Rough Sleeping in Inner London:

A Cluster Analysis using the Foursquare.com data

Banu Simmons-Sueer

Identifying the Research Problem:

- Rough sleeping or homelessness is a growing social problem in central London.
- Our focus of analysis here: why certain London boroughs are where homeless people congregate most?
- The Westminster borough has by far the highest number of rough sleeping people in London, followed by Camden.
- Westminster is also one of the most touristic boroughs with many theatres, hotels, cafes and restaurants.
- Westminster and Camden boroughs are the assumed stakeholders of our study.
- In local policy-making to reduce the number of rough sleepers one has to gain insights to:
  - The socio-economic characteristics of these boroughs
  - Why rough sleepers are attracted to them in the first place (nearby venues & amenities)

Research strategy and the hypotheses to be tested:

- Our research attempts provide information to the local policy makers using the following methods:
  - Descriptive analysis of the London inner borough socio-economic data.
  - Spatial analysis of Westminster and Camden via the unsupervised machine learning method K-Means Cluster analysis using the Foursquare.com data.
- The hypotheses to be tested:
  - Socio-economic characteristics such as income or unemployment in these boroughs are unlikely be explanatory factors.
  - Is there any relation between rough-sleeper rates and criminality rates?
  - Spatial characteristics of these boroughs:
    - 'Shelter effect' of certain venues (theatres with big illuminated entrances, underground stations)
    - Hotels that guarantee all day circulation of affluent tourists.
    - Parks & gardens with bathroom facilities and providing privacy in summer.
    - Coffee shops/cafes (especially the chain ones) that facilitate the donation of excess fresh daily produced food.

Data 1:

Table 3: Inner London boroughs by number of rough sleepers, 2018				
Local Authority	**2017**	**2018**	**Change on 2017**	**% change on 2017**
Westminster	217	306	89	41%
Camden	127	141	14	11%
City of London	36	67	31	86%
Lambeth	34	50	16	47%
Southwark	44	47	3	7%
Islington	27	43	16	59%
Wandsworth	13	25	12	92%
Hackney	18	23	5	28%
Kensington and Chelsea	20	20	0	0%
Hammersmith and Fulham	5	12	7	140%
Tower Hamlets	21	10	-11	-52%
Greenwich	8	7	- 1	13%
Lewisham	22	5	-17	-77%
Total (Inner London)	**592**	**756**	**164**	**28%**

Official statistics of the UK Ministry of Housing, Communities and Local Government, we use the rough sleeper statistics for 2018 as shown above

Data 2:

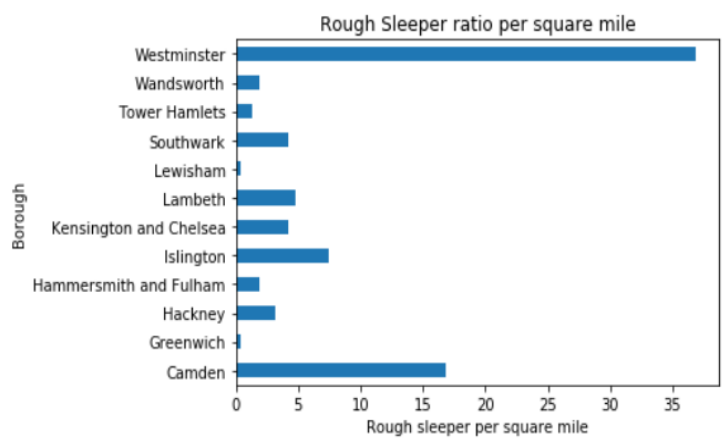
- The rest of the data sources are as follows:
- For socio-economic characteristics of inner London boroughs:
  - https://data.london.gov.uk/dataset/london-borough-profiles/london-borough-profiles.csv
- For borough locations, post codes, areas we scraped data via Beautifulsoup library from the following sources:
  - https://en.wikipedia.org/wiki/List_of_London_boroughs
  - https://en.wikipedia.org/wiki/List_of_areas_of_London
- For exploration of the venues in Camden and Westminster:
  - API queries via https://foursquare.com/

