

Applied Data Science Capstone Project

London Restaurants

Samir Shah

Contents	Page
1. Background	2
2. Hypothesis & Problem Statement	2
3. Data Source	2
4. Methodology	3
5. Results	5
6. Discussion	12
7. Conclusion	12
8. Appendix	13

1. Background

Over the last half century, London's restaurant scene has evolved into one of the most diverse and dynamic on the planet and is often quoted as "the food capital of the world".

London is firmly established as the world's most exciting restaurant scene. With 17,000 restaurants, it may not have as many as Paris (40,000) or New York (45,000) but with a current rate of about 200 new non-chain openings a year, London's dynamism, depth, breadth and diversity is unmatched. Authentic takes on every cuisine are available, from Senegalese to Szechuan, and some of the finest chefs in the world (both home-grown and from abroad) ply their trade in the capital. Concepts created in the city have been exported around the globe.

One growing trend is for healthier eating. In particular, demand for meat-free food increased by 987% in 2017 in the UK, and going vegan was predicted to be the biggest trend in 2018.

2. Hypothesis & Problem Statement

Whilst many restaurants cater for vegetarians and vegans, there are still relatively very few restaurants in London that are 100% vegetarian/vegan. Thus, there is a potential for a business to gain entry into a growing market with little competition. As a potential restaurant owner, it would be important to understand the dynamics of such a market including factors such as population growth, demographics and of course how saturated or not the London market is with respect to vegetarian or vegan restaurants.

3. Data Sources

For the purpose of this project data will be sourced and used as follows:

- 1) Demographic and population trend data from the London Datastore, an open data portal made available from the Greater London Authority. This data will be used to assess the potential for growth for London and will be used to chart (e.g. histograms) to show trends such as which parts of London have the highest workday population. I will also source profiling data on vegans in the UK.
- 2) Foursquare is a location provider with more than 105 million venues mapped around the world. Its location data will be used to assess neighborhoods and to determine the current number and location of vegetarian/vegan restaurants using k-means clustering and what might be a suitable location for a new vegetarian/vegan restaurant. I will utilise the Folium library in Python to map the locations.

4. Methodology

The following analysis was undertaken:

- 1) Obtained population historical and forecast growth data from 1950 to 2030 from the London Datastore for major global cities. The data was read into a dataframe and cleansed. It was then sliced to pull out data on London, which was the focus of this analysis. A scatter plot was then plotted on the London population data. This showed that data would require a polynomial regression and a line of best fit was obtained:

$$y=1x^2-5062x+5001621$$

- 2) Age profiling data was obtained on vegans in the UK. The data was read into a dataframe, cleansed, and a histogram plotted showing the percentage of the vegans by age range in the UK.
- 3) UK population data incorporating age ranges was then obtained from the London Data Store. The data was cleansed and sliced to focus on London. This data had the population of London split by age range, but in different age ranges to that obtained on vegans in the UK. The London age range data was subsequently wrangled to match the UK vegan age range. As the data was actual population data as opposed to percentages, calculations were performed to obtain the percentages. A histogram was then plotted.
- 4) Restaurants and cafes cater for meals at all times of the day, and it was therefore important to understand the workday population of London. UK workday population data was obtained from the London Data Store, cleansed and sliced to focus on London. Similarly, to the general London population data, the workday population data had to be wrangled to match the UK vegan age range.

The data was then sorted by London district in descending order of workday population. It was felt at this stage of the analysis to initially focus on the top 5 London districts, as these would, based simply on size, provide the most opportunity for new restaurant opportunities.

The results of each of these districts was then read into separate dataframes, calculations performed to obtain percentages and histograms plotted.

- 5) The age range profile data for UK vegans, London population and the 5 top workday population districts of London were then combined into a single dataframe. Based on the comparison, Westminster was chosen as the district for further analysis using Foursquare.
- 6) Geographical coordinates of the Westminster neighborhoods were obtained and inserted into a dataframe.

A map of Westminster with markers for the neighborhoods was produced using Folium. Top 100 venues were obtained. The json file was then cleansed and fit into a pandas dataframe.

A function was created to repeat the same process for all the Westminster neighborhoods. Number of venues was returned for each neighborhood as well as unique venue categories.

