Stefano Di Pierro.

Project: Building a Mall in Singapore

Photo by Stefano Di Pierro of Singapore's Skyline.

I have been asked to analyze the pros and cons of building a new shopping mall in Singapore. Singapore is a highly populated city with high prices in real estate and a minimum wage salary that is very high compared to the rest of Asia and even Europe.

In a very busy environment as Singapore is, citizens as tourists and business travelers tend to enjoy Shopping Malls during their spare time. People spend time in Shopping Malls for many reasons like eating lunches between working hours, dining out, doing shopping, have a business meeting in a coffee shop or just have a cool break during a warm day. For retailers Malls are a good opportunity as well to sell their products and have a good window to publicize themselves. Many Property owners are seeing the pros since many Shopping malls have been built and many more are being built while we are talking because it allows property owners to earn steady rental income.

Singapore is unique in having more dining outlets in malls than retail ones because of the culture here of eating out and socializing rather than eating at home. It was one of the first things that I noticed when I first came here six years ago. The food courts and centers along with more interesting gourmet dining options anchoring malls are often as important as the big retail malls anchoring them.

Building a Mall is a serious matter and needs a deeper analysis.

Business Problem.

Where to open a new Mall? Thanks to Data Science and Machine Learning we will determine where is better to open a new Shopping Mall. This project goal is to give an answer to investors and stakeholders at the question: Where we should open a new Mall?

General Data About the Situation.

There are 103 malls in Singapore, that's one for every 53,000 people living in Singapore. Is that sustainable? The evidence from malls and retailers would suggest not.

This project so is timely because there is an oversupply of the Shopping Mall at the moment.

Comparing with Data from Mainland China we can see that on average, mall visitors made nine purchases or transactions over the past six months, with total annual spending in excess of Rmb700bn ($113bn). I would wager that this figure will be less in Singapore as many people go to window shop, enjoy the free kids' activities, and most of all for the free air conditioning to escape the humidity and heat outside.

China accounted for 44 percent of total global shopping mall completions in 2014, according to CBRE, with more mall space added in the central city of Wuhan alone (993,000 square meters) than in the whole of the Americas (800,000 sqm). So how many malls we have in Singapore and where are they located?

How to organize Data.

We need the following data: a list of neighborhoods of Singapore.

Latitude and Longitude of these neighborhoods

Venues data, especially Shopping Malls to perform clustering on the neighborhoods.

Sources and methods.

I will use this webpage (https://en.wikipedia.org/wiki/Postal_codes_in_Singapore) to get the data.

I will scrap data from this webpage using beautifulSoup and python requests than I will use python geocoder to request coordinates of those neighborhoods.

I will follow up using my personal Foursquare API to request venues data for these neighborhoods. Foursquare contains numerous data and I am particularly interested in data regarding Shopping Malls. Then I will use the Folium package to visualize data in a map. Summering the item I will make use of many skills that I learned in past courses:

Web Scraping with BeautifulSoup

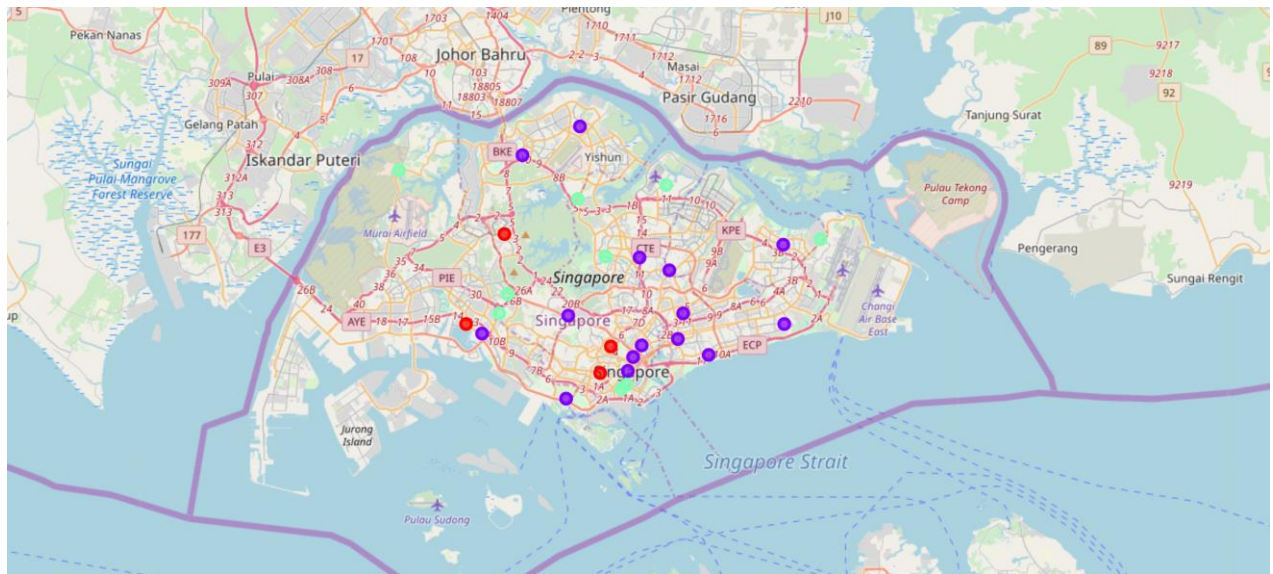Data frames building with pandas

Vectors functions with numpy

API handling

Machine learning method as KMeans (clustering)

Map visualization with folium.

CONCLUSIONS.

It was possible thanks to Folium to visualize the clustering and the shopping mall on a detailed map.

So I requested Foursquare to give me all venues in Singapore than I kept just the venues having to belong to the category Shopping Mall. Than our clustering resulted in 3 clusters.

Cluster 0 represented in violet. Cluster 0 is the more populated one in terms of Shopping Malls

Cluster 1 is represented in light green and is medium populated

Cluster 2 represented in red has just a few Shopping malls and it is situated on the west part of the map

IN EASY TERMS.

From this report, we can see 2 things, first one that cluster 2, the red one, contains less shopping Malls, so if I were a property investor I would consider to build one in this area.

Second thing, most Malls are situated in the south part of the city, close to the coast, there is probably a good reason for that, but as a new investor, I would avoid that area as it looks there is an oversupply of Malls.

WHAT COULD BE IMPROVED.

First I reckon 3 clusters, Kmeans could be reconsidered to find a better fit for the data since results are a bit overlapping in my opinion.

Second I would definitely consider other factors than merely coordinates data, such as the density of population, average income per area and costs of housing per neighborhood.

These factors are important in such a difficult choice.

CONCLUSION.

After clustering data of existing venues, they have been assimilated trough machine learning in 3 clusters, associating them for similar characteristics.

The results showed us the less populated and so less competitive and probably more appetible cluster resulted in being cluster number 2.