The Battle of the Wards (Neighborhoods) in London

Opening a new Luxury Wine Bar in London's ward.

Applied Data Science Capstone by IBM/Coursera - Capstone Project

Sebastian Nastuta

Applied Data Science Capstone by IBM/Coursera - Capstone Project

This project is part of the **IBM Applied Data Science Capstone** on Coursera. The main objectives of the project were to define a business problem (where to invest), to identify data in the web and use *Foursquare location API* to compare different neighbourhoods in terms of suitability of investment.

Table of Contents

1. Business Problem and Background

2. Methodology

3. Data access

4. Analysis: Mapping London Wards (neighborhoods)

4.1. Create map of all wards in London

4.2. Explore Wards in London through Venue Clustering

4.3. Clustering the Wards/ Neighborhoods in London

4.3.1. London Wards distribution according to venue Clustering

4.4. Description of the Cluster Segments

4.5. Demographic description of London Boroughs

4.5.1. Aspects correlated with Household Income in London Wards

5. Results and Discussion

5.1. Identification of High-Income Wards as potential places of opening a Luxury Wine Bar

5.2. Overall discussion of the results

6. Recommendations

1. Business Problem and Background

Project structure

The project will evaluate in which **wards** of London would be preferable to invest for opening a new **Wine Bar** or a chain of wine bars. The project will describe all analytical steps from data access, data analysis and visualisation to final recommendation of 5 to 10 wards in which is recommendable to open a new wine bar.

In this project, we will explain in a step by step manner from *business problem and background*, to *data access* and preparation, from *methodology* to *statistical data analysis* and conclusions. A set of recommendations will be proposed to the business stakeholders to make their decisions related to feasibility of investment.

Although neighbourhoods are proposed as units of analysis the wards will be analysed as these administrative territorial units of London are easily identifiable by people and they are big enough as population size to be relevant from business perspective.

Business problem

Having a stable population of 8.6 million people and a daytime population of over 10 million people (**London Datastore**), London is the biggest city in Europe. London has a relatively young population, of people who mainly come here to study and work. London is extremely diverse from cultural and ethnic point of view has a strong hospitality sector consisting mainly in restaurant, hotels, pubs, bars and coffee shops. Especially pubs are extremely popular in London as they represent an important socialising place for Londoners and tourists. According to *2016 Pub Audit*, there were **3615** pubs in London a decline from 4835 pubs back in 2001 (**Estate agent London News** , 2018). Although many pubs might closed their doors due to high rents and taxes and strong competitive environment, another important factor can be the change in consumer behaviour as a considerable segment of consumers prefer to switch from drinking beer which is the dominant product sold in pubs to drinking wines which tends to be bought for domestic consumption but also consumed in wine bars. As millennials seem to be spend more money on good quality wines, opening a chain of wine bars in London can potentially answer to an increasing need to educate consumers in this direction. This report can be of interest for investors who might be interested in creating an alternative gathering place to pubs, coffee shops and restaurants. Although wines can be served in multiple places it is mainly bought by Londoners from retails stores and consumed at home. Wine bars can potentially offer a more sophisticated place to spend time in spaces less crowded such as pubs.

Socio-demographic profile of wine consumers

According to a study proposed by **Duke University** (2014) wine consumers can overall be classified as **overwhelmed (23%)** by the available choices on store shelves, **Image Seekers (20%)** who use wine as a symbol of status, **traditionalists (16%)** who have clear established preferences, the **Savy Shoppers (16%)** who "enjoy shopping for wine and discovering new varietal s on their own", **Satisfied Sippers (14%)** who simply like to drink and **wine enthusiasts (12%)** who "consider themselves knowledgeable about wine". Although this study was not performed in London, it is reasonable to assume the profiles can easily fit London wine market.

2. Methodology

Several Python packages and statistical procedures will be involved in analysing the two datasets.

- Pandas **Crosstabulation** will be used to perform exploratory data analysis (**EDA**) on association between boroughs and cluster or between venues and clusters when the content of clusters will be described.
- Pandas package was used also for separating a string variable containing both Borough and Ward name into two separate columns, for excluding or keeping specific variables. Joining separate datasets was also performed using Pandas library.
- **K-means clustering** will be used to allocate each of London's wards into a cluster based on what venues they include,
- Matplotlib **histograms** and **scatter-plot graphs** will be used to plot the data and the association between demographic data,
- **Descriptive statistics** (means, etc) will be performed while grouping by boroughs to document the proposal we need to make regardinng the wards where to open the wine bars. Following the 1st stage in which some candidate wards will be identified based on their componence in the 2nd stage of the analysis the socio-demographic profiles of the wards will be analysed.

In the 1st step the venue cluster will be created. Secondly it will be decided which wards are more attractive for potential investments in luxury wine bars based on their demographic profile. Next the cluster of each selected wards will be identified, and a recommendation will be made where would be recommendable to open the wine bars.

3. Data access

Two datasets will be analysed in order to develop and propose some strategic recommendation regarding the feasibility of opening some Wine Bars in some of London wards.

3.1. London Wards geolocation dataset

The first dataset was accessed from **Data.Gov.UK** . From this webpage which offers geolocation data for all *8668 wards in the UK*, I already excluded the wards which are not part of Greater London Area which is our market of interest. Further down variables were renamed, excluded and cleaned of missing values. Finally only 4 variables such as Borough, Ward (Neighbourhood), Latitude and Longitude for **654 wards in Greater London Area** were kept for analysis.
Using the **Foursquare location API** the venues of interest in these wards will be identified and a KNN segmentation will be performed to clasify the wards based on the componence of their venues. Once created the profiles for each cluster will be developed using Pandas Crosstabulations.

