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.

METHODOLOGY

The Wikipedia page is scraped and downloaded into the notebook with the use of Beautiful Soap library. But then it is just a list of names with no meaningful insight. Then we use the python geocoder package to extract the latitude and longitude of the neighborhood extracted. We then make use of the folium plot to visualize the neighborhoods for sanity check. Next we use foursquares api to get the venue data. The calls can be made by using a client ID, Client key which can be obtained by simply registering as a developer with Foursquares.

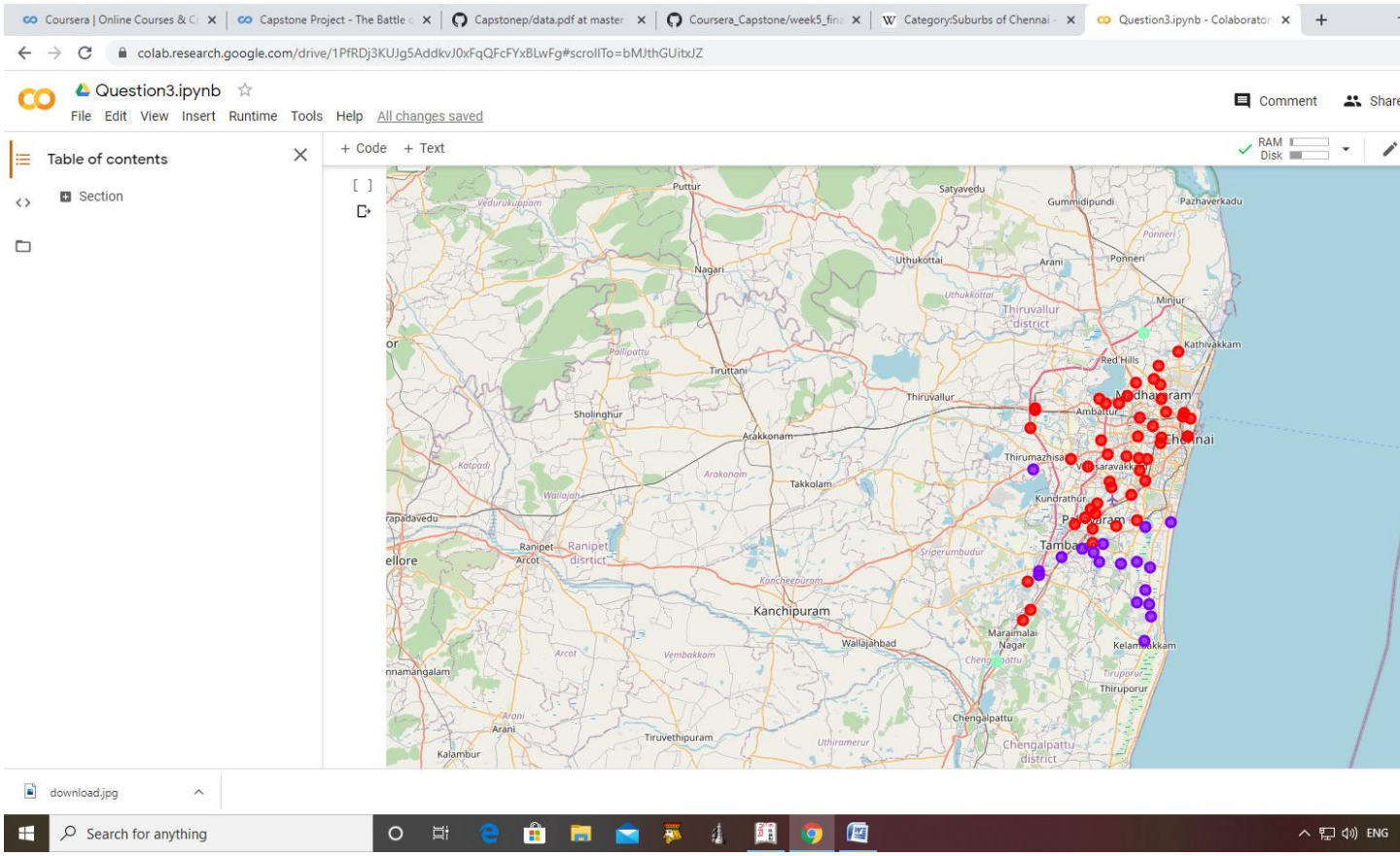
Then we make calls to obtain various venue data near the specified neighborhoods from which we filter out the shopping mall category. The presence of shopping mall in each location is done by one hot encoding and we find if the given neighborhood as shopping mall or not. Then we cluster the different neighborhoods with the help of KMeans clustering . We visualize the clustered neighborhoods to analyze a pattern. Based on the clusters it would be useful for us to determine a place to open a new shopping mall.

RESULTS

The results show the division of neighborhood into three clusters based on the frequency of shopping malls. The green clustered shows that the neighborhoods are outside Chennai and frequency is high.

The red colored neighborhood shows that they are centered in the city and have a decent frequency. The malls are already spaced out in the central part of the city.

The violet colored neighborhoods are suburbs and they contain very low frequency to malls. In some cases there are no malls present.



DISCUSSIONS

AS we see the above depicted diagram, the violet clustered neighborhoods give a bright chance to open a new mall in the city of Chennai as the frequency of the malls is relatively low. The red clustered are city-centric and there will be a heavy competition to build a brand new shopping

mall there. The green colored clustered neighborhoods already have shopping malls and also are out of scope for building new ones as it is outside the suburbs of the city.

LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This project concentrates only on the frequency of the malls to cluster them, but various other factors like population, average salary can be taken into account for taking decisions.

The unavailability of those data for the time being compromises various factors that might determine the new location. I have used regular calls for the foursquare api. We can use premium account to make use of various other objects that is returned by the api.

CONCLUSION

I have scraped data from the webpage, added longitude and latitude locations for each neighborhood using the geocoder python package. Then we use regular calls to get location venues near the referenced location using foursquare api. We only choose shopping malls from categories. We then cluster the neighborhoods based on the frequency of shopping malls. We analyse the clusters and it is finally decided that Violet colored neighborhoods are the best place to build a new shopping mall in the city of Chennai. The best place would be NAVALUR.