Predict Potential Business Services in Vietnam

Le Huy Hoang

July 05, 2021

Outline

- Introduction
- Data
- Methodology & Analysis
- Results
- Conclusion

Introduction

- Vietnam is a country in Southeast Asia with a population of over 96 millions.

- With rapid urbanization and the emerging of middle class, Vietnam has become a very attractive investment destination.

- By predicting development trends in Service Sector, the project will help
  - Investors to make their decisions.
  - Urban planners to optimize the effectiveness of land use and infrastructure.

Data

- We are going to analyze services' structures in the two cities
  - o Shanghai, China
  - o Singapore

Then compare their common characteristics with the status of Hanoi, Vietnam to assess the accuracy.

- The list of districts in each city will be extracted from Wikipedia (except Singapore)

- Locations of districts will be obtained from OpenStreetMap

- Venues in the cities will be queried from Foursquare Database

Data – Shanghai, China

- Districts with locations

Area	Latitude	Longitude
Huangpu, Shanghai	31.2322758	121.4692071
Xuhui, Shanghai	31.163698	121.4279938
Changning, Shanghai	31.2092762	121.3899859
Jing'an, Shanghai	31.2297756	121.44306
Putuo, Shanghai	31.2513263	121.3912291
Hongkou, Shanghai	31.266703	121.501751
Yangpu, Shanghai	31.2620112	121.5214305
Pudong, Shanghai	31.2217826	121.5387401
...	...	...

Data – Shanghai, China

- Venues

Area	Venue	Venue Latitude	Venue Longitude	Venue Category
Huangpu, Shanghai	Shanghai Grand Theater	31.2310298	121.4672625	Theater
Huangpu, Shanghai	Le Royal Club Lounge	31.2364040	121.4713638	Lounge
Huangpu, Shanghai	Lobby Bar @ 38th floor - JW Marriott	31.2324214	121.4655531	Hotel Bar
Huangpu, Shanghai	Jing'an Sculpture Park	31.234794	121.463911	Sculpture Garden
Huangpu, Shanghai	BON VIVANT	31.2259758	121.4709393	French Restaurant
...	...	...	...	...

Data – Singapore

- Districts with locations

Area	Latitude	Longitude
Raffles Place, Singapore	1.2835417	103.851460232669
Anson, Singapore	1.2737957	103.843473
Queenstown, Singapore	1.2946226	103.8060366
Telok Blangah, Singapore	1.27102005	103.809694766557
Pasir Panjang, Singapore	1.27620135	103.79147582342
High Street, Singapore	1.290383	103.8497316
Middle Road, Singapore	1.2986301	103.8538615
Little India, Singapore	1.30684265	103.849273551709
…	…	…

Data – Singapore

- Venues

Area	Venue	Venue Latitude	Venue Longitude	Venue Category
Raffles Place, Singapore	The Fullerton Bay Hotel	1.28387752161848	103.853314336212	Hotel
Raffles Place, Singapore	Ritual Gym	1.28596464274139	103.848650631204	Gym
Raffles Place, Singapore	The Fullerton Hotel	1.2862	103.85298	Hotel
Raffles Place, Singapore	Amoy Hotel	1.28311832260938	103.848538559423	Hotel
Raffles Place, Singapore	Sabaai Sabaai Traditional Thai Massage	1.28696389067267	103.8495122821	Massage Studio
...	...	...	...	...

Data – Hanoi, Vietnam

- Districts with locations

Area	Latitude	Longitude
Ba Đình, Hanoi	21.0340746	105.8145271
Bắc Từ Liêm, Hanoi	21.0698605	105.7573392
Cầu Giấy, Hanoi	21.02916475	105.803437672196
Đống Đa, Hanoi	21.0129196	105.8271961
Hà Đông, Hanoi	20.97026	105.775000621979
Hai Bà Trưng, Hanoi	21.0059701	105.8574845
Hoàn Kiếm, Hanoi	21.0285237	105.8507155
Hoàng Mai, Hanoi	20.9745977	105.8637067
…	…	…

