Business Expansion

Shubham Jain May 15, 2020

I. Introduction

1. Background

Growth is crucial to the long term survival of a business. It makes it easier to acquire assets, attract new talent, and fund investments. It also drives business performance and profit. Growth helps to respond to market demand, allowing the company to increase its market share and capitalize on its growing brand. It often spurs innovation, helping the company to differentiate in the market and stave off competition. Growth can also boost the business' credibility, allow the companies to broaden its supply base, and increase stability and profits. However, to be successful and sustainable, growth has to be strategic and has to happen for the right reasons. One of the most common examples of growth is the expansion of Restaurants in various cities.

2. Problem

Whether a person owns one restaurant or 20, expansion takes careful planning and consideration. There has to be a considerable strategy that'll help set the next location up for success. Any brand of the restaurant while expanding in a particular location desires that is a particular location has a similar or a more promising market environment than the previous one.
After deciding the location which is most similar to the previous location, it is also desired to know about the most competing restaurants in that area so that we can study them and build a strategy to become better than them.

3. Interest

This project aims at any company or business, which might be thinking that it's time to expand their business into another city for growth.

II. Data

1. Data Requirements

To find a solution to the problem and build a recommender model that outputs the location where a Restaurant can expand its business, we need the following information about the restaurants located in various locations in a city.

a) Its geographical coordinates (latitude and longitude) to find out where exactly it is located.

To study a restaurant's neighborhood, we need to access its location i.e., Latitude and Longitude are to be known so that we can point at its coordinates.

b) The population of the neighborhood where the restaurant is located.

The population of a neighborhood is a very important factor in determining a restaurant's growth and amount of customers who turn up to eat. Logically, the more the population of a neighborhood, the more people will be interested to walk openly into a restaurant and the less the population, the less number of people frequently visit a restaurant. Also, if more people visit, the more successful will be the restaurant.

c) The average income of the neighborhood to know how much is the restaurant worth.

The income of a neighborhood is also a very important factor as the population was. Income is directly proportional to the richness of a neighborhood. If people in a neighborhood ear more than an average income, then it is very much possible that they will spend more however not always true with very little probability. So a restaurant's assessment is proportional to the income of a neighborhood.

2. Data Sources

a) A list of the neighborhood is scrapped from Wikipedia.

b) Used Google Map API to fetch Latitude and Longitude for the scrapped neighborhood.

	Borough	Neighborhoods	Latitude	Longitude
0	SouthernSuburbs	Anjanapura	12.8604	77.5612
1	SouthernSuburbs	Arekere	12.8875	77.5970
2	Southern	Banashankari	12.9255	77.5468
3	NorthEastern	Banaswadi	13.0120	77.6471
4	Southern	Basavanagudi	12.9421	77.5754

c) Population data for some of the cities were collected from here. For the cities not available in the link, the population is assumed and maybe inaccurate but since this is a demonstrating project, the main idea is to get the working model. The data frame for Bangalore neighborhood Population looks like:

	Borough	Neighborhoods	Population	Normalized_Population
0	SouthernSuburbs	Anjanapura	940039	0.955027
1	SouthernSuburbs	Arekere	138760	0.128036
2	Southern	Banashankari	810407	0.821235
3	NorthEastern	Banaswadi	632031	0.637135
4	Southern	Basavanagudi	426903	0.425425

d) Income for the main city of Bangalore can be found here. Again for the neighborhood income, it is assumed and maybe inaccurate but since this is a demonstrating project, the main idea is to get the working model. The data frame for Bangalore neighborhood Income looks like:

	Borough	Neighborhoods	AverageIncome	Normalized_Income
0	SouthernSuburbs	Anjanapura	44218.92255	0.655989
1	SouthernSuburbs	Arekere	29378.71663	0.406048
2	Southern	Banashankari	57524.20953	0.880079
3	NorthEastern	Banaswadi	53349.70118	0.809771
4	Southern	Basavanagudi	63161.96222	0.975031

3. Foursquare API

The use of foursquare is focused to fetch the nