Coursera Capstone Project

Battle of Neighborhoods

Overview


Introduction


Business Problem


Data


Methodology


Discussion


Results


Discussion and Conclusion

Introduction

Bangalore is the largest and the capital city of the state of Karnataka, India. Touted as the ‘Garden City’ for its beautiful parks and green landscapes, it is now regarded as one of the largest IT hubs in the country

Various neighborhoods in Bangalore are slowly becoming hotspots for culinary indulgences and for a person looking to open a food joint or a café, making an informed decision is necessary to reap maximum benefits in terms of customer loyalty and revenue.

Business Problem

IT boom in the recent years has led to an increasing number of office spaces thereby leading to an increase in the number of food services in and around such spots.

The main aim of this project is to find such ideal spots across the city to open a café or a restaurant and gain maximum profit by catering to the needs of this demographic

Data

01

List of neighborhoods in Bangalore :

Web scraping Wikipedia page

02

Latitudes and Longitudes of neighborhoods :

Nominatim Geocoding API

03

Venues in and around neighborhoods :

Foursquare API

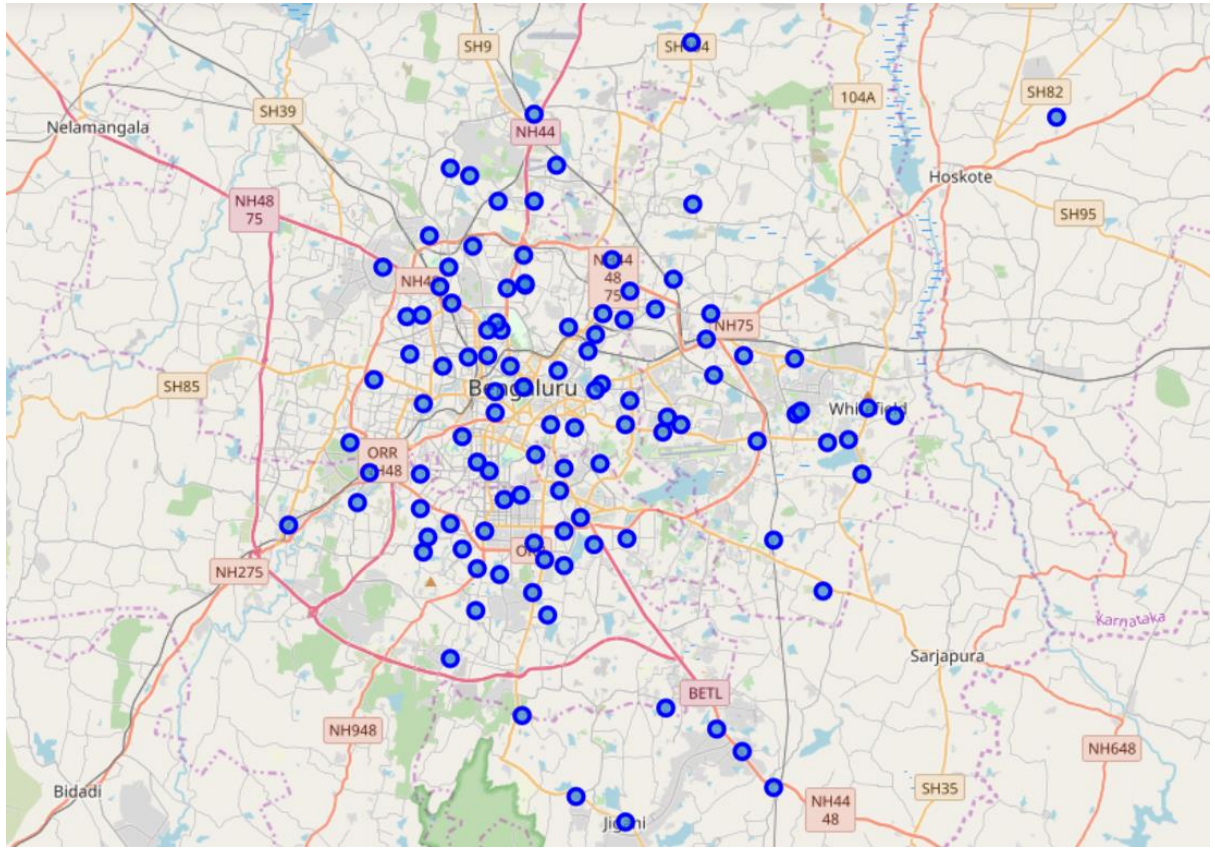
Methodology

Using Folium to create a map of Bangalore

```
# create map of Bangalore using latitude and longitude values
map_bangalore = folium.Map(location=[lat, long], zoom_start=11)   #Passing location info of Bangalore

# adding markers to map
for lat, lng, label in zip(df['Latitude'], df['Longitude'], df['Neighbourhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_bangalore)

map_bangalore
```



