

Applying Data Science in identifying the best place to setup a cafe in Ho Chi Min City

Vigneshkumar Sivakumaran

Business Problem:

If you have a chance to visit Ho Chi Minh City (HCMC), you will see that it's one of the most active cities in the world. People here love to talk, to chat everywhere. The catering services in this city are on high demand.

We need to clarify the differences between a *coffee shop* and a *cafe*. Coffee shop has no similar connotations. From personal experience in the United States, a cafe serves meals, while a coffee shop usually just sells snacks (muffins, scones, shortbread). This is not strictly the case, and both usually serve coffee. In this project, we suppose to work only on the cafe.

Although there are already a lot of cafes in HCMC, their density between districts is not uniform. There are some districts containing too many cafe while there are less in some others. If we have some knowledge about the *population density*, the *housing price* in each district coupling with an overview of *the number of cafe*, we can have a better idea to set up a new business there.

If we think of it by an investor, we expect to choose a place where the population density is high but less competitors. If the housing price in that place is low, it's more attractive to us.

By using Data Science and some geometric factors about the relation between districts in HCMC, we can give good answers of following questions to the investors so that they can have a better vision about not only the cafe but also about other venues in HCMC.

1. *How many venues in each district? Answering this question gives us a better understanding about the dynamic level of a district.*
2. *How many categories in each district? Answering this question helps us know about the diversity in business of a district.*
3. *How many venues in each category? This question shows the magnitude of a category in a district*
4. *What are the most popular categories in each district? If investors change their mind to focus on other commercial fields instead of opening a cafe.*
5. *How many clusters we can use to categorize the districts based on the popularity of cafes?*
6. *In which districts, the average housing price is low and the number of cafes is low also?*
7. *Where there are many people but less cafes?*
8. *Visualize all information on the map so that we can have a better look on what we want to find the answers*