

# Find Location to Build New Movie Theater in Chennai, India

Capstone Project - The Battle of Neighborhoods

## Introduction

**Recommend a location in Chennai to open a new movie theatre.**

My Client, the stakeholder wants to open a new cinema as his company's new business.

He explains that watching movie is a part of whole afternoon or night activities. Cinema should have 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 requested to compare against top favorite movie theatres with foursquare online traffic (tips & rating).

He selected 5 possible locations to build the cinema. Which location should be suggested to the stakeholder?

Options	Area
Option1	George Town, Chennai
Option2	Anna Nagar, Chennai
Option3	Nandanam, Chennai
Option4	Nungambakkam, Chennai
Option5	Saligramam, Chennai

# Data

## Geographic coordinate of Chennai movie theatres

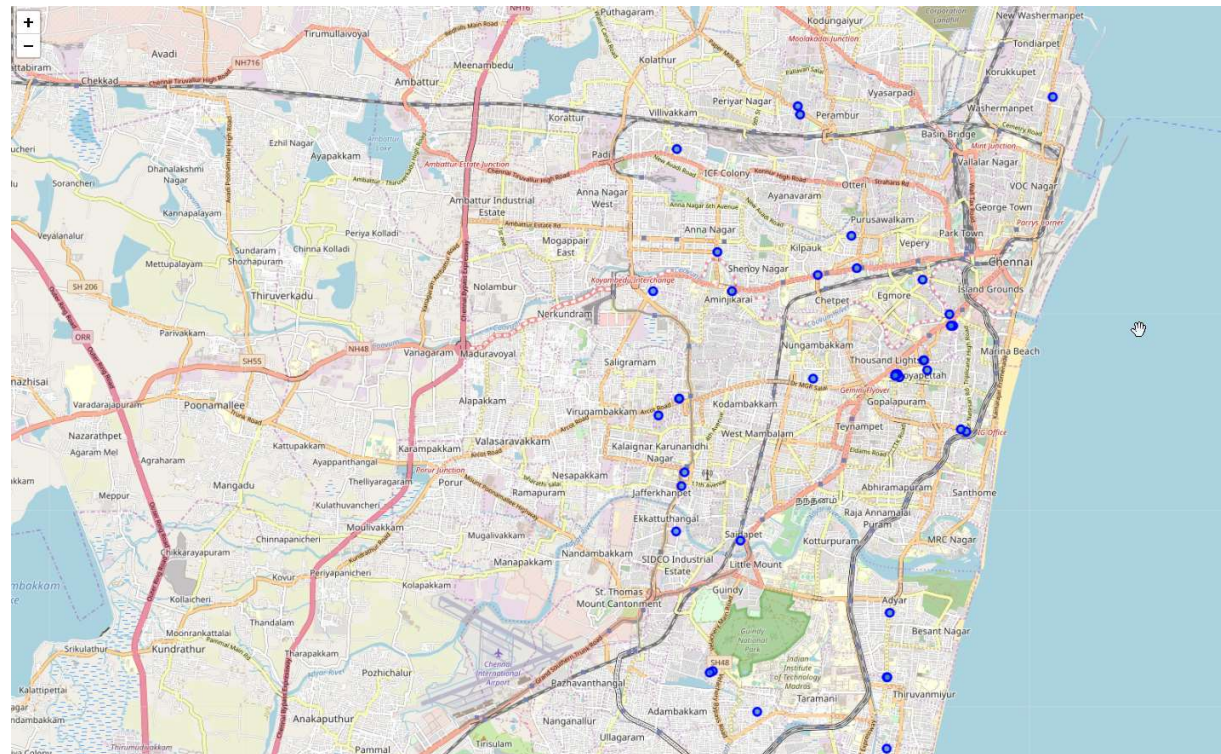
I need to compare 5 possible locations with current cinemas in Chennai. Therefore, I need to find a list of Chennai movie theatre and its geographic coordinates. I have extracted this data using Foursquare API, by search query 'Movie'.

