



# **CLUSTER DISTRICTS IN HANOI BASED ON THE SIMILARITY OF ART VENUES USING K-MEANS**

Truong Son NGUYEN



# CONTENT

- I. Introduction
- II. Business problem
- III. Data
- IV. Methodology
- V. Result
- VI. Conclusion



# I. INTRODUCTION

- Hanoi is the capital of Vietnam, including 30 urban districts. There are many works of art, famous museums, and leisure activities for tourists
- Problem : travelers can't get enough information about the location offer which types of art options



## II. BUSINESS PROBLEM

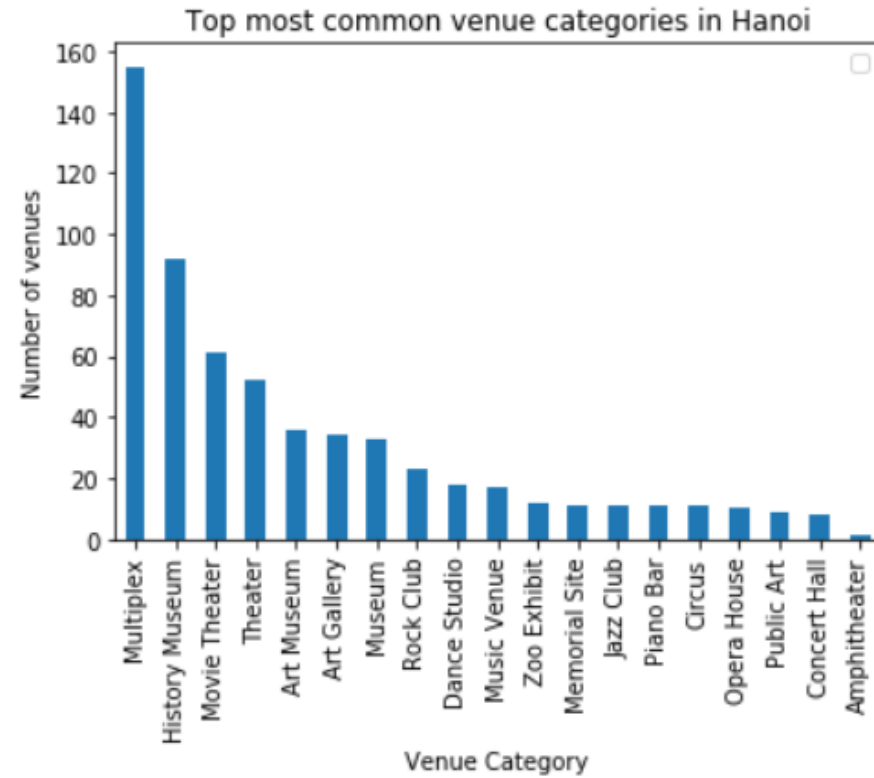
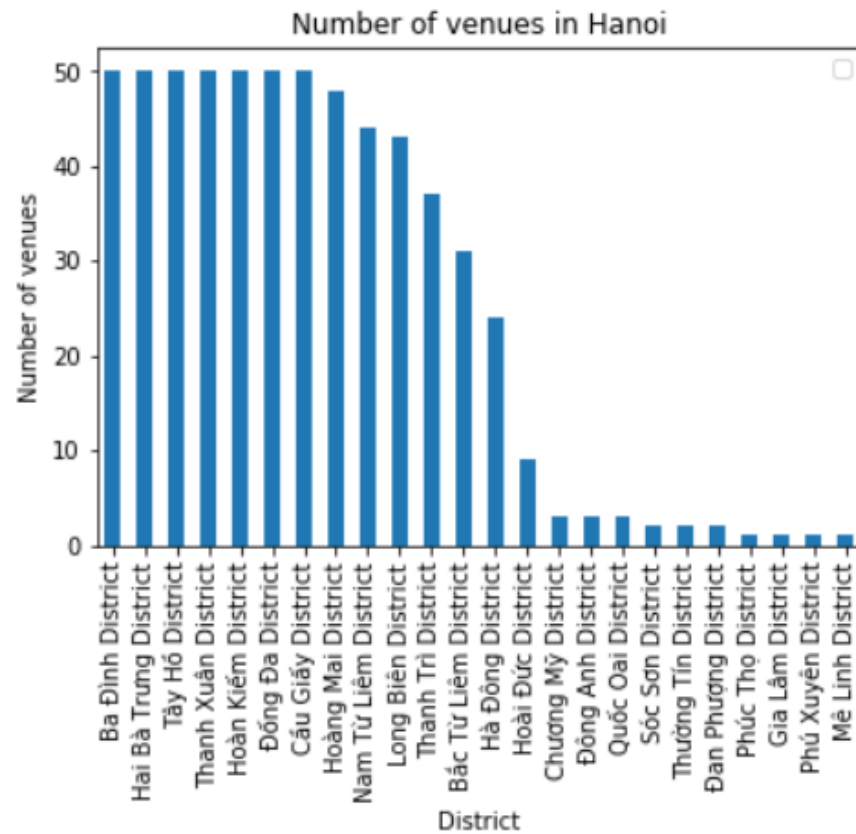
- Business owners, artists or art fans can find the result of this project useful in their decision-making process. Business owners can see areas that have many art venues and come up with further decisions. Tourists with great interest in Vietnamese artworks can look at the project's result to choose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



