

Hanoi Art District Clustering

*Clustering districts in Hanoi based on the similarity of art venues using
k-means*

Luu Minh Sao Khue
01/22/2020

Introduction

Background & Business Problem

Hanoi, capital of Vietnam, is a dynamic and multicultural city. Spread across 30 districts, there are a variety of entertainment options for both locals and tourists to enjoy. Culture and art lovers can spend a whole day in museums to learn about Vietnamese history and traditions or visit various art galleries to get themselves authentic artworks from talented local artists. At night, music fans can listen to the glamorous sound of the saxophone in a cozy jazz club, or watch the latest blockbusters in modern movie theaters. Every year, millions of tourists from around the world, as well as a large number of people from different provinces, come to Hanoi, looking for opportunities and experience. This is a paradise for tourists and art lovers to explore one of the greatest distinct cultures in Asia, as well as a prospective market for entrepreneurs to open their businesses, especially those with art-related. However, there might not be enough information on which location offer which types of art options. Therefore, this project was created to solve that problem.

In other words, the objective of this project is to recommend the best district in Hanoi to open an art gallery based on the interest of people in that location. Statistical techniques and a machine learning algorithm are used to analyze and cluster districts in Hanoi based on their similarity in the number of art venues. From there, I can check which cluster has the most art venues and galleries, which means there are potential customers usually go shopping in the district, and that would be a great place to open an art gallery.

Who would be interested?

Business owners, artists or art fans can find the result of this project useful in their decision-making process. Business owners can see areas that have many art venues and come up with further decisions. Tourists with great interest in Vietnamese artworks can look at the project's result to choose their destination for their next trip to Hanoi.

Data

Describe data

To solve the project problem, I need to collect these data:

- The list of district names in Hanoi (scraped from Wikipedia using BeautifulSoup)
- The latitude and longitude of these neighbourhoods (Geocoder package)
- Venues data related to the art section in each district to help find the location with more interest in art galleries (from Foursquare API)

How it is used to solve the problem?