Here is the list of Movie theatres (sample) in Chennai,

*The geographical coordinate of Chennai are 13.0518752, 80.2830143027039.  
43 Movie Theatres were returned by Foursquare.*

name	categories	address
Escape Cinemas	Multiplex	Express Avenue
Sathyam Cinemas	Movie Theater	8, Thiru-vi-ka Road
Six Degree	Movie Theater	Sathyam Cinemas
Carve at Escape Cinemas	Movie Theater	Express Avenue
Devi Cineplex	Multiplex	Mount Road

Marked the movie theatres in Chennai map,



### 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

Options	Area	Latitude	Longitude
Option1	George Town, Chennai	13.092316	80.282559
Option2	Anna Nagar, Chennai	13.089136	80.209562
Option3	Nandanam, Chennai	13.030105	80.241425
Option4	Nungambakkam, Chennai	13.059537	80.242479
Option5	Saligramam, Chennai	13.054605	80.199798

### Get the list of Popular Movie theatres

By getting the ratings & tip counts for all the movie theatres, I can **use it as profile to select the best location.**

name	address	lat	lng	rating	tips
Sathyam Cinemas	8, Thiru-vi-ka Road	13.055688	80.258096	9.3	197
Escape Cinemas	Express Avenue	13.058746	80.264170	9.2	112
Luxe Cinemas	Phoenix Market City	12.991041	80.216962	8.8	96
PVR	Ampa Skywalk Mall	13.073799	80.221392	7.0	66
S2 Cinemas	L.B. Road,	12.989660	80.256037	6.5	51

**Merging with original list & selecting the top theatres based on rating & tips count,**

name	rating	tips
Sathyam Cinemas	9.3	197
Escape Cinemas	9.2	112
Luxe Cinemas	8.8	96
PVR	7.0	66
S2 Cinemas	6.5	51
Palazzo	8.8	17

## 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 1000 meters, which is about 10 minutes walking distance.

Following type of venue category will be used to search,

**College, Juice, Bus Stop, Food, Coffee, Restaurant**

Let's use Foursquare API to query some venues around the first cinema ('Sathyam Cinemas') in the cinema list. Only first 5 venues in the category are shown below.

- Restaurant

Name	Latitude	Longitude
Sowbakiya Fast Food & Restaurant	13.057573	80.262873
Krishna Restaurant - Woodlands	13.045185	80.262107
Kurinchi restaurant	13.050201	80.262098
مطعم الزيتون Zaitoon restaurant	13.059848	80.251993
Sea Shell Family Restaurant	13.057127	80.255069

- Bus Stop

Name	Latitude	Longitude
Thousand Light Bus Stop	13.056020	80.255940
Thousand Light Mosque Bus Stop	13.054891	80.254884
Royapettah Police Station Bus Stop	13.051099	80.263225
Semmozhi Poonga Bus Stop	13.050546	80.252385
Gemini Bus Stop	13.052785	80.250697

- College

Name	Latitude	Longitude
New College	13.053608	80.260541
Meston College of Education	13.058036	80.263504
Ethiraj College For Women	13.063878	80.259383
college side juice shop	13.054087	80.261276
Indian Overseas Bank, The New College	13.055510	80.259561

## Methodology

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 Chennai theatres, a matrix which captured characteristic of venues nearby cinema are built. Popular theatres 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 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.

### Data Cleansing and Preparation

Some duplicated cinemas are found. These cinemas are some "Special House" inside Multiplexes or listed in different names and should be removed

Sample Duplicate records,

	name	address	lat	lng	rating	tips
	Escape Cinemas	Express Avenue	13.058746	80.264170	9.2	112
	Carve at Escape Cinemas	Express Avenue	13.058681	80.264379	7.2	1
	Luxe Cinemas	Phoenix Market City	12.991041	80.216962	8.8	96
	IMAX®	Phoenix Market City	12.990639	80.216310	7.6	3
	Six Degree	Sathyam Cinemas	13.055605	80.258115	7.6	3
	Seasons	Sathyam Cinemas	13.055444	80.257860	6.5	3
	Santham	Sathyam Cinemas	13.055136	80.258913	NaN	0
	Studio 5	Sathyam Cinemas	13.055472	80.257928	NaN	0

Finally, there are 24 cinemas in Chennai will be used to resolve the problem.

