and west of St James's', 'Whitechapel, Spitalfields, Shadwell, Limehouse, Stepney, Portsoken'.

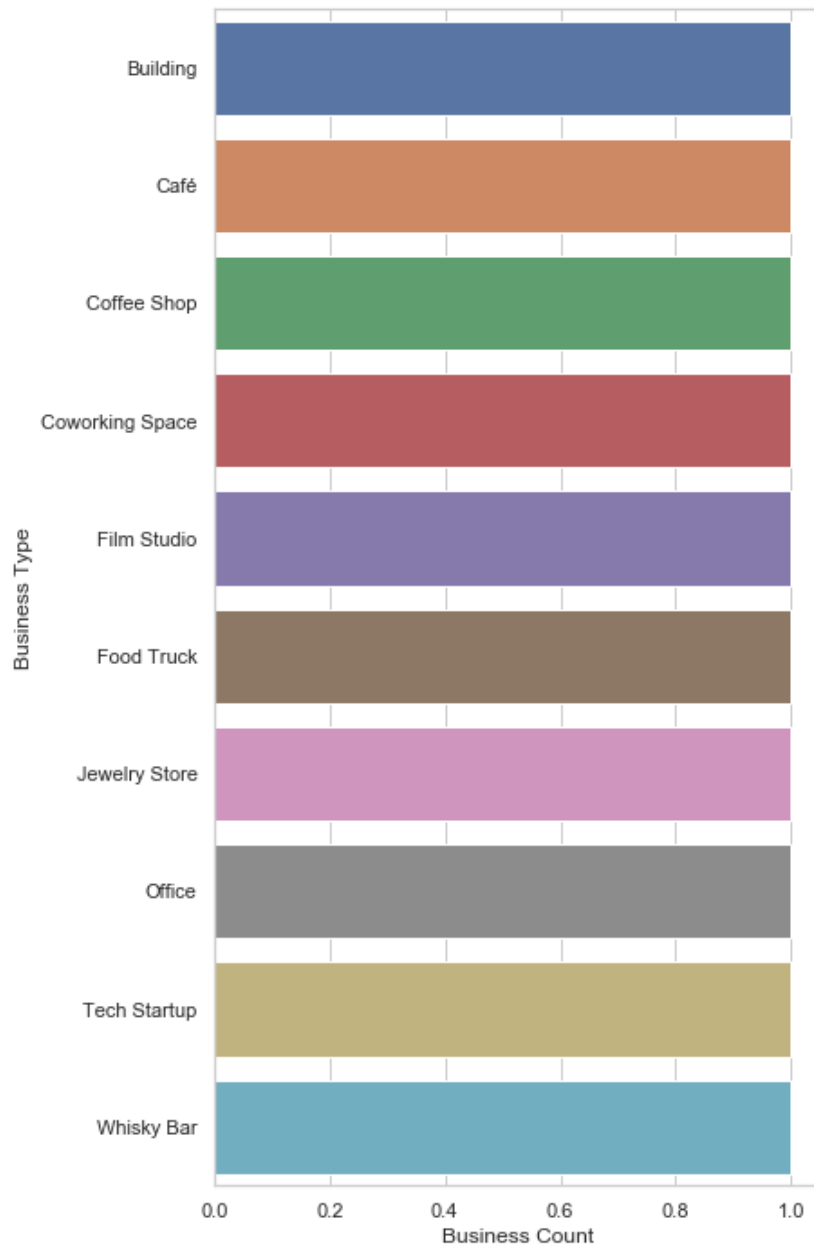
Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.



3.6.2.4. Cluster 4

Cluster 4 has 1 neighborhood within it, it is Hatton Garden.

Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.