Obtaining Venue details

Foursquare API is used to obtain venue details

```
explore_df_list = []

for i, nbd_name in enumerate(df3['Neighbourhood']):
    try :

        nbd_name = df3.loc[i, 'Neighbourhood']
        nbd_lat = df3.loc[i, 'Latitude']
        nbd_lng = df3.loc[i, 'Longitude']
        radius = 1000
        LIMIT = 30
        url = 'https://api.foursquare.com/v2/venues/explore?client_id={} \
        &client_secret={}&ll={},{}&v={}&radius={}&limit={}'\
        .format(CLIENT_ID, CLIENT_SECRET, nbd_lat, nbd_lng, VERSION, radius, LIMIT)

        results = json.loads(requests.get(url).text)
        results = results['response']['groups'][0]['items']
        nearby = json_normalize(results)

        filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
        nearby = nearby.loc[:, filtered_columns]
        columns = ['Name', 'Category', 'Latitude', 'Longitude']
        nearby.columns = columns
        nearby['Category'] = nearby.apply(get_category_type, axis=1)

        for i, name in enumerate(nearby['Name']):
            s_list = nearby.loc[i, :].values.tolist()
            f_list = [nbd_name, nbd_lat, nbd_lng] + s_list
            explore_df_list.append(f_list)

    except Exception as e:
        pass
```

One hot encoding

One-hot encoding is performed to normalize data and pivot the categories to enable easy analysis for clustering.

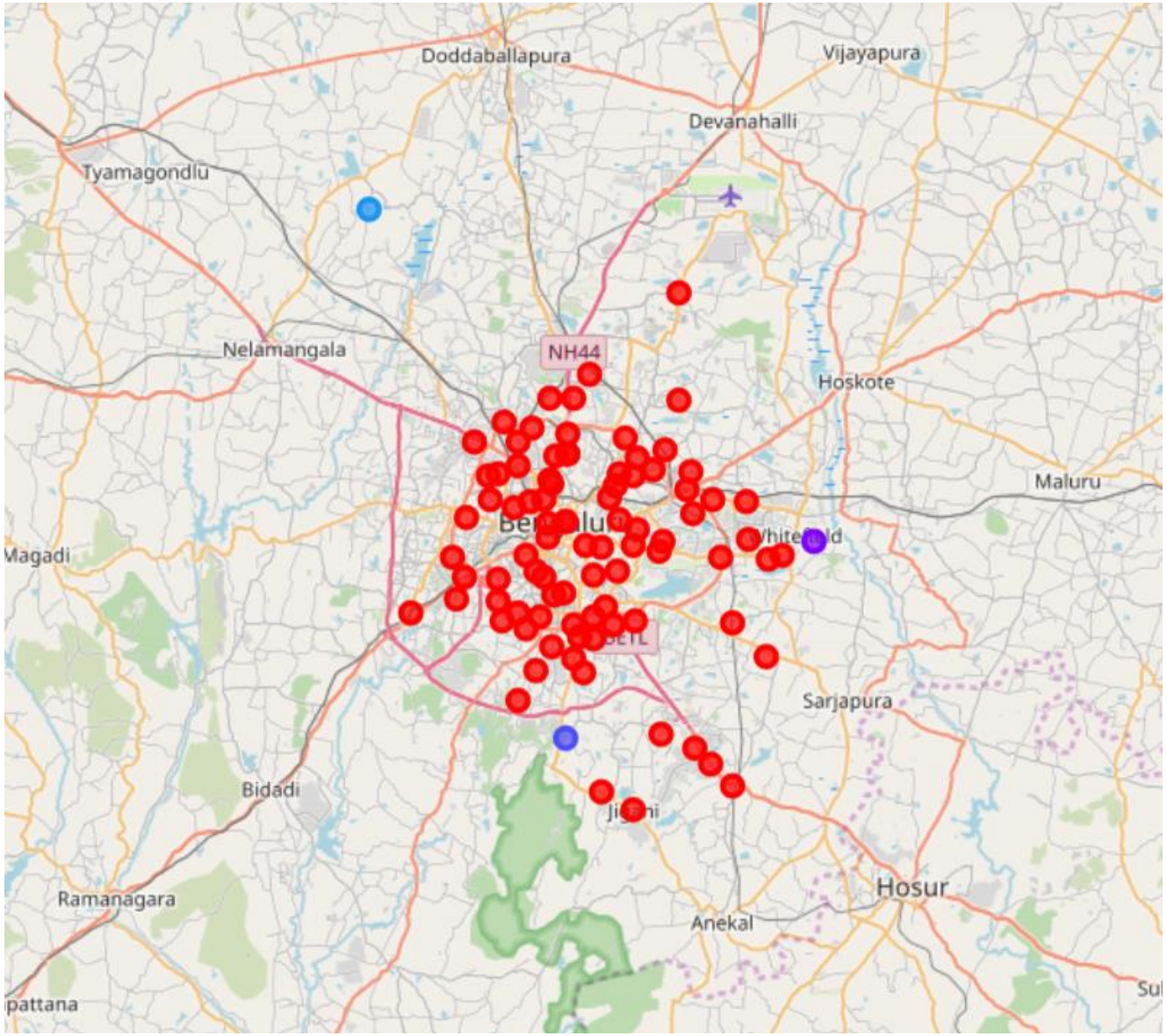
	Neighbourhood	ATM	Accessories Store	American Restaurant	Andhra Restaurant	Animal Shelter	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	...	Tennis Court	Theater	Toy / Game Store	Trail	Train Station	Ud Restaur
0	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
1	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
3	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
4	Adugodi	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	