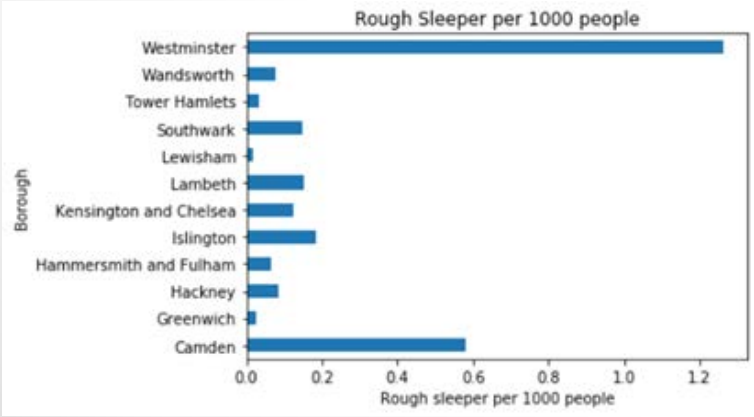
Methodology:

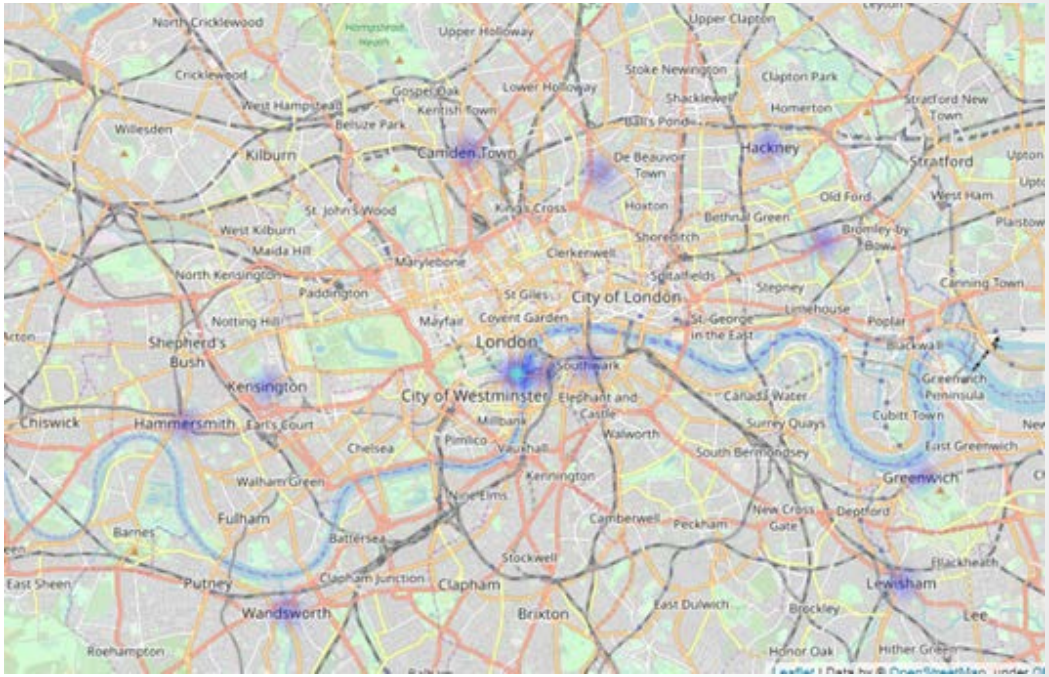
- Preliminary analysis of socio-economic characteristics of inner London boroughs and rough sleeper ratios.
- Analysis of the foursquare.com venue categories for the locations of Camden and Westminster that consists of:
  - API queries of the venues via Foursquare.com
  - One-hot encoding of the venue categories to make the data binary.
  - Search for optimal number of (k) clusters using Elbow, average silhouette and the gap-statistic
  - K-means cluster analysis with K= 6.
  - The resultant clusters are superimposed on the London map

