Capstone Project — the Battle of Neighbourhoods in Patna

Introduction

Patna capital of Bihar state, northern India. Patna is a riverside city that extends along the south bank of the Ganges. Total population of Patna is more than 6.5 Million. It has total 40 neighborhood. It seems that over last decade its continuously grow as a result it grab 33rd spot in the Ease of living index 2020 released by the Union ministry of housing and urban affairs. The official language of Patna and the one that is most widely spoken is Hindi. However, English is also spoken as a formal language within businesses and government agencies. With its diverse culture, comes diverse food items. There are many restaurants, Food Truck and Cafe's in Patna.

Aim

This project aims to find the best location to open a Café, Chinese Restaurant, Furniture store, American Restaurants in the city of Patna, India, to maximize the profit of the Client and give him best recommendations. The target audience for this project is the people who want to start these above business. We will provide him location where he can make good money.

Data Description

Dataset containing Neighbourhoods of Patna, the dataset holds the names Neighbourhood in Patna extracted from Wikipedia. Here is a link. https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Patna.

I used python library geopy.geoder to find the coordinates of neighborhoods in the data frame. After that we visualize using folium. Then used Foursquare API to explore the most common places of a neighborhood in the form of a JSON file.

Methodology

In this section, I will describe the data analysis and how I used the data to yield the results.

I am using BeautifulSoup to scrap the data from the mentioned URL and take the fields which I will be working on. Then, I enabled geopy functions by installing the conda-forge geopy package. I used the nominatim function to add geospatial data to the data frame that is the latitude and the longitude. Using the folium package and my data frame, I then created a map with the forty one city districts on it.

Now, foursquare data comes into play. Then, retrieved the foursquare data for all venues on foursquare with a distance of less than 3000 meters from each center of each city district, as indicated as blue dots in the map above. The result was a list of 75 venues all over Patna City. There are 35 unique categories. There are different type of Restaurant, cafe and Furniture store.

we get Buddha Colony and Serpentine Road returned the highest number of venues i.e. 21. With the help of One hot encoding on the obtained data set i use it find out the 10 most common venue category in each neighborhood.

There are many techniques are available in Data science field. For my project, I am going to use clustering. Particularly, I am going to use K-Means clustering. So, I need to know how many optimal number of cluster could be better for my data. To find the optimal number of clusters we will use silhouette score metric technique. Now we have cleaned data and found all the prerequisite to model our data from the graph it looks cluster 4 could be the better one.so, let's use number of clusters as 4.

Add the cluster labels to the neighborhoods_venues_sorted dataframe. And let's create a new dataframe kl_merged which has the neighborhood details, cluster labels and the 10 most common venues in that neighborhood. we see in the table are the city districts and their most common venues, and they now have been assigned five different cluster labels from 0 to 3.

We can now use the cluster labels to show the city districts marked with a cluster-specific color on a map, again using folium: As in the map every cluster had given different colors. so there are total 4 cluster of different color cluster 1

color is Purple, Cluster 2 color is sky blue, Cluster 3 color is red and cluster 4 color is Fade yellow.

Result

We now can show four clusters with the most common venues

Cluster 1

```
ATM                   13
Chinese Restaurant    10
Bus Station            8
Boat or Ferry          8
Café                   8
Bar                    8
American Restaurant    8
Golf Course            7
Train Station          7
Accessories Store      7
dtype: int64
```

According to our Question we are looking for Cafe, Chinese Restaurant, Furniture store, and American Restaurants In above cluster there are 8 Cafes ,8 American Restaurant, 10 Chinese Restaurant and No Furniture store.

Cluster 2

```
Fast Food Restaurant    19
Café                    17
Hotel                   17
Indian Restaurant       16
Park                    15
Department Store        14
Pizza Place             14
Train Station           14
Hotel Bar               12
Juice Bar               12
dtype: int64
```

Cluster 3

```
Fast Food Restaurant       2
Food Court                 2
Furniture / Home Store     2
Platform                   2
Women's Store              2
Cosmetics Shop             2
Chinese Restaurant         2
Cricket Ground             2
Department Store           2
Clothing Store             2
dtype: int64
```

In above cluster there are 2 Chinese Restaurant , 2 Furniture store and No cafes and American Restaurant

