Capstone Project - The Battle of Neighbourhoods Introduction

Introduction where you discuss the business problem and who would be interested in this project.

"Would you recommend a location in Hong Kong to open a new cinema?"
My boss, the stakeholder wants to open a new cinema as company's new business.

He explains that watching movie is a part of whole afternoon or night activities. Cinema should has many restaurants and shopping places nearby. Transportation is also an important factor. Customer can walk to cinema within 5 minutes from public transport facilities is perfect.

He wants me concentrated on selection of cinema location according to its nearby environment. Cinema facility and rental price is not my concern. He lists out his top 10 favorite cinemas in Hong Kong with rating.

I work with my teammates and select 5 possible locations to build the cinema. Which location should be suggested to the stakeholder?

Data

Data where you describe the data that will be used to solve the problem and the source of the data.

According to the question, following data are required.

1. Geographic coordinate of Hong Kong cinemas

I need to compare 5 possible locations with current cinemas in Hong Kong. Therefore, I need to find a list of Hong Kong cinema and cinemas' geographic coordinates. Luckily, I can find the list and coordinates from the website https://hkmovie6.com/cinema .

2. Geographic coordinates of 5 possible cinema addresses

Geographic coordinates of 5 possible cinemas are required and I can use Google Map API to find this information

Data-frame of 5 target locations with geographic coordinates information
Location AddressLatitude Longitude

- L1
- L2
- L3
- L4
- L5

Sau Mau Ping Shopping Centre, Sau Mau Ping
Tuen Mun Ferry, Tuen Mun 22.371780
Un Chau Shopping Centre, Cheung Sha Wan
Prosperity Millennia Plaza, North Point 22.291698 114.208168 Tsuen Fung Centre Shopping Arcade, Tsuen Wan 22.372112

22.319503 113.966039 22.337280

114.232187 114.156457

114.119317

3. Favourite cinema list of stakeholder

The favourite cinema list of stakeholder is an important information that I can use it as profile to select the best location.

Name Rating

- Broadway Circuit - MONGKOK 4.5
- Broadway Circuit - The ONE 4.5
- Grand Ocean 4.3

- The Grand Cinema 3.4
- AMC Pacific Place 2.3
- UA IMAX @ Airport 1.5

4. Eating, Shopping and Public transportation facility around cinema

The recommended cinema location needs to have many eating and shopping venues nearby. Convenient public transport is also required.

These data can be found by using FourSquare API to find these venues around the location. The radius of exploration distance is set to 500 meters, which is about 5 minutes walking distance.

Following type of venue category will be used to search

Methodology
Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

With above data, I can use content-based recommendation technique to resolve the problem.

Combine with FourSquare API which provides how many venues in different category of Hong Kong cinemas, a matrix which captured characteristic of venues nearby cinema are built. Stakeholder's favourite list is the profile to combine with the matrix to become a weighted matrix of favourite cinema.

The weighted matrix can be applied on 5 target locations with venues information to generate a ranking result. The the top one on the ranking list can be recommended to the stakeholder.

Before building the matrix, I have to prepare the required data and apply some data analysis.

Machine Learning

Now, let's use Content-Based or Item-Item recommendation systems. In this case, I am going to try to figure out the boss's favourite new cinema location by counting number of nearby venues and ratings given.

Results
Results section where you discuss the results.

With the boss's profile and the complete list of cinemas and their venues count in hand, I am going to take the weighted average of every location based on the profile and recommend the top location that most satisfy it.

Discussion

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

I should contact local commercial property agents to find more suitable locations. Moreover, FourSquare is not popular in Hong Kong, the data maybe out-dated or unreliable, the report should gather more data from other location data source such as Google Place API.

Conclusion
Conclusion section where you conclude the report.

The stakeholder's problem is resolved. Stakeholder wants to find the best place to build a new cinema in Hong Kong, and the factors of "best location" is based on the number of venues in eating, shopping, transportation category around the location. Stakeholder also provide his favorite list of cinema to further explain what the "best location" is. Content-based filtering machine learning technique is the most suitable method to resolve the problem. It combines stakeholder's preference and cinema profile to make the recommendation result.

The 5 target locations of new cinema may not be a good choices. As the weighting matrix is developed, I can quickly pick other locations and make the recommendation again.