Battle of the Neighborhoods
- report -

Introduction/Business Problem

The local public authorities in Paris, France, plan to provide financial support for the opening of new arts venues across the city. One of the challenges is to use that investment so as to ensure a diversity of arts venues in all the neighborhoods of the city. For example, in any given neighborhood they wouldn't want to invest in types of venue that are already the most common ones available. For that, they need an analysis that would show them which types of arts venues are already available and are most common in the different neighborhoods of the city.

Data

Since the project is about Paris, France, and its neighborhoods, it requires a dataset containing all of those neighborhoods. Such a dataset is available on the Open Data Paris website, provided by the Paris Mayor's Office. The dataset is called 'Quartiers administratifs', and it contains a list of all the neighborhoods in Paris, including their geographical coordinates.

Using the list of Paris neighborhoods and their coordinates and the Foursquare API, I will get the top 5 most common arts venues for each neighborhood. After that, I will cluster the neighborhoods into 5 clusters, using k-means. Such clustering will provide the basis for analysis and recommendations.

Methodology

After downloading the dataset containing all the neighborhoods of Paris, France, I checked that it included all 80 of them, and then I visualized them superimposed over the map of Paris.

Before requesting the data for all neighborhood, from Foursquare, I first tested one neighborhood, to check the information delivered by the Foursquare API. To do that, I first obtained the latitude and longitude for the first neighborhood in the list (Enfants-Rouges), and then I requested all the arts venues for that neighborhood. I also checked the categories returned.

After that, I requested all arts venues for all neighborhoods in Paris, and in the process, I also displayed the name of each neighborhood in order to monitor that the request was working properly. I checked the size of the dataframe and the top 5 rows, to see what the structure of the columns looked like.

I used one hot encoding to map the different types of arts venue for each neighborhood. The result was a dataframe that used 1 to mark the arts venues present in a neighborhood and 0 for the ones which were absent. After that, I grouped the new dataframe by neighborhood, and also calculated the frequency for each specific type of arts venue for each neighborhood.

As an extra check, I then displayed the top 5 arts venues for each neighborhood. I created a new dataframe with the top 5 arts venue for each neighborhood. The columns were ordered from 1st most common arts venue first to 5th most common arts venue last.

Using k-means, I grouped the neighborhoods of Paris into 5 clusters, based on the top 5 most common arts venues in each neighborhood. That resulted in a dataframe with an additional column for the cluster labels.

The next step was to visualize the 5 clusters among the neighborhoods superimposed over the map of Paris. The different colors clearly identified the 5 clusters. After that, I proceeded to examine each cluster separately, by displaying the neighborhoods and their top 5 most common arts venues for each cluster.

Results

The analysis of the five clusters showed that art galleries and theaters were in general the most common arts venues among the top 5 arts venues across Paris. In certain clusters, they were the most common ones among the top 2 arts venues.

However, there were also important differences between the clusters. In cluster 1, the top most common venue was the theater, by a significant difference. In that cluster, art galleries were the most frequent among the second most common venues.

In cluster 2, the first and second most common venue categories were dominated by art galleries and theaters, but the former were more frequent among the first most common venues.

In cluster 3, theater was the most common venue among the first most common venues, but among the second most common venues the more frequent categories were museums.

Clusters 4 and 5 were smaller than the other three. Cluster 4 had theaters as the first most common venues, while cluster 5 had a more balanced group of first most common venues. In both clusters, the group of second most common venues was not dominated by any particular category.

Discussion

The results have shown that art galleries and theaters are the most common arts venues across Paris, suggesting that investment in new arts venues should target in particular other categories. That applies to almost all clusters, except clusters 4 and 5. Cluster 5 in particular did not have a clearly dominant arts venue category, so in that cluster investments could be made available for any type of arts venue. In cluster 4, theaters would not be an appropriate investment, but any other category would be suitable, including art galleries.

In the case of cluster 3, new investments should also exclude museums, which were the second most common venues in that cluster. Cluster 3 was a large cluster, and the category of first most common venues was relatively comparatively dominated by theaters and art galleries. However, the category of second most common venues in that cluster was dominated by museums, so museums should be added to the group of venues that are not a priority in cluster 3.

Conclusion

The dataset containing the names and coordinates of all the neighborhoods in Paris has enabled a visualization of their distribution across the map of Paris. When combined with the information on most common arts venues from Foursquare, the resulting data has made it possible to carry out an analysis of the distribution of such venues across Paris. The clustering of the neighborhoods by the top 5 most common arts venues has provided the basis for recommendations aimed at ensuring that investment in opening new arts venues

will also serve the purpose of achieving a more uniform distribution of the different categories of arts venues in the different neighborhoods of Paris.