Each neighborhood was analysed and details obtained of each category of venue present in that neighbourhood. These were then grouped to show the mean of the frequency of each category. The top 5 venues by frequency of occurrence was obtained for each neighborhood.

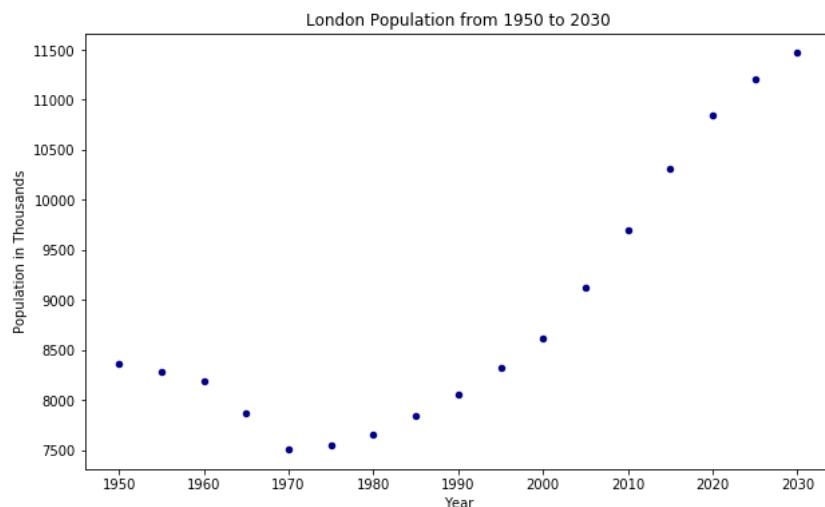
The top ten most common venues for each neighborhood was obtained.

K-Means clustering was then undertaken, and individual clusters examined further to determine the discriminating venue categories that distinguish each cluster.

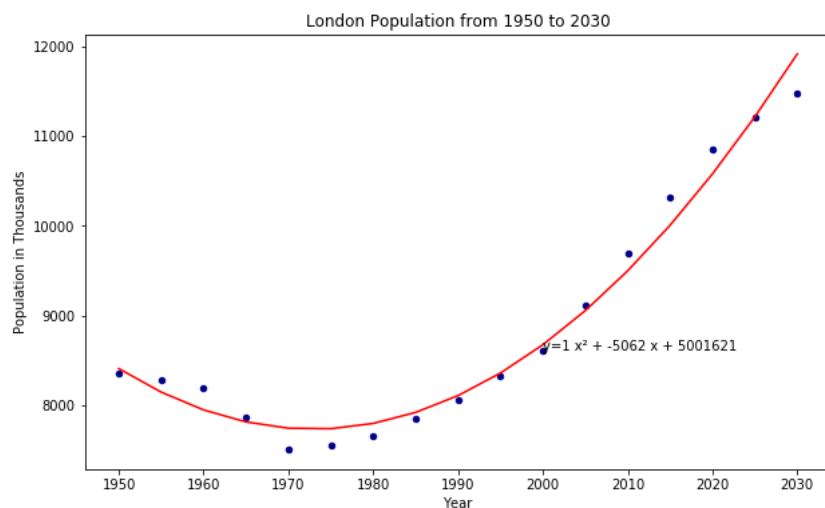
5. Results

London Population Growth

The London population having declined from 1950 to 1970 has subsequently increased, with considerable growth in the last two decades. Projected growth indicates that the population is expected to reach 11.5million by 2030.

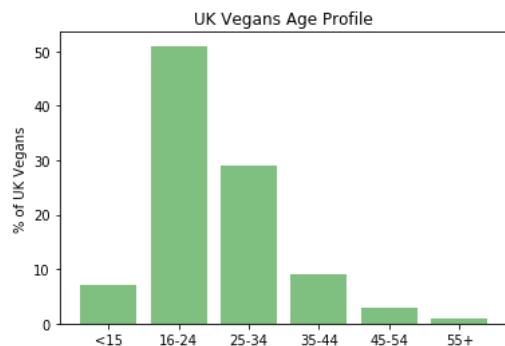


Applying polynomial regression gave the line of best fit as $y=1x^2-5062x+5001621$



UK Vegan Age Profile

It would appear that in the UK, the highest percent of UK Vegans are in 16-24 age range followed by the 25-34 range.