Data – Hanoi, Vietnam

- Venues

| Area	Venue	Venue Latitude	Venue Longitude	Venue Category
Ba Đình, Hanoi	Sky Walk Lotte Centre	21.0321312412437	105.812428062158	Scenic Lookout
Ba Đình, Hanoi	Polygon Music	21.029921758545	105.822861692114	Rock Club
Ba Đình, Hanoi	Cup of Tea Cafe & Bistro	21.0330838747582	105.810379214208	Tea Room
Ba Đình, Hanoi	Carambola Cafe	21.0334450711338	105.81612380212	Café
Ba Đình, Hanoi	Bornga - Original Korean Taste	21.0315121955838	105.812574921882	Korean Restaurant
…	…	…	…	…

Methodology & Analysis

- Collect information of venues around the divisions of the cities
- Remove divisions with less than 10 venues
- Encode venues' categories using One-hot Encoding
- Use K-means Clustering to segment divisions in each city based on theirs encoded venues' categories
- To find the best number of clusters, Elbow Method & Silhouette Score are used
- The most common venues in each cluster will be summarized
- The set of business services which are common in both Shanghai & Singapore will be extracted as prediction results
- The results will be compared with the service structure of Hanoi, Vietnam to evaluate the prediction 's performance

Methodology & Analysis

- Shanghai, China

Best number of Clusters: 5

Clustering Evaluation

Changning, Shanghai - Cluster 0
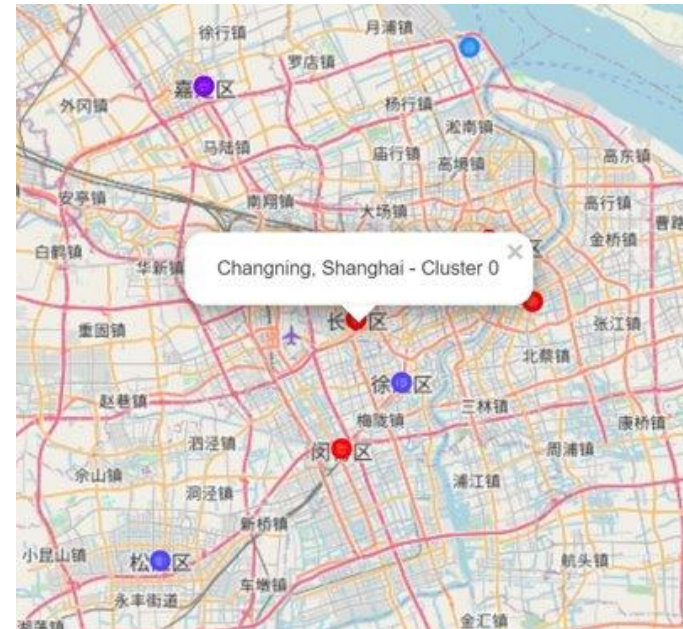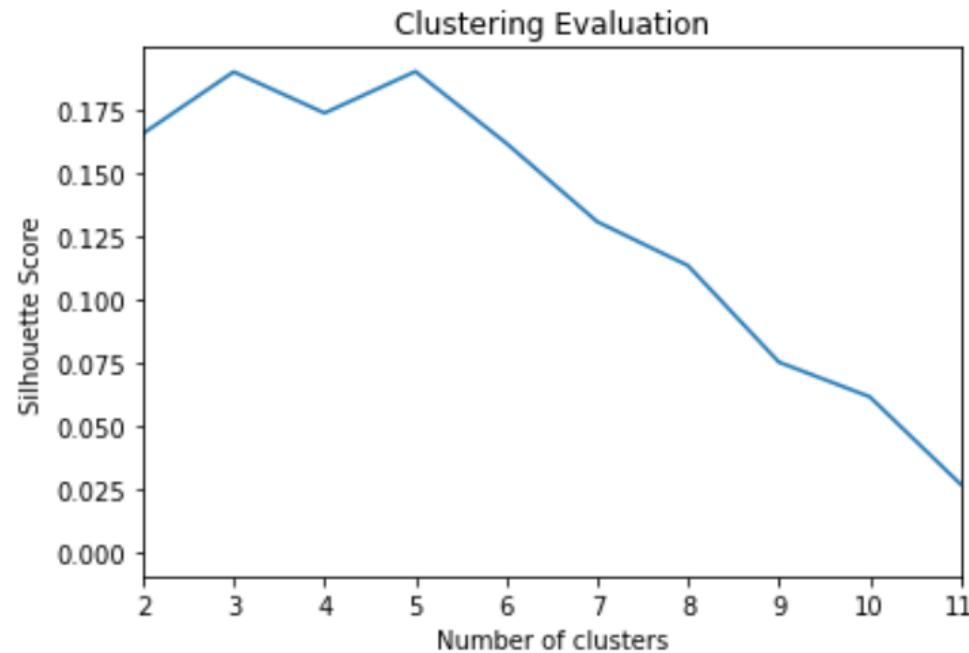
Methodology & Analysis

