Hanoi venues and real estate correlations analysis

Vi Thanh Dat

March 2019

1 Introduction

1.1 Description of the problem and background

Hanoi is the capital of Vietnam, the political, economic, cultural, commercial and tourist center of the country. Currently, It is one of the most fast-growing cities in the world with over 8 million in population and population density of 2,392 people per square kilometer. It is the 3rd most dynamic city in the world according to [1]. I decided to use Hanoi in my project since I myself am a resident of the city.

For such a dynamic city, sometimes it's really hard to choose the area where apartments price is low. And at the same time, geographically convenient. I think this is the problem that many parties interested in such as investors, city's residents and even the government.

My plan in this project is to create a correlation graph between apartments price and venues distribution in each district, and cluster districts then examine the correlation between each clustered and the price range.

1.2 Data preparation

Data used for the project:

1. Hanoi's 12 urban districts and their coordinates: I scraped the data from [3] and it's coordinate section.
2. I used Foursquare API to get common venues in each districts.
3. I scraped the web [2] to get the price of about 10000 apartments. And then I take the average per square price for each district.

References

[1] Chinadaily. World's top 10 most dynamic cities, 2019.

[2] https://batdongsan.com.vn. Vietnamese real estate website.

[3] Wikipedia. Hanoi, 2019.