



IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

Recommend the best location to start a
business in Hanoi

ABSTRACT

Hanoi, a thousand-year-old capital city of Vietnam, is well-known for its historical heritage and beautiful landscape. Thanks to that, investors have found Hanoi as an attractive destination, thus, more and more businesses are opened in this city. Considering this trend, the aim of the project is to sort out the best location to start a business in Hanoi.

LINH THUY PHUNG

June, 2020

Introduction

As one of the most ancient cities in the world that is filled with rich culture and beautiful lakes, Hanoi is the capital of Vietnam that is situated in the northern part of the country. For more than a thousand year, Hanoi has always been the center of culture and political, creating a deep connection with government and large network for investors to share information and experiences. Moreover, the city is chosen as headquarter of many educational institutions, combined with a high proportion of youngster, it naturally provides a large number of high-quality young labors. Especially, the inexpensive labor costs is particularly advantageous to small businesses in saving costs and boosting revenues. In addition to that, Vietnamese government has made a lot of efforts on the openness of foreign investment, reformed policies, welcoming business environment, and simplified process to lower the barriers for both domestic and foreign investors. For these reasons, it is no doubt that Hanoi has become the hot spot for investment and starting new business.

Business problem

Starting a new business is never easy as it requires a thorough plan with so many factors to be taken into consideration. Among them, choosing a location is one of the most important decision. It especially arduous for foreigners who have little knowledge of the local destination. Therefore, the objective of this capstone project is to solve that specific task. In this project, we shall analyze data to select the most suitable location in Hanoi to open a business. Using data science knowledge, the project aims to explore geographical data of Hanoi, make visualization with map chart and marker and then, cluster venues into groups to find out which group is the best to locate a new business.

Target audience of the project

As previously mentioned, the main user of this project will be investors who are interested in opening a business in Hanoi, Vietnam. Among them, foreigners would be more keen on the results due to their unfamiliarity with the local background. Moreover, the results are not solely restricted to one particular viewer like investor, it could also yield benefits for many other subjects

Local authorities could use it to understand which part of the city is the most occupied with businesses to manage. Builders can seek the information of the most popular area to construct new building for hiring. Even students can find it useful to find the location of potential employer.

Data section

Introduction of data to be used to solve the business problem

A list of all neighborhoods in Hanoi. The scope of this project therefore will be restricted to only Hanoi city of Vietnam, since it is the special location to attract investment and is currently home of many businesses

The coordinate data of Hanoi city, including specific latitude and longitude of all neighborhood in the above list. This information will eventually be used to plot a map chart to depict and visualize the location of each neighborhood

Data of venues around the neighborhoods in Hanoi. At the end of the project, these venues shall help us cluster the neighborhoods into different groups

Data source and methodology to retrieve data

Wikipedia is a multilingual online encyclopedia created and maintained as an open collaboration project by a community of volunteer editors using a wiki-based editing system. It is the largest and most popular general reference work on the World Wide Web. Hence, For the list of neighborhoods in Hanoi, we will use the data extracted from Wikipedia. The list contains 30 neighborhoods could be retrieved from this link: https://en.wikipedia.org/wiki/Category:Districts_of_Hanoi. Then, we shall scraping data from this website using BeautifulSoup.

After that, we install the Geocoder package of Python. With the names of all neighborhoods in Hanoi city, Geocoder could help identify the geographical coordinates, including both latitude and longitude.

Finally, Foursquare API could be used in this last step to gather all venues around the city. It has one of the largest databases of information about more than 105 millions locations, thus, it becomes really famous and frequently used by many developers around the world. So, in this project, we also use Foursquare API to retrieve data of venues located in Hanoi. The detailed explanation on methodology will be discussed in the next section.