Capstone Project - A 'New-style' Classical Music Festival

Applied Data Science Capstone by Coursera/IBM

Table of contents

1. Introduction
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion

1. Introduction

The benefits of listening to classical music are endless. It can improve your mood, lower stress levels, increase productivity, spark creativity, and studies have shown it can even boost your brain power by improving memory. However, classical music has a diversity problem. It is far too white and far too inaccessible to a lot of people. More needs to be done to take classical music out of the traditional concert hall and brought to a much wider range of communities, to make it more accessible to all, and to break down the barriers and stereotypes surrounding the genre.

Through combining demographic data collected by Greater London Authority in 2016, and Foursquare location data, this project aims to highlight potential areas and venues in London that could host a 'New-style' Classical Music Festival. It should appeal to any philanthropist wanting to increase the outreach of classical music.

2. Data

- Kaggle data set on London Boroughs from study conducted by GLA in 2016 which can be found here.
- Geopy library to get the latitude and longitude values of London Boroughs
- Foursquare API to get location and venue data for potential concert venues

2.1 Cleaning the Data

London Boroughs Kaggle data set

After loading the London Boroughs csv file into a pandas df, I could see that there were many unnecessary features that were not relevant to the project such as election data, happiness and anxiety scores, health data etc. I therefore sliced the df by selecting only the columns that could be of interest. Not only in this project, but also that could be relevant for future analysis when approaching this problem. The df then looked as follows:

	Area name	Inner/ Outer London	GLA Population Estimate 2016	Average Age, 2016	Proportion of population of working-age, 2016	Proportion of population aged 65 and over, 2016	% of population from BAME groups (2016)
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	City of London	Inner London	8,548	42.9	90.6	9.4	27.5
2	Barking and Dagenham	Outer London	205,773	32.9	86.1	13.9	49.5
3	Barnet	Outer London	385,108	37.2	83.3	16.7	38.7
4	Bexley	Outer London	243,303	38.9	89.0	11.0	21.4

The first row was all missing values as this must have been a blank row in the csv file. I also saw that some missing values were stored as '.'. I therefore wrote code to convert any string 'NaN' and '.' To numpy NaN values for easy manipulation. I then needed to see where all the NaN values were to know how to treat them so created the following df with all of the entries containing NaN:

	Area name	Inner/ Outer London	GLA Population Estimate 2016	Average Age, 2016	Proportion of population of working-age, 2016	Proportion of population aged 65 and over, 2016	% of population from BAME groups (2016)
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
34	Inner London	NaN	3,494,269	34.5	91.0	9.0	43.1
35	Outer London	NaN	5,256,484	36.8	86.8	13.2	42.1
36	London	NaN	8,750,753	35.9	88.5	11.5	42.5
37	England	NaN	54,316,618	39.6	32.2	17.7	NaN
38	United Kingdom	NaN	64,596,752	39.8	32.2	17.6	NaN
39	National comparator	NaN	54,316,618	39.8	32.2	17.7	NaN

As all of these entries were not individual London Boroughs, it was safe to drop all entries with NaN values from the df, so I wrote code to do that.

The next problem with the data was that the columns with numbers had been uploaded as type 'object' so these needed to be converted to float and string types for data manipulation.

Geopy Library

I knew that it would be important to the project for mapping the boroughs that the data frame was appended with the latitude and longitude coordinates for each London Borough. I therefore wrote code to loop through each of the area names in the df and use the geopy library to search for their corodinates. I appended the results to the df to create this:

	Inner/ Outer London	Population	Average Age	% working-age	% 65+	% BAME	latitude	longitude
Area name								
City of London	Inner London	8548	42.9	90.6	9.4	27.5	51.515618	-0.091998
Barking and Dagenham	Outer London	205773	32.9	86.1	13.9	49.5	51.554117	0.150504
Barnet	Outer London	385108	37.2	83.3	16.7	38.7	51.653090	-0.200226
Bexley	Outer London	243303	38.9	89.0	11.0	21.4	51.441679	0.150488
Brent	Outer London	328568	35.5	82.5	17.5	64.9	51.563826	-0.275760