	Neighborhood	Borough	Longitude	Latitude
0	Abbey	Barking and Dagenham	0.081291	51.539822
1	Alibon	Barking and Dagenham	0.150987	51.545921
2	Becontree	Barking and Dagenham	0.116912	51.552601
3	Chadwell Heath	Barking and Dagenham	0.138596	51.583420
4	Eastbrook	Barking and Dagenham	0.173453	51.555191

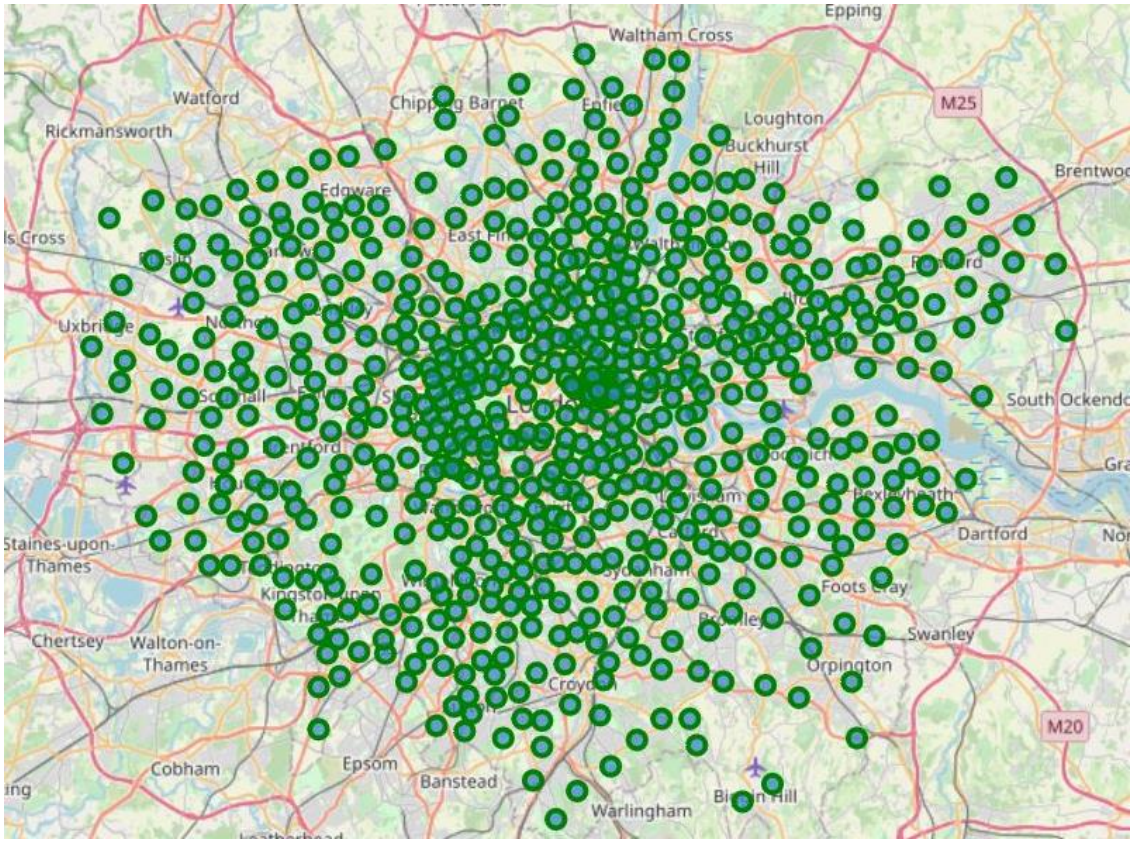
neighborhoods dataset will be used in sections **4.1**, **4.2** and **4.3.** of this report.

3.2. London wards socio-demographic profiling

Another dataset containing several social and demographic details at wards level was accessed from **London Datastore**.
This dataset contains details about population size, age structure, mean age per ward, population density, proportion of Black and Ethnic Minorities according to 2011 Census; children obesity, employment indicators, education, access to state benefits. The proportion of rented and owned houses in the ward or the council tax bands would offer more indications about the financial afluence of those living in a wards. Conbining the initial cluster structure with the socio-demographic profiling it will be possible to recommend in which wards is most recommended to open one of more Wine Bars.

	BOROUGH	Ward	ID_case	Ward name	Population	% Children 0-15	%_Working_age	Mean Age	Median Age	Area - Square Kilometres	...	% semi-detached houses - 2011	% terraced houses - 2011	Flats	% Owners	% Social Rents	F
0	Barking and Dagenham	Abbey	1	Barking and Dagenham - Abbey	14750	25	69	29	29	1	...	7	22	66	32	26	
1	Barking and Dagenham	Alibon	2	Barking and Dagenham - Alibon	10600	25	64	33	33	1	...	18	63	14	45	36	
2	Barking and Dagenham	Becontree	3	Barking and Dagenham - Becontree	12700	25	65	33	32	1	...	24	52	19	46	29	
3	Barking and Dagenham	Chadwell Heath	4	Barking and Dagenham - Chadwell Heath	10400	24	61	36	34	3	...	29	32	34	54	32	

4. Analysis: Mapping London Wards (neighborhoods) ¶



Using the Foursquare API, a new set of 12192 venues from London was generated.

