Battle of the Neighborhoods

Exploring Art & Entertainment Neighborhoods in Atlanta, Georgia

Introduction

Atlanta is the capital and most populous city of the U.S. state of Georgia. According to the 2019 estimation, the city has 506,811 inhabitants, making the city the 37th most populous city in the United States. The capital serves as the Atlanta metropolitan area cultural and economic hub, home to more than 6 million people and the nation's 9th metropolitan area [1]. The Atlanta History Center, the Martin Luther King Jr. National Historic Site dedicated to the African-American leader's life and times, the Downtown Centennial Olympic Park comprises the massive Georgia Aquarium has been a glamorous lure for visitors.

The city's historical and cultural assets play a crucial role in its economic development. The U.S. Bureau of Economic Analysis reported that all arts and cultural production accounts contributed over $22 billion in Georgia, or 4.1 % of the total gross state product, which is more than the construction, education, agriculture, and mining sectors contribute. This economic contribution supports more than 140,000 jobs, and the Georgia Council for the Arts reports that the nonprofit arts and culture sector represents over 3,000 organizations, over 31,000 jobs, and revenue of more than $1.3 billion with an overall economic impact of $2.2 billion [2].

In order to enhance this contribution of the art and entertainment sector, the Georgians for the Arts, a 501c4 established in 2019, with a mission to provide vision, leadership, and resources that ensure the growth, prosperity, and sustainability of arts and culture in Georgia [3]. A study by Americans for the Arts (2017) on the arts and economic prosperity in the Metro Atlanta area found that communities supporting the arts and culture are investing in an industry that supports jobs and generates government revenue.

However, with limited public funds available and a decrease in donors since the recession, the sector finds it difficult to grow their investment funds and increase donations. Considering arts as an investment that delivers both communities' wellbeing and economic vitality, one of the policymakers' challenges is balancing arts funding with the demand to support jobs and grow their economy.

We hope that this geospatial analysis will be a valuable resource for stakeholders for promoting the state as a destination for general arts and entertainment events location for film, music, and digital entertainment projects and planning and mobilizing state resources for economic development. Therefore, we will use our data science powers to generate Atlanta based arts and entertainment neighborhoods using machine learning tools

Data Source

The data frame of neighborhoods in Atlanta, Georgia, is built using REST API from City of Atlanta - Planning GIS Open Data (https://opendata.atlantaregional.com/datasets/297d3d69d8ab4c6ba5f9264ad5e75c0a_3?geometry=-91.170%2C32.119%2C-77.701%2C35.316). The JSON file has comprised 244 neighborhoods in 25 Neighborhood Planning Unit (NPU) of Atlanta, Georgia including geographical coordinates of the neighborhoods.

From snapshot form the JSON file shows the most relevant variables we need for the analysis are found under *feature* key.

Get Data from opendata API

```
url = 'https://opendata.arcgis.com/datasets/297d3d69d8ab4c6ba5f9264ad5e75c0a_3.geojson'
results = requests.get(url).json()
results
```


```
{'type': 'FeatureCollection',
 'name': 'Neighborhoods',
 'crs': {'type': 'name',
  'properties': {'name': 'urn:ogc:def:crs:OGC:1.3:CRS84'}},
 'features': [{'type': 'Feature',
   'properties': {'OBJECTID': 148,
    'LOCALID': None,
    'NAME': 'Atlanta University Center',
    'GEOTYPE': 'Neighborhood',
    'FULLFIPS': None,
    'LEGALAREA': None,
    'EFFECTDATE': '1970/01/01 00:00:00+00',
    'ENDDATE': '1970/01/01 00:00:00+00',
    'SRCREF': None,
    'ACRES': 331.57,
    'SQMILES': 0.52,
    'OLDNAME': 'Atlanta University',
    'NPU': 'T',
    'CREATED_USER': None,
    'CREATED_DATE': '1970/01/01 00:00:00+00',
```


The next task is essentially transforming this data of nested Python dictionaries into a pandas DataFrame. The transformation results a DataFrame that has 244 neighborhoods located in 25 Neighborhood Planning Unit (NPU) of Atlanta with their geographical location. The first five neighborhoods are presented below.

```
neighborhoods.head()
```