- Shanghai, China

Summary of clusters with most common venues

Cluster	Most Common Venues		
	1st	2nd	3rd
0	{'Coffee Shop'}	{'Japanese Restaurant', 'Hotel', 'Fast Food Restaurant'}	{'Italian Restaurant', 'Hotel', 'Fast Food Restaurant', 'Shopping Mall'}
1	{'Hotel'}	{'Garden'}	{'Food'}
2	{'Hotel', 'Fast Food Restaurant'}	{'Hotel', 'Fast Food Restaurant', 'Coffee Shop'}	{'Metro Station', 'Hotel', 'Coffee Shop'}
3	{'Shopping Mall'}	{'Port'}	{'Boat or Ferry'}
4	{'Hotel'}	{'Dumpling Restaurant', 'French Restaurant'}	{'Chinese Restaurant', 'Hotpot Restaurant'}

Methodology & Analysis

- Singapore

Best number of Clusters: 13
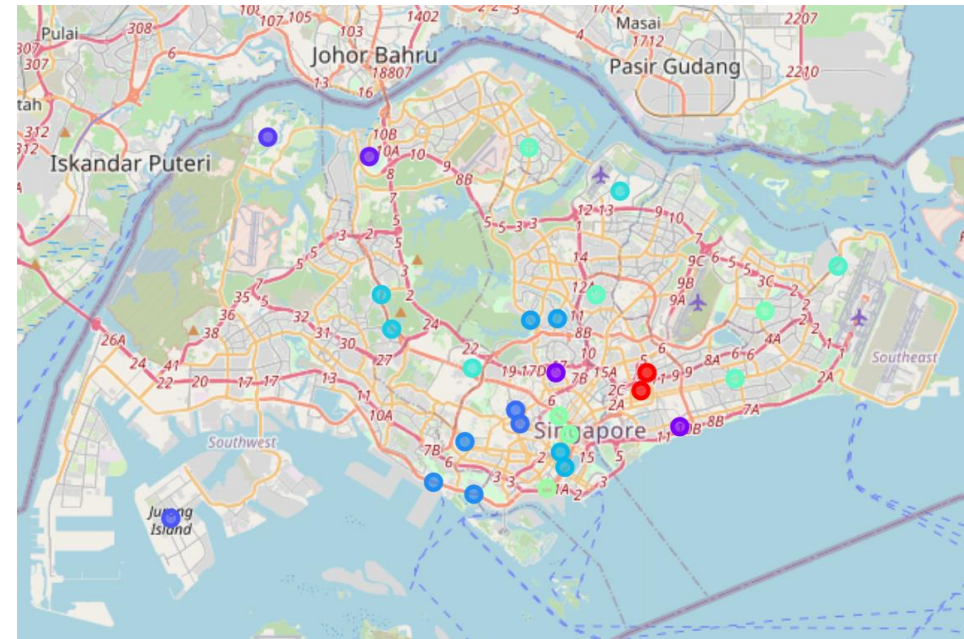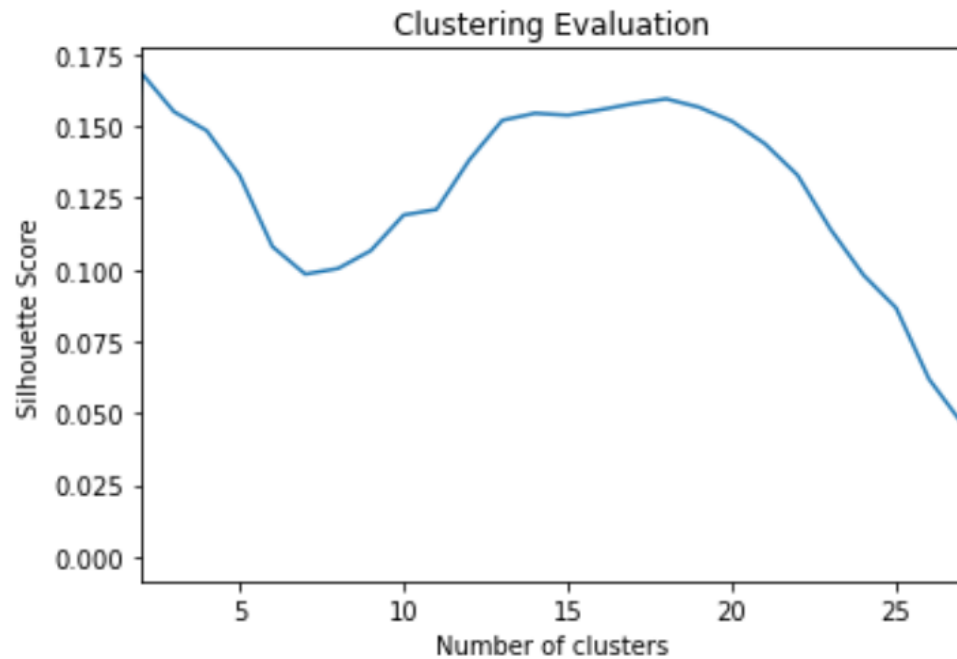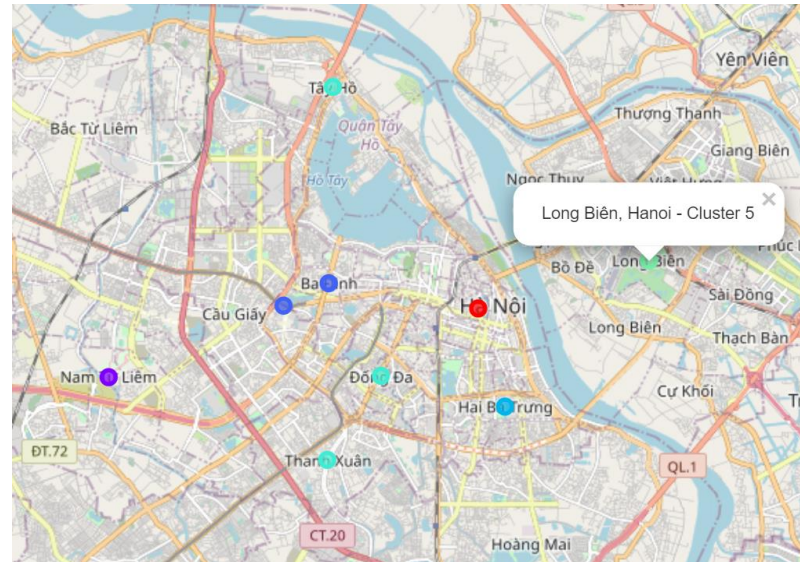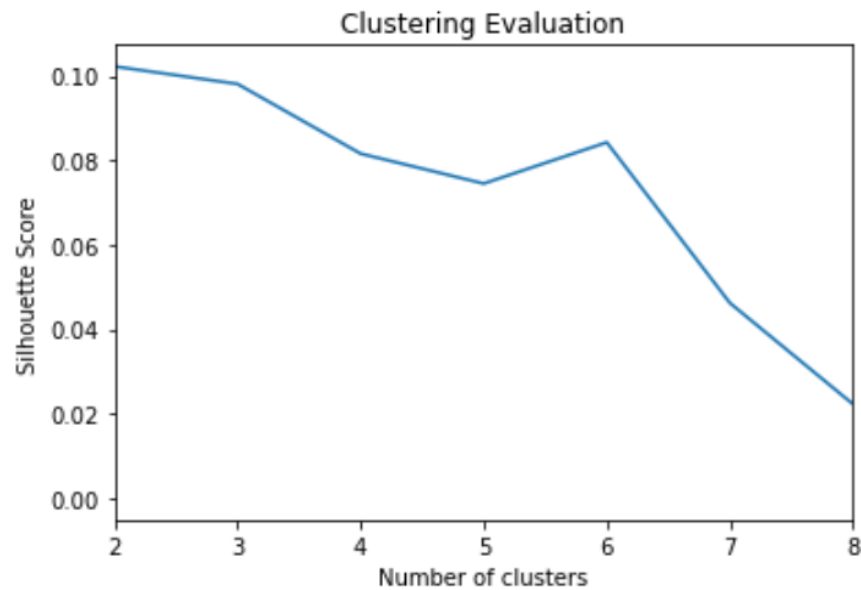
Methodology & Analysis

