The Battle of Neighborhood

Problem Background:

Kolkata is one of the populous city in India. It is diverse and is the capital of West Bengal (A state of the eastern part of India).It is also known as commercial and financial hub of East and North-East India. The city has an overall population of 1.49 crores as per census data of 2020. It is also major center of banking of finance, retailing, tourism, entertainment and advertising of India. Indian movie industry is well famous in the world and earn a huge revenue yearly.

As Indian movie industry market is highly competitive, the business is very risky and cost of doing business on movie is also very high. Thus, any startup on this sector needs to be analyzed carefully. The insights derived from the analysis will give a good understanding of the business environment, which helps in strategically targeting the market. This will help in the reduction of risk in business and make the business profitable.

Problem Description:

ABC Leisure Limited is a famous multiplex chain, which operate mainly in western and northern part of India, wants to start it's business in the eastern part of the country. As Kolkata is a major city of eastern part of India, the company want to start it's business from here. The target of the company is to find out a suitable place in Kolkata where it can start 1st movie theatre. If the business is successful, it will further plan to extend its market in the eastern India.

Description of Data:

The city will be analyzed in the project is Kolkata.

The area of Kolkata district is divided among various small parts, which are in the jurisdiction of different post offices. All post office area has a unique pin code, which can be used to identify the post office/the area under the post office. This dataset is available on the following link,
https://finkode.com/wb/kolkata.html

Foursquare API is used to explore other venues/competitors in the locality.

From the above web reference, we can fetch the details of all sub divisions of Kolkata city. Foursquare is used to explore the nearest venues and find out the competitors in this business.

Methodology:

In order to get a suitable place to open a movie theater, we will explore the Kolkata city (sub area wise), cluster and segment the area based on top 10 venues (according to the average number of different venue category). The following process/method is followed,

1. We import all required libraries to complete the task.
2. We scraped the data from https://finkode.com/wb/kolkata.html in order to start understanding how the city is divided among sub-areas. All sub-areas have a unique pin code which basically represent the area/the post office of this area.
3. Our next step is to store the data in pandas data frame, so that we can analyze the data easily. It seems the entire city has 81 unique pin codes, which represent 81 unique areas.

	Pincode	District	Post Office
0	700001	Kolkata	Council House Street S.O, Customs House S.O, K...
1	700002	Kolkata	Cossipore Gun Factory S.O, Cossipore H.O, Indi...
2	700003	Kolkata	Amrita Bazar Partika S.O, Baghbazar S.O, Giris...
3	700004	Kolkata	R.G.Kar Medical College S.O, Shyambazar Mail S...
4	700005	Kolkata	Ahritola S.O, Hatkhola S.O
...	...	...	...
76	700095	Kolkata	Golf Green S.O
77	700099	Kolkata	Kalikapur S.O, Mukundapur S.O
78	700107	Kolkata	E.K.T S.O
79	700108	Kolkata	ISI PO S.O
80	700139	Kolkata	Daulatpur B.O, Memanpur Chandannagar B.O

4. Next, we find the latitudes and longitudes of all areas. We have used module *pgeocode* for this purpose.

|	Pincode	District	Post Office	Latitude	Longitude
0	700001	Kolkata	Council House Street S.O, Customs House S.O, K...	22.7101	88.3167
1	700002	Kolkata	Cossipore Gun Factory S.O, Cossipore H.O, Indi...	22.7101	88.3167
2	700003	Kolkata	Amrita Bazar Partika S.O, Baghbazar S.O, Giris...	22.7101	88.3167
3	700004	Kolkata	R.G.Kar Medical College S.O, Shyambazar Mail S...	22.7101	88.3167
4	700005	Kolkata	Ahritola S.O, Hatkhola S.O	22.7101	88.3167
...	...	...	...	...	...
76	700095	Kolkata	Golf Green S.O	22.4485	88.3883
77	700099	Kolkata	Kalikapur S.O, Mukundapur S.O	22.5553	88.3558
78	700107	Kolkata	E.K.T S.O	22.9660	88.2036
79	700108	Kolkata	ISI PO S.O	22.9660	88.2036
80	700139	Kolkata	Daulatpur B.O, Memanpur Chandannagar B.O	22.1169	88.4016

5. We then use foursquare API to explore each subareas. We perform onehot coding on the dataframe and finally find out the top ten venues(calculating the average number of venue category in each area),

	Pincode	Top 1st Venue	Top 2nd Venue	Top 3rd Venue	Top 4th Venue	Top 5th Venue	Top 6th Venue	Top 7th Venue	Top 8th Venue	Top 9th Venue	Top 10th Venue
0	700001	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market	Food
1	700002	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market	Food
2	700003	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market	Food
3	700004	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market	Food
4	700005	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market	Food
5	700006	ATM	Pharmacy	Food Court	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market
6	700007	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market	Food
7	700008	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant	Field	Flea Market	Food

6. We have used here KMeans clustering to cluster and segment the areas. This is an unsupervised learning as no previous data set is available to train the model. After merging all the information and cluster label altogether, we have the following data frame,