	Neighborhood	Geotype	NPU	Latitude	Longitude
0	Atlanta University Center	Neighborhood	T	33.7546	-84.4028
1	Hunter Hills	Neighborhood	K	33.7636	-84.433
2	Historic Westin Heights/Bankhead	Neighborhood	K	33.7756	-84.4174
3	English Avenue	Neighborhood	L	33.7813	-84.4145
4	Adair Park	Neighborhood	V	33.7393	-84.4082

Our additional data source is Foursquare API. In order to find the available art and entertainment centers in Atlanta, we used Forsquare API to get the most common venues of given neighborhoods. In addition, the coordinate of Atlanta obtained using Google geocoding API. The essential aspects of the dataset include the number of existing art and entertainment centers in the neighborhood, the number and types of similar art and entertainment in each avenue of the nearby neighborhoods, which is

obtained from Foursquare API. This project is mainly focused on geospatial analysis of the art centers of Atlanta City to discover Atlanta Neighborhoods made for Art Lovers and stakeholder for the development of the sector.

Methodology

Our first important aspect of this project is to acquire an official dataset that comprises relevant variables for the analysis of at least all the neighborhoods and their geographical locations of the target areas of study. After all the data is acquired, we perform cleansing and transform the dataset into panda DataFram. The transform DataFram is used primarily to visualize and map the target location using python *folium* library and then to get all the list of Art & Entertainment venues across Atlanta city using Foursquare API.

The second essential aspect of this project is applying one of the most cluster methods of unsupervised learning. We use K-means clustering to partition the art and entertainment venues that we get from Foursquare API into K clusters in which each venue location belongs to the cluster with the nearest cluster centers.

Analysis

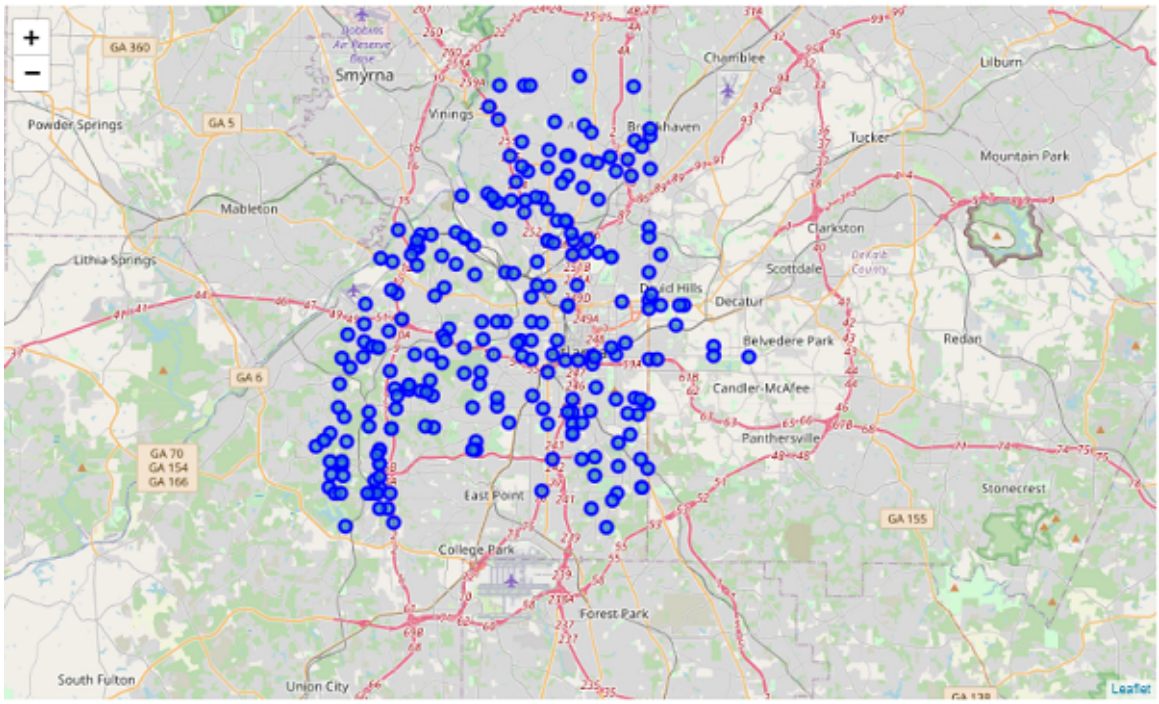
We use our first transformed DataFrame to visualize the geographic location of the 244 neighborhoods of Atlanta using Python **folium** library. We used the *geopy* library to get the latitude and longitude values of Atlanta. In order to define an instance of the *geocoder*, we need to define a *user_agent* and the latitude and longitude values of the neighborhoods used to get a map of Atlanta.

```
address = 'Atlanta, Georgia'

geolocator = Nominatim(user_agent="atl_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Atlanta, Georgia are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Atlanta, Georgia are 33.7489924, -84.3902644.
```