	name	categories	lat	lng
0	Nando's	Portuguese Restaurant	51.539780	0.082297
1	Cristina's	Steakhouse	51.536523	0.076672
2	The Gym London Barking	Gym	51.536193	0.078601
3	Subway	Sandwich Place	51.538000	0.081319
4	Costa Coffee	Coffee Shop	51.539272	0.081341
5	B&M Store	Discount Store	51.537903	0.081301
6	Domino's Pizza	Pizza Place	51.539833	0.082223
7	Boots	Pharmacy	51.538888	0.080724
8	Lidl	Supermarket	51.535761	0.081843
9	Asda	Supermarket	51.537603	0.077375
10	Tesco Express	Grocery Store	51.536931	0.080411
11	Travelodge London Barking	Hotel	51.537277	0.079857
12	Barking Dog (Wetherspoon)	Pub	51.539659	0.081583
13	Tesco Express	Grocery Store	51.539637	0.084760
14	KFC	Fast Food Restaurant	51.539124	0.080955
15	Platform 4	Platform	51.539748	0.080395
16	Platform 1	Platform	51.539979	0.080547
17	Relish@BLC	Café	51.536312	0.079394

4.2. Explore Wards in London through Venue Clustering

The above created function was applied to repeat the same process to all the neighborhoods (**wards**) in London to identify all 12192 places of interest. While grouping the venues by **Venue Category** and sorting it descending by number it can be observed there are **841** Pubs, **710** Coffee shops, **630** Cafe and **438** Hotels. This indicate a strong demand for such leisure places which indicates potentially a solid investment opportunity. Interestingly, the histogram chart in this section indicates that out of 654 wards there are 400+ wards which have between 0 to 20 venues. This suggests these might be excluded from the list of potential investment places, reducing the number of neighborhoods of interest to around 200.

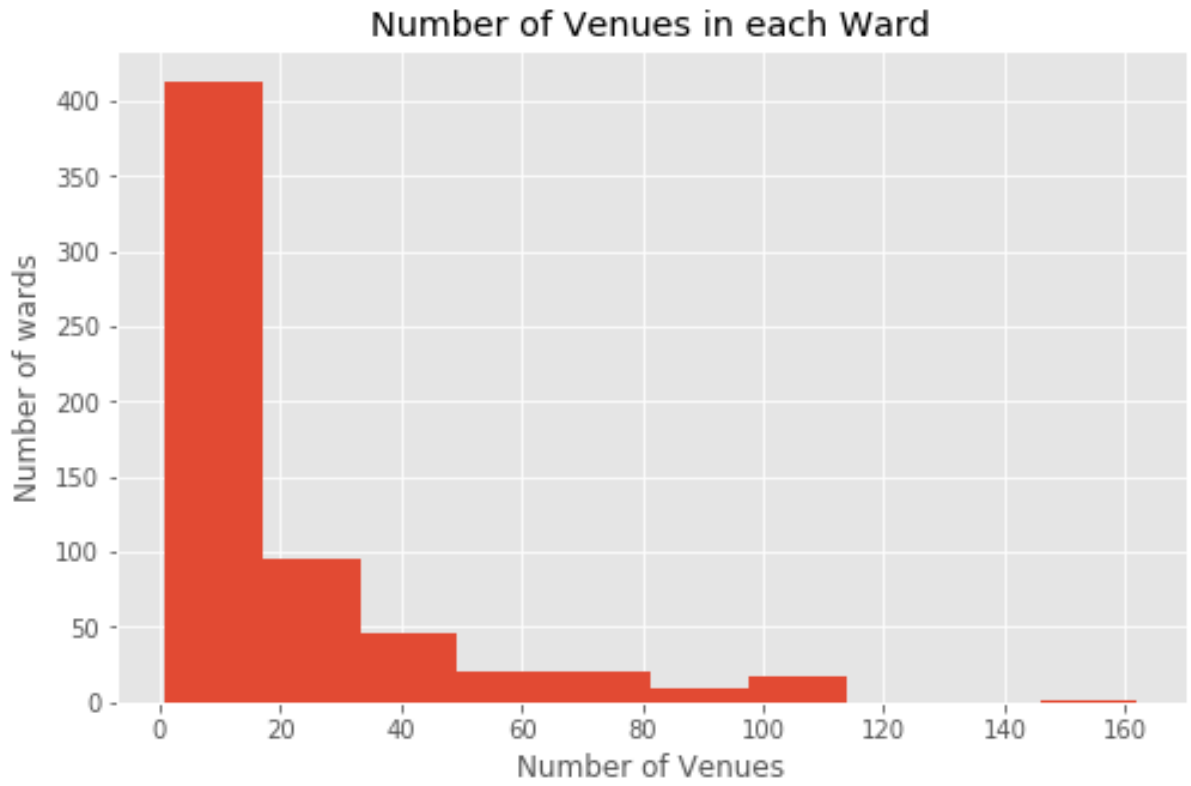


Figure 1. Histogram of Number of Venues in each neighborhood.

The histogram above indicates that out of 654 wards there are 400+ wards which have between 0 to 20 venues. This suggests these might be excluded from the list of potential investment places. Overall, there are already at least 12192 venues identified in London wards. The count table below suggest the pubs, coffee shops cafe-s, hotels and grocery store are the most popular venues as identified by the Foursquare API from the total of 425 unique categories.

For the next sequence of the analysis dummy variables from each venue were created increasing the number of variables to 426. Then a function was created to identify the top 5 places from each ward. Finally, a new dataset **neighborhoods_venues_sorted** was created and this included strictly the name of wards and the columns with the top 10 venues in each ward.