Display first 5 records of Chennai Cinemas after cleansing.

name	address	lat	lng	rating	tips
Sathyam Cinemas	8, Thiru-vi-ka Road	13.055688	80.258096	9.3	197
Escape Cinemas	Express Avenue	13.058746	80.264170	9.2	112
Luxe Cinemas	Phoenix Market City	12.991041	80.216962	8.8	96
PVR	Ampa Skywalk Mall	13.073799	80.221392	7.0	66
S2 Cinemas	L.B. Road,	12.989660	80.256037	6.5	51

Now I can use the Foursquare API to explore nearby venues of Chennai cinemas. Total 1128 venues are found from Foursquare.

Display first 5 records as example,

Category	Cinema Name	Latitude	Longitude	Name
College	Sathyam Cinemas	13.053608	80.260541	New College
College	Sathyam Cinemas	13.058036	80.263504	Meston College of Education
College	Sathyam Cinemas	13.063878	80.259383	Ethiraj College For Women
College	Sathyam Cinemas	13.054087	80.261276	college side juice shop
College	Sathyam Cinemas	13.055510	80.259561	Indian Overseas Bank, The New College

Number of venues in each category,

	Count
<b>Restaurant</b>	274
<b>Food</b>	216
<b>Bus Stop</b>	209
<b>Coffee</b>	164
<b>Juice</b>	139
<b>College</b>	126

Explore nearby venues of 5 possible/target locations.

	Count
<b>Restaurant</b>	70
<b>Food</b>	50
<b>Bus Stop</b>	47
<b>Coffee</b>	35
<b>Juice</b>	27
<b>College</b>	21



Finally, 1128 venues are found in Chennai cinema list and 83 venues are found in target locations list.

Category	Bus Stop	Coffee	College	Food	Juice	Restaurant
Options						
Option1	7.0	1.0	2.0	4.0	5.0	10.0
Option2	8.0	11.0	5.0	18.0	6.0	23.0
Option3	5.0	5.0	3.0	8.0	3.0	15.0
Option4	17.0	13.0	11.0	10.0	8.0	16.0
Option5	10.0	5.0	0.0	10.0	5.0	6.0

## Data Analysis

Check the data type of variables

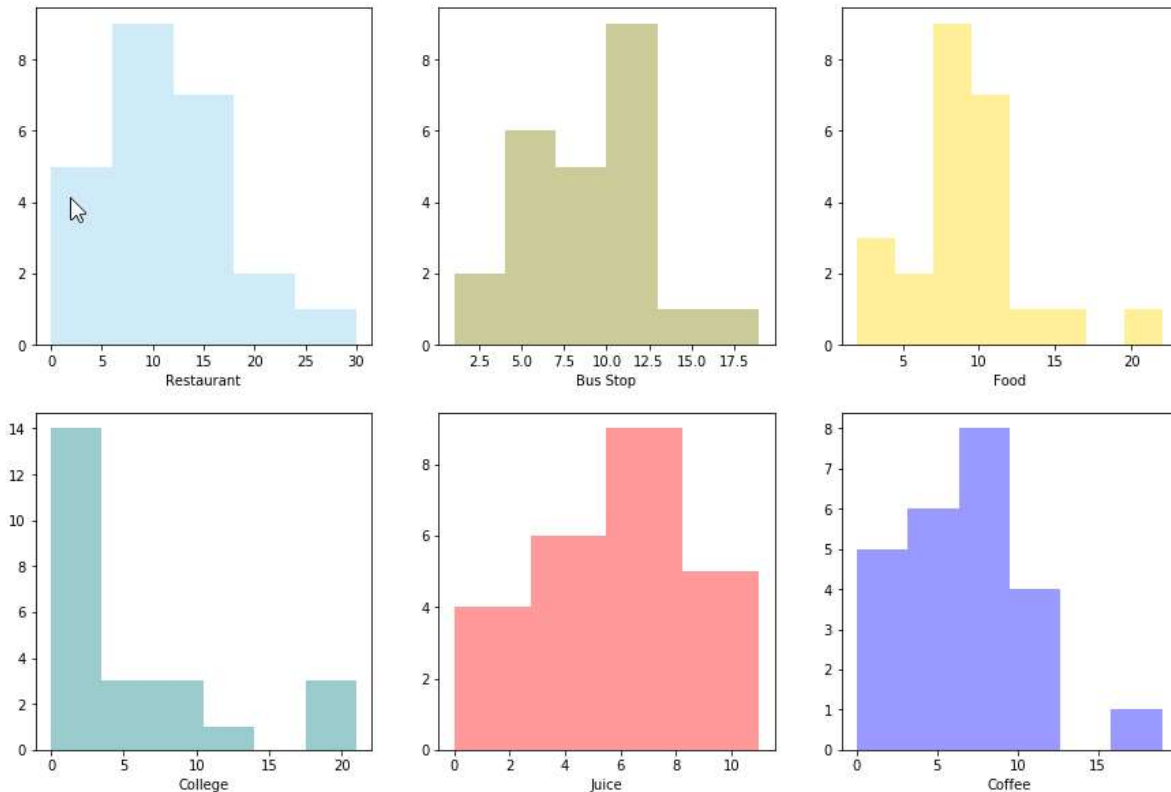
Data Type	
Category	
Bus Stop	float64
Coffee	float64
College	float64
Food	float64
Juice	float64
Restaurant	float64