Cluster 4

```
Fast Food Restaurant       2
Food Court                 2
Furniture / Home Store     2
Department Store           2
Women's Store              2
Cosmetics Shop             2
Cricket Ground             2
Golf Course                2
Train Station              2
Garden                     2
dtype: int64
```
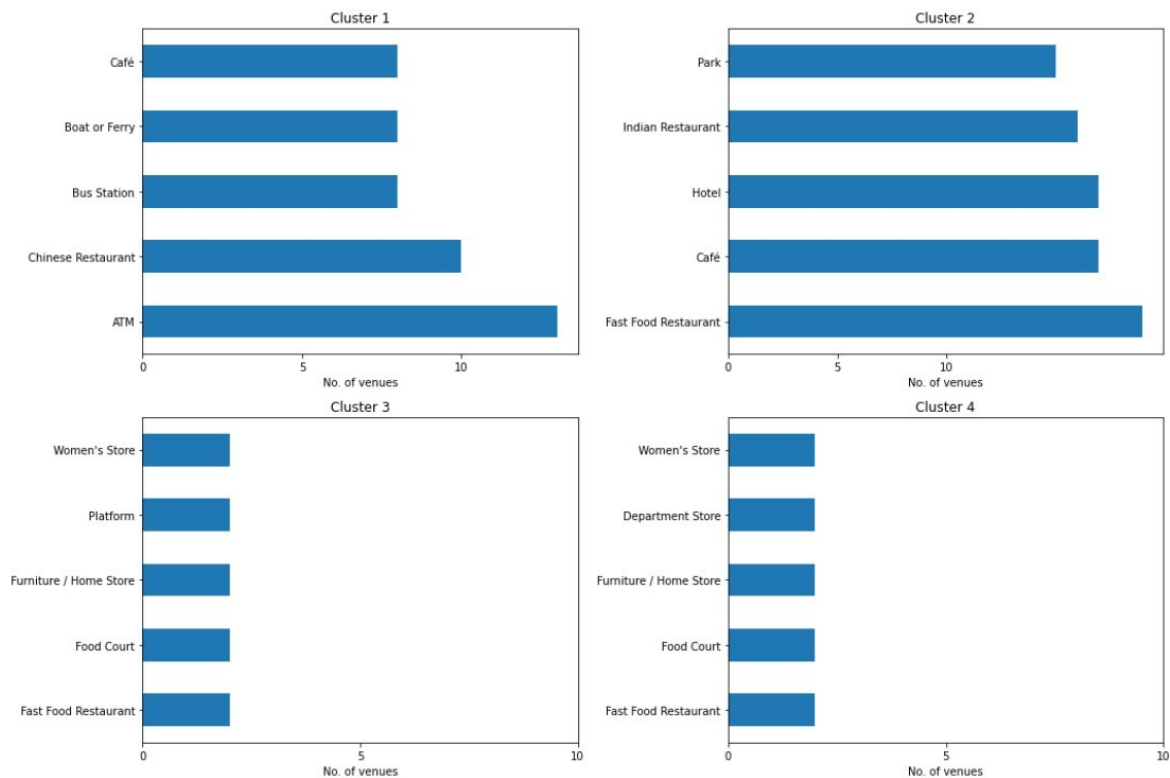
In above cluster there is only 2 Furniture store and no cafes , Chinese and American restaurant.

Discussions

Best Location to open Chinese Restaurant

Kankarbagh , Boring Road, Frazer Road , Rajender nagar are the best neighborhoods for Chinese Restaurant as in Cluster 2 and Cluster 4 has no such Restaurant so if anyone wants to open Chinese Restaurant then the business be in profit.

Best Location to open Cafes

Bhita , Masaurhi have no Cafe to chill out so the people who live there would covered big distance if they want to go to Cafe.As those area are still not well developed so if new concept like cafe will start then there is a high probability to make good money

Best Location to open American Restaurant

People of Patna also like American style food like Fries , Burgers etc and also there is not so many american restaurant so if anyone wants to open American Restaurant then it will be a game changer for stakeholders. Places like Exhibition Road , New Karbhigya ,Gandhi maidan marg are well developed and good for business

Best Location to open Furniture store

After Food the most growing business in Patna is Furniture Stores. Places like Digha , Buddha Colony , Danapur are good to start.

Conclusion

Purpose of this project was to analyse the neighborhoods of Patna and create a clustering model to suggest places to start a new business based on the category. The neighborhoods data was obtained from Wikipedia and the Foursquare API was used to find the major venues in each neighborhood. Then locations were used to create a clustering model. The best number of clusters i.e. 4 was obtained using the silhouette score. Each cluster was examined to find the most venue categories present, that defines the characteristics for that particular cluster. A map showing the clusters has been provided. Both of these can be used by stakeholders to decide the location for the required type of business.