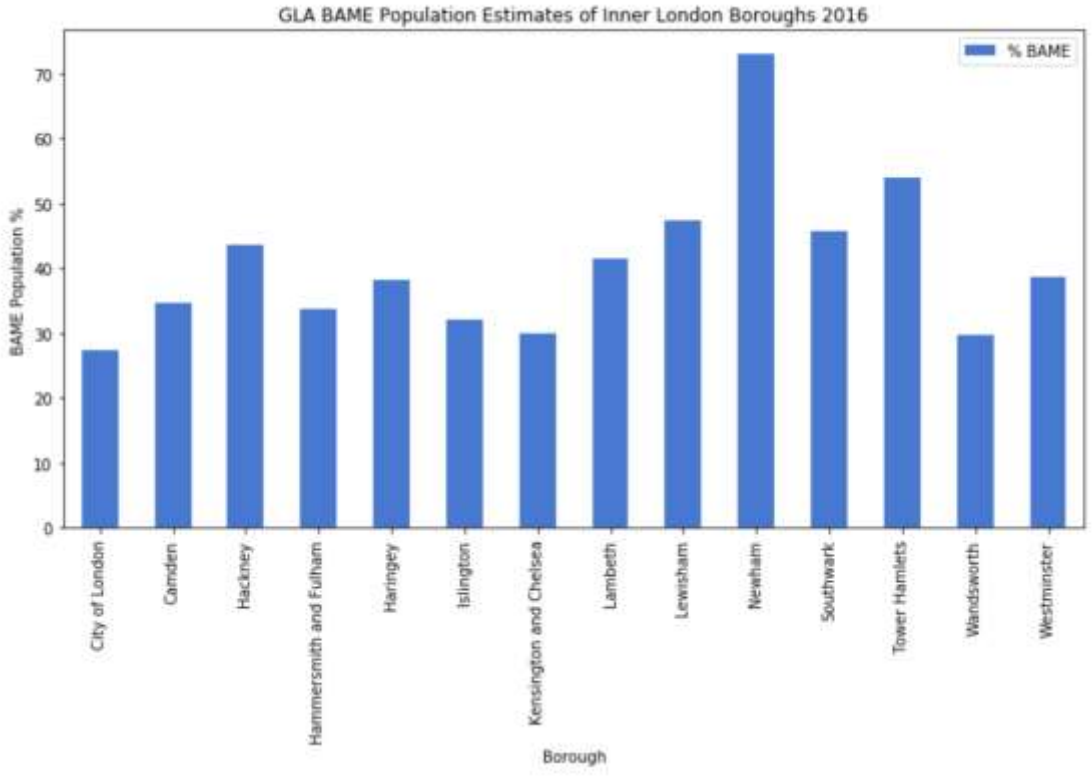
3. Methodology

In this project, I chose to focus on boroughs that are within ‘Inner London’ and have a high BAME population percentage. I chose to focus on Inner London only as it is more accessible via public transport and will therefore likely increase attendance figures, and I chose to focus on areas with a high BAME population to address the diversity problem that classical music has.

