Introduction/Business Problem:

In Today's busy world, Entertainment is very important for the people in order to come out of the daily work pressure particularly in metropolitan cities. As a result of this, Multiplexes became one of the popular business opportunities for the Builders. But choosing the best location for the construction could be a challenge for the builders. This project is to address the problem and provides the inputs about the multiplexes density across the city.

Interest:

Being a demanding and profitable business there are already lot of multiplexes in the city. Hence, there is huge competition in the market for new builders to start new multiplex. The analysis done in this project would be more helpful for New Builders by providing insights of the multiplex locations data where there are already good number of multiplexes in place. Based on this analysis, new builders can make decisions on the locality to start the construction.

Data:

In order to perform the Analysis and provide solution, I have considered several data points mentioned as below.

1. Popular locations in Bangalore. Can be mentioned as Neighbors of the Bangalore.

   Data Source: Wikipidea https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore

2. Longitude and Latitudes of the locations.

   Source: From Geocoder pacakges.

3. Number of Multiplexes in the neighbors.

   Source: Extracted from Foursquare.

   https://foursquare.com/explore?mode=url&near=Bangalore%2C%20Karn%C4%81taka%2C%20India&nearGeoId=72057594039205269&q=Multiplex

4. Venues in the neighbors.

   Source: Provided by Foursquare API.

Methodology:

- **Data Scraping and cleansing:**

I extracted the data from Wikipedia page, the data is not ready for analysis and also in-sufficient. Wikipedia data shows all the areas and neighborhoods in Bangalore city. I have extracted the data from wiki page using the BeautifulSoup package and converted the data into tabular with Rows and columns which is more readable for Analysis. Just getting the location does not fully sufficient to build Machine learning algorithms. So I have brought the respective longitude and latitudes using the Geocoder package. These are required to to locate the neighborhoods in the maps.

	Neighborhood	Latitude	Longitude
0	List of areas in Bangalore Cantonment	12.97568	77.60539
1	List of areas in Bengaluru Pete	12.96053	77.64385
2	List of neighbourhoods in Bangalore	12.88813	77.57735
3	Adugodi	12.94402	77.60800
4	Agara, Bangalore	12.84283	77.48759

There are total 139 Neighborhoods provided by Wikipedia.

- **Data Analysis and Visualization:**

I used Folium library to generate the map of Bangalore with the locations extracted from Wikipedia. This is required to visualize the data extracted with locations in the Bangalore city. There are 139 data points in the map.
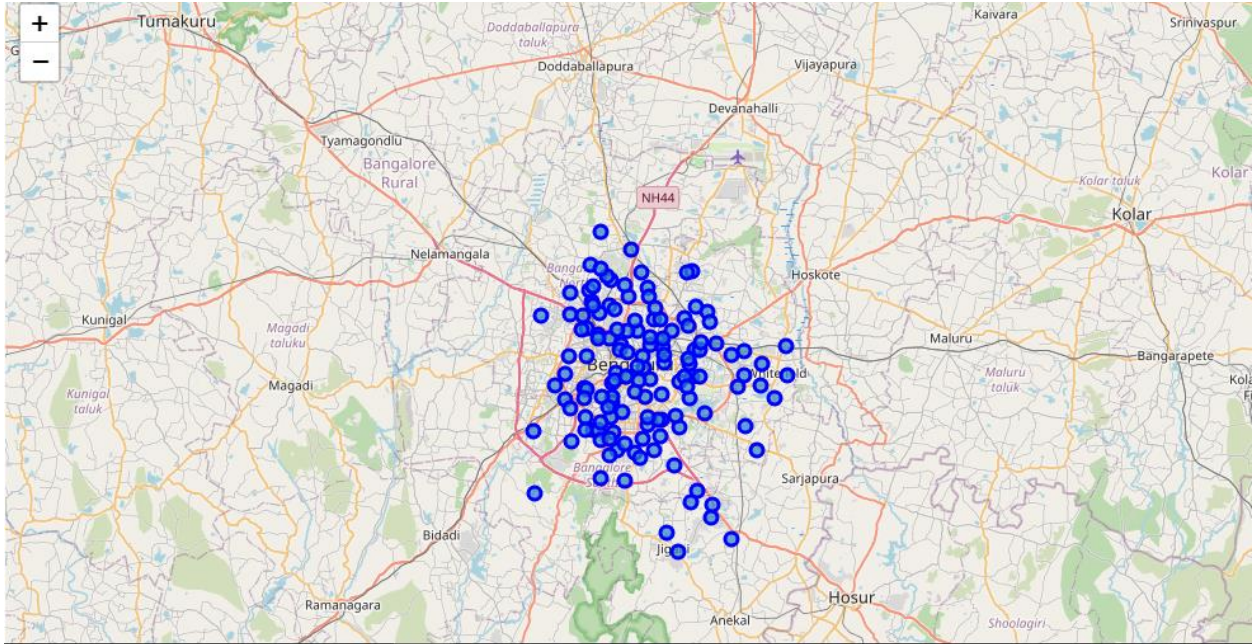
 As the idea is to find the Multiplexes in the locations, I extracted all the multiplexes from the locations in Bangalore. I used Foursquare website to get all the multiplexes with help of latitudes and longitudes of the locations. Used Foursquare API to further explore the data and API returned all the popular venues 9584 in the neighborhood.

