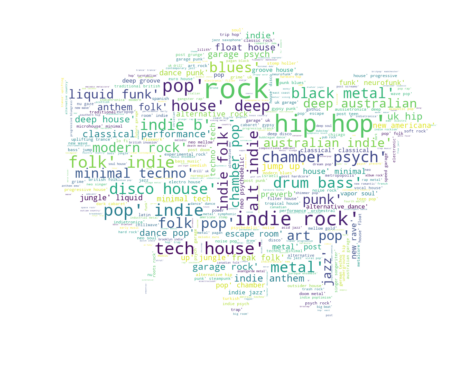
****Fuinki City

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

Through urban computing and analytics, a lot of effort is being directed towards producing a Digital City, were most of its components can be sensed, analysed, predicted and thus induced (Batty et al, 2012). For cultural and societal information this is a more nuanced and complex process.

In this context, UK has identified that its live music scenery generates a distinctive urban personality, expressing social and cultural values and making cities more attractive, and profitable. In recent years, there has been a trend towards attempting to quantify the economic impact of the music industry in a city or region (Homan, 2008; Makkonen, 2014; Titan Music Group, 2015; Boston Consulting Group, 2017; UK Music, 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, the political leadership in London clearly deems 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 their characteristics. Similar GIS mapping techniques have been used to allow users to find out about the geographic element of music (Homan, 2008; Cohen, 2012b), but they have all been static datasets, not taking into account the dynamic nature of live music. 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 these 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 Python’s scikitlearn: 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 led to this clustering.

Moreover, from Regional and Economics theory, there is evidence that supports and explains why some business and economic activities tend to cluster themselves (Krugman, 1991). As he explains in his pioneering work, as more businesses of the same kind start to agglomerate, they start to enjoy a set of benefits and spill overs that generate positive economics of scale. In particular these are known as Agglomeration Economics. Although, at this point there is no evidence to support that musical events are benefiting from clustering, to understand these drivers is a complex research question out of the scope of this study.

Having said this, the first step towards researching in this direction, is to detect the existence of any spatial distribution.

Spatial autocorrelation might be detected using the Local Indicator of Spatial Interaction (LISA) Moran’s I detailed in Aneslin (1995). The statistic will be estimated using GeoDa® software. Among the different vast amount of work done trying to identify spatial patterns, this research project will adapt and follow the methodologies used to explain how the space could affect the risk of obesity (Huang et al, 2015). This is because the clarity and the fact that restaurants, and venues hosting gigs, might share some similarities. In this research it is shown how Moran’s I can be comparable to spatial scan and how the strength of spatial relation loses power as other variables as taken into account.

The work done in Indonesia to explore the spatio-temporal relationship of the manufacturing industries provides useful insights on how the temporal component could be relevant in identifying the strength of the relationship (Rothenberg et al, 2017).

Finally, Guo et al. (2013) present a novel methodology to establish the spatial relationship between point and polygons. The methodology presented in this case could have the potential to expand the research presented in this paper.

# Web Development

……………………………………

……………………………………….

# Analysis

## The data set

As described above, the main source of information of the realization of this project was SongsKick, which was accessed using an API. From this site we were able to retrieve information about the venues and the bands that were playing in that week. Although, there was information about the gig, it was complemented by using Spotifys, Last.fm and FourSquare APIs. Consequently, the data from SongKick, could be significantly enriched to satisfy our previously stated objectives. In the Appendix a complete description of the final dataset can be found.

By developing and deploying a set of routines to be executed every day for a month, we were able to build a rich dataset, of the gigs in London and other major cities in the world, such as Tokyo, Buenos Aires, Berlin or Liverpool. As SongKick, Spotify, FourSquare and Last.fm, are services working globally, these dataset could be retrieved for a different set of cities. This information and all the processes here explained and used form London could be deployed for any other geographic region. Of course the quality of the information will vary from country to country, or even among cities, the methodologies developed for these objectives are systematic and scalable.

## London’s music environment

Between the 13th of April and the 26th of May, a total of 1863 gigs were recorded and as expected, as the weekday approaches to the weekend, the average amount of gigs increases. On average, Fridays host the most events (100), followed by Saturdays with 96. On average, Sunday were less active than the Thursdays by almost 10 gigs. The following graph shows these results.

|  |
| --- |
| Figure : Distribution of Gigs by days |
|  |
| Source: https://fuinki.netlify.com/about |

Regarding the different characteristics that the music played by these bands exhibit, it was possible to start analysing the ‘vibe’ of the city and its venues. Using the previously described variables retrieved from Spotify, Last.fm and SongKick, the following spider graph was created. The lines in the graph represent the standardized values of the average values of the different characteristics of the songs played during these bands. Although, the understanding of the units in the graph might be difficult, the main idea is to compare among the different days of the week. For instance, it’s seem that the bands playing on Fridays have the most followers and play the most Live music. Saturdays, are the most energetic days and Sundays seems the day to listen to more lyrics and dance. The most popular bands are playing during Tuesdays. Finally, by looking at the distributions of these variables, it is possible to identify high dispersion and extreme values. As a result, the median was chosen as the statistic to study.

|  |
| --- |
| Figure : Distribution of Gigs by days |
|  |
| Source: https://fuinki.netlify.com/about |

Before, moving to a spatial analysis it was also necessary to study how frequently the different genre were labelled. As the spatial analysis will be divided into different types of music, it was important to study the distribution of the different music types. The most frequent labelled genre is Pop music and it was followed by Rock with almost 600 appearances.