K- means clustering

K-means clustering is employed to garner insights about similar groups of neighborhoods - finding groups of neighborhoods that share certain characteristics

An optimal value of K was found by iterating from 2 to 20 and using silhouette score as a parameter for accuracy

```
K = 4
bgc = blr_grouped_clustering
kmeans = KMeans(n_clusters = K, init = 'k-means++', random_state = 0).fit(bgc)
```



Results

The neighborhoods are segmented into K clusters depending on the value of K assigned while training the model

This can be visualized on a map where each cluster is indicated by a different color

Discussion

After a thorough analysis of the clusters, cluster 0 seems to be the best fit containing the target venue category – Café.

	Neighbourhood	Latitude	Longitude	Cluster Labels	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 M Comm Ver
26	Cooke Town	13.002785	77.624747	0	Café	Bakery	Department Store	Fast Food Restaurant	Shopping Mall	Chinese Restaurant	Indian Restaurant	Clothing Store	Cof Sh
30	Devarachikkanahalli	12.902105	77.601220	0	Café	Multiplex	Pizza Place	Indian Restaurant	Italian Restaurant	Coffee Shop	Clothing Store	Mediterranean Restaurant	Garc
33	Ejipura	12.945245	77.626914	0	Café	Pizza Place	Indian Restaurant	Gym / Fitness Center	Clothing Store	Ice Cream Shop	Department Store	Fast Food Restaurant	Fc Cc
47	HSR Layout	12.911623	77.638862	0	Café	Indian Restaurant	Ice Cream Shop	Pizza Place	Snack Place	Coffee Shop	Kebab Restaurant	Liquor Store	Bur Jc
48	Hulimavu	12.877349	77.602803	0	Café	Movie Theater	Multiplex	South Indian Restaurant	Department Store	Badminton Court	Lake	Fast Food Restaurant	Bowl A
101	Sadashivanagar	13.007708	77.579589	0	Café	Coffee Shop	Indian Restaurant	Department Store	Dessert Shop	Gym	Ice Cream Shop	Women's Store	Fi
105	Shikaripalya	12.835985	77.657181	0	Café	Indian Restaurant	Fast Food Restaurant	Hotel	Coffee Shop	Juice Bar	Department Store	Pizza Place	Chine Restaur

Conclusion

The organized food service industry will continue to flourish given the penchant for exploring various cuisines amongst the residents of metropolitan cities

This project has successfully devised a quantitative approach for a potential café owner to make an informed decision about the location to start their business

This could save a lot of costs and help the business thrive when a consistent customer base has developed.