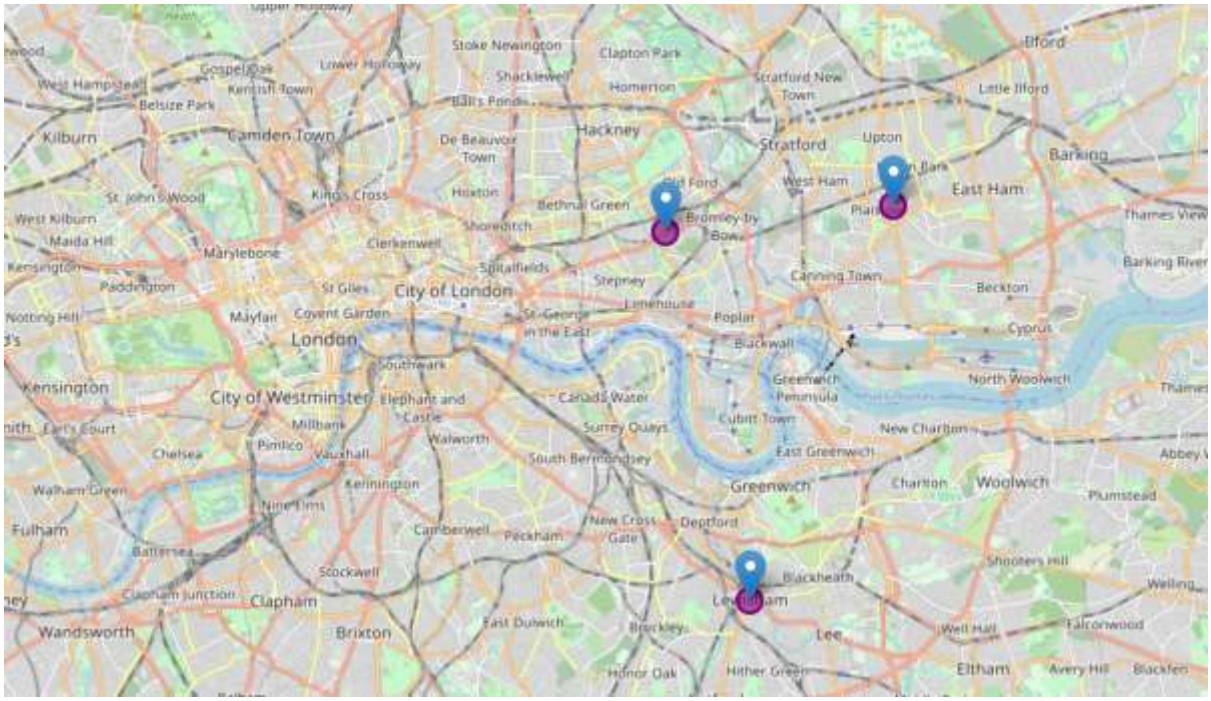
Within these boroughs, I searched for venues that fall under the category ‘event space/community center/church’ and are described as community centres/halls, to highlight potential performance venues that wouldn't necessarily be usually considered as classical music performance spaces. Thereby, potentially reaching new audience members and breaking down barriers and stereotypes of where classical music can be performed.

First, I did some data visualisation to explore the dataset and highlight the areas of interest.

Here is a bar chart displaying the BAME population percentage estimates from the GLA 2016 survey, for the Inner London Boroughs:



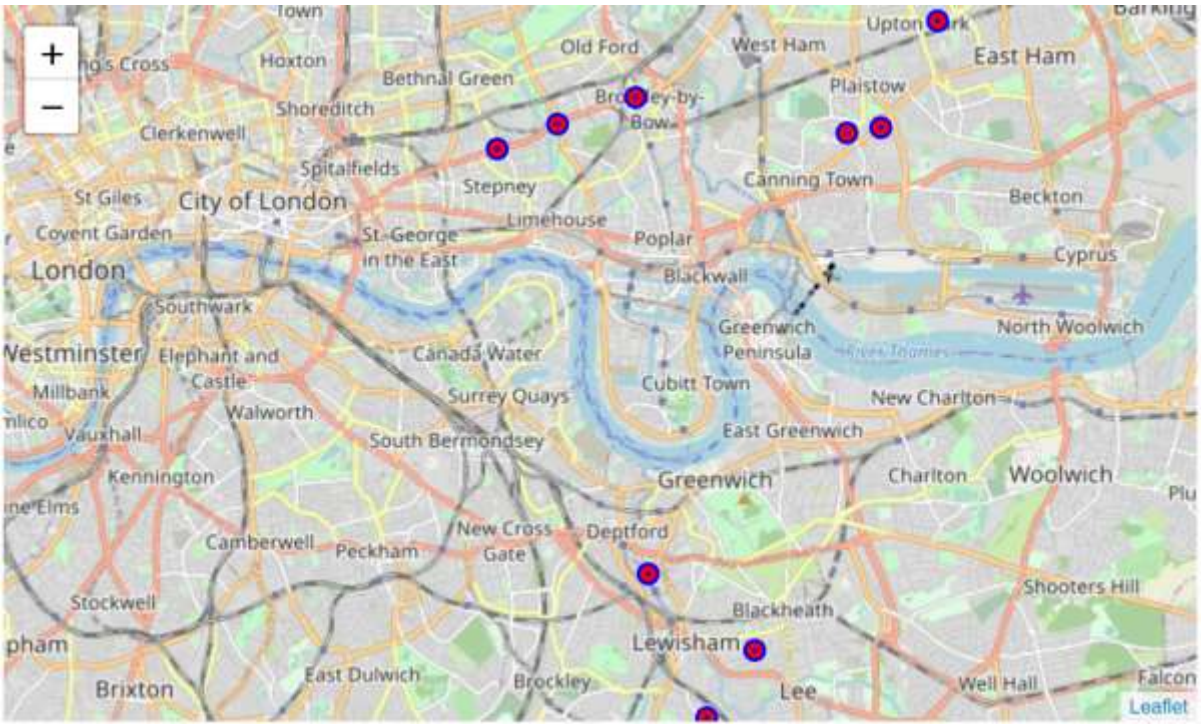
It was clear therefore that Newham, Tower Hamlets and Lewisham are the top 3 boroughs with the highest BAME population, so I chose these as the focus areas. Here they are plotted on the map using Folium:



Taking each Borough in turn, I used the Foursquare API to search for nearby potential concert venues and compiled the top 3 from each borough into the following data frame:

	name	categories	address	lat	lng
0	Olga turl memorial hall	Event Space	Barking Rd, London Borough of Newham, London E13	51.524294	0.027506
1	St Cedd's Hall	Community Center	Foster Road	51.523727	0.020876
2	Full Gospel Hall	Church	NaN	51.537255	0.038296
3	Bromley Public Hall	City Hall	Bow road	51.527971	-0.020063
4	LSE Rosebery Hall	None	90 roseberry avenue	51.521775	-0.046649
5	All Hallows church	Church	Bow	51.524831	-0.034932
6	Kingswood Halls	Event Space	Kingswood Pl., Dacre Pk.	51.461579	0.003035
7	Mehfil Hall	Event Space	NaN	51.470608	-0.017528
8	St Swithuns Hall	Event Space	St Swithuns Rd	51.453275	-0.006207

I then plotted the venues onto this map:



4. Results

This project has highlighted the boroughs in Inner London with the highest BAME population and has resulted in a list of 9 possible venues in these areas combined which could host a 'New-Style' Classical music festival.

The London Boroughs were Newham, Tower Hamlets and Lewisham.

The list of venues was as follows:

- Olga turl memorial hall, Newham
- St Cedd's Hall, Newham
- Full Gospel Hall, Newham
- Bromley Public Hall, Tower Hamlets
- LSE Rosebery Hall, Tower Hamlets
- All Hallows church, Tower Hamlets
- Kingswood Halls, Lewisham
- Mehfil Hall, Lewisham
- St Swithuns Hall, Lewisham

5. Discussion

In this project I analysed the demographic data from the Kaggle data set and made use of the Foursquare API to get location data of potential venues within the areas of focus. However, there is lots of information needed when making a decision on a concert venue that is not provided by foursquare. For example, how large the venue is and how many people it could hold. This could be determined by further research into each venues website for example.

6. Conclusion

The aim of this project was to suggest possible venues for a classical music festival which could help increase outreach of the genre, make it more accessible and break down barriers. It is difficult to know from the list alone which of these venues would be suitable for a classical music concert, so further research would definitely need to be done, but it certainly provides a useful starting point. Further research could include site visits as the next step. A successful concert venue would be well situated in amongst lots of other types of venues. It would also be good to do further analysis into the surrounding areas of each venue to see what other kind of venues are in the near vicinity, such as bars/pubs and restaurants.