Title: To analyze suitable place for opening New Shopping Mall

Introduction :-

Shopping malls are preferred by peoples these days because of multifunction availabilities. Peoples can have grocery shopping, accessories shopping and at the same time there are fun activity corners for children's also.

For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the city nearby IT Hubs and many more are to be built. Opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. So here, location plays a crucial role which will decide whether the opening of mall will be profitable or failure.

Business Problem:-

The objective of this capstone project is to analyze and select the best locations in the city of Pune, India to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city , if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Data :-

To solve the problem, we will need the following data:

1. List of neighborhood's in Pune city.
2. Latitude and longitude coordinates of those neighborhood's. This is required in order to plot the map and also to get the venue data.
3. Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them:-

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Pune) contains a list of neighborhood's in Pune, with a total of 46 neighborhood's which comes under Pune Municipal Corporation. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup packages. Then we will get the geographical coordinates of the neighborhood's using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.