Capstone Project – IBM Course

Dec 2018

# Introduction

My boss wants to purchase/invest in a new cinema, and undertake an analysis of potential cinema location according to its nearby environmental factors. Cinema facility and rental price are not part of the study. He lists out his top 10 favourite cinemas in Hong Kong with rating to guide me with his preferences.

He explains that watching movie is a part of whole afternoon or night activities. Cinema should be close to a high number of restaurants and shopping places. 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 should not be part of this study at the moment.

# Data Sourced

## 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.

## 3. Favourite cinema list of boss

The favourite cinema list of boss is an important information to be used for analysis.

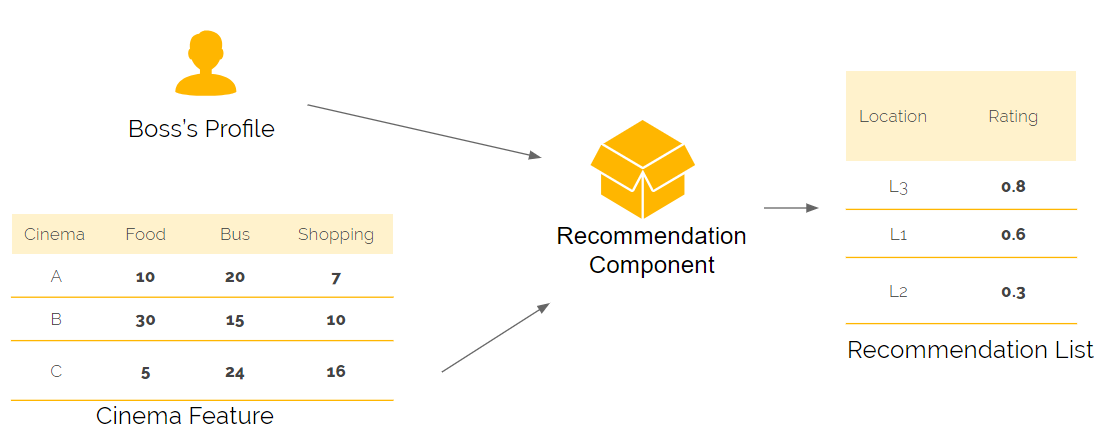
## 4. Eating, Shopping and Public transportation facility around cinema

The recommended cinema location needs to have as many eating and shopping venues nearby as possible. Convenient public transport is also an important factor.

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.