	Pincode	District	Post Office	Latitude	Longitude	Cluster Labels	Top 1st Venue	Top 2nd Venue	Top 3rd Venue	Top 4th Venue	Top 5th Venue	Top 6th Venue	Top 7th Venue
0	700001	Kolkata	Council House Street S.O, Customs House S.O, K...	22.7101	88.316700	2	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant
1	700002	Kolkata	Cossipore Gun Factory S.O, Cossipore H.O, Indi...	22.7101	88.316700	2	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant
2	700003	Kolkata	Amrita Bazar Partika S.O, Baghbazar S.O, Giris...	22.7101	88.316700	2	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant
3	700004	Kolkata	R.G.Kar Medical College S.O, Shyambazar Mail S...	22.7101	88.316700	2	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant
4	700005	Kolkata	Ahritola S.O, Hatkhola S.O	22.7101	88.316700	2	ATM	Frozen Yogurt Shop	Department Store	Dessert Shop	Dhaba	Electronics Store	Fast Food Restaurant
5	700006	Kolkata	Beadon Street S.O, Bidhan Sarani S.O, Jorasank...	22.7101	88.322286	2	ATM	Pharmacy	Food Court	Department Store	Dessert Shop	Dhaba	Electronics Store

7. We plot the cluster in Map. Folium library is used to plot the map here.
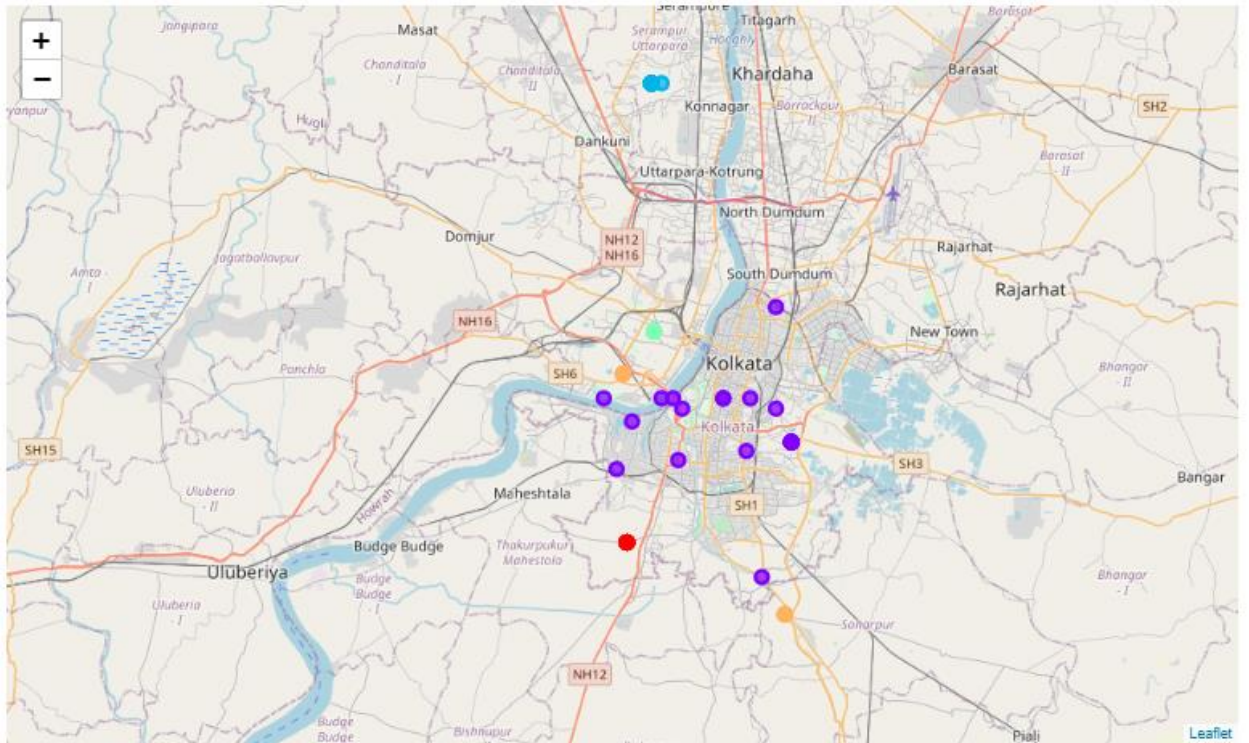
8. After clustering and segmenting the areas, we analyze each cluster to find out the best place to open a movie theater.

Results:

The number of area cluster_wise,

```
No of areas in cluster_1 7
No of areas in cluster_2 21
No of areas in cluster_3 9
No of areas in cluster_4 6
No of areas in cluster_5 9
```

Pin code of area in cluster 1 where Movie Theater is available,

```
700061
700062
700065
700066
700067
700068
700069
```

Pin code of area in cluster 2 where movie theater is available,

```
700037
```

No movie theater is available in cluster 3, cluster 4 and cluster 5.

Discussion:

The total number of areas in cluster_1 and the total number of movie theater in this cluster is same which is 7.It seems if a company starts its new business in any area on cluster_1 it will be very competitive and difficult to get the success.Cluster_2 has total 21 areas and only one movie theater exists.Cluster_3, Cluster_4 and Cluster_5 has no movie theater but the number of area covered are 9, 6, 9 respectively. The size of cluster_2 is almost 2.5 times of cluster_3 and cluster_5.Big cluster size means population is also very high, hence though cluster_2 has one movie theater it will be good to start a new theater in this cluster if we also consider cluster size and population is proportional. Second choice will be cluster_3 or cluster_5 as no movie theater exists and size is also comparatively bigger than cluster_4.Third or last choice will be cluster_4, which contains least number of area.Cluster_1, is not recommended due to high level of competition.

Conclusion:

The purpose of the project is to identify which area is in Kolkata is best to open a new movie theater for ABC Leisure Limited. As Kolkata is very high-populated city, we assume that if the size of area is big, population of this area is also high comparatively. In order to aid the company to look the optimal location of a movie theater, we used foursquare API to explore different areas of Kolkata. Final decision

on optimal location of a new movie theater will be made by stakeholders based on specific characteristics of location and other factors like convenience of each location (proximity to station), availability of nearest city center etc.