Now we create a map of Atlanta, Georgia with neighborhoods superimposed on top.



Foursquare venues

Next, we will start utilizing the Foursquare API to explore the art venues of the neighborhoods and segment them. The Foursquare API allows us to explore venues in neighborhoods of Atlanta that comprise venues of the art and entertainment, music, and Performing Arts. Forusquare API is set to 500 venues limit and radius of 5000 meters for each neighborhood from a given latitude and longitude. We specify a top-level of a category of Art & Entertainment to the categoryID in which all sub-categories will also match the query.

```
#https://developer.foursquare.com/docs/resources/categories
#Arts & Entertainment = 4d4b7104d754a06370d81259
art_venues = getNearbyVenues(names=neighborhoods['Neighborhood'],
                             latitudes=neighborhoods['Latitude'],
                             longitudes=neighborhoods['Longitude'],
                             radius=5000,
                             categoryIds= '4d4b7104d754a06370d81259')
```

```
art_venues.shape
```

```
(11434, 7)
```

```
art_venues.head(5)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Atlanta University Center	33.754649	-84.402799	Mercedes-Benz Stadium	33.755298	-84.400942	Football Stadium
1	Atlanta University Center	33.754649	-84.402799	Revery: VR Bar	33.774272	-84.371114	Bar
2	Atlanta University Center	33.754649	-84.402799	Federal Penitentiary	33.714119	-84.368650	History Museum
3	Atlanta University Center	33.754649	-84.402799	O4W TARDIS	33.752836	-84.374104	Street Art
4	Atlanta University Center	33.754649	-84.402799	John Lewis Mural	33.755666	-84.380313	Public Art

We extracted a total of 11,434 venue categories directly or indirectly related to art and entertainment for our 244 neighborhoods. Foursquare API provides us latitude and latitudinal values for both neighborhoods and venues and the venue name and their category. All these venues belong to 81 venue categories, and 73 of them are labeled as art and entertainment venue. All these figures from the Foursquare API are the basis for our subsequent clustering analysis.

Analyzing Each Neighborhood

To facilitate our clustering analysis using a k-means algorithm, we need to transform our categorical data from Foursquare API into numerical data using the One-Hot encoding technique. The technique is often useful to transform categorical features into vectors so that you can do vector operations on our clustering analysis.

Analyzing Each Neighborhood

In [25]:
```
# one hot encoding
atl_onehot = pd.get_dummies(art_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
atl_onehot['Neighborhood'] = art_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [atl_onehot.columns[-1]] + list(atl_onehot.columns[:-1])
atl_onehot = atl_onehot[fixed_columns]

atl_onehot.head()
```

Out[25]:

	Neighborhood	American Restaurant	Amphitheater	Aquarium	Arcade	Art Gallery	Art Museum	Arts & Entertainment	Bar	Baseball Stadium	Basketball Stadium	Betting Shop	Bookstore	Bowling Alley	Bre
0	Atlanta University Center	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Atlanta University Center	0	0	0	0	0	0	0	1	0	0	0	0	0	
2	Atlanta University Center	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Atlanta University Center	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Atlanta University Center	0	0	0	0	0	0	0	0	0	0	0	0	0	

In [26]:
```
atl_onehot.shape
```

Out[26]: (11434, 82)

The next step is to group rows by neighborhood and mean of each venue category's frequency of occurrence.

In [27]:
```
atl_art_grouped = atl_onehot.groupby('Neighborhood').mean().reset_index()
print(atl_art_grouped.shape)
atl_art_grouped.head()
```