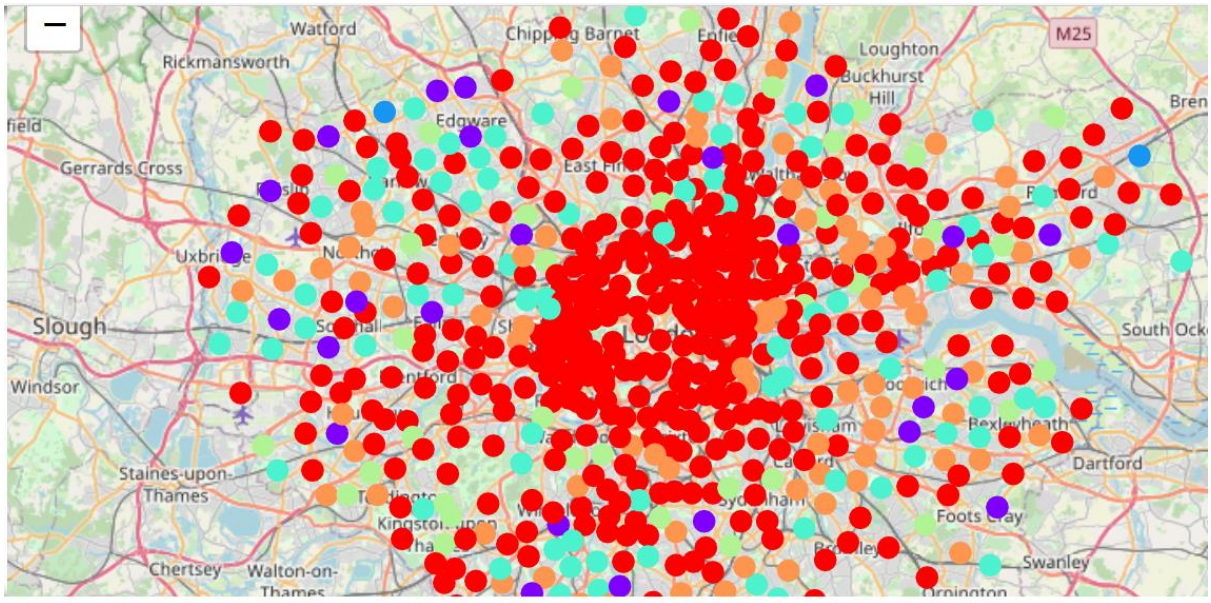
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbey	Grocery Store	Pub	Coffee Shop	Bus Stop	Supermarket	Bar	Pizza Place	Lebanese Restaurant	Discount Store	Steakhouse
1	Abbey Road	CafÃ©	Recording Studio	Historic Site	Metro Station	Gift Shop	Persian Restaurant	Garden	Grocery Store	Gym / Fitness Center	Boat or Ferry
2	Abbey Wood	Grocery Store	Playground	Zoo Exhibit	Field	Event Service	Event Space	Exhibit	Fabric Shop	Factory	Falafel Restaurant
3	Abingdon	Hotel	Pub	Italian Restaurant	CafÃ©	Pizza Place	Breakfast Spot	Persian Restaurant	Sushi Restaurant	Supermarket	Japanese Restaurant
4	Acton Central	Grocery Store	Hotel	Breakfast Spot	Train Station	Indian Restaurant	Park	Bed & Breakfast	Zoo Exhibit	Farm	Falafel Restaurant
5	Addiscombe	Breakfast Spot	Grocery Store	Pub	Fish & Chips Shop	CafÃ©	Filipino Restaurant	Event Space	Exhibit	Fabric Shop	Factory
						Italian		Chinese		Thai	

A dataset with the **top 10 most common venues in the ward** was created.

4.3. Clustering the Wards/ Neighborhoods in London

Using the above dataset and the combinations of venues in each ward a total of 6 clusters was created.

4.3.1. London Wards distribution according to venue Clustering



Map 2. Representation of all London Wards as Venue Clusters.

4.4. Description of the Cluster Segments

The crosstabulation between Cluster and Venues Ranking from **1st Most Common Venue** to **5th Most Common Venue** allows us to describe the 6 segments. Segment 2 is rather a registration errors as in a populated area which I know only 1 location appears to be registerd.

Cluster 0 which includes 408 wards, cluster 3 which includes 82 wards and cluster 5 including 84 wards are have relativelly distinctive profiles.

Before describing each cluster, it is interesting to observe Pubs represent the most common venue in 103 (16%) of wards, mainly in clusters 0 and 4. Cluster 2 will be ignored as it includes only 2 cases.

- **Cluster 0** can be labeled as **Gathering Places** as they include mainly pubs (18%), cafe-s (11%) and coffee shops (10%), hotels and indian restaurants.
    - Such wards are massivelly located in Central London in City of London and Westminster in which people come mainly for their work rather than living there.
    - Southwark, Hackney, Camden are strongly gentrified area in which younger people live and work or live closer to the corporate areas where they work. Hackney and Camden are strong touristic attractions also.
    - While Kensington and Chelsea is a more "posh" part of London having museums, restaurants and expensive residential area, Tower Hamlet is rather a place where people live and work.
- **Cluster 1** can be labelled as **Affluent Living Areas** cluster as it includes parks, yoga studios as 1st to 4th ranked venues. They are located at about 30-40 minutes from central London and they represent quiet areas where people mainly live.
- **Cluster 3** can be labelled as **Affordable Living Areas** cluster. This includes building sites probably for making new flats or office areas. These wards are in proximity with cluster 1.
    - These are in Croydon, Harrow, Hillingdon, Brent or Bexley.
- **Cluster 4** can be labelled as **Have a Pint on the Way to Home** as they include mainly pubs (60% as 1st venue). This are in zones 3-4 where people rather live than work and they presumably drink a pint of beer and meet some friends before going home.
    - These are spread in small numbers in most boroughs suggesting people gather in some wards before going home in the evening.
- **Cluster 5** can be labelled as **Shopping areas** as they include mainly grocery stores and supermarkets.