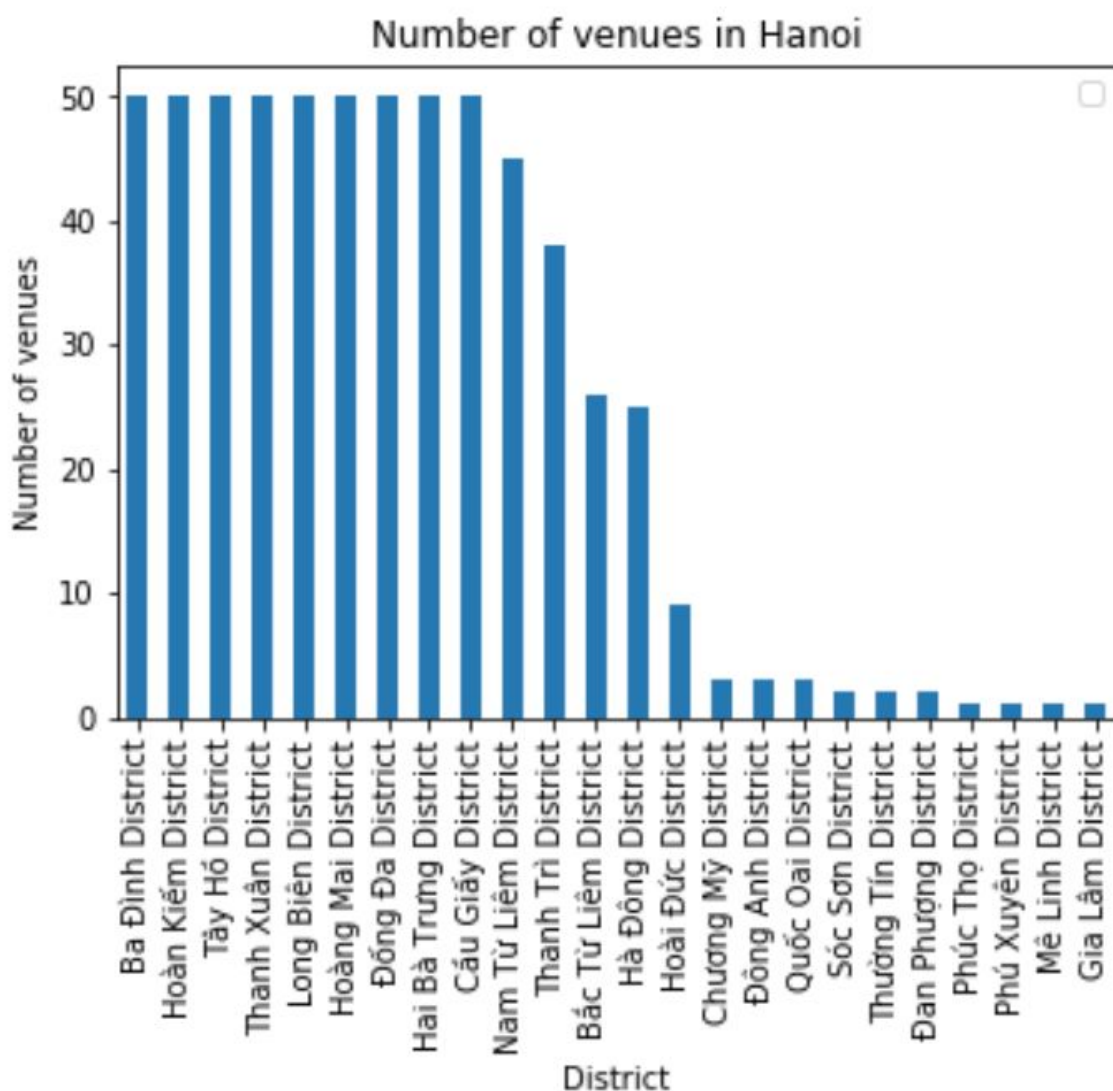
From the coordinates, I searched for 50 art-related venues within 10 kilometers of each district, then use k-means clustering technique to cluster neighbourhoods into 5 categories that represent similarity of the number of Vietnamese art venues in that district.

Methodology

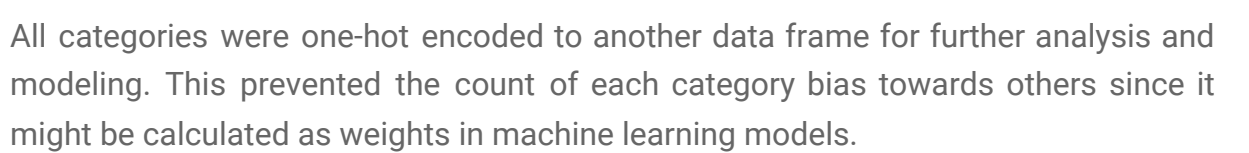
Exploratory Data Analysis

Data from different sources is firstly integrated into a single data frame. Then I check the unique values of the venue categories and found that there are 18 of them, which are Art Gallery, Museum, Movie Theater, Multiplex, Jazz Club, Music Venue, Opera House, Art Museum, History Museum, Rock Club, Theater, Concert Hall, Memorial Site, Performing Arts Venue, Public Art, Dance Studio, Zoo Exhibit, and Amphitheater.

Among the districts, Gia Lam, Me Linh, Phu Xuyen, and Phuc Tho have the least art venues. There is only one venue found for each district within 10 kilometers from the district center. Other districts near the city centers gather more venues.



Among the venue categories, 'Movie Theater' is the most popular with 199 venues, followed by 'History Museum' (107 venues). 'Theater' and 'Art Gallery' are the third and fourth popular venues in Hanoi, with 62 and 46 found venues respectively. Only a few venues are categorized into other categories.

[illegible]

After that, data were grouped by District and the mean of all counts for each category in a district was calculated. The result is as follows.