Results:

- Preliminary Analysis:

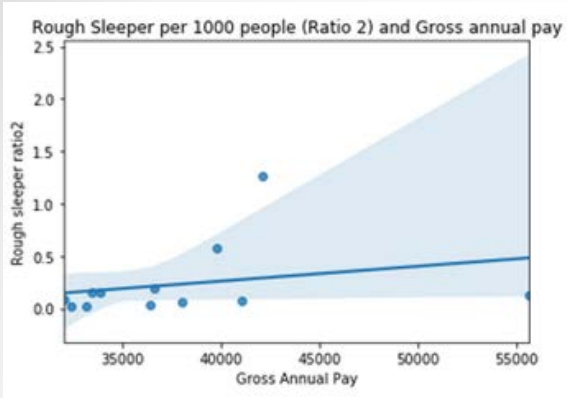


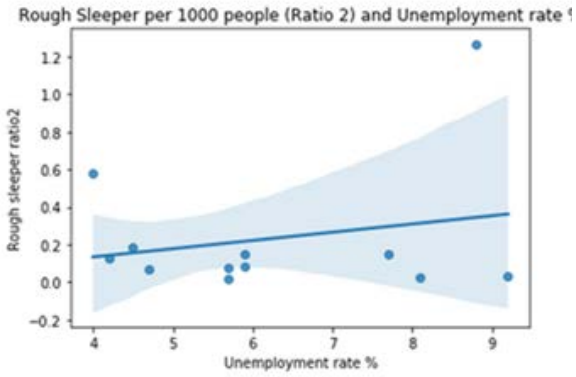


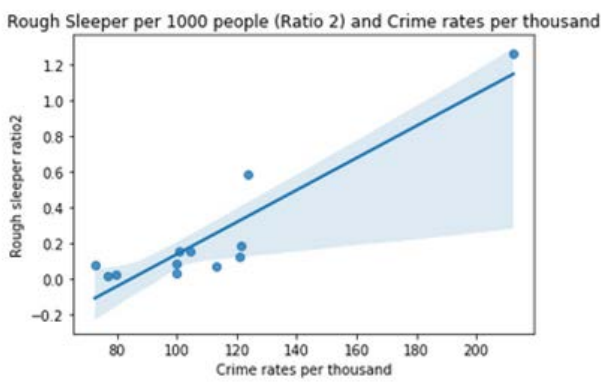


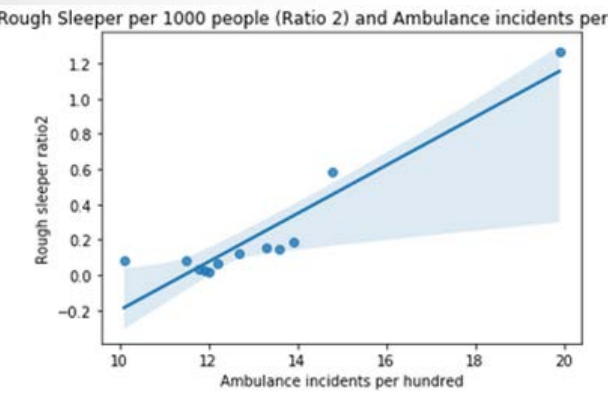
Irrespective of the rough sleeper ratio (per square mile or per 1000 people, Westminster is the worst affected borough as shown in the heat map

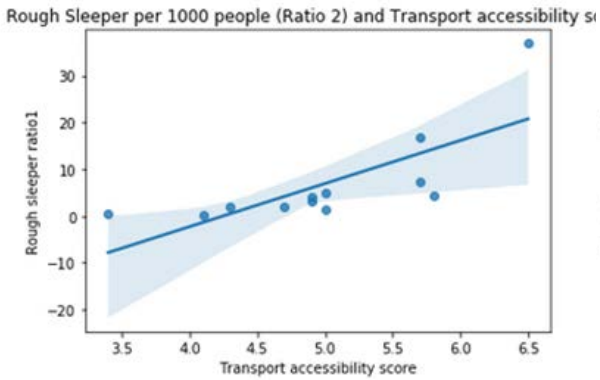
Descriptive Analysis:

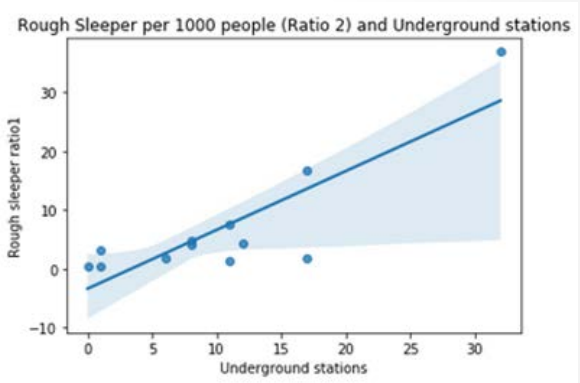








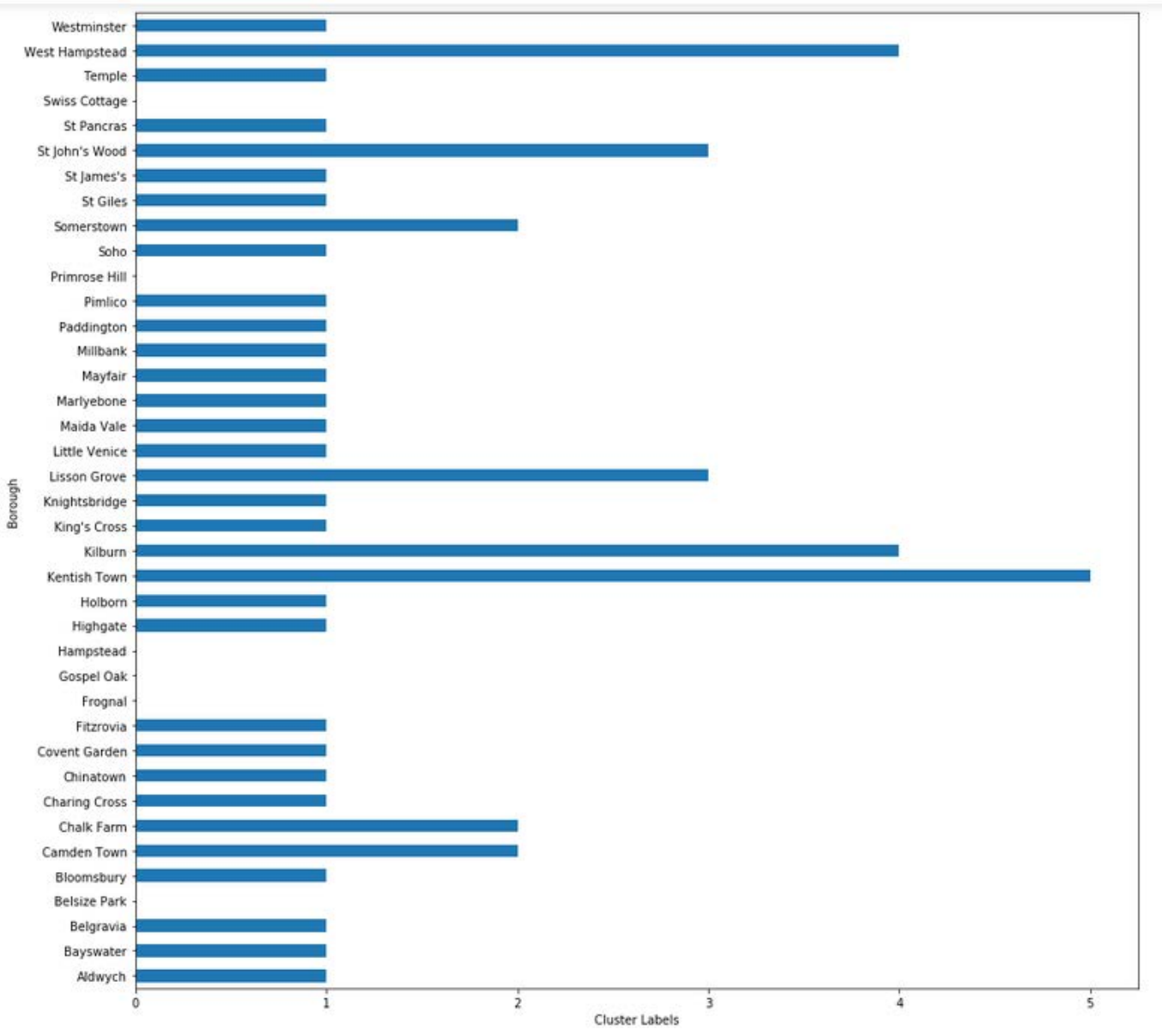




Variables:	Correlation with rough sleeper per 1000:
Unemployment rate %	0.222480
Gross Annual Pay	0.259476
Transport accessibility score	0.714759
Underground stations	0.848301
Crime rates per thousand	0.916834