Preliminary observations

- As an observation borough as Westminster, City of London, Camden and Hackney have more than 90% of their wards in cluster 0 (see Bar Chart).
- Considering the density of such venues in these borough they can be top of the list for placing a new Wine Bar.
- Most people work in these boroughs and most tourists travel around these areas.
- However, such a decision can be affected by the huge competition from other alternative places.

1st Most Common Venue	C0	C1	C2	C3	C4	C5	All
All	408	28	2	82	45	84	649
Pub	71	1	0	1	27	0	100
Grocery Store	12	0	0	0	0	45	57
CafÃ©	51	0	0	0	0	0	51
Coffee Shop	38	0	0	0	0	1	39
Indian Restaurant	18	1	0	4	0	5	28
Hotel	26	0	0	1	0	0	27
Park	3	17	0	6	0	1	27
Construction & Landscaping	1	0	0	23	0	0	24
Convenience Store	15	0	0	4	2	3	24
Gym / Fitness Center	10	0	0	4	1	3	18

Table 1. 1st Most Common Venue by cluster

2nd Most Common Venue	C0	C1	C2	C3	C4	C5	All
All	408	28	2	82	45	84	649
Pub	58	0	0	1	11	7	77
Coffee Shop	38	0	0	0	1	2	41
CafÃ©	34	0	0	1	0	4	39
Grocery Store	17	2	0	2	1	15	37
Park	9	7	0	10	1	2	29
Zoo Exhibit	7	6	2	4	5	1	25
Indian Restaurant	18	1	0	3	0	2	24
Italian Restaurant	18	0	0	0	0	1	19
Supermarket	12	0	0	1	1	4	18
Hotel	16	0	0	0	0	2	18

Table 2. 2nd Most Common Venue by cluster

3rd Most Common Venue	C0	C1	C2	C3	C4	C5	All
All	408	28	2	82	45	84	649
Pub	31	3	0	4	6	7	51
Coffee Shop	36	1	0	1	2	3	43
Park	8	4	0	21	2	5	40
CafÃ©	29	0	0	3	2	1	35
Zoo Exhibit	14	8	0	1	3	3	29
Grocery Store	14	0	0	0	0	8	22
Hotel	13	1	0	1	1	2	18
Italian Restaurant	16	0	0	0	1	1	18
Indian Restaurant	11	0	0	1	0	5	17
Pizza Place	9	0	0	3	2	1	15

Table 3. 3rd Most Common Venue by cluster

4th Most Common Venue	C0	C1	C2	C3	C4	C5	All
All	408	28	2	82	45	84	649
Zoo Exhibit	8	8	0	8	7	6	37
Coffee Shop	29	0	0	1	1	2	33
Park	11	0	0	14	3	3	31
CafÃ©	19	1	0	1	2	3	26
Grocery Store	19	0	0	1	1	5	26
Event Service	10	0	0	7	5	1	23
Gym / Fitness Center	16	0	0	0	0	3	19
Hotel	12	1	0	3	0	2	18
Pizza Place	15	0	0	0	0	3	18
Pub	11	0	0	1	1	3	16