```
(244, 82)
```

Out[27]:

	Neighborhood	American Restaurant	Amphitheater	Aquarium	Arcade	Art Gallery	Art Museum	Arts & Entertainment	Bar	Baseball Stadium	Basketball Stadium	Betting Shop	Bookstore	Bowling Alley
0	Adair Park	0.0	0.0	0.02	0.000000	0.060000	0.020000	0.0	0.02	0.0	0.02	0.0	0.0	0.00
1	Adams Park	0.0	0.0	0.00	0.020000	0.220000	0.020000	0.0	0.00	0.0	0.02	0.0	0.0	0.04
2	Adamsville	0.0	0.0	0.00	0.000000	0.044444	0.000000	0.0	0.00	0.0	0.00	0.0	0.0	0.00
3	Almond Park	0.0	0.0	0.00	0.022222	0.155556	0.022222	0.0	0.00	0.0	0.00	0.0	0.0	0.00
4	Amal Heights	0.0	0.0	0.00	0.020000	0.160000	0.000000	0.0	0.02	0.0	0.02	0.0	0.0	0.00

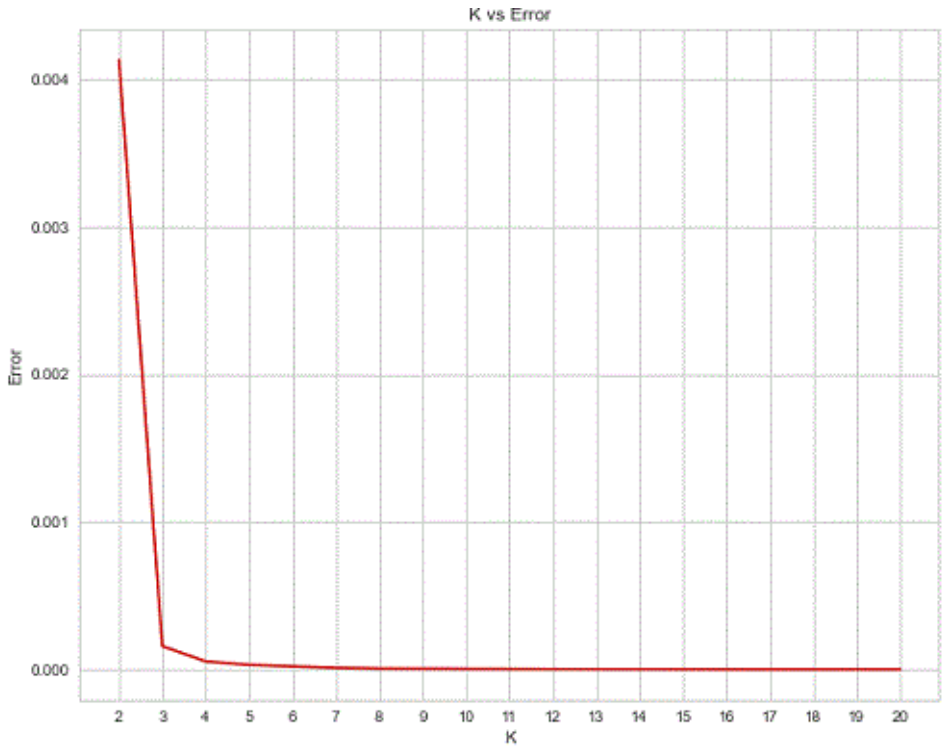
In [28]:
```
atl_art_grouped.shape
```

Out[28]:
```
(244, 82)
```

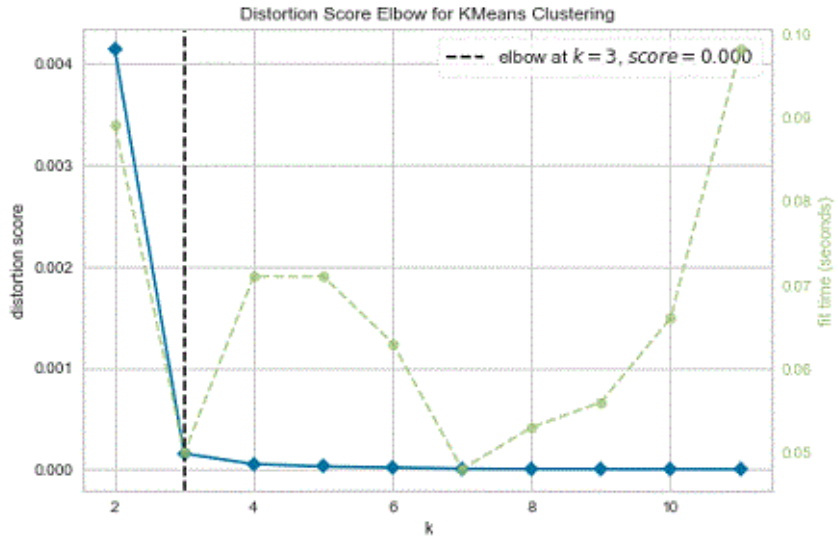
K-means Algorithm

K-means is a simple unsupervised machine learning algorithm that groups data into a specified number (k) of clusters. We used an unsupervised learning K-means algorithm to cluster the neighborhoods based on neighborhoods with similar averages of Art and Entertainment centers in that neighborhood. We used the Elbow Point Technique to get our optimum K value to avoid over-fitting or under-fitting the model. In this technique, we ran a test with different K values, measured the accuracy, and then chose the best K value.

We determined the best value of K when the line has The best K value is selected from a point in which the line has the sharpest turn. In our case, we had the Elbow Point at K = 3 that indicates we should apply a total of 3 clusters.



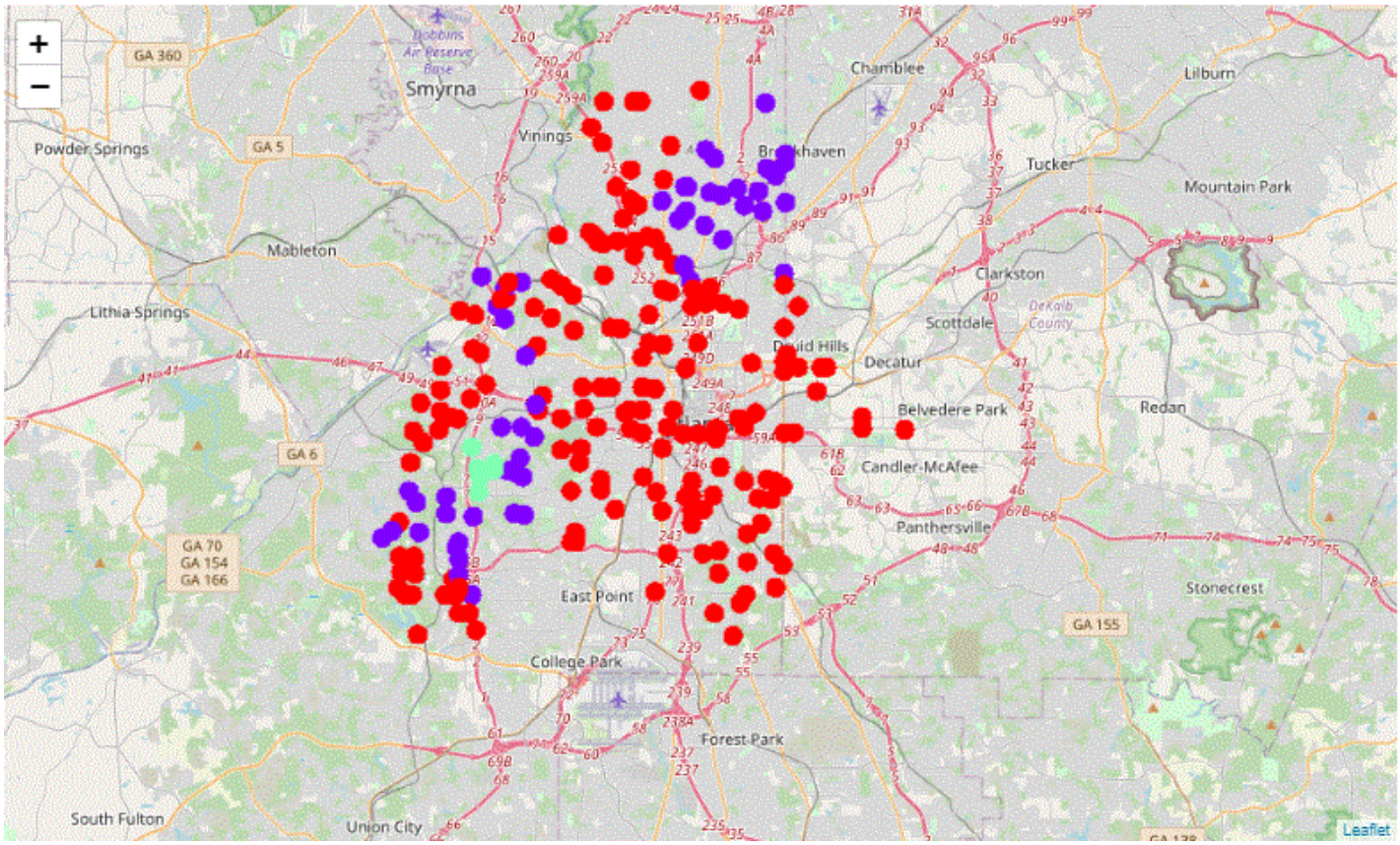
