Capstone Project - The Battle of Neighborhoods

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 .

There are 68 cinemas in Hong Kong

First five records of Hong Kong cinemas

	Name	ChiName	AddressLatitude	Longitude
0	Emperor Cinemas - Entertainment Building	英皇戲院 - 娛樂行	3/F, Emperor Cinemas Entertainment Building, 3...	22.281453	114.154230
1	The Coronet @ Emperor Cinemas - Entertainment ...	The Coronet @ 英皇戲院 - 娛樂行	3/F, Emperor Cinemas Entertainment Building, 3...	22.281453	114.154230
2	Emperor Cinemas - Tuen Mun	英皇戲院 - 屯門新都商場	3/F, New Town Commercial Arcade, 2 Tuen Lee St...	22.390776	113.975983
3	Broadway Circuit - CYBERPORT	百老匯戲院 - 數碼港	Shop L1 - 3, Level 1, The Arcade, 100 Cyberpor...	22.261067	114.129825
4	Broadway Circuit - PALACE IFC	百老匯戲院 - PALACE IFC	Podium L1, IFC Mall, 8 Finance Street, Central	22.285545	114.157979

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

Dataframe of 5 target locations with geographic coordinates information

	Location	AddressLatitude	Longitude
0	L1	Sau Mau Ping Shopping Centre, Sau Mau Ping	22.319503	114.232187
1	L2	Tuen Mun Ferry, Tuen Mun	22.371780	113.966039
2	L3	Un Chau Shopping Centre, Cheung Sha Wan	22.337280	114.156457
3	L4	Prosperity Millennia Plaza, North Point	22.291698	114.208168
4	L5	Tsuen Fung Centre Shopping Arcade, Tsuen Wan	22.372112	

114.119317

3. Favorite cinema list of stakeholder

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

	Name	Rating
0	Broadway Circuit - MONGKOK	4.5
1	Broadway Circuit - The ONE	4.5
2	Grand Ocean	4.3
3	The Grand Cinema	3.4
4	AMC Pacific Place	2.3
5	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 favorite list is the profile to combine with the matrix to become a weighted matrix of favorite 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 favorite 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 lcoation 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.