IBM Capstone Project by Coursera

- Written by Stefano Di Pierro

- Project: Segmenting Singapore
- Why? Opening a Shopping Mall
- How Data Science and Machine Learning could help us in this decision

IBM CAPSTONE PROJECT BY COURSERA

IBM Capstone Project by Coursera

- 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?

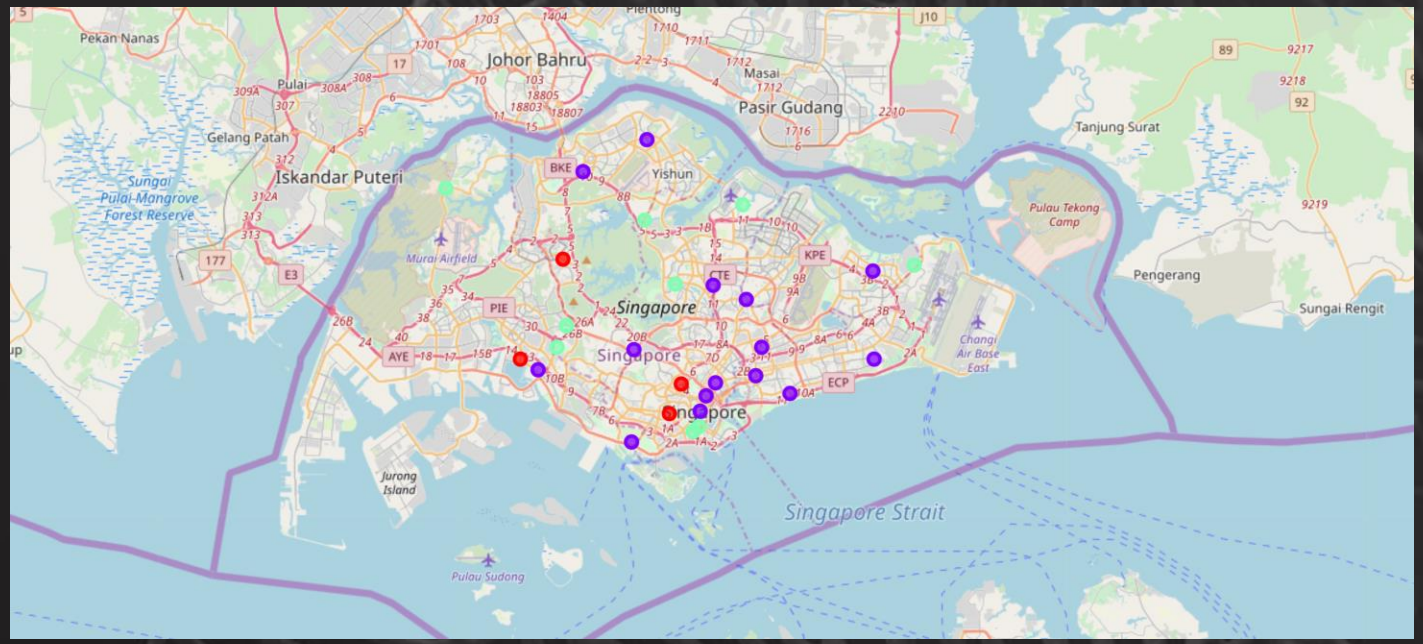
- China accounted for 44 percent of total global shopping mall completions in 2014, according to CBRE, with more mall space added in the central city of Wuhan alone (993,000 square meters) than in the whole of the Americas (800,000 sqm). So how many malls we have in Singapore and where are they located?

IBM Capstone Project by Coursera

- Procedure.
- We need the following data: a list of neighborhoods of Singapore.
- I will use this webpage (https://en.wikipedia.org/wiki/Postal_codes_in_Singapore) to get the data.
- Scraping with BeautifulSoup
- Data frames building with pandas
- Vectors functions with numpy
- API requests sent to Foursquare
- Machine learning method as KMeans (clustering) applied
- Map visualization with folium.

IBM Capstone Project by Coursera

- The MAP visualized



IBM Capstone Project by Coursera

- 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

IBM Capstone Project by Coursera

- 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.

IBM Capstone Project by Coursera

- 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 in come per area and costs of housing per neighborhood.

- These factors are important in such a difficult choice.

IBM Capstone Project by Coursera

- CONCLUSION.

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

THE END THANKS FOR REVIEWING