(7345, 7)

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	List of areas in Bangalore Cantonment	12.975680	77.605390	The 13th Floor	12.975364	77.604995	Lounge
1	List of areas in Bangalore Cantonment	12.975680	77.605390	Blossom Book House	12.975042	77.604813	Bookstore
2	List of areas in Bangalore Cantonment	12.975680	77.605390	Brik Oven	12.974757	77.605441	Pizza Place
3	List of areas in Bangalore Cantonment	12.975680	77.605390	M.G Road Boulevard	12.975771	77.603979	Plaza
4	List of areas in Bangalore Cantonment	12.975680	77.605390	The Entertainment Store	12.975413	77.603045	Toy / Game Store
5	List of areas in Bangalore Cantonment	12.975680	77.605390	Matteo	12.974496	77.607115	Café
6	List of areas in Bangalore	12.975680	77.605390	Hard Rock Cafe Bengaluru	12.976389	77.601468	American Restaurant

I have further explored the data and found 209 unique categories. From the total venues data extracted from Foursquare, there are 59 Multiplexes in the data.

Number of Multiplexes: 59

- Model identification:

As I need to get the best locations for new multiplexes, I wanted to identify the areas where there are very less number of multiplexes in the city so that new multiplex in those locations would be profitable. It is good idea to divide the city neighborhoods into clusters and provide the number of Multiplexes in each cluster so that the builder can make decision in which cluster he can start the venture. To achieve this, I used K-means segmentation and Clustering Unsupervised machine learning algorithm to find the better segments for new Multiplex.

Results:

I Used the best value 3 of K to get the K-means segmentation and clustering model which resulted Bangalore neighborhoods with multiplexes as below. Created cluster labels for Neighborhoods and number of multiplexes in each cluster. The below represents number of multiplexes in a cluster of neighborhoods. For example Neighborhood Adugodi is in cluster 2 has 2 multiplexes.

	Neighborhood	Multiplex	Cluster Labels
0	Adugodi	2	2
1	Agara, Bangalore	0	1
2	Ananthnagar	2	2
3	Anjanapura	0	1
4	Arekere	4	0
5	Austin Town	0	1
6	BTM Layout	4	0
7	Babusapalya	0	1
8	Bagalur, Bangalore Urban	0	1
9	Bahubalinagar	1	2

We would like to see the result in map. Merged the longitude and latitude for each neighborhood.

	Neighborhood	Multiplex	Cluster Labels	Latitude	Longitude
138	Yeswanthpur	3	0	13.029540	77.540220
76	Koramangala	3	0	12.966200	77.649820
54	Immadihalli	3	0	13.038700	77.661920
53	Hulimavu	3	0	12.991980	77.715060
51	Hoodi	3	0	12.830840	77.680340
84	Madiwala	3	0	12.920520	77.620900
86	Mahalakshmi Layout	3	0	13.016350	77.544810
87	Malleswaram	3	0	13.006320	77.568405
90	Mathikere	3	0	13.032350	77.558660
91	Milk Colony	4	0	13.009660	77.556830
42	Garudacharpalya	3	0	13.026430	77.617510

