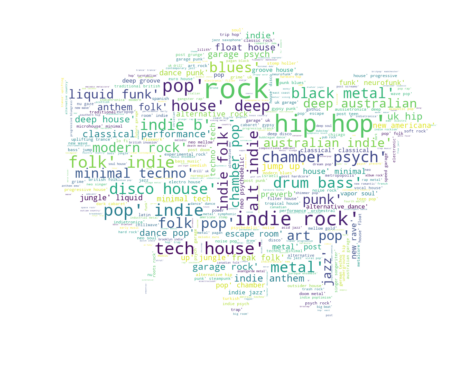
****Fuinki City

雰囲気

Spatial Data Capture, Storage and Analysis

**Msc. Smart Cities and Urban Analytics**

**Word Count: 5000 XXXXX**

**Year 2018**

**Ilja Panic**

**Soma Suzuki**

**Alfred Long**

**Jonathan Cohen**

# Introduction

Cities have been one the most important inventions of humanity. There is no doubt that since humanity started to settle and increase in density, life was affected in the most unexpected ways. Throughout all human history, cities have been called the attention, not only of researchers, but also citizens and policy makers. The sense of belonging and the everydayness to a specific location, produces a high level of interest and engagement.

In recent years, with the increase in computing power and storing capacity, and the introduction of several disruptive technologies, discussions about Cities, and specially Smart Cities seem to be dominating the agenda of researchers, governors and citizens.

Through urban computing and analytics, a lot of effort is being allocated to produce a Digital City, were most of its components can be sensed, analysed, predicted and thus induced.

In this context, UK has identified that its live music scenery generates a distinctive urban personality, making cities more attractive. As a result in recent years, there has been a trend towards attempting to quantify the economic impact of the music industry in a city (Titan Music Group, 2015; Boston Consulting Group, 2017). A recent census in the UK has focused solely on the live music industry, showing among other facts that these events enhanced social capital, mood and contribute to identity formation (Webster et al, 2017).

In this context, the Greater London Authority is undertaking a “Rescue Plan for London’s Grassroots Music Venues” to ensure that London’s vibrant personality regarding its musical environment is protected and promoted (Greater London Authority, 2017). This report looked into the cultural and societal services provided by “grassroots venues” and found that “gig listings for grassroots music venues are patchy. Buying tickets often requires knowledge of the music scene that many tourists don’t have” (Greater London Authority, 2017: 9). With this report and the creation of the role of Night Czar, political leadership in London clearly deem the cultural and societal role of live music to be important.

Another avenue into the study of the societal and cultural value of music in a city has been qualitative ethnographic studies (Cohen, 2012a; Cohen, 2012b). These studies have focused on Liverpool and the history of music and live music performance in the city. An attempt was made to create a GIS map of historic existing, and no longer existing music venues in the city. The map was created to invite reflection on “why live music might thrive in a particular urban area” (2012b:  591) amongst other questions. One of the difficulties come up against in this research was “given the absence of archival documents it was difficult to trace the history or even the precise location of such venues” (2012b: 593). The research did “prompt reflection on the scope and distribution of live music venues and their embedding in the dynamics of space and time” (2012b, 595). A key point brought out by this paper is that “music venues often provided a physical idiom for defining a particular social group and the relationships involved” (2012b: 595). Another writer on this subject, Malcom Miles, writes that events such as gigs provide the “means to articulate the implicit values of a city when its users occupy the place of determining what the city is” (Miles, 1997, p. 59).

Having identified the lack of a web platform, that provides the user a simple and useful interface of the musical events happening in their city (gigs), the present project will develop a user friendly, but useful interactive, tool which will allow the users to navigate through the gigs and its characteristics. Approaching this objective in a systematic way, generates the perfect opportunity to study these events. This project is a milestone in this research domain and therefore will contribute in generating a detail dataset of the locations of the gigs and analysing its spatial distribution.

In the following section a brief literature review of the importance and how other researches have intended to study the subject, will be found. Section 2 will clearly state the objectives of the present research. In the following section a Literature review will provide the necessary state of the art regarding the importance of cultural events and how to measure spatial clustering. Section 4 and 5, will describe the web development and the results of the clustering analysis. Finally, last two chapters will highlight the main findings, future steps and the shortcomings encountered as researching.

# Research Objectives

The project pursues two main objectives. As stated above, there is an opportunity to allow residents of and visitors to a city to explore and discover the music scene. This spatial display of events has not been widely available before. On one hand the project is focused on creating a simple, but direct visualization of London’s live music scene, easily navigable for the lay user.