All datatype is numeric

Generates descriptive statistics that summarize the central tendency, dispersion and shape of a dataset's distribution

Category	Bus Stop	Coffee	College	Food	Juice	Restaurant
count	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000
mean	8.708333	6.833333	5.250000	9.000000	5.791667	11.416667
std	4.196056	4.018075	6.066659	4.075377	2.873858	7.156126
min	1.000000	0.000000	0.000000	2.000000	0.000000	0.000000
25%	6.000000	4.000000	1.000000	7.000000	4.000000	6.750000
50%	8.000000	7.000000	3.000000	9.000000	6.000000	10.000000
75%	11.250000	9.000000	7.000000	11.000000	8.000000	17.000000
max	20.000000	19.000000	21.000000	22.000000	11.000000	30.000000

Plot the distribution of all the variables,



The distribution of all the variables are quite similar. Now check their **Pearson Correlation**

Category	Bus Stop	Coffee	College	Food	Juice	Restaurant
Category						
<b>Bus Stop</b>	1.000000	0.712231	0.518184	0.574541	0.526192	0.649852
<b>Coffee</b>	0.712231	1.000000	0.699183	0.615991	0.546583	0.763099
<b>College</b>	0.518184	0.699183	1.000000	0.298954	0.494391	0.620420
<b>Food</b>	0.574541	0.615991	0.298954	1.000000	0.389788	0.642545
<b>Juice</b>	0.526192	0.546583	0.494391	0.389788	1.000000	0.465282
<b>Restaurant</b>	0.649852	0.763099	0.620420	0.642545	0.465282	1.000000

It seems that 'Bus Stop', 'Coffee' & Restaurant' category are highly correlated.  
Find **P-Value** of the variables

By convention, when the p-value is:

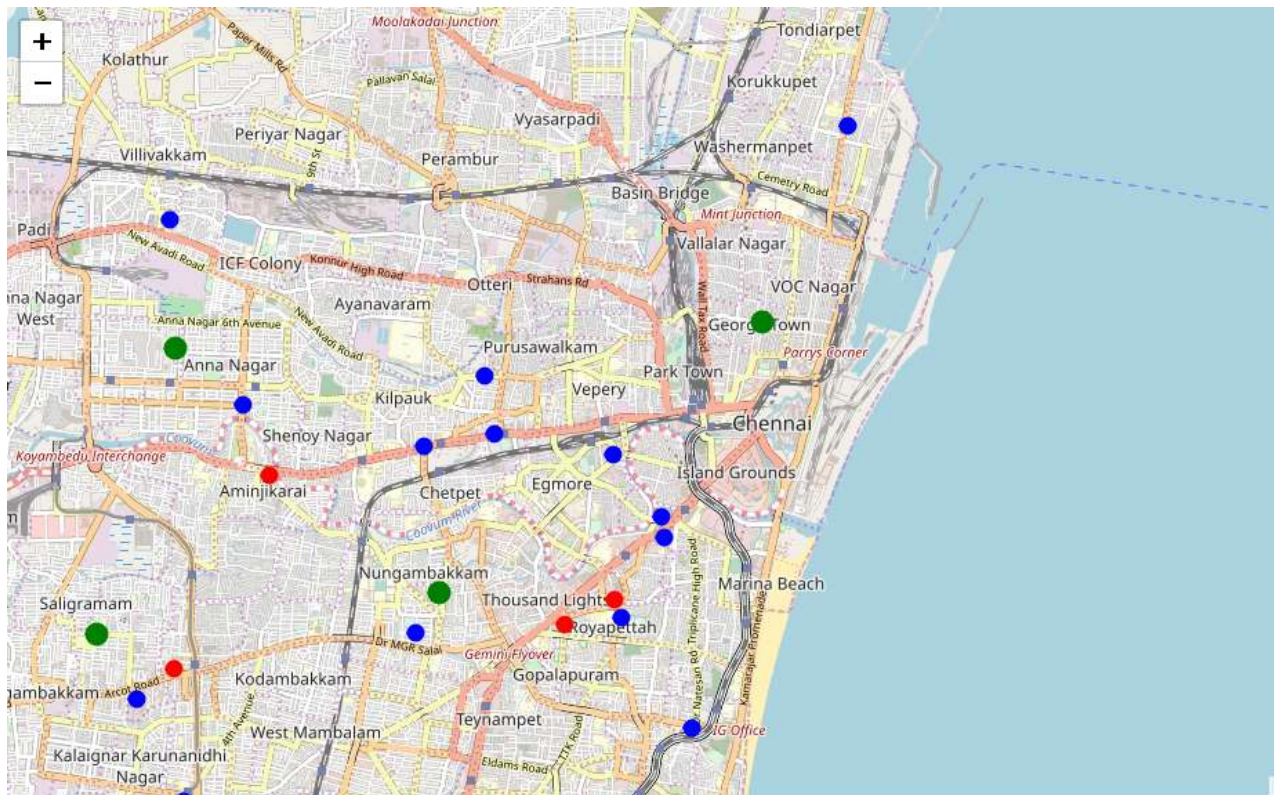
- $< 0.001$  we say there is strong evidence that the correlation is significant,
- $< 0.05$ ; there is moderate evidence that the correlation is significant,
- $< 0.1$ ; there is weak evidence that the correlation is significant, and
- is  $> 0.1$ ; there is no evidence that the correlation is significant.



Category	Bus Stop	Coffee	College	Food	Juice	Restaurant
Bus Stop	strong	strong	moderate	moderate	moderate	strong
Coffee	strong	strong	strong	moderate	moderate	strong
College	moderate	strong	strong	no	moderate	moderate
Food	moderate	moderate	no	strong	weak	strong
Juice	moderate	moderate	moderate	weak	strong	moderate
Restaurant	strong	strong	moderate	strong	moderate	strong

The correlation between 'Bus Stop', 'Coffee' & Restaurant' are statistically significant, and the coefficient of  $> 0.5$  shows that the relationship is positive

**Visualize the location of cinemas, target location and top cinemas on the map,**



Most of Chennai movie theatres (blue circle) and popular cinemas (red circle) location are built near main road, and centralized in urban area of Chennai. The target locations (green circle) of new cinema.

## 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 new cinema location by counting number of nearby venues and ratings given.