Similarly, as we demonstrate in the following figure, the KElbowVisualizer fits the KMeans model for a range of K values from 2 to 12 on a sample two-dimensional dataset with three random clusters of points. When the model fits 3 clusters, we can see a line annotating the “elbow” in the graph to be the optimal number.



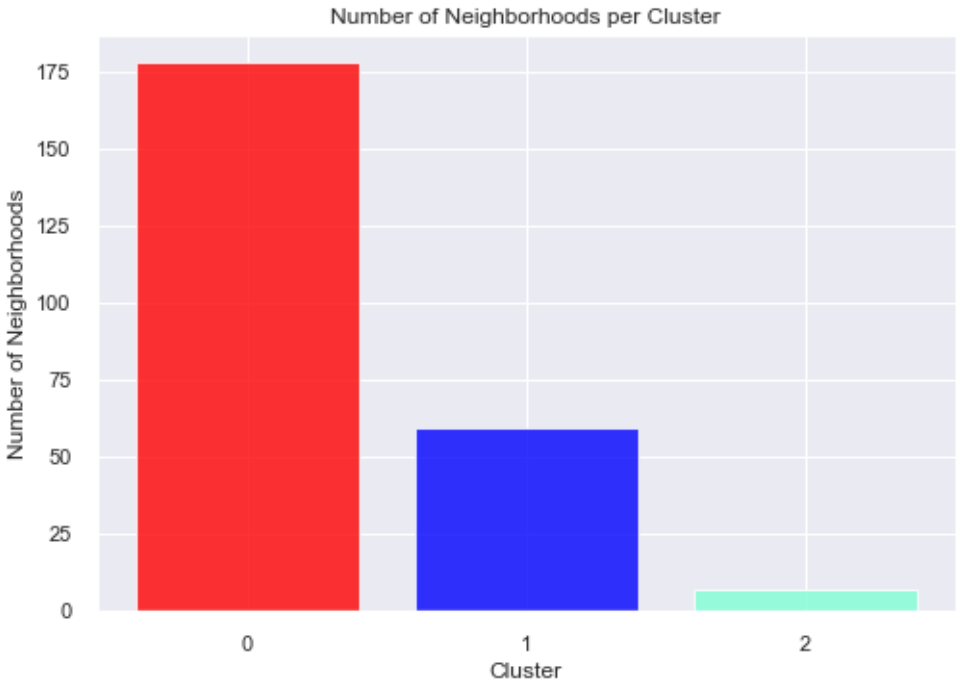
From the above two figures, we see that the Elbow is at 3; therefore, all the 244 Atlanta neighborhoods with a similar mean frequency of Art & Entertainment centers are grouped into three clusters. We labeled each of these clusters from 0 to 2, as we see in the following figure.

	Neighborhood	Arts & Entertainment	Cluster Labels
0	Adair Park	0.0	0
1	Adams Park	0.0	0
2	Adamsville	0.0	0
3	Almond Park	0.0	0
4	Amal Heights	0.0	0

By merging the above labeled data with our Atlanta venues dataset to add the locational information, we create a cluster map using the folium package. The following map depicts the three clusters that have a similar mean frequency of Art & Entertainment venues. The first cluster (cluster 0), the second, and the third clusters are depicted on the map by red, blue, and aquamarine colors, respectively.



The above maps represent the 3 clusters on which the first cluster (cluster = 0) comprises the classifications' significant proportion. We can see how many neighborhoods are in each cluster on the flowing figure.



We see that the first cluster (cluster = 0) comprises a significant proportion of the art and entertainment labels. We plot a chart below to depict how many neighborhoods are in each cluster. Of the 244 neighborhoods, 178 of them labeled in cluster 0. Cluster 1 identified by blue color comprises 59 neighborhoods, and cluster 2 identified by aquamarine color contains 7 neighborhoods.