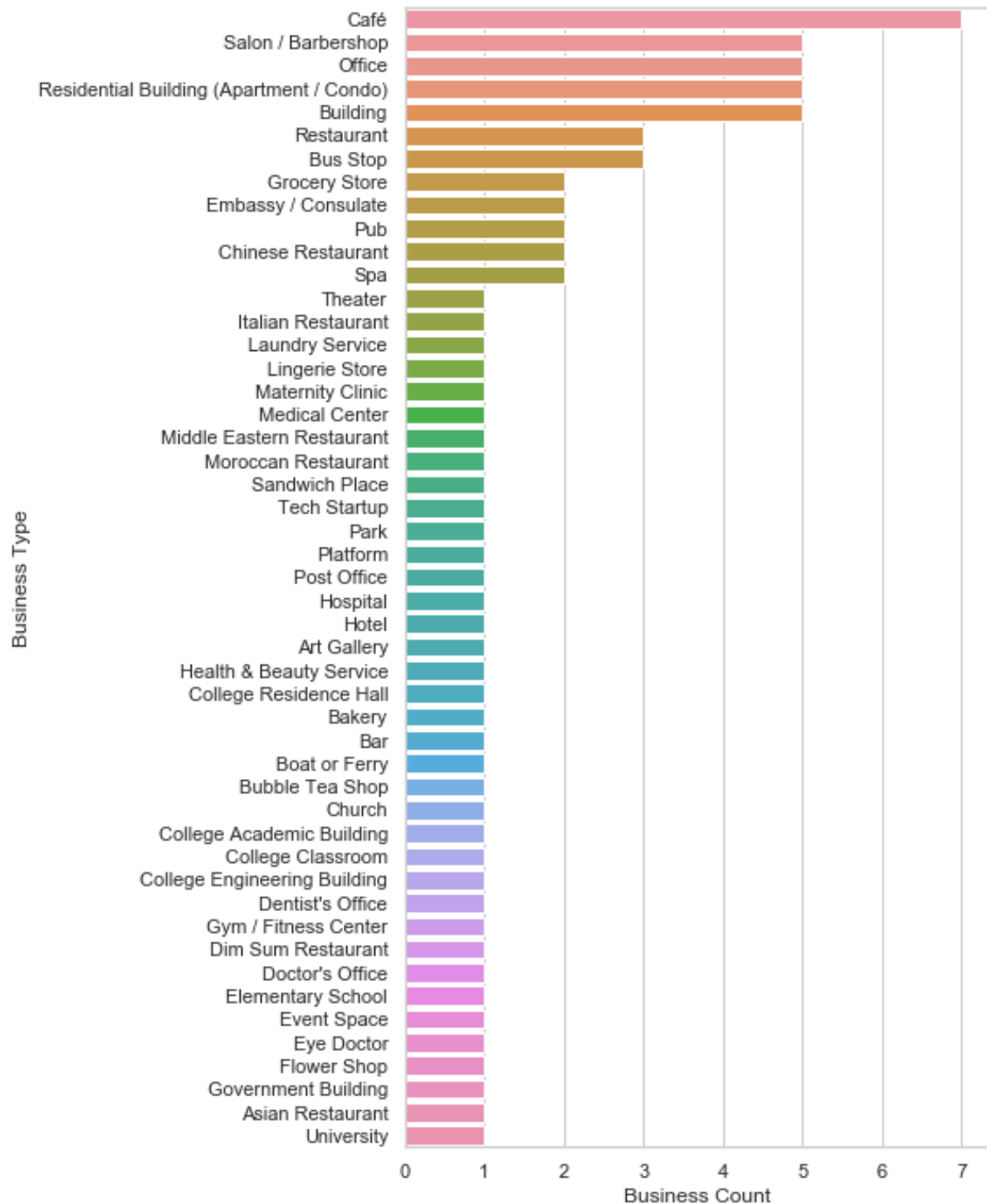


3.6.2.5. Cluster 5

Cluster 5 has 8 neighborhoods within it, these are 'Rest of Belgravia (north of Eaton Square), Knightsbridge (eastern half) and Chelsea (a corner)', 'Soho (south east); Chinatown, Soho Square', 'South Kensington district: South Kensington, Knightsbridge (part)', 'Maida Hill district: Maida Hill, Maida Vale, Little Venice (part)', 'Portland Place, Regent Street', 'St John's Wood, Primrose Hill

(south), Marylebone (north), Lisson Grove (north)", "Kilburn, Brondesbury, West Hampstead, Queen's Park", "North Kensington, Kensal Town, Ladbroke Grove (north), Queen's Park (part)"

Based on search of venues conducted for specific categories we identified venues and the barchart below provides an overview of the same.



