

CAPSTONE PROJECT

IBM APPLIED DATA SCIENCE CAPSTONE

INTRODUCTION:

Bangalore, officially known as Bengaluru is the capital of the Indian state of Karnataka. It has a population of over 10 million, making it a megacity and the third most populous urban agglomeration in India. It is located in Southern India, on the Deccan Plateau at an elevation of over 900m (3000ft) above sea level, which is the highest among India's major cities. It is multi-ethnic, multi-religious and cosmopolitan in nature.

BUSINESS PROBLEM:

I am currently pursuing my engineering from RNS Institute of Technology, which is a college located in Bangalore. Lots of student studying in this college stay away from their homes during their 4 years of tenure of study. Being a student, I would like to reside in the neighborhood of the college. The main aim of this Capstone Project is to explore the venues of each neighborhood of this college and look for the neighborhoods having ample hostels, malls, restaurants, departmental stores, book stores and transportation etc. This analysis may help students who wish to reside nearby the college campus.

DATA:

I downloaded the required data from a website published by Government of India. Here is a link to the website: <https://data.gov.in/resources/all-india-pincode-directory-contact-details-along-latitude-and-longitude>

This csv file contains the list of pin codes of all cities of India, officenames (post office), the head office (HO) names of each officename etc. The Latitude and Longitude columns are incomplete for most areas. So, I will try to fill it with the help of Geopy. I am using Foursquare API to get the venues of each neighborhood of the institution.

I have downloaded the required csv file and loaded this data into the project to be used as a data asset. I have then converted the csv file to pandas dataframe and displayed the head of the dataframe below. Further data formating steps are given below under the heading of "Formating and Cleaning Data".

METHODOLOGY:

The first step in implementing the given problem and analyzing is to format and clean the data. The steps for formating and cleaning the data is given below:

FORMATING AND CLEANING DATA:

1. The data pertaining to Districtname='Bangalore' is needed, so I will filter only the rows having Districtname = 'Bangalore'.
2. Since, RNSIT's Related HeadOffice is "Basavanagudi", only the data pertaining to this condition is kept.
3. Next, we will keep only the columns officename, pincode, longitude, latitude.

4. Multiple occurrences of same pincode are checked.
5. Rename the column 'officename' as 'neighborhood' and combine the neighborhoods having the same pincode with a comma (,).

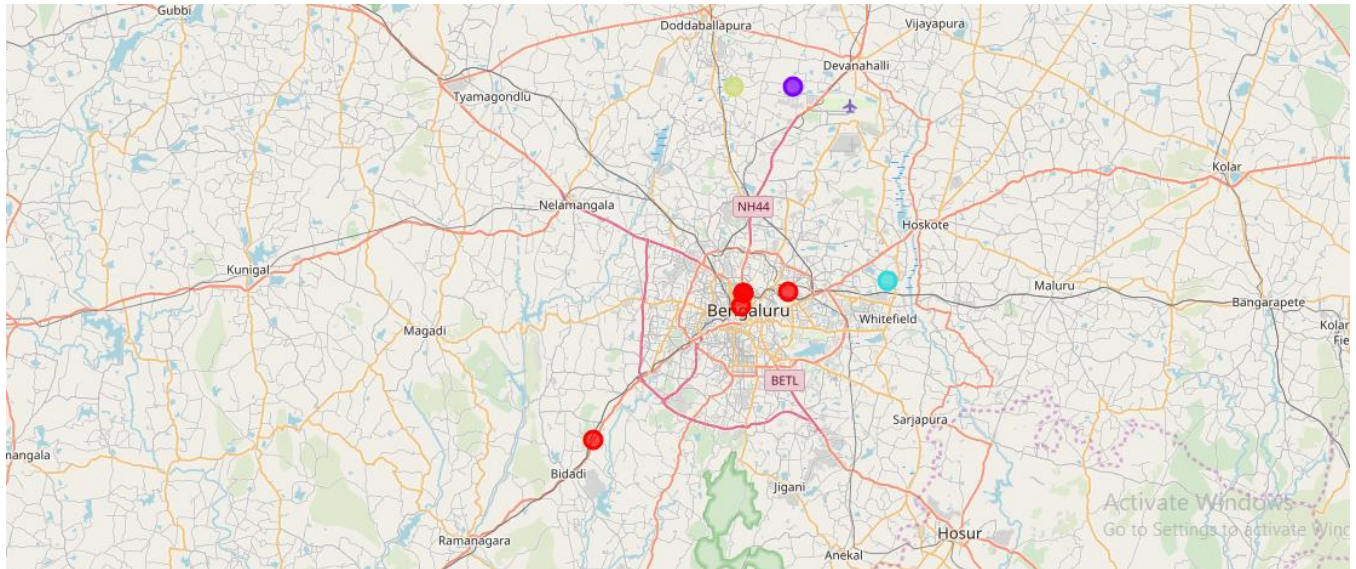
After forming and cleaning the data Geopy will be used for finding the latitude and longitude of all the neighborhoods because some of the neighborhoods are not provided with latitude and longitude. Before adding the latitude and longitude 'latitude' and 'longitude' columns will be added to the dataframe for further operations on the dataframe. After making few adjustments to the dataframe the dataframe is mapped.