Table 4. 4th Most Common Venue by cluster

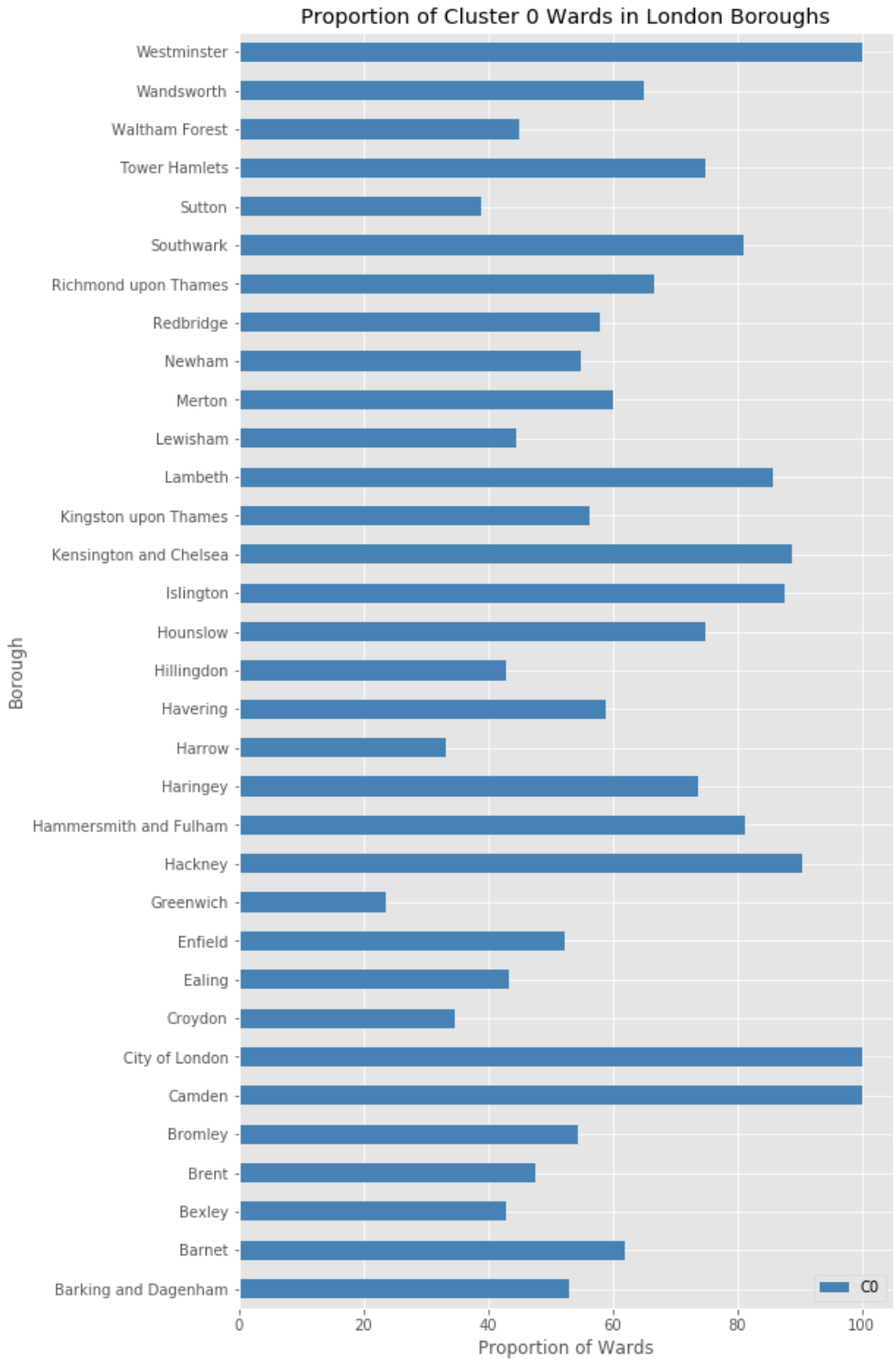


Figure 2. Percentage of **Gathering Places Wards** (Cluster 0) within London's Boroughs

4.5. Demographic description of London Boroughs

In the previous part of the report it was shown that London wards are segmented in 5 clusters out of which Cluster 0 (**Gathering Places**), Cluster 1 (**Affluent Living Areas**), Cluster 4 (**Have a Pint on the Way to Home**) and Cluster 5 (**Shopping areas**) are more interesting are people seem to gather there for recreational purposes. In this part of the report we will operate a selection of the most affluent wards in terms of income to decide where to recommend the opening of the **Luxury Wine Bars**.

Using the second dataset which includes the accessible demographic variable at ward level, they were grouped by Borough to understand if higher income is associated with a tendency to **conspicuous consumption** which is a required behaviour when attracting people to a luxury wine bar.

4.5.1. Aspects correlated with Household Income in London Wards

Several variables were compared between Boroughs in order to decice for which aspects people in different region of London behave differently. Only the variables listed below seem to differ strong enough between boroughs. Scatter plot charts were created to observe which other aspects are correlated with the Median income from London wards.

- In wards with higher median income the **proportion of house owners is higher** (Figure 3).
- In wards with higher median income the **proportion of private rents is higher** (Figure 8) while the **proportion of social rents is lower** (Figure 6).
- In wards with higher median income the **median house price is higher** (Figure 9) and also the **proportion of flats is increasing** (Figure 7).
- In wards with higher median income the **amount of properties on sale** is higher (Figure 5).
- Finally, it was observed that wards with higher income have smaller population size (Figure 4).