## III. DATA

- The list of district names in Hanoi (scraped from Wikipedia using BeautifulSoup)
- The latitude and longitude of these neighborhoods (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)

## IV. METHODOLOGY



- All categories were one-hot encoded to another data frame

[illegible]



- The mean of all counts for each category in a district

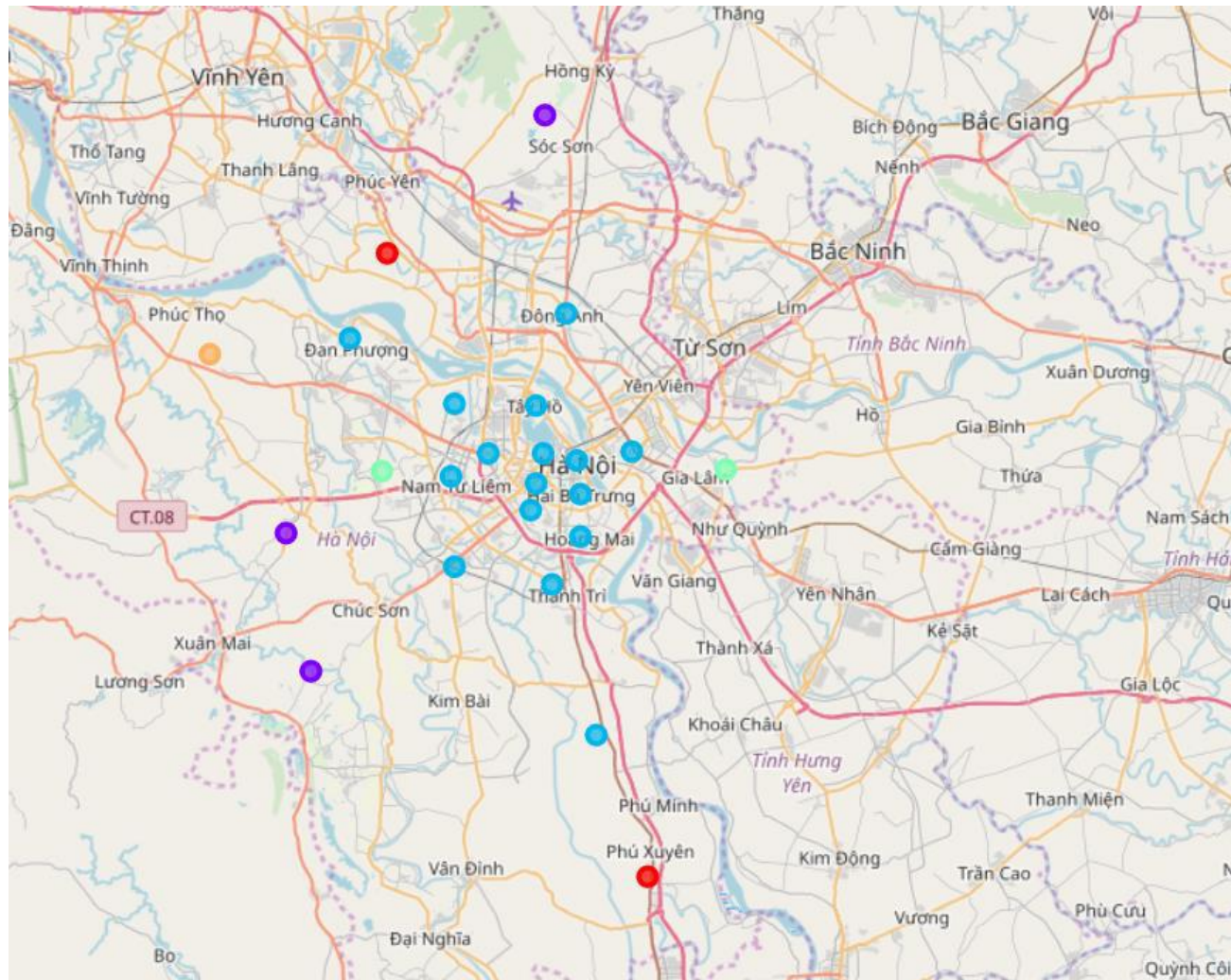
	District	Amphitheater	Art Gallery	Art Museum	Circus	Concert Hall	Dance Studio	History Museum	Jazz Club	Memorial Site	Movie Theater	Multiplex	Museum
0	Ba Đình District	0.000000	0.040000	0.060000	0.020000	0.000000	0.020000	0.160000	0.020000	0.020000	0.100000	0.260000	0.060000
1	Bắc Từ Liêm District	0.000000	0.129032	0.096774	0.000000	0.000000	0.032258	0.129032	0.000000	0.032258	0.064516	0.290323	0.032258
2	Chương Mỹ District	0.000000	0.333333	0.000000	0.000000	0.000000	0.333333	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000
3	Cầu Giấy District	0.000000	0.060000	0.060000	0.020000	0.020000	0.020000	0.160000	0.020000	0.020000	0.080000	0.240000	0.060000
4	Gia Lâm District	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
5	Hai Bà Trưng District	0.000000	0.040000	0.060000	0.020000	0.000000	0.020000	0.160000	0.020000	0.020000	0.100000	0.260000	0.060000
6	Hoài Đức District	0.000000	0.000000	0.111111	0.000000	0.000000	0.111111	0.111111	0.000000	0.000000	0.000000	0.666667	0.000000
7	Hoàn Kiếm District	0.000000	0.060000	0.060000	0.020000	0.000000	0.020000	0.160000	0.020000	0.020000	0.100000	0.240000	0.060000