The next step is to use the Foursquare API to get the nearby venues. After that for clustering and visualizing neighborhoods, k-means clustering method is used. To find the optimal k in k-means clustering method, Silhouette method is used. After find the optimal k, the no. of clusters is specified.

The next step is to visualize the clusters and examining the clusters.

RESULTS:

Visualization: Given below is the map formed out of RNSIT, Bangalore neighborhoods



EXAMINING CLUSTERS:

The dataframe of each cluster is given below:

Cluster 0:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Mavalli, Basavanagudi H.O, Pampamahakavi Road,....	Hotel	Indian Restaurant	Vietnamese Restaurant	Office	Coffee Shop	Gym	Hotel Pool	Juice Bar	Hotel Bar	Racetrack
1	Deepanjalinagar, Nayandahalli, Governmemnt Ele...	Clothing Store	Shopping Mall	Multiplex	Fast Food Restaurant	Vietnamese Restaurant	Gym	Art Gallery	Asian Restaurant	Boutique	Café
5	Chickpet	Café	Indian Restaurant	Boutique	Hotel	Dance Studio	Pub	Andhra Restaurant	Restaurant	Art Gallery	Soccer Field
6	Bengaluru Vishwavidyalaya, Mallathahalli	Café	Indian Restaurant	Boutique	Hotel	Dance Studio	Pub	Andhra Restaurant	Restaurant	Art Gallery	Soccer Field
7	Rv Niketan	Café	Indian Restaurant	Boutique	Hotel	Dance Studio	Pub	Andhra Restaurant	Restaurant	Art Gallery	Soccer Field

Cluster 1:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Agara, Koramangala, Koramangala I Block, St. J...	Asian Restaurant	Vietnamese Restaurant	Hotel	Art Gallery	Boutique	Café	Chinese Restaurant	Clothing Store	Coffee Shop	Convenience Store

Cluster 2:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Carmelram	Convenience Store	Vietnamese Restaurant	Hotel	Art Gallery	Asian Restaurant	Boutique	Café	Chinese Restaurant	Clothing Store	Coffee Shop

Cluster 3:

00c[.58]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Adugodi	Farm	Vietnamese Restaurant	Hotel	Art Gallery	Asian Restaurant	Boutique	Café	Chinese Restaurant	Clothing Store	Coffee Shop

DISCUSSIONS:

After examining the areas in Cluster 0, we can see that they are more common in Hotel, Clothing Store, Café, Indian Restaurant etc. whereas cluster 1 is more popular for Asian Restaurant and Hotels.

So, students who prefer Hotel and Clothing store or café is more likely to stay in Neighborhoods of cluster 0 .i.e. Chickpet, Basavanagudi, Rv Niketan, Bengaluru Vishwa Vidyalaya, Mallathahalli etc.

On the other hand if the students prefer Convenience stores more near their neighborhoods, students are more likely to prefer to stay in Neighborhoods of cluster 2 . i.e., Carmelram neighborhood.

CONCLUSION:

The result of this report provides a general overview of ideal neighborhoods around an institution named RNS Institute of Technology. This report is helpful for students or people who wish to reside nearby the institute.