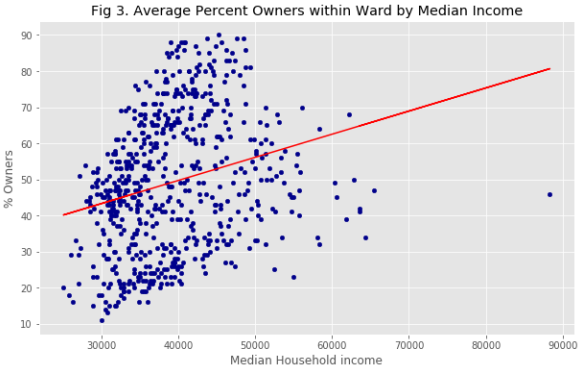


Fig 3. Average Percent Owners within Ward by Median Income

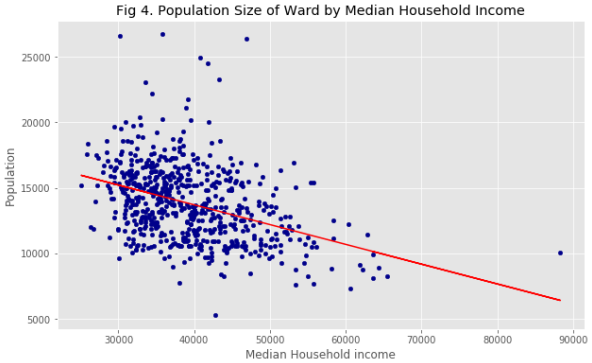
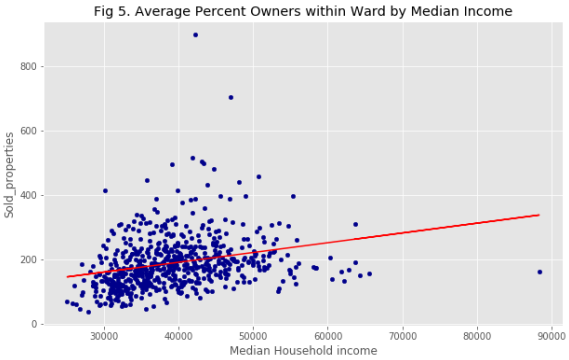
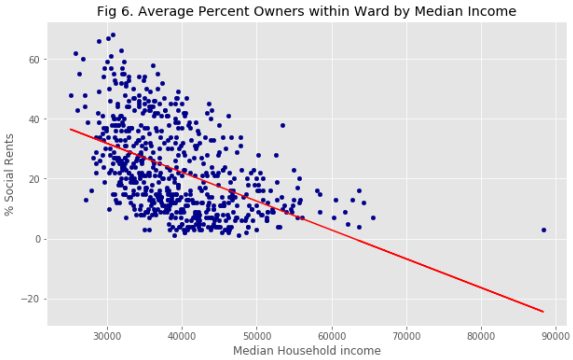
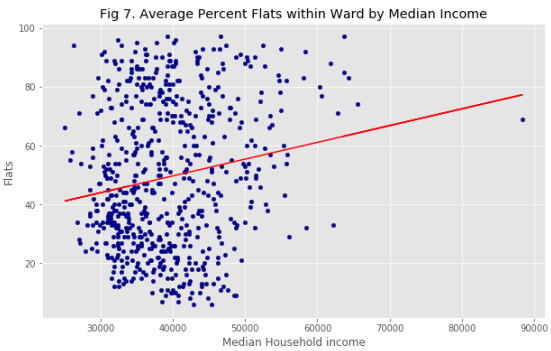
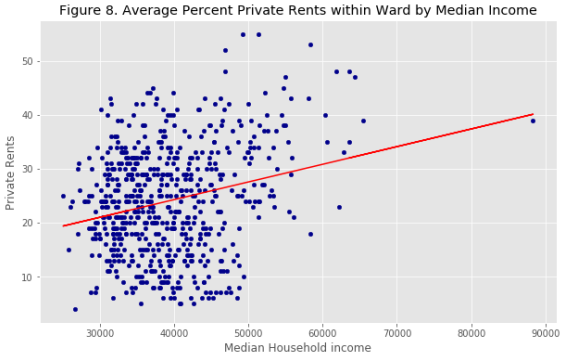


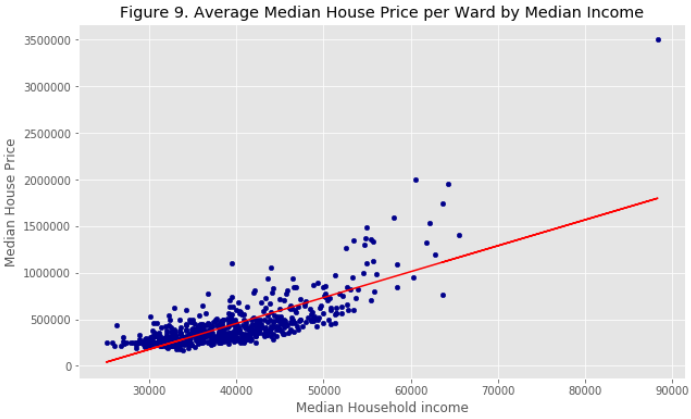
Fig 4. Population Size of Ward by Median Household Income











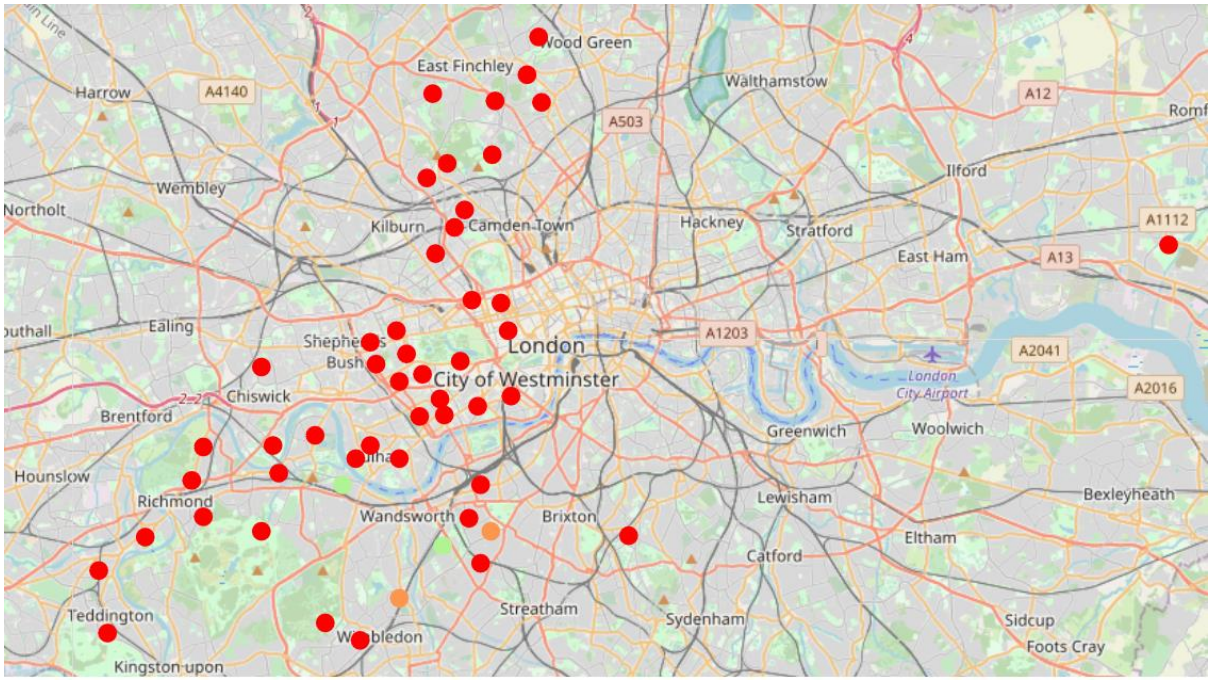
The visual inspection of Figures 3 to 9 revealed that wards with median income higher than 50.000£ seem to behave differently. Select the wards where the median income is above 50000 £ per year.

