COURSERA CAPSTONE

IBM APPLIED DATA SCIENCE CAPSTONE

Opening a café near train stations (MRT) and shopping malls in Singapore

By: Lee Jun Yi

February 2020

BUSINESS PROBLEM

- The objective of this Capstone project is to enable users to select the best locations in the city of Singapore for opening a shop in the vicinity of a shopping mall beside the train station.
- Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to the business need of finding the most suitable location with high human traffic in Singapore.
- In essence, the question is : where should a business owner in Singapore open a shop?

DATA

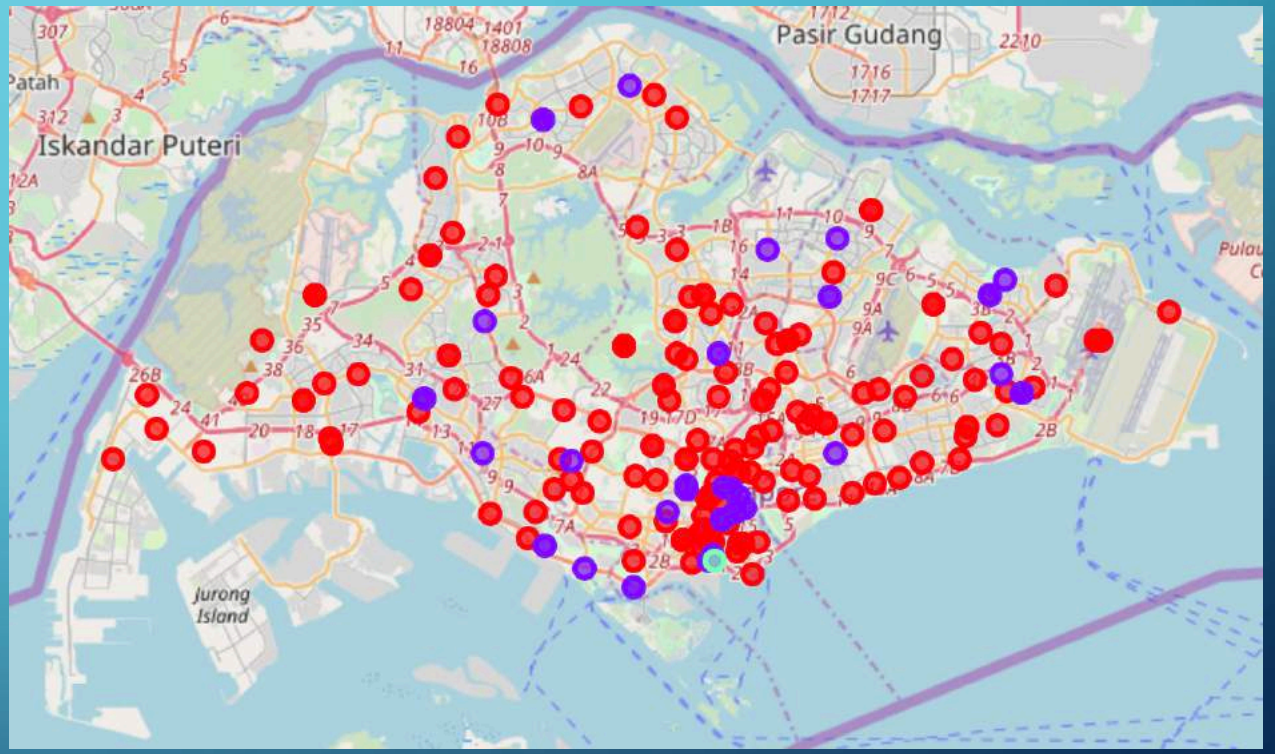
- Data required
  - List of train stations in Singapore
  - Latitude and longitude coordinates of every station — to plot the map and also get the venue data
  - Venue data, relating to shopping malls – to perform clustering of the neighborhoods around the train stations
- Sources of data
  - Wikipedia page for train stations in Singapore (https://en.wikipedia.org/wiki/List_of_Singapore_MRT_stations)
  - Geocoder package for latitude and longitude coordinates
  - Foursquare API for venue data

METHODOLOGY

- Web scraping Wikipedia page for neighbourhoods list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data
- Group data by neighbourhood and taking the mean of the frequency of occurrence of each venue category
- Filter venue category by Shopping Mall
- Perform clustering on the data by using k-means clustering
- Visualize the clusters in a map using Folium

RESULTS

- Categorized the neighbourhoods into 3 clusters :
  - Cluster 0: very few to no shopping mall near the train stations
  - Cluster 1: moderate number of shopping malls near the train stations
  - Cluster 2: high concentration of shopping malls near the train stations



DISCUSSION

- Most of the shopping malls are concentrated in the city at the South area of Singapore
- There is highest concentration of shopping malls near train stations in Cluster 2, followed by moderate number in Cluster 1
- Cluster 0 has very few to no shopping mall near the train stations

RECOMMENDATIONS

- Areas in Cluster 1 presents great opportunity and high potential for opening new cafes as there is very little to no competition from existing malls, which typically would have coffee chains and fastfood outlets to serve the needs
- Therefore, this project recommends business owners who want to open a cafe to tap on these findings to select a location that fits their cafe's target segment in cluster 1 with moderate competition, yet benefitting from the convenience and traffic of being close to a mall
- Shops in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From a business perspective, a cafe startup should not choose this area unless the cafe has extremely appealing Unique Selling Point
- Cafe owners with unique themes or menu can also open at the areas in Cluster 0 with minimal competitions and gain first-mover advantage at the neighborhood

CONCLUSION

- Answering the business problem: The neighborhoods near the train stations in Cluster 1 are the most preferred locations to open a new cafe
- Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shop
- We can observe the higher concentration of shopping malls in the central to southern area of the country, which signifies the key office areas that coincides with the Central Business District having the heaviest traffic on average daily
- Lastly, this finding can also help business owners who have a shopfront experiencing intense competition to find alternatives for relocation in other area

THANK YOU!