4. Results

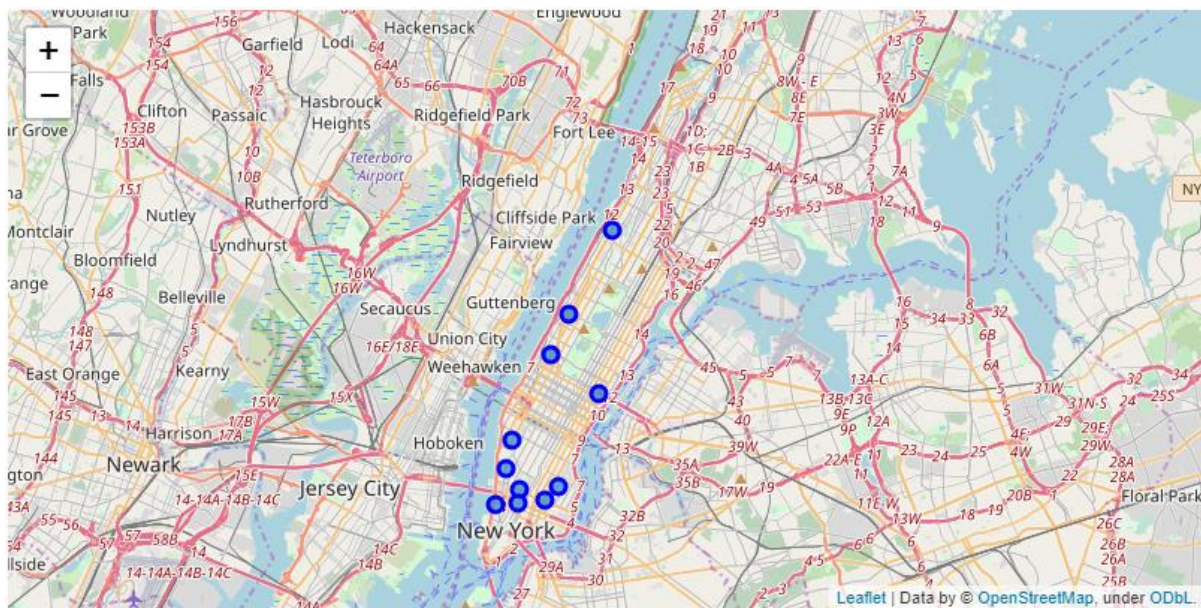
4.1. Recommended Neighborhoods in Manhattan

Cluster 1 from the K-means of both the Explore Query and the Search Query were identified as the most suitable. The neighborhoods from both clusters were put in sets.

An intersection of both sets is the set of recommended neighborhoods which the target audience should consider should they want to relocate to Manhattan. A total of 11 Neighborhoods were included.

Neighborhoods recommended are 'Lincoln Square', 'Sutton Place', 'Greenwich Village', 'East Village', 'Soho', 'Tribeca', 'Manhattanville', 'Upper West Side', 'Noho', 'West Village', 'Chelsea'.

These are visualized on the folium map below.



4.2. Recommended Neighborhoods in City of London & City of Westminster

Cluster 1 from the Search Query results and Cluster 5 from the Explore Query results were identified as the most suitable. The neighborhoods from both clusters were put in sets.

An intersection of both sets is the set of recommended neighborhoods which the target audience should consider should they want to relocate to the City of London & City of Westminster. A total of 9 neighborhoods were identified.

Neighborhoods included are 'Barbican', 'Between Vauxhall Bridge, Grosvenor Bridge and Victoria Station, includes Pimlico', 'St Bartholomew's Hospital', 'Shoreditch', 'Mansion House', 'Broadway, Queen Anne's Gate and Old Queen Street', 'Great Portland Street, Fitzrovia', 'Euston, Regent's Park, Baker Street, Camden Town, Somers Town', 'Bankside, South Bank, Southwark'.

These are visualized on the folium map below.



5. Observations

The explore query returns recommended venues and these are then selected. As part of this project we will be identifying the intersection of the selected 'Search' query cluster and the largest 'Explore' query cluster to recommend neighborhoods that offer a certain level of features uniformly. This will be the best fit result.

The limitation is because we can have intersections between the selected 'Search' query cluster, and any 'Explore' query cluster. Assuming the intersection is not a null set it's a potential target neighborhood to consider that offers a different level of features uniformly.

This will depend on individual target audience and for the purpose of this report the other options are ignored.

6. Conclusion

In this report we analyzed the neighborhoods of Manhattan and the City of London and City of Westminster. We used Foursquare API's Explore and Search to get location specific data. The results we got from the query became the data set for further analysis. We used K-means clustering as a way of identifying neighborhoods with similar features and used the criteria of evaluation set to shortlist key cluster. Set theory was applied on the neighborhoods and the intersection of the 2 sets became the recommended cluster.