Consequently, a new dataset including only 55 wards with the highest median income was created.

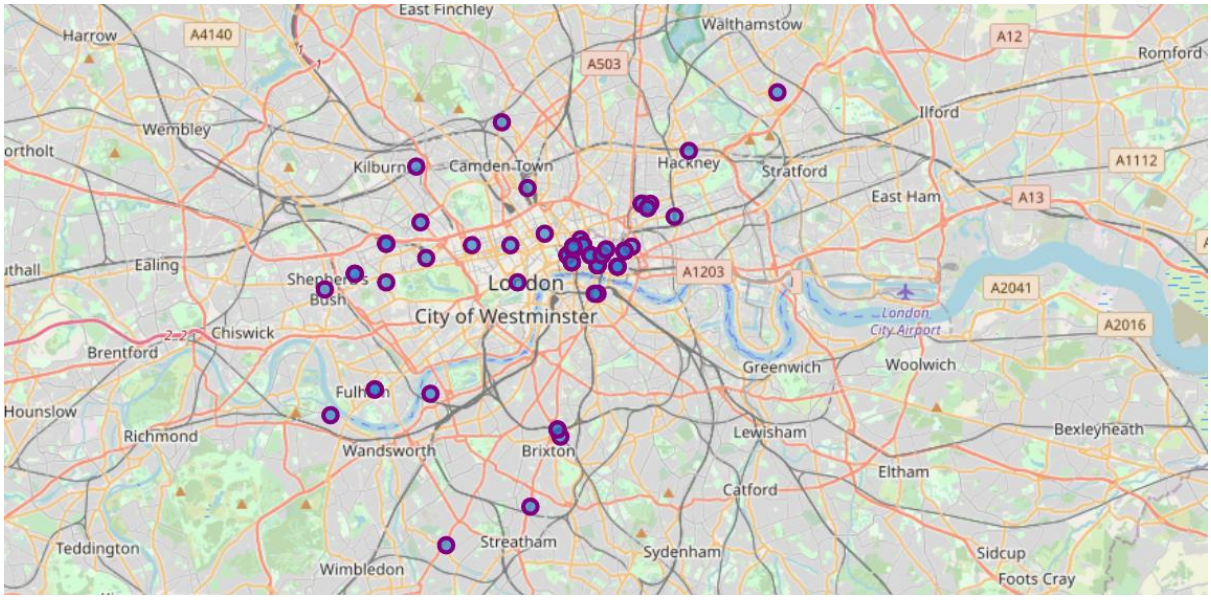
5. Results and Discussion

5. Identification of High-Income Wards as potential places of opening a Luxury Wine Bar

The names of these wards were used as a list to select these wards from the dataset including the Venue Clusters. This dataset included a list of 52 wards which will be mapped as clusters in Map 3 below.



Map 3. Representation of High-Income Wards as Venue Clusters.



Map 4. Representation of existing Wine Bars as provided by Foursquare API.

5.2. Overall discussion of the results

Once the 52 list of wards was selected, we created another map (**Map 4**) of the already existing Wine bars. Most of them are, interestingly, concentrated in City of London and otherwise scattered in different wards. Many of the wards we identified as potential location for luxury wine bars do not seem to have such a dedicated venue although they definitively have restaurants or other luxury places. The association table between Boroughs and clusters within the High-income dataset suggests the Boroughs in which the investment in some new luxury wine bars are Kensington and Chelsea, Richmond upon Thames, Wandsworth, Westminster, Camden, Haringey and Hammersmith and Fulham.

6. Recommendations

- Corroborating maps 3 and 4 the 1st venue where is feasible to open a luxury wine bar would be **Richmond upon Thames** which has 9 high income clusters but no dedicated wine bar.
- Although **Kensington and Chelsea** and **Westminster** have few wine pubs, the affluence of the area and the abundance of tourists allows the opening of another 1 or 2 wine bars. Wandsworth in the South part of London and Haringey in North are other two affluent areas which can support the opening of new luxury wine bars. Wards such as Alexandra (near Alexandra's Palace) or Hampstead Town in Haringey/Camden, West End in Westminster seem to be recommended for such investment as they attract not only tourists but they have affluent Londoners who might be targeted by such a new venue. Opening a wine bar in more isolated wards such as Teddington, Twickenham Riverside, Kew or Parsons Green and Walham would require to rely more on local people but seasonally on tourists also who might come in the area for tennis or rugby events.