Tabulation of Each Cluster

The tables presented below show the highly uneven distribution of the art and entertainment centers in Atlanta. The first table of cluster 0 shows a zero mean frequency of art and entertainment venues for Atlanta's 178 neighborhoods. On the other hand, cluster 1, with a total of 59 neighborhoods, has a 0.021277 mean frequency. The art venues' highest mean frequency is found in cluster 2 with 0.046512 in only 7 neighborhoods. This tabular analysis shows the uneven concentration of the art and entertainment centers in some downtown areas of the city.

Cluster 0 (Red Color)

	Neighborhood	NPU	Arts & Entertainment	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Atlanta University Center	T	0.0	0	33.754649	-84.402799	Mercedes-Benz Stadium	33.755298	-84.400942	Football Stadium
1	Atlanta University Center	T	0.0	0	33.754649	-84.402799	Revery: VR Bar	33.774272	-84.371114	Bar
2	Atlanta University Center	T	0.0	0	33.754649	-84.402799	Federal Penitentiary	33.714119	-84.368650	History Museum
3	Atlanta University Center	T	0.0	0	33.754649	-84.402799	O4W TARDIS	33.752836	-84.374104	Street Art
4	Atlanta University Center	T	0.0	0	33.754649	-84.402799	John Lewis Mural	33.755666	-84.380313	Public Art
5	Atlanta University Center	T	0.0	0	33.754649	-84.402799	Ponce City Market Rooftop - Skyline Park	33.772956	-84.366158	General Entertainment
6	Atlanta University Center	T	0.0	0	33.754649	-84.402799	San Francisco Coffee Roasting Co.	33.787317	-84.355444	Coffee Shop
7	Atlanta University Center	T	0.0	0	33.754649	-84.402799	Venkman's	33.766817	-84.363354	Restaurant