- Top 6 most common venues of each district

	District	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	12th Most Common Venue	13th Most Common Venue
0	Ba Đình District	Multiplex	History Museum	Movie Theater	Theater	Art Museum	Museum	Rock Club	Art Gallery	Jazz Club	Circus	Dance Studio	Zoo Exhibit	Memorial Site
1	Bắc Từ Liêm District	Multiplex	Art Gallery	History Museum	Art Museum	Movie Theater	Rock Club	Music Venue	Piano Bar	Museum	Theater	Memorial Site	Dance Studio	Concert Hall
2	Chương Mỹ District	Art Gallery	Dance Studio	History Museum	Zoo Exhibit	Memorial Site	Art Museum	Circus	Concert Hall	Jazz Club	Movie Theater	Theater	Multiplex	Museum
3	Cầu Giấy District	Multiplex	History Museum	Theater	Movie Theater	Art Gallery	Art Museum	Museum	Rock Club	Circus	Concert Hall	Dance Studio	Zoo Exhibit	Jazz Club
4	Gia Lâm District	Multiplex	Zoo Exhibit	Memorial Site	Art Gallery	Art Museum	Circus	Concert Hall	Dance Studio	History Museum	Jazz Club	Movie Theater	Theater	Museum
5	Hai Bà Trưng District	Multiplex	History Museum	Movie Theater	Theater	Art Museum	Museum	Rock Club	Art Gallery	Jazz Club	Circus	Dance Studio	Zoo Exhibit	Memorial Site
6	Hoài Đức District	Multiplex	Art Museum	Dance Studio	History Museum	Zoo Exhibit	Memorial Site	Art Gallery	Circus	Concert Hall	Jazz Club	Movie Theater	Theater	Museum

# V. RESULT





## VI. CONCLUSION

- 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.