- **Normalize** the values of venues dataframe by using MinMaxScaler method and display the first 5 records

	Category	Bus Stop	Coffee	College	Food	Juice	Restaurant
Cinema Name							
	AGS Navalur	0.333333	0.421053	0.095238	0.65	0.454545	0.566667
	AGS Royal Cinemas	0.111111	0.052632	0.000000	0.00	0.181818	0.166667
	Abirami Mega Mall	0.333333	0.210526	0.190476	0.25	0.545455	0.233333
	Albert Theatre	0.222222	0.368421	0.047619	0.40	0.545455	0.700000
	Avm Rajeswari	0.388889	0.526316	0.047619	0.35	0.636364	0.266667

- **Top Theatres** - Join above data with favorite list.

Bus Stop	Coffee	College	Food	Juice	Restaurant
1.000000	1.000000	1.000000	0.55	0.909091	1.000000
0.500000	0.526316	0.857143	0.40	0.818182	0.566667
0.277778	0.473684	0.095238	0.25	0.272727	0.333333
0.277778	0.315789	0.238095	0.20	0.545455	0.100000
0.777778	0.263158	0.000000	0.45	0.545455	0.400000
0.500000	0.473684	0.095238	0.45	1.000000	0.300000

- **Recommendation Component** - Dot product to get the weight of rating on each category according to stakeholder's favorite list.

Category	Weighted Rating
Bus Stop	27.744444
Coffee	26.4
College	20.528571
Food	19.28
Juice	34.545455
Restaurant	23.386667

- **Normalize** the values of target venues by using same MinMaxScaler

Category	Bus Stop	Coffee	College	Food	Juice	Restaurant
Options						
Option1	0.333333	0.052632	0.095238	0.1	0.454545	0.333333
Option2	0.388889	0.578947	0.238095	0.8	0.545455	0.766667
Option3	0.222222	0.263158	0.142857	0.3	0.272727	0.500000
Option4	0.888889	0.684211	0.523810	0.4	0.727273	0.533333
Option5	0.500000	0.263158	0.000000	0.4	0.454545	0.200000

- **Recommendation List** - With the top theatres 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. Following table shows the estimated rating of 5 target locations

Options	Area	Latitude	Longitude	Rating
Option4	Nungambakkam, Chennai	13.059537	80.242479	0.650405
Option2	Anna Nagar, Chennai	13.089136	80.209562	0.547507
Option5	Saligramam, Chennai	13.054605	80.199798	0.322029
Option3	Nandanam, Chennai	13.030105	80.241425	0.282742
Option1	George Town, Chennai	13.092316	80.282559	0.250313

## Results

I should recommend the location "Option4" of area "Nungambakkam, Chennai" to the stakeholder. The result is reasonable. Location '**Option4 - Nungambakkam**' has the most number of venues in category "Bus Stop", "College", "Coffee" and "Juice".

Category	Bus Stop	Coffee	College	Food	Juice	Restaurant
Options						
Option1	7.0	1.0	2.0	4.0	5.0	10.0
Option2	8.0	11.0	5.0	18.0	6.0	23.0
Option3	5.0	5.0	3.0	8.0	3.0	15.0
Option4	17.0	13.0	11.0	10.0	8.0	16.0
Option5	10.0	5.0	0.0	10.0	5.0	6.0

Moreover, these categories are most concerned by the stakeholder according to profile rating,

Juice	34.545455
Bus Stop	27.744444
Coffee	26.400000
Restaurant	23.386667
College	20.528571
Food	19.280000

Therefore, Location "**Option4 - Nungambakkam**" should be recommended to the stakeholder for the new movie theatre.

## Discussion

Number of venues of 5 target locations are actually better than the average

Average count of venues in Chennai cinema location,

Average Count	
Category	
Bus Stop	8.500000
Coffee	6.833333
College	5.250000
Food	9.000000
Juice	5.791667
Restaurant	11.416667

Average count of venues in 5 target locations,

Average Count	
Category	
Bus Stop	9.4
Coffee	7.0
College	4.2
Food	10.0
Juice	5.4
Restaurant	14.0

I should contact local commercial property agents to find more suitable locations. Moreover, Foursquare is not popular in India, the data maybe outdated 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 Chennai, India, 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. Another key things to consider could be traffic & real estate value. As the weighting matrix is developed, I can quickly pick other locations and make the recommendation again.