```
# counting unique values NPU
count_npu = len(pd.unique(df_cls_0['NPU']))
print('Cluster Labels 0 is located in {} Neighborhood Planning Unit (NPU)'.format(count_npu))
print('There are {} neighborhoods Labeled in cluster 0'.format(len(df_cls_0.Neighborhood.unique())))
```

```
Cluster Labels 0 is located in 24 Neighborhood Planning Unit (NPU)
There are 178 neighborhoods Labeled in cluster 0
```

Cluster 1 (Blue Color)

	Neighborhood	NPU	Arts & Entertainment	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Laurens Valley	R	0.021277	1	33.712420	-84.474553	Warehouse 2800/campbellton Road Community Center	33.699789	-84.479998	Music Venue
1	Laurens Valley	R	0.021277	1	33.712420	-84.474553	body control dance studios	33.737713	-84.432663	Dance Studio
2	Laurens Valley	R	0.021277	1	33.712420	-84.474553	Cascade Skating Center	33.755308	-84.496814	Roller Rink
3	Laurens Valley	R	0.021277	1	33.712420	-84.474553	ballethnic school of dance	33.684692	-84.442884	Dance Studio
4	Laurens Valley	R	0.021277	1	33.712420	-84.474553	T. Lang Dance | The Movement Lab	33.739358	-84.434157	Dance Studio
5	Laurens Valley	R	0.021277	1	33.712420	-84.474553	McGhee Tennis Center	33.732901	-84.447511	Stadium
6	Laurens Valley	R	0.021277	1	33.712420	-84.474553	Exhibit Works	33.762937	-84.516152	Art Gallery
7	Laurens Valley	R	0.021277	1	33.712420	-84.474553	Make Music	33.722450	-84.464310	Concert Hall
8	Laurens Valley	R	0.021277	1	33.712420	-84.474553	Buku Event Center	33.704523	-84.507282	General Entertainment