Ambulance incidents per hundred	0.940492

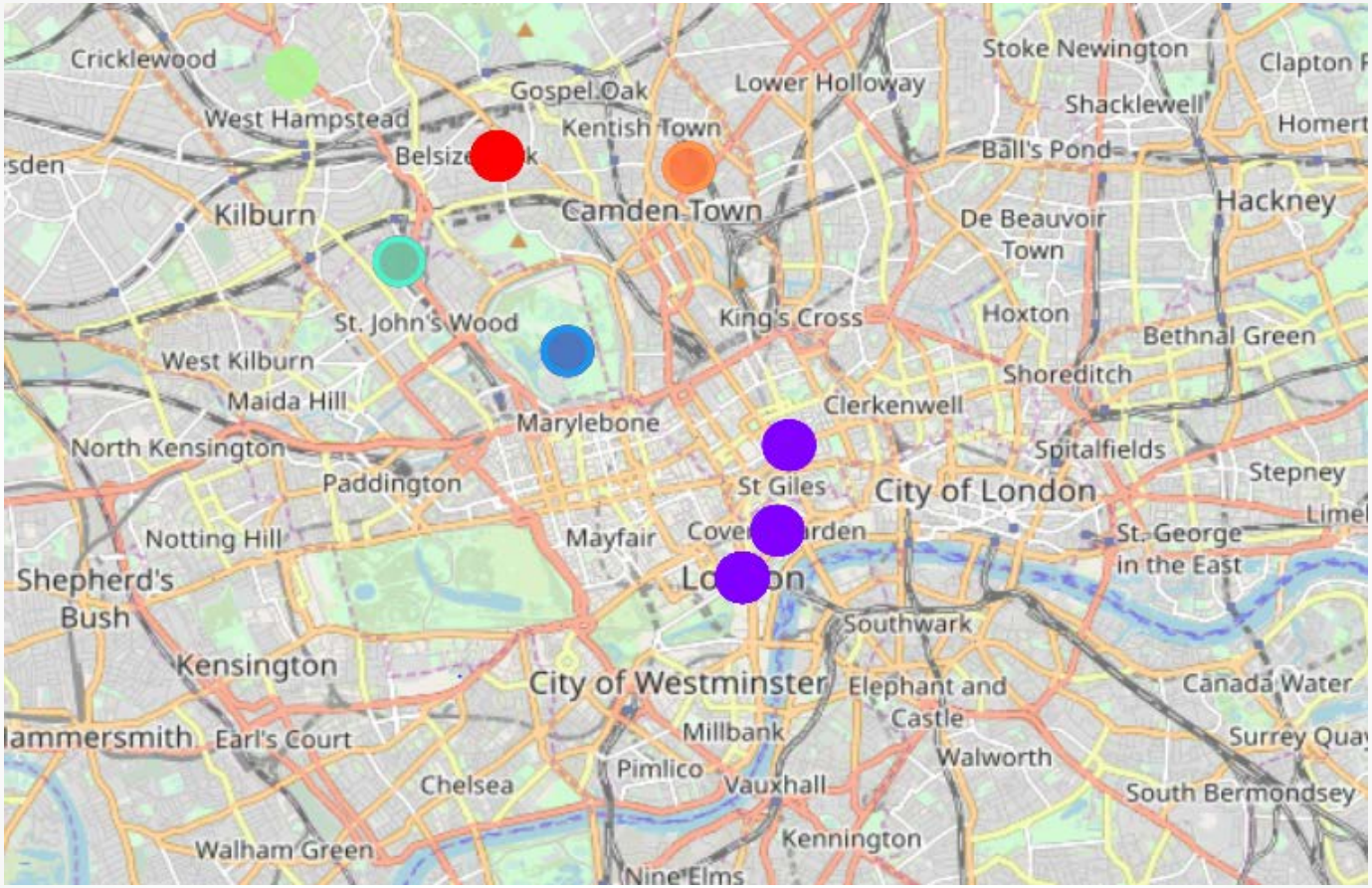
Low correlation of rough sleep rates with socio- economic variables (unintuitive sign in the case of Gross Annual Pay) High correlation with underground stations, criminality and ambulance incidents.