- Singapore

Cluster	Most Common Venues		
	1st	2nd	3rd
0	{'Resort', 'Park'}	{'Café', 'Trail'}	{'Scenic Lookout', 'Indian Restaurant'}
1	{'Chinese Restaurant', 'Food Court', 'Thai Restaurant', 'Coffee Shop'}	{'Chinese Restaurant', 'Thai Restaurant', 'Food Court', 'Coffee Shop'}	{'Indian Restaurant', 'Chinese Restaurant', 'Food Court', 'Coffee Shop'}
2	{'Hotel'}	{'Bakery'}	{'Japanese Restaurant', 'Clothing Store'}
3	{'Harbor / Marina'}	{'Beach'}	{'Resort'}
4	{'Hotel'}	{'Japanese Restaurant', 'Waterfront'}	{'Performing Arts Venue', 'Waterfront'}
5	{'Farm'}	{'Campground'}	{'Harbor / Marina'}
6	{'Chinese Restaurant', 'Café', 'Airport'}	{'Chinese Restaurant', 'Café', 'Park'}	{'Italian Restaurant', 'Café', 'Coffee Shop'}
7	{'Korean Restaurant', 'Coffee Shop'}	{'Indian Restaurant', 'Nature Preserve'}	{'Café', 'Food Court'}
8	{'Grocery Store'}	{'Food Court'}	{'Pizza Place'}
9	{'Japanese Restaurant', 'Bakery', 'Coffee Shop'}	{'Asian Restaurant', 'Italian Restaurant', 'Coffee Shop'}	{'Supermarket', 'Chinese Restaurant', 'Spanish Restaurant'}
10	{'Asian Restaurant', 'Chinese Restaurant'}	{'Asian Restaurant', 'Chinese Restaurant'}	{'BBQ Joint', 'Noodle House'}
11	{'Coffee Shop'}	{'Park'}	{'Supermarket'}
12	{'Hotel'}	{'Café', 'Indian Restaurant'}	{'Japanese Restaurant', 'Café'}

Methodology & Analysis

- Hanoi, Vietnam

Best number of Clusters: 6

Methodology & Analysis

- Hanoi, Vietnam

Summary of clusters with most common venues

Cluster	Most Common Venues		
	1st	2nd	3rd
0	{'Hotel'}	{'Vietnamese Restaurant'}	{'Coffee Shop'}
1	{'Café'}	{'Coffee Shop'}	{'Bakery'}
2	{'Café', 'Coffee Shop'}	{'Japanese Restaurant', 'Hotel'}	{'Vietnamese Restaurant'}
3	{'Vietnamese Restaurant'}	{'Café'}	{'Japanese Restaurant'}
4	{'Vietnamese Restaurant', 'Noodle House', 'Coffee Shop'}	{'Multiplex', 'Vietnamese Restaurant', 'Coffee Shop'}	{'Café', 'Bakery'}
5	{'Vietnamese Restaurant'}	{'Auto Garage'}	{'Shopping Mall'}

Results

- In both Shanghai, and Singapore, Restaurant, Coffee Shop, and Hotel are the most popular services

- In Shanghai, Fast Food is the most popular type of restaurants, while in Singapore, people seem to prefer Foreign Cuisines

- From the similarities between Shanghai, and Singapore, we could predict that Restaurant, Coffee Shop, and Hotel will be the most promising business services in other cities in the nearby regions

- The prediction is matched with the service structure in Hanoi, one of the biggest cities in Vietnam

Conclusion

- By analyzing the similarities in the business structures of Shanghai, China, and Singapore, a set of potential business services in Vietnam has been extracted.

- The accuracy of the prediction has been proved by comparing to the data of Hanoi, one of the biggest cities in Vietnam

- Improvement ideas
  o To dig further into the densities of venues in each city to find out saturation thresholds
  o To utilize historical aspects to customize the prediction for each city