COURSERA CAPSTONE PROJECT

IBM DATASCIENCE

BY : RAAGHUL VISWANATH

OPENING A NEW SHOPPING MALL IN CHENNAI, INDIA.

INTRODUCTION

For many shoppers visiting shopping malls is a great way to spend their leisure time during weekends. The different kinds are stupid gigantic, windowless urban, after-hours pedestrian, impressive urban, smooth, snow-covered, successful regional, massive mega, unfortunate modern, expansive pedestrian, massive regional, flashy corporate, big, faceless, dizzying underground and has to be high end pedestrian. It consists of restaurants ,spa ,kids-zones, theatres and arcades. It houses people of all age groups . It is a growing demand to build shopping malls in a developing country like India. Chennai , the capital of southern state Tamil Nadu is a fast growing economy and is the perfect time for builders to build shopping malls here. But the location of a mall is important in determining it's success.

BUSINESS PROBLEM

The objective of this capstone project is to analyze and select the location for a new shopping mall in the city of Chennai using data science methodology. It provides the answer to the business question , which area is the best location to build a new shopping mall. If a property developer is looking to open a new shopping mall, where would you recommend to open it?

TARGET AUDIENCE

This project is targeted towards real estate builders and promoters in the city of Chennai. The project is timely as the city is in the process of westernization as the number of malls are also significant. The new shopping mall can be opened by the real estate promoters based on the project findings ie the optimum location to open the mall.

DATA NEEDED

To solve the problem , we need the following data,

1. List of neighborhoods in the city of Chennai.
2. Location of neighborhoods found.
3. Venue data related to shopping malls. This can be later used for clustering.

The Wikipedia page https://en.wikipedia.org/wiki/Category:Suburbs_of_Chennai contains a list of suburbs of Chennai. HTML Parsing is done using module such as Beautiful soup . Once scraping of web page is done to extract the required data, python module Geocoder is used to extract the latitude and longitude of the location

The foursquares api is used to get the venue data of shopping malls in each location. Foursquares api will provide name, location and categories of venue data near the given location after we specify the radius and limits. We will use the Shopping Mall category. We will use Kmeans clustering to cluster the locations with the shopping mall category and see the best cluster to build a new mall.