# Methodology

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

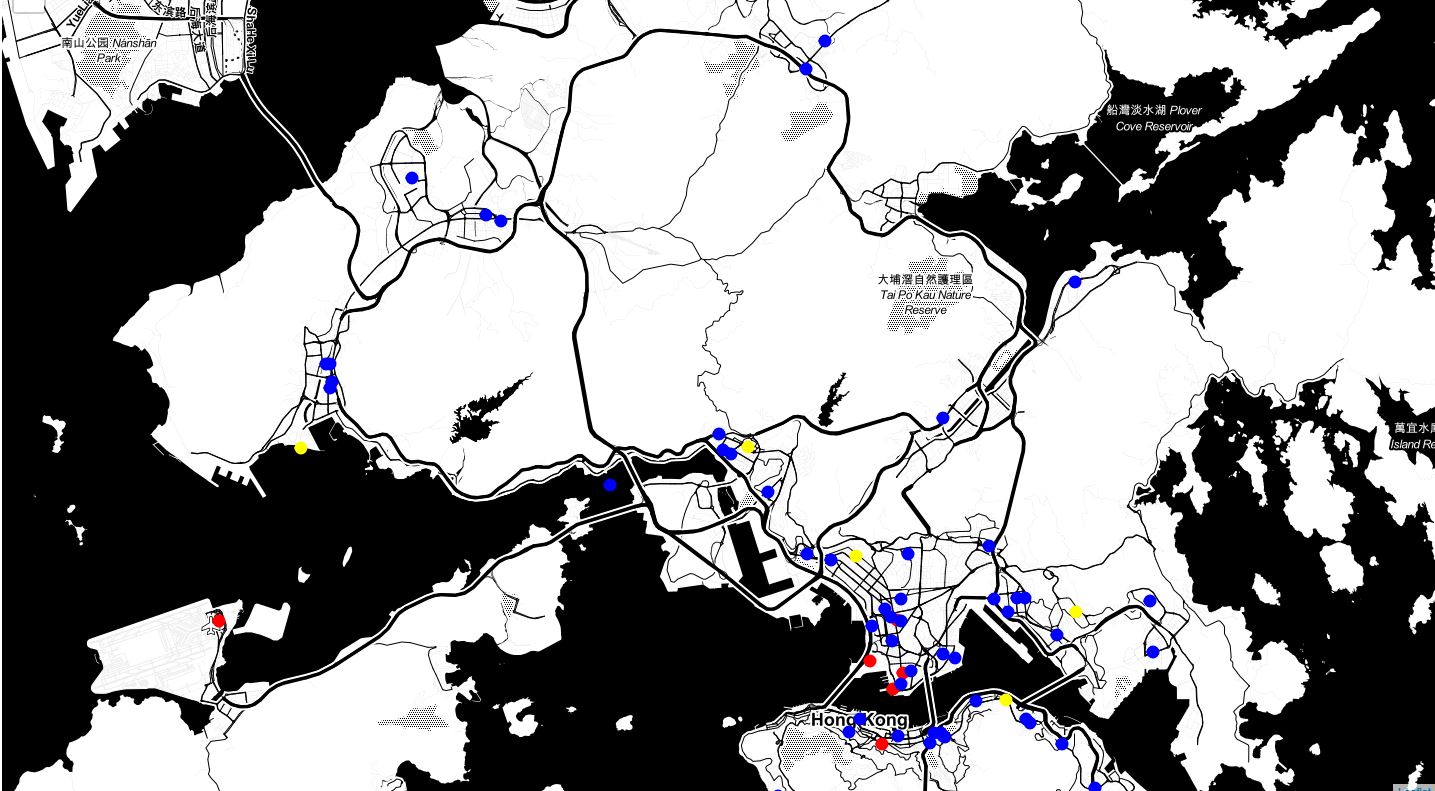


The detailed methodology and the processes of data analytics can be found at my analytics file (https://github.com/Edthenerdy/IBMcap/blob/master/Capstone%20Project%20-%20The%20Battle%20of%20Neighborhoods%20-%20Final.ipynb). In summary, I

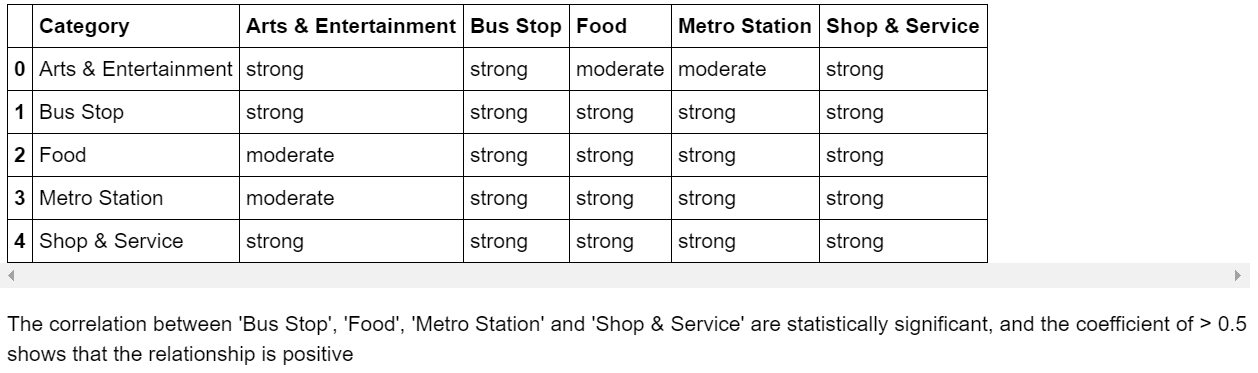
* Prepared and cleansed the required data and applied some preliminary data techniques to understand the cinema data better.
* Used the FourSquare API to explore nearby venues of Hong Kong cinemas
* Imported the favourite cinemas of my boss and identify key aspects of his preferences using FourSquare API data.
* Identified Bus Stop', 'Food', 'Metro Station' and 'Shop & Service' are statistically significant preferences
* Used Sklearn (MinMaxScaler) to find a recommended cinema new cinema location by counting number of nearby venues and ratings given and assigning a score based on a weighted score on cinema feature (food, bus and shopping locations) and boss’s preference.

# Results and Discussions

Most of Hong Kong cinemas (blue circle) and boss's favourite cinemas (red circle) location are built near main road, and centralised in urban area of Hong Kong. The target locations (yellow circle) of new cinema are not near to main road.



The result is reasonable. Location "L5" has the most number of venues in category "Bus Stop", "Food", "Metro Station" and "Shop & Service".



Therefore, Location "L5" should be recommended to the boss.

However, it appears that FourSquare is not popular in Hong Kong and hence the data maybe out-dated or unreliable.

Further location data sources such as Google Place API may help improve the results and accuracy of the analyses.

# Conclusion

A recommendation has been put forward to the boss, based on his inherent preference of cinemas (as per the analysis, the number of venues in eating, shopping, transportation category near the cinema).

This technique to identify individual preferences (i.e. they choices) and using these preferences to identify inherent traits for choices seems to work well in my organisation for analysis. As per previous sessions, the data sources are probably outdated and any new data may help improve the accuracy of the results.

The technique is also blunt. It doesn’t take into account other factors that influence individual choices (e.g. where the boss lives, and his lifestyle factors), which may impact on his choices. A detailed, more qualitative and quantitative research may reveal further preferences and allow us to recommend a commercial location for a cinema better.

In addition, this analysis doesn’t take into account any commercial aspect of a decision (e.g. rent, commercial availability..etc).