K-Means Cluster Analysis: Locations by cluster labels



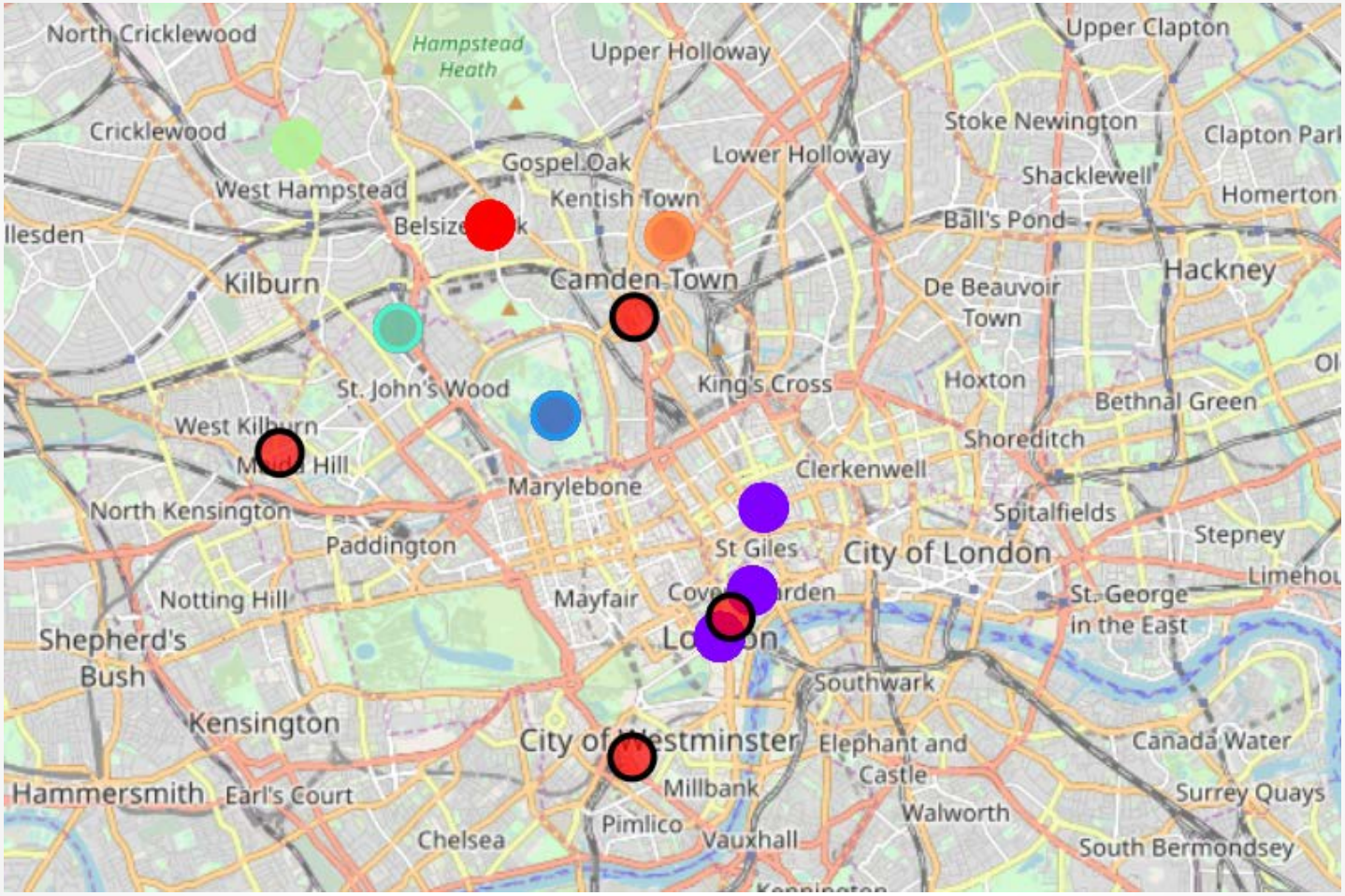
We chose K=6, on the basis of various methods and the plausibility of the resultant clustering. The chart on the left shows the cluster labels (0 to 5) of various Camden and Westminster locations on the basis of the venues they contain. There are 106 unique categories.