Finally, to begin our spatial analysis of how the gigs are spatially distributed, in the about page there will be a public map, produced WITH Folium technology will show the point densities of the gigs. As this is a visual representation of the density of the gigs, we also included the following bar chart to show the amount of gigs per Boroughs.

|  |
| --- |
| Figure : Distribution of Gigs by days |
|  |
| Source: https://fuinki.netlify.com/about |

Regarding For more information about descriptive statistics about the distribution of the gigs during the week, type of music or the spatial distribution, visit <https://fuinki.netlify.com/about>.

# 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) LISA Moran’s I and (ii) DB Scan for cluster detection.

## LISA: Moran’s I

In order to work with this LISA statistic, the gigs location was joined to the map of the wards, then as a result that after a radius of 10 km from the City of London, there were almost no records, the city was sub-seated. Once we had an selection of the entire set, different ardency matrix were created and tested. As we were trying to capture the local spatial correlation, we decided to choose 15 Nearest neighbours. This ensured that all the adjacent wards were taken into account. The Moran’s I was executed for all the gigs.

We found out that Moran’s I index was 0.14, indicating that the spatial interaction for the amount of gigs at a ward level, was not strong as expected. Although, it was not estimated, we can assume that there is a high correlation between the following maps and the ones of population, density, income or employment.

|  |  |
| --- | --- |
| Figure 4: Moran’s I – 15 Nearest Neighbours | |
| *Moran’s I INDEX* | *Cluster Map* |
|  |  |
| *Significance Map* | *Local G Cluster Map* |
|  |  |
| *Source: Own production* | |

## 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.

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

|  |
| --- |
| Figure 5: Distribution of Gigs by days |
|  |
| *Source: Own production* |

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.

|  |
| --- |
| Figure 6: Silhouette Metrics at different Eps Values |
|  |
| Source: https://fuinki.netlify.com/about |

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.

|  |
| --- |
| Figure 7: Number of cluster at different EPS Values |
|  |
| Source: https://fuinki.netlify.com/about |

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.

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.

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.

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

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.

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.

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.

|  |
| --- |
| Figure 8:These results are mapped in the following table of images. Figure 8: Cluster analysis by genre |
| Classical Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059091485_classical-cluster.png |
| Folk Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059375619_folk-cluster.png |
| House Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059578848_house-cluster.png |

|  |
| --- |
| Jazz Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059788195_jazz-cluster.png |
| Punk Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527059903036_punk-cluster.png |
| Reggae Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527060643328_reggae-cluster.png |
| Soul Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527060786710_soul-cluster.png |
| Techno Cluster |
| https://d2mxuefqeaa7sj.cloudfront.net/s_177EA3C7AA9AD0FBCB783ED748A1C4C186C1AF3B720331E87665D8BF754708E1_1527060934082_techno-cluster.png |
| Source: https://fuinki.netlify.com/about |

# Conclusions

Each of the genre types that have been picked out show a distinctive clustering profile that would easily allow further interrogation either through qualitative or quantitative means.

The collection and structuring of data has already addressed the shortcomings in data pointed out in Cohen’s paper (2012b). This provides a way of analysing Miles’ (1997) claim that live music is one avenue for the assertion of a city’s implicit values. These musical communities can now be compared and contrasted with any other form of community sharing the same urban space.

The app has provided one route to overcoming the difficulty in getting to know a music scene by providing a dynamic immersive experience in which a user can navigate using the feel of the music rather than having to know specific artists or venues. This directly addresses the point made in the “Rescue Plan for London’s Grassroots Music Venues” report (Greater London Authority, 2017) about the impenetrability of the local music scene.

By providing a database of gigs in a way that makes quantitative analysis possible it would be very easy to create metrics to improve the musical city algorithm proposed by Baker (2017), this would also be of interest to the initiators of the UK Live Music Census and the institutions behind the economic analyses of the music industry in cities and regions.

# Limitations

The first input source used is the SongKick website. This means that all the data we have collected is initially restricted to those events registered on the SongKick platform. The SongKick data, once collected, was used to call the APIs of Spotify, Lastfm, and Foursquare. Genre information was taken from both Spotify and LastFm, this means that the data was in fact subset again, to those events with artists for which Spotify or LastFm had matching records.

As with any study based on internet sources, what is being analysed is collected events, what is actually happening in the city may be quite different. Many of the most informal events would not be posted on SongKick, perhaps not on any online platform. This may mean our analysis is affected by the digital divide, it is restricted to the kind of music produced and played by people with enough time, education, and money to exploit these kinds of digital platforms.

While analysing the results of Moran’s I and DB Scan, the need to understand the factors influencing the location of the gigs became clear. In other words, this research can be taken as an exploratory exercise of the agglomeration of live music events. The following questions to be answered in the research agenda go in to different directions by understanding the:

1. Effects (positive and negative- direct and indirect) of these activities over the city, to individuals and society.
2. Consequences causes behind the agglomeration of these activities and introducing more sophisticated methodologies that will take into account other demographic, temporal and built environment variables.

# Future steps

Historic data could provide an opportunity to generate spatio-temporal analysis of the musical events in London and any other city. Retrieving this information could be vital to understanding how the dynamics of gentrification are affecting London's night environment.

Moreover, the effects of the night tube, over the spatial distribution of the venues could be studied in greater detail if historical data was provided.

Finally, the tool and the analysis generated for this specific project can potentially be expanded to other cities. Doing this will contribute to gain comprehension of how the music environment is in other cities. This will allow to make a comparison between different urban areas.

# Bibliography

# Bibliography