```
# counting unique values NPU
count_npu = len(pd.unique(df_cls_1['NPU']))
print('Cluster Labels 1 is located in {} Neighborhood Planning Unit (NPU)'.format(count_npu))
print('There are {} neighborhoods Labeled in cluster 1'.format(len(df_cls_1.Neighborhood.unique())))
```

```
Cluster Labels 1 is located in 10 Neighborhood Planning Unit (NPU)
There are 59 neighborhoods Labeled in cluster 1
```

Cluster 2 (Aquamarine Color)

	Neighborhood	NPU	Arts & Entertainment	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mays	H	0.046512	2	33.739551	-84.500243	The 166ATLery	33.752270	-84.501675	Art Gallery
1	Mays	H	0.046512	2	33.739551	-84.500243	Southwest Art Center	33.729624	-84.541587	Theater
2	Mays	H	0.046512	2	33.739551	-84.500243	Cascade Skating Center	33.755308	-84.496814	Roller Rink
3	Mays	H	0.046512	2	33.739551	-84.500243	McGhee Tennis Center	33.732901	-84.447511	Stadium
4	Mays	H	0.046512	2	33.739551	-84.500243	Buku Event Center	33.704523	-84.507282	General Entertainment
5	Mays	H	0.046512	2	33.739551	-84.500243	Exhibit Works	33.762937	-84.516152	Art Gallery
6	Mays	H	0.046512	2	33.739551	-84.500243	Make Music	33.722450	-84.464310	Concert Hall
7	Mays	H	0.046512	2	33.739551	-84.500243	Kill Island	33.765543	-84.551901	Theme Park Ride / Attraction
8	Mays	H	0.046512	2	33.739551	-84.500243	Douglass Theater	33.761703	-84.468288	Theater

```
# counting unique values NPU
count_npu = len(pd.unique(df_cls_2['NPU']))
print('Cluster Labels 2 is located in {} Neighborhood Planning Unit (NPU)'.format(count_npu))
print('There are {} neighborhoods Labeled in cluster 2'.format(len(df_cls_2.Neighborhood.unique())))
```

```
Cluster Labels 2 is located in 2 Neighborhood Planning Unit (NPU)
There are 7 neighborhoods Labeled in cluster 2
```

Result

From our analysis, we find out the existence of extremely uneven distribution of art and entertainment centers in Atlanta, Georgia. Most of Atlanta's neighborhoods (178 our 244) in cluster 0 have a zero mean frequency in our k-mean clustering analysis. These group of neighborhoods are in almost all Neighborhoods Planning Unit (NPU). Atlanta University Center, English Avenue, Historic Westin Heights/ Bankhead, and other neighborhoods are among cluster 0 neighborhoods with the lowest art and entrainment venues of the cluster and outside of downtown.

On the other side, cluster 1, represented by blue color, contains 59 neighborhoods in 10 Neighborhood Planning Unit. These neighborhoods in cluster 1 are located in the North-East and South-West part of the city. Some of the neighborhoods in the South-West side of Atlanta are adjacent to the Downtown. The mean frequency of this cluster group is 0.021277.

The third group of clusters, marked by aquamarine color, are located only in two Neighborhood Planning Unit (NPU). There are only seven neighborhoods Labeled in cluster 2 in the Downtown area of the city. This cluster has the highest mean frequency of art and entertainment venues equal to 0.046512.

In conclusion, we have found that the art and entertainment centers are unevenly distributed across Atlanta City. In consideration of art and entertainment as an essential part of the city's social and economic development, policymakers and stakeholders should try to address this uneven distribution of the centers to ensure the growth, prosperity, and sustainability of the arts and cultural aspects in Atlanta, Georgia.

Reference

[1] https://en.wikipedia.org/wiki/Atlanta (https://en.wikipedia.org)
[2] https://www.prnewswire.com/news-releases/georgians-for-the-arts-presents-cultural-capitol-2020-a-celebration-of-georgias-arts-and-culture-300985801.html (Georgians for the Arts*)
[3] https://georgiansforthearts.org/2020/01/27/cultural-capitol/ (Georgians for the Arts**)
[4] Copyright 2017 by Americans for the Arts, 1000 Vermont Avenue NW, Sixth Floor, Washington, DC 20005. Arts & Economic Prosperity is a registered trademark of Americans for the Arts. Reprinted by permission. Printed in the United States

(mailto:zekariasmulu@gmail.com) (https://github.com/zekariasm)