On the other hand, manipulating the data collected allows a better understanding of the characteristics and spatial distribution of the different events happening in London (GLA). This study aims to be a milestone in the research of cultural events, not only by mapping but also by making a cluster analysis of the location of the gigs.

In sum, by providing a systematic approach towards the data collection, visualization and analysis, this research pretends to create a baseline towards a better understanding of the importance of the live events.

# Literature Review

The cluster analysis of the gig distribution by music genre will use a similar method to Dennett and Page’s (2017) paper. Similarly to that paper we will use the density-based cluster analysis method DBSCAN (Ester et al, 1996). It will be implemented using the sklearn.cluster.DBSCAN package in Python (Pedregosa et al, 2010). DBSCAN is useful as cluster analysis for theses purposes as the number of clusters does not need to be specified beforehand. DBSCAN does take parameters of EPS, the maximum distance between two points before they are assumed to be in the same cluster, and minimum samples, the minimum amount of points to be in a group before it is labelled as a cluster. How these are calibrated will be covered in the methodology.

The correctness of the clusters, a rating of their density and tightness will be found using the sklearn.metrics silhouette score function (Rousseeuw, 1987). This will be used for calibration of the DBSCAN EPS parameter. As in Dennett and Page’s (2017) paper, strong and significant clustering of events in certain locations will indicate that further geographic or location specific factors must have lead to this clustering.

# Web Development and Potential Users

# Purpose and Potential Users

# Technical Details

# Site Preview

# Analysis

# Data Set

The main source of information for the project is SongKick. As described above the site information was accessed using an API. For a month, on daily basis, data from this site was stored in a database. Using the venue and artist information, the database was enriched by retrieving information about the venues from foursquare, and Last.fm and Spotify for the artists.

Any musical event, which is specified in SongKick will be defined as a gig.

From SongKick the following set of variables were be accesed:

In the Appendix, it will be found a comple list and description of all the variables retrived from Last.fm and Spotify.

1. Location

2. Name of the artists including

3. Name of the venue

As a result a data set of XXXXX gigs and information regarding the character of the music played was created. The next section will describe the main findings of exploring this dataset that corresponds between Apr 13th and May 26th.

# London’s gigs

# A total XXXX of musical events were retrived for London and as expected, there

# Methodology

In order to detect the existence of Spatial Correlation among the location of the gigs during this period of time, two methods are applied: (i) Anselin’s Moran’s I and (ii) DB Scan for cluster detection.

# Moran’s I

characteristics and spatial

This spatial statistic was calculated at a ward level, therefore the gigs dataset were assigned to each polygon. Once the process was finished, different weight matrices were calculated for all the GLA. As seen before, after a radius of 10 km, there were almost no gigs; thus the area of study was reduced (see Appendix). With the new set of wars, the Moran’s I was re estimated, for different weight matrixes and the results showed a significant, but weak spatial relationship.

**asdsa**

# Clustering

Subsets of the gig database were created by searching for a genre term, for example “reggae”, and returning any event which had an artist playing described in either Spotify or LastFm as a reggae artist. The events were compiled into dataframes for each of the genres.

The minimum sample parameter will be set at 5, assuming that any smaller sample size is not significant enough to be considered a cluster of gigs.

It will be necessary to assess the strength of the clusters overall at different EPS DBSCAN parameters. This will be done before plotting them spatially. Different EPS distances will be set, every 10 meters from 1m to 2000m. 2000m is chosen as at this point the neighbourhood communities searched for would become too large and instead start to represent the centres of London as a City. [DO I NEED TO CITE?]

The silhouette coefficient and number of clusters for DBSCAN results will be measured for each genre, for each of the different EPS distances. The silhouette coefficient plot will be used to assess which EPS distance to set for the DBSCAN clustering to create the cluster maps. It will also show which of the genres show strong clustering at different EPS distances. The plot of the number of clusters will show how many “communities” there are of each genre.

After the DBSCAN EPS parameter has been decided the clusters will be plotted on a map and compared against the Moran’s I analysis, this will show what kinds of communities or clusters are picked up by either technique.

If strong clustering relationships occur that are different for different genres, this study allows a starting point for investigating how these dynamic musical communities relate to the other communities sharing the city.