Clusters k=6



Red: Cluster 0 (Cafes) Purple: Cluster 1 (Theaters and Hotels)
Blue: Cluster 2 (Parks & Gardens) Turquoise: Cluster 3 (Café/Yoga studio) Light green: Cluster 4 (Gyms/Fitness centre) Orange: Cluster 5 (Skate park, various stores)

Homeless Centres and Cluster Labels



Black circles with red filling indicate the main Homeless Help Centres in Westminster and Camden. Not surprisingly, they are close to Cluster 1 and Cluster 2, where rough sleeper density is particularly high.

Discussion:

- The purpose of this analysis was to gain some insights to the rough-sleeping problem in Westminster and Camden
- The rough sleep statistics are only available at borough level (not at location level)
- Nevertheless, based on anecdotal evidence (personal and journalism-based) we hypothesized that rough sleepers tend to concentrate in the vicinity of theatres, hotels, parks & gardens and cafes.
- Our cluster analysis indeed backs up these hypotheses by identifying these as predominant clusters in the area.
- The optimal cluster was chosen as K=6, but the sensitivity of the results to a more coarse clustering (K=3) was also analyzed.
- The K-means procedure identified , Parks & Gardens, Theatres & Hotels as distinct clusters even when K was set as 3.

Conclusions:

- Our spatial analysis of Westminster and Camden via K-means clustering using the Foursquares.com data revealed that these boroughs contain attractive venues for the rough-sleepers.
- The socio-economic characteristics of these boroughs do not really explain the high ratio of rough sleepers.
- When the cluster labels and the homeless help centres are superimposed on the London map, it is clear that these centres are especially close to Cluster 2 (Parks & Gardens) and Cluster 1 (Theatres and Hotels).
- It was reported in newspapers such as the Guardian that soup kitchens in Westminster act like a magnet for not just vagrants but also drug dealers. Our analysis indeed confirms high correlation between rough sleeping and criminality rates.

Conclusions cont.

- Our analysis concludes that at local council level, there are not many policy alternatives as 'not helping' is not a humanitarian option.
- As the attractiveness of the venues cannot be changed, the only option to reduce the number of rough sleepers in these particular boroughs is to move the homeless help centres elsewhere.
- At national level, according to a report by BBC on Feb 26th, the UK government pledged to allocate 236 million GBP to help rough sleepers (https://www.bbc.com/news/uk-politics-51653744). Clearly, a national policy is a more effective way of tackling the problem.