A map is generated with the latitude and longitude information to see the clustered data. There are 3 clusters in the map, Cluster 0,Cluster 1, Cluster 2. From the above, the distinct count of multiplexes are 0,1,2,3. Higher the concentration(density) of the cluster will have the lower multiplexes.
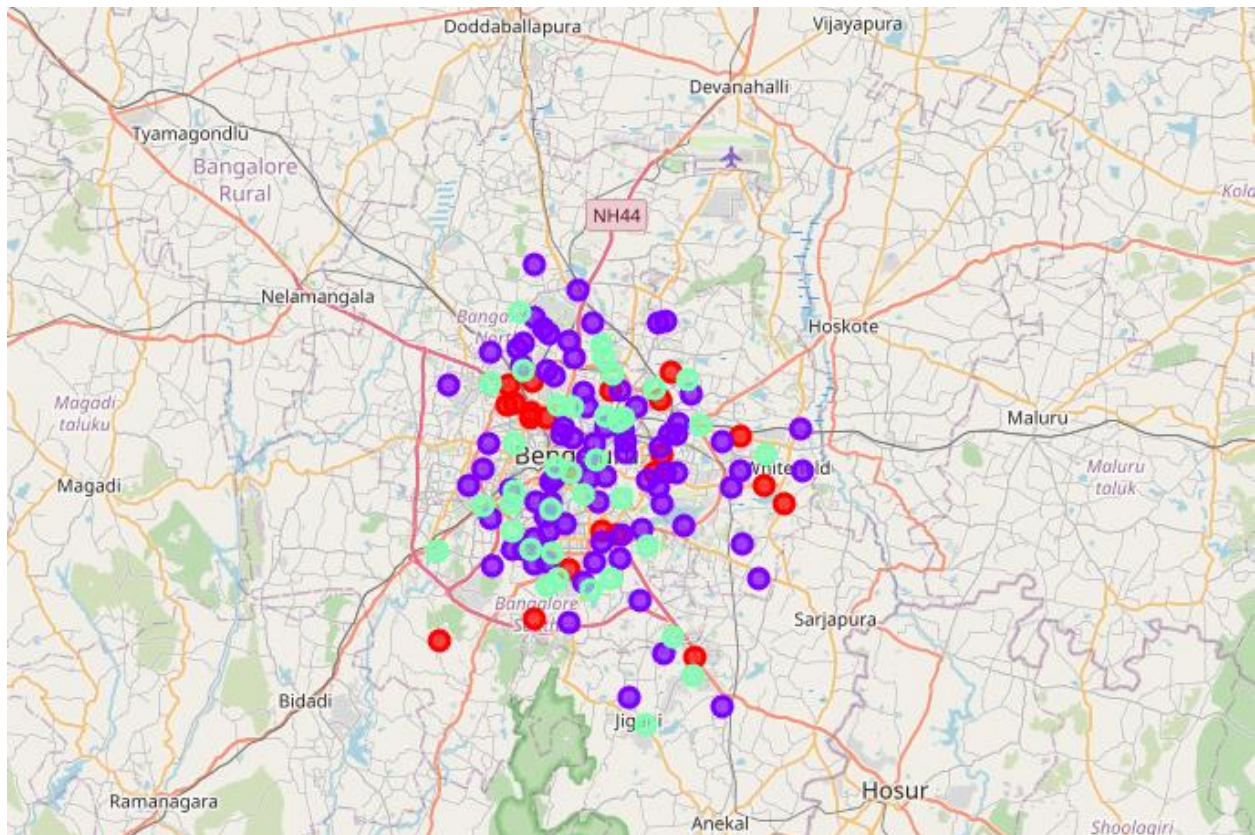
Calculated the Clusters which has 0 multiplexes,1 multiplex,2 and 3 multiplexes.

```
Multiplex  Cluster Labels
0          1                80
1          2                28
2          2                 9
3          0                16
4          0                 6
Name: Cluster Labels, dtype: int64
```

Cluster 1 has 80 data points(Neighborhoods) with 0 multiplexes.

Cluster 2 has 28 data points(Neighborhoods) with 1 multiplexes and 9 datapoints(Neighborhoods) with 2 Multiplexes

Cluster 0 has 16 data points(Neighborhoods) with 3 multiplexes and 6 Data points with 4 Multiplexes.

As per the above data, Cluster 1 has least number if multiplexes and cluster 0 has higher number of multiplexes.(74) and Cluster 2 has 18 multiplexes.

As there are very less number of Multiplexes in Cluster 1, this could be the best for new Multiplex.

Discussion:

Bangalore is a big city with a high population. There are several neighborhoods in Bangalore and population differs in each. Being a IT hub there is huge demand for Entertainment and one of them is Multiplex. As the City grows, the demand increased by should understand the better place of construction to make more profitable.

In order to analyze and come-up with better solution to start with new Multiplex, I used the K-means Segmentation and Clustering as there are many neighborhoods in Bangalore. Considered big radius 3000 to get the enough number of coverage. I set the optimum k value to 3 for number of clusters and used foursquare to fetch the information on the multiplexes data.

Provided visualization of Bangalore map with clusters which would ultimately derive which cluster among the three is best to start with the business.

Conclusion:

To conclude, there is lot of scope for business in Bangalore with respect to Multiplexes. This Analysis can be a primary and provide most of the criteria. However, we should also consider the additional parameters in the specific neighborhood. For example, though there are 80 neighborhoods with no multiplex, all the 80 neighborhoods may not be suitable considering the space, lifestyle of the people living there etc. With this analysis, not only the real estate people but also the govt can make use for study like to find out if there are any specific reasons for not building Multiplexes.