# Results

As is visible in the plot below, the silhouette scores of dbscans for clustering of different genres, Reggae is not deemed to be strongly clustered. Techno clusters silhouette scores rise as the EPS rises from 500m to 1000m as a group of clubs around Elephant and Castle, and another around Angel become clusters. Most genres reach an early peak at an EPS of 600m. Almost all of the genres either level off or dip after this point with House, Techno and Reggae as exceptions. House and Techno continue to rise to a peak at an EPS of 1000m, and Reggae only levels off at an EPS value of 1500. All the other genres rise again by the EPS of 1500m, at this point the clusters are growing so large in scale that it becomes less meaningful when working on a city scale. The minimum samples parameter was set to 5 gigs.

|  |
| --- |
|  |
|  |
|  |

This relationship is also shown by the graph of EPS against the number of clusters by genre, at an EPS of 1500m the largest number of clusters for any single genre included is 7, for Punk and Techno. With EPS values this large it stops becoming meaningful to look for clusters that represent communities or neighbourhoods.

|  |
| --- |
|  |
|  |
|  |

This relationship is also shown by the graph of EPS against the number of clusters by genre, at an EPS of 1500m the largest number of clusters for any single genre included is 7, for Punk and Techno. With EPS values this large it stops becoming meaningful to look for clusters that represent communities or neighbourhoods.

By plotting the clusters as well as the Moran’s I high-high wards, it becomes possible to see the spatial trends of gigs of particular genres. These can then be viewed as musical communities forming/being formed.

See the plots below:

|  |
| --- |
| classical |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059091485_classical-cluster.png |
|  |

Here the gigs are all very centrally located. The DBSCAN clustering picks out clusters in the West End, many gigs at a single venue on the Southbank, a few around Kings Cross, the City of London and in Camden. These are well correlated with the Moran’s I highlighting of wards. Due to the relatively low number of gigs, Moran’s I found high-high relationships where the DBSCAN minimum sample size of 5 did not produce clusters. This shows that classical events are happening mainly in central London, though non-clustered gigs are distributed far further both east and west, though not many in South London.

|  |
| --- |
|  |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059375619_folk-cluster.png |
|  |

Folk gigs are shown to be very prevalent with many spread over many different areas of london. Lots of distinct clusters form in North London, especially to the east. Interestingly Moran’s I does not pick out the large cluster around Mayfair, this is due to the large gap of any gigs between this cluster and further west.

|  |
| --- |
|  |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059578848_house-cluster.png |
|  |

The clustering of House gigs shows more in areas that had not been populated by Folk or Classical gigs, Southwark and Brixton. Again Moran’s I does not pick out the strong clusters further out. This suggests that these areas must either have a very interested local audience, or pull performers and audience members from their surrounding wards [THIS IS BULLSHIT].

Interestingly Moran’s I highlights wards far out from the centre of London without many gigs in them [WHAT DOES THIS MEAN?]

|  |
| --- |
|  |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059788195_jazz-cluster.png |
|  |

Jazz gigs are highly clustered, with distinct clusters far from one another. Moran’s I also represents this disparate clustering, though highlighting some other areas.

|  |
| --- |
|  |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059903036_punk-cluster.png |
|  |

Punk gigs form many distinct clusters in North London, especially in the east.

|  |
| --- |
|  |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527060643328_reggae-cluster.png |
|  |

Reggae did not have enough events close together to show any particular strong clustering, as shown by the EPS analysis earlier, there is one distinct cluster in Camden but overall it is remarkably evenly distributed. Moran’s I picks out Camden, but also Shepherd’s Bush and Holloway and Archway.

|  |
| --- |
|  |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527060786710_soul-cluster.png |
|  |

Soul shows a very large amount of gigs with the majority of clusters in north and central London. Moran’s I highlights wards that correlate very well with the larger gig clusters.

|  |
| --- |
|  |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527060934082_techno-cluster.png |
|  |

Techno gigs show a few well defined clusters in Camden, Kings Cross, and around Elephant and Castle. Fabric fittingly provides a cluster from the gigs that have happened there alone.

# Conclusions

# Future Steps and Limitations

# Bibliography

# Appendix

|  |
| --- |
| Form Last.fm the following set of variables can be accesed: |
| 1. Genre - folksonomy |
|  |

|  |
| --- |
|  |
| From Spotify the following set of variables can be accesed:  1. Genre  2. Popularity  3. Followers  4. Danceability  5. Energy  6. Key  7. Loudness  8. Mode  9. Speechiness  10. Acousticness  11. Instrumentalness  12. Liveness  13. Valence  14. Tempo  15. Duration\_ms |
|  |

|  |
| --- |
|  |
| From foursquare the following set of variables were accesed:  1. Venue rating  2. Venue tips - visitor reviews |
|  |

|  |
| --- |
|  |
|  |
|  |