	District	Amphitheater	Art Gallery	Art Museum	Concert Hall	Dance Studio	History Museum	Jazz Club	Memorial Site	Movie Theater	Museum	Music Venue	Opera House	Performing Arts Venue	Public Art
23	Đống Đa District	0.000000	0.080000	0.060000	0.040000	0.020000	0.180000	0.020000	0.0	0.300000	0.04	0.060000	0.020000	0.02	0.0
21	Đan Phượng District	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.500000	0.50	0.000000	0.000000	0.00	0.0
9	Hà Đông District	0.040000	0.040000	0.040000	0.040000	0.080000	0.080000	0.000000	0.0	0.560000	0.00	0.000000	0.000000	0.00	0.0
11	Mê Linh District	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.00	1.000000	0.000000	0.00	0.0
17	Thanh Trì District	0.026316	0.052632	0.052632	0.052632	0.052632	0.157895	0.026316	0.0	0.342105	0.00	0.078947	0.026316	0.00	0.0

Machine Learning Model

I filtered the top 15 most common venues in each District and get them with k cluster equal 5.

	District	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
0	Ba Đình District	21.036305	105.828986	2.0	Multiplex	History Museum	Theater	Movie Theater	Art Gallery	Music Venue	Art Museum	Concert Hall	Museum	Jazz Club	Dance Studio
2	Bắc Từ Liêm District	21.071570	105.760472	2.0	Multiplex	History Museum	Art Museum	Art Gallery	Movie Theater	Rock Club	Concert Hall	Music Venue	Museum	Theater	Memorial Site
3	Cầu Giấy District	21.036308	105.786075	2.0	Multiplex	History Museum	Theater	Art Gallery	Movie Theater	Art Museum	Music Venue	Museum	Jazz Club	Concert Hall	Dance Studio
4	Chương Mỹ District	20.878474	105.649250	1.0	Art Gallery	Dance Studio	History Museum	Zoo Exhibit	Theater	Art Museum	Concert Hall	Jazz Club	Memorial Site	Movie Theater	Multiplex
5	Đan Phượng District	21.119627	105.678469	4.0	Movie Theater	Museum	Memorial Site	Art Gallery	Art Museum	Concert Hall	Dance Studio	History Museum	Jazz Club	Zoo Exhibit	Theater
6	Đông Anh District	21.136736	105.846359	2.0	Rock Club	Art Gallery	Multiplex	Zoo Exhibit	Memorial Site	Art Museum	Concert Hall	Dance Studio	History Museum	Jazz Club	Movie Theater
7	Đống Đa District	21.014488	105.823588	2.0	Multiplex	History Museum	Theater	Art Gallery	Movie Theater	Art Museum	Music Venue	Concert Hall	Museum	Jazz Club	Dance Studio
8	Gia Lâm District	21.023748	105.970415	3.0	Multiplex	Zoo Exhibit	Memorial Site	Art Gallery	Art Museum	Concert Hall	Dance Studio	History Museum	Jazz Club	Movie Theater	Theater
9	Hà Đông District	20.953590	105.759971	2.0	Multiplex	Movie Theater	Dance Studio	History Museum	Art Gallery	Art Museum	Concert Hall	Theater	Zoo Exhibit	Rock Club	Amphitheater
10	Hai Bà Trưng District	21.006518	105.857671	2.0	Multiplex	History Museum	Theater	Art Gallery	Movie Theater	Music Venue	Art Museum	Concert Hall	Museum	Jazz Club	Dance Studio

This project aims to cluster 30 districts of Hanoi into 5 areas based on the districts common in art venues. By scraping district info and coordinates from the internet and using data analysis and machine learning models, the project finds that districts in the city center are more art-oriented with more art galleries, whereas the districts that are near rural areas have fewer options. Though the findings only based on the

art venues similarity, I hope this brings values to people to have a first overview of the art and entertainment market in Vietnam.