London Population Age Profile

However, London's population itself is split fairly evenly across age ranges, with the highest being in the 25-34 group.



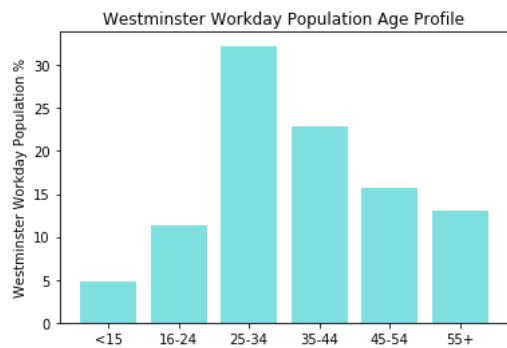
London's Workday Population

Analysis indicated that 5 inner districts of London, had the highest workday populations:

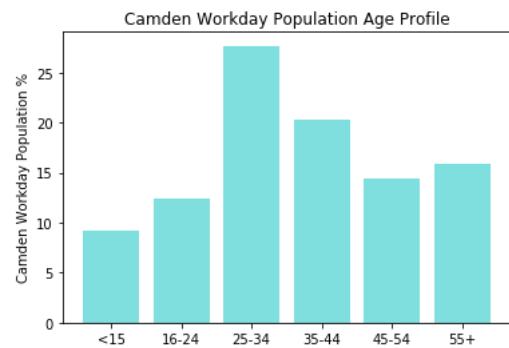
Area	All categories: Age	Age						
		<15	16-24	25-34	35-44	45-54	55+	
0	Westminster	689572	32584	78035	222510	157219	108701	90523
1	Camden	384107	35465	47738	106369	78150	55583	60802
2	Tower Hamlets	368200	50143	45739	104970	78444	48571	40333
3	City of London	360075	620	29683	138140	102606	60674	28352
4	Southwark	324494	53382	41015	74004	59282	47397	49414
5	Barnet	314492	74234	37036	46760	42295	37772	76395
6	Croydon	310641	79110	37106	40792	39875	39272	74486
7	Hillingdon	308668	56893	37987	50835	49609	45939	67405
8	Ealing	306007	68877	34993	53307	45920	38206	64704
9	Enfield	280224	70001	34359	38905	37567	35139	64253

Focusing further analysis on the top 5 districts showed the age profiles of each of these districts as follows:

Westminster:



Camden:



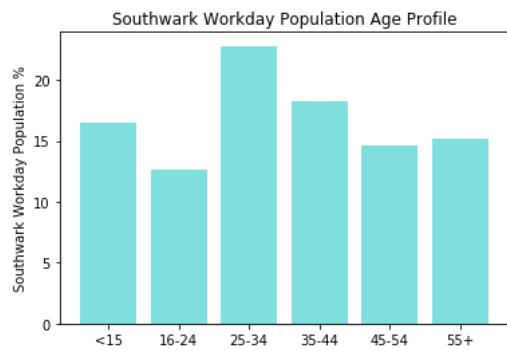
Tower Hamlets:



City of London:



Southwark:



