Introduction/ Business Problem:

Ho Chi Minh City, abbreviated as HCMC, is the most populous city in Vietnam with a population of 9 million as of 2019. Located in southeastern Vietnam, the metropolis surrounds the Saigon River and covers about 2,061 square kilometers (796 square miles).

Being the financial center of Vietnam, HCMC has the headquarters of many national and international banks and companies. As a major gateway to Vietnam, the city received over 8.6 million international visitors in 2019, of which 4.1 million were overnight visitors.

That is the reason why Ho Chi Minh city is the most developed city in Vietnam opening many opportunities for Small and Medium Enterprises (SMEs) to start a new business. Apart from online business, companies also need stores or shops for displaying and selling their products. Therefore, in current years, the real estate market in HCMC is hot with high prices and high demand.

As an investor or an owner of an enterprise, we expect to buy or rent a place with a low price and easy to approach the customers. Especially in the Covid-19 period, the price of renting is falling but how much and what is the price offered by the owners? I will try to answer these questions in this project by analyzing the HCMC neighborhood. The purpose is to give the SMEs an overview of the housing and rental price in each district of this city to help them have a better decision to choose a place depending on their budget.

HCMC consist of 24 districts including 19 urban districts and 5 rural districts. Because there is not much data for rural districts, this project will focus only on 19 urban districts of HCMC. We will create a map and information chart to illustrate the current situation of the real estate market in HCMC and cluster districts according to the venue density and housing and renting price.

Data Description:

To analyze the Business Problem above, I will collect data from sources as below:

- Data of Housing Average Price of each District per square meter scraped from a real estate broker website (https://mogi.vn/gia-nha-dat). It has 2 columns: District name and Average price.
- Also, in that website (https://mogi.vn/ho-chi-minh/thue-mat-bang-cua-hang-shop), I will take the data of 600 latest article of office, store, shop rental with the information of address, area and price of monthly rental and convert it to a data frame to calculate the average renting price per square meters in each District.

(These 2 data frames need to be prepared including extracting District name from full address, delete the unit and convert string to the number.)

- Foursquare API to explore the most common venues in each District
- To create the map, I need to use Geocoder to get the latitude and longitude. However, some districts are large, and the coordinates generated by Geocoder are not the good ones (not many venues surrounding the point), I would use Google Map to adjust the coordinates of 3 districts (District 7, District 9 and District Binh Tan).

With the data above, I can cluster the districts by K-means Clustering base on the most common venues, housing price, and renting price.