**Coursera**

**Capstone Project**

**Opening a Shopping Mall in Hyderabad, India**

**A picture containing outdoor, building, road, clock

Description automatically generated**

**Introduction:**

Shopping Malls are great places to have fun and pass time. They are like a one stop destination for all types of shoppers. Grocery shopping, Electronics, Personal items, restaurants, movie theatres are present in most of the shopping malls. Also, from a business perspective shopping mall are a great way to attract the customers as well as act as a source for new companies/ industries to advertise themselves. In Hyderabad city, there are many shopping malls at many different locations. Location is one of the most important things in establishing a shopping mall. This helps in the success or failure of the shopping malls.

The objective of this project is to select the best locations in Hyderabad to open a new shopping mall.

**Data:**

Here in this project, we are going to use the following as data:

* List of neighborhoods in Hyderabad, India.
* Geographical coordinates and data of the city Hyderabad and the surrounding neighborhoods.
* Venue data which is related to shopping malls.

We get the

List of neighborhoods data from (<https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India>). We use web scraping technique to extract the data from Wikipedia page in the form of text using python packages. We also will get the geographical data from Python Geocoder package. This package contains the latitudes and longitudes of the neighborhoods. Once we get these data, we are ready to use Foursquare API to get the venue data of the neighborhoods. In this project we are using K-means Clustering algorithm for the analysis.

**Methodology:**

Firstly we get the neighborhoods data from (<https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India>). We use web scraping technique to extract the data from Wikipedia page in the form of text using python packages. We also will get the geographical data from Python Geocoder package. This package contains the latitudes and longitudes of the neighborhoods. After doing this, we convert the data into a pandas dataframe and then visualize using the folium package.

Once we install all the packages into the machine, we use Foursquare API to get venue details. Here, we need the Client ID and Secret key from our foursquare API. Then Foursquare gives us the data which contains venue name, category, latitude and longitude. From this we can see how many venues are shown and the unique categories. In order to do clustering, we will analyze each neighborhood by grouping the rows and taking the mean of the frequency of occurrence of each venue. As we are focused on shopping malls, we use shopping mall as venue category. Now we will perform K means clustering algorithm which will identify k number of centroids and then assigns nearest clusters. It is a unsupervised machine learning algorithm which we are using in this project. Here, we are using 3 clusters for the algorithm. We can see the areas of neighborhoods which has the high or least amount of shopping malls. By this we can see which areas has the highest or least number of shopping malls and it can be used in identifying the location in establishing the new shopping mall.

**Results and Conclusion:**

As we are using 3 clusters in the K means algorithm the following results are generated.

* Cluster 0: This cluster consists the neighborhoods with moderate number of shopping malls.
* Cluster 1: This cluster contains the zero to low number of shopping malls.
* Cluster 2: This cluster contains the high number of shopping malls.

![A picture containing text, map

Description automatically generated]()

This is the visualization of the clusters in which Cluster 0 is shown by red, Cluster 1 by purple and Cluster 2 by Green color.

As we can see that there are many shopping malls in the middle of Hyderabad city. The highest number in the cluster 2 shows us the information. Also we can see from Cluster 1, it has only one number. We can see that in the cluster 2 there are many shopping malls and there will be more competition in that area. So, we can try to think about opening a new shopping mall in the region of Cluster 1 as there will be little competitors. We can also try to look at the area of Cluster 0 in addition with Cluster 1 to open in new areas.

**References:**

Category: Neighborhoods in Hyderabad. Wikipedia. Retrieved from

<https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India>

List of Shopping malls data. Retrieved from

<https://list.fandom.com/wiki/List_of_shopping_malls_in_Hyderabad>

Python documentation for packages. Python. Retrieved from

<https://docs.python.org/3/>

Foursquare documentation. Foursquare. Retrieved from

<https://developer.foursquare.com/docs/>