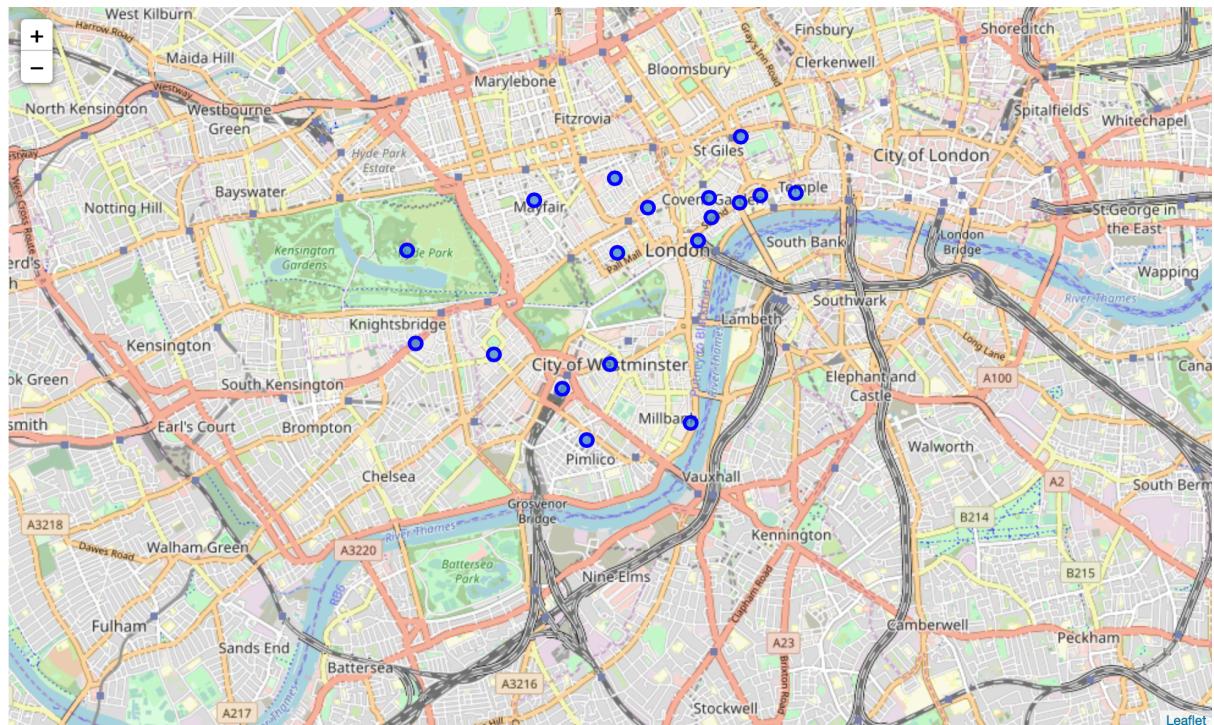
Common to each was the fact that age ranges 25-34 was the highest, followed by 35-44.

When combining the data into a single table as per below, it was evident that there was no perfect match for age profile range of UK vegans. The closest appeared to be Westminster, based on the relative portion of 16-24- and 25-34-year old's.

Age Range	UK Vegans	London	Westminster	Camden	Tower Hamlets	City of London	Southwark
0 <15	7	19.0	5.0	9.0	14.0	0.0	16.0
1 16-24	51	13.0	11.0	12.0	12.0	8.0	13.0
2 25-34	29	20.0	32.0	28.0	29.0	38.0	23.0
3 35-44	9	16.0	23.0	20.0	21.0	28.0	18.0
4 45-54	3	12.0	16.0	14.0	13.0	17.0	15.0
5 55+	1	20.0	13.0	16.0	11.0	8.0	15.0

Utilising Foursquare to Find Venues

Westminster districts were plotted using Folium:



Venue information was obtained using Foursquare:

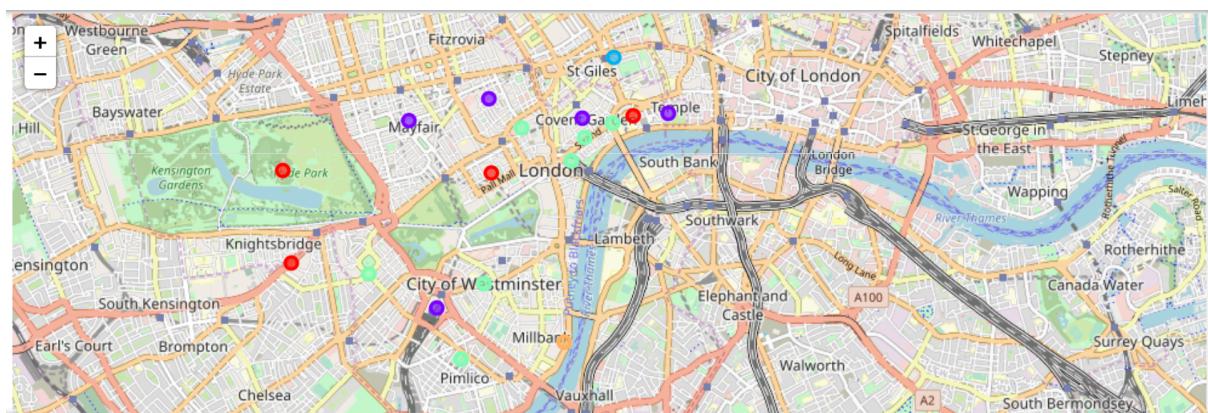
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Adelphi	100		100	100	100	100
Aldwych	100		100	100	100	100
Belgravia	58		58	58	58	58
Charing Cross	100		100	100	100	100
Chinatown	100		100	100	100	100
Covent Garden	100		100	100	100	100
Holborn	100		100	100	100	100
Hyde Park	22		22	22	22	22
Knightsbridge	100		100	100	100	100
Mayfair	100		100	100	100	100
Millbank	54		54	54	54	54
Pimlico	63		63	63	63	63
Soho	100		100	100	100	100
St James	100		100	100	100	100
Strand	100		100	100	100	100
Temple	81		81	81	81	81
Victoria	100		100	100	100	100
Westminster	75		75	75	75	75

