**Coursera**

**Capstone Project - The Battle of Neighborhoods**

* A description of the problem and a discussion of the background.

Kochi is a vibrant and upcoming city in the heart of Kerala , a state in India . It is fast developing and shows promise for great business prosperity in the food & beverage industry .

Business Group A plans to establish a coffee shop in the region and has plans to expand its reach in Kochi in the future based on its success .

In order to start its “pilot project” the company has decided to get help from a data analyst in order to decide on the location that is most profitable and suitable to set up shop .

The business problem is to find a location with prospective customers and a well established neighborhood, that at the same time is not too crowded with similar shops to establish a coffee shop . In order to determine popularity of a location ,various facilities and shops set up there are to be analyzed and studied.

This project helps by reducing manual search needed to find a location and provides the client with an easier and efficient alternative.

* Approach\Methodology :

Kochi being a fairly sized city, has to be segmented based on the popular and prominent locales. Ones the data regarding the locations is collected by scrapping the internet, it is filtered to provide uniformity. Once the core locations have been found, shops and other amenities near it are discovered. This information is analysed usinga K means Clustering technique to study various regions and provide a conclusive result that gives a suitable location to set up shop.

* Data & Tools :

Wikepedia :

https://en.wikipedia.org/wiki/Neighbourhoods\_of\_Kochi

Details about the various regions in the city was scrapped from the above mentioned page.

Foursquare API :

It is the prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

* Libraries Used:

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

* Result :

The devised result for the above mentioned problem, can be observed in the notebook attached. The result highlights the various locations in and around Kochi that have potential for setting up a coffee shop with the main competitors being cafes and also those locations which provides least chances of success. From the 4 developed clusters consisting of 38 data points choosing a region with high potential for customer in flow is essential, also giving on par importance to the competition in the region. The various shops in the clusters can be observed to see the common features of each clusters. This can be used to make informed decisions. Certain clusters have high customer inflow but would give newcomers a hard time. But these locations can be part of an expansion after successful initial deployment.

* Discussion:

The main obstacles faced was the lack of accurate data on Kochi, on the major locations and the shops nearby. With more data, a better model could be devised. The clustering could also have been based on more constraints to get a more accurate result, like the transportation features and population density near the analysed locations.

* Conclusion :

In the above study, we have scrapped the geo location of various hotspots in Kochi and have used this to find the various shops near them. This information has been used to cluster the data points into 4 clusters. These clusters have been utilized to decide on a feasible and profitable location to open a coffee shop in Kochi.