The data showed there were 201 unique categories of venues in the Westminster area. The shape of the datafile was 1553 x 202. The data showed that across the Westminster districts there were a limited number of vegetarian/vegan restaurants – located in Millbank, Temple and Soho:

	Neighborhood	Accessories Store	African Restaurant	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	... Udon Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	
0	Adelphi	0.01	0.00	0.01	0.00	0.000000	0.000000	0.000000	0.000000	0.010000	...	0.00	0.000000	0.00
1	Aldwych	0.00	0.00	0.02	0.00	0.010000	0.010000	0.010000	0.000000	0.000000	...	0.00	0.000000	0.00
2	Belgravia	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.00
3	Charing Cross	0.01	0.01	0.01	0.00	0.000000	0.010000	0.010000	0.020000	0.010000	...	0.00	0.000000	0.00
4	Chinatown	0.00	0.00	0.01	0.00	0.000000	0.010000	0.010000	0.010000	0.000000	...	0.01	0.000000	0.01
5	Covent Garden	0.01	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.010000	0.010000	...	0.00	0.000000	0.00
6	Holborn	0.00	0.00	0.00	0.01	0.010000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.00
7	Hyde Park	0.00	0.00	0.00	0.00	0.000000	0.045455	0.000000	0.000000	0.000000	...	0.00	0.000000	0.00
8	Knightsbridge	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.010000	...	0.00	0.000000	0.00
9	Mayfair	0.01	0.00	0.00	0.00	0.000000	0.060000	0.000000	0.000000	0.010000	...	0.00	0.000000	0.00
10	Millbank	0.00	0.00	0.00	0.00	0.000000	0.111111	0.000000	0.000000	0.018519	...	0.00	0.018519	0.00
11	Pimlico	0.00	0.00	0.00	0.00	0.015873	0.000000	0.000000	0.015873	0.000000	...	0.00	0.000000	0.00
12	Soho	0.00	0.00	0.01	0.00	0.000000	0.010000	0.000000	0.010000	0.000000	...	0.00	0.010000	0.00
13	St James	0.01	0.00	0.00	0.00	0.010000	0.030000	0.010000	0.000000	0.010000	...	0.00	0.000000	0.00
14	Strand	0.00	0.00	0.03	0.00	0.000000	0.010000	0.010000	0.000000	0.000000	...	0.00	0.000000	0.00
15	Temple	0.00	0.00	0.00	0.00	0.012346	0.012346	0.012346	0.000000	0.012346	...	0.00	0.012346	0.00
16	Victoria	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.010000	0.000000	...	0.00	0.000000	0.00
17	Westminster	0.00	0.00	0.00	0.00	0.000000	0.013333	0.000000	0.000000	0.000000	...	0.00	0.000000	0.00

Undertaking K-Means clustering (using 5 clusters resulted in the following analysis and map. None of the districts had a vegetarian or vegan restaurant within the top 10 most common venue:

Neighborhood		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
0	Adelphi	Burger Joint	French Restaurant	Ice Cream Shop	Coffee Shop	Dessert Shop	Theater	Hotel	Wine Bar	Cocktail Bar	Italian Restaurant
1	Aldwych	Theater	Hotel	Pub	Coffee Shop	Cocktail Bar	Italian Restaurant	French Restaurant	Burger Joint	Restaurant	Dessert Shop
2	Belgravia	Hotel	Café	Restaurant	Hotel Bar	Boutique	Italian Restaurant	Plaza	Shoe Store	Clothing Store	Gastropub
3	Charing Cross	Hotel	Theater	Pub	Ice Cream Shop	French Restaurant	Wine Bar	Garden	Plaza	Steakhouse	Pizza Place
4	Chinatown	Theater	Italian Restaurant	Cocktail Bar	Restaurant	Ice Cream Shop	Hotel	Japanese Restaurant	Seafood Restaurant	Lounge	Ramen Restaurant
5	Covent Garden	Burger Joint	French Restaurant	Ice Cream Shop	Dessert Shop	Wine Bar	Clothing Store	Theater	Coffee Shop	Sushi Restaurant	Indian Restaurant
6	Holborn	Coffee Shop	Pub	Restaurant	Café	Hotel	Theater	Japanese Restaurant	Korean Restaurant	Bakery	Bookstore
7	Hyde Park	Bar	Scenic Lookout	Modern European Restaurant	Bakery	General Entertainment	Monument / Landmark	Café	Fountain	Outdoor Sculpture	Boat or Ferry
8	Knightsbridge	Café	Boutique	Italian Restaurant	Hotel	Japanese Restaurant	Clothing Store	Seafood Restaurant	Tea Room	Middle Eastern Restaurant	Restaurant
9	Mayfair	Boutique	French Restaurant	Art Gallery	Clothing Store	Coffee Shop	Hotel	Steakhouse	Italian Restaurant	Lounge	Hotel Bar
10	Millbank	Art Gallery	Café	Hotel	Park	Coffee Shop	Sandwich Place	Garden	Restaurant	Plaza	Bar
11	Pimlico	Hotel	Pub	Italian Restaurant	Sandwich Place	Café	Turkish Restaurant	Indian Restaurant	Park	Thai Restaurant	Chinese Restaurant
12	Soho	Coffee Shop	Tapas Restaurant	Cocktail Bar	Italian Restaurant	Bakery	BBQ Joint	Record Shop	English Restaurant	Indian Restaurant	Pizza Place
13	St James	Hotel	Indian Restaurant	Clothing Store	Boutique	Cocktail Bar	Dessert Shop	Bookstore	Lounge	Art Gallery	Italian Restaurant
14	Strand	Theater	French Restaurant	Coffee Shop	Cocktail Bar	Dessert Shop	Burger Joint	American Restaurant	Hotel	Italian Restaurant	Restaurant
15	Temple	Pub	Sandwich Place	Coffee Shop	Restaurant	Building	Italian Restaurant	Hotel	Bar	Scenic Lookout	Japanese Restaurant
16	Victoria	Coffee Shop	Italian Restaurant	Hotel	Café	Sandwich Place	Theater	Pub	Chinese Restaurant	Gym / Fitness Center	Turkish Restaurant
17	Westminster	Hotel	Sandwich Place	Coffee Shop	Italian Restaurant	Theater	Sushi Restaurant	Pub	Café	Juice Bar	Hotel Bar



Each of the clusters were diagnosed further to determine the discriminating venue categories for each:

Cluster 1

Latitude	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	
2	51.5121	Theater	Hotel	Pub	Coffee Shop	Cocktail Bar	Italian Restaurant	French Restaurant	Burger Joint	Restaurant	Dessert Shop
8	51.5073	Bar	Scenic Lookout	Modern European Restaurant	Bakery	General Entertainment	Monument / Landmark	Café	Fountain	Outdoor Sculpture	Boat or Ferry
9	51.4991	Café	Boutique	Italian Restaurant	Hotel	Japanese Restaurant	Clothing Store	Seafood Restaurant	Tea Room	Middle Eastern Restaurant	Restaurant
13	51.5070	Hotel	Indian Restaurant	Clothing Store	Boutique	Cocktail Bar	Dessert Shop	Bookstore	Lounge	Art Gallery	Italian Restaurant

Cluster 2

Latitude	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	
6	51.5118	Burger Joint	French Restaurant	Ice Cream Shop	Dessert Shop	Wine Bar	Clothing Store	Theater	Coffee Shop	Sushi Restaurant	Indian Restaurant
10	51.5116	Boutique	French Restaurant	Art Gallery	Clothing Store	Coffee Shop	Hotel	Steakhouse	Italian Restaurant	Lounge	Hotel Bar
15	51.5136	Coffee Shop	Tapas Restaurant	Cocktail Bar	Italian Restaurant	Bakery	BBQ Joint	Record Shop	English Restaurant	Indian Restaurant	Pizza Place
16	51.5123	Pub	Sandwich Place	Coffee Shop	Restaurant	Building	Italian Restaurant	Hotel	Bar	Scenic Lookout	Japanese Restaurant
17	51.4952	Coffee Shop	Italian Restaurant	Hotel	Café	Sandwich Place	Theater	Pub	Chinese Restaurant	Gym / Fitness Center	Turkish Restaurant

Cluster 3

Latitude	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	
7	51.5172	Coffee Shop	Pub	Restaurant	Café	Hotel	Theater	Japanese Restaurant	Korean Restaurant	Bakery	Bookstore

Cluster 4

Latitude	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	
0	51.4973	Hotel	Sandwich Place	Coffee Shop	Italian Restaurant	Theater	Sushi Restaurant	Pub	Café	Juice Bar	Hotel Bar
1	51.5101	Burger Joint	French Restaurant	Ice Cream Shop	Coffee Shop	Dessert Shop	Theater	Hotel	Wine Bar	Cocktail Bar	Italian Restaurant
3	51.4982	Hotel	Café	Restaurant	Hotel Bar	Boutique	Italian Restaurant	Plaza	Shoe Store	Clothing Store	Gastropub
4	51.5081	Hotel	Theater	Pub	Ice Cream Shop	French Restaurant	Wine Bar	Garden	Plaza	Steakhouse	Pizza Place
5	51.5110	Theater	Italian Restaurant	Cocktail Bar	Restaurant	Ice Cream Shop	Hotel	Japanese Restaurant	Seafood Restaurant	Lounge	Ramen Restaurant
12	51.4907	Hotel	Pub	Italian Restaurant	Sandwich Place	Café	Turkish Restaurant	Indian Restaurant	Park	Thai Restaurant	Chinese Restaurant
14	51.5114	Theater	French Restaurant	Coffee Shop	Cocktail Bar	Dessert Shop	Burger Joint	American Restaurant	Hotel	Italian Restaurant	Restaurant

Cluster 5

Latitude	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	
11	51.4922	Art Gallery	Café	Hotel	Park	Coffee Shop	Sandwich Place	Garden	Restaurant	Plaza	Bar

6. Discussion

Based on the analysis undertaken, it is worth noting the following:

1. The London population has seen considerable growth and is expected to grow over the next few years to reach 11.5 million by 2030, creating opportunities for businesses.
2. Whilst the age profile of the London neighborhoods does not match 100% the age profile of UK vegans, opportunity exists to target the 25-34 age group and also the younger age bands. In particular, the Westminster area of London had the largest workday population with age range 25-34.
3. Reviewing the type of venues in the Westminster area clearly showed that restaurants, cafes, burger joints, pubs, etc. were some of the top venues – this is not surprising as the Westminster area also caters heavily for tourists. What was also clear was that there were very few restaurants which could be classified as 100% vegan or vegetarian, and where they did exist, they were located in only 3 districts of Westminster – Millbank, Temple and Soho (in that order).

There exist opportunities for 100% vegetarian / vegan restaurant in the Westminster area. However, any such venture would require further analysis and research such as understanding the costs of opening and running a restaurant in London, availability of staff, demand (based on for example consumer surveys), etc.

7. Conclusion

This report was produced utilizing third party data sources on the population of London, age-profile of UK vegans and workday populations of London neighborhoods.

The analysis undertaken indicates that opportunities exist for 100% vegetarian/vegan restaurants in the Westminster area of London.

However, it should be noted that this analysis does not equate to viability of such a venture as further analysis as has been outlined in the Discussion section would be required to be undertaken.

8. Appendix

This section of the report contains the Python code written to undertake the above analysis. The code was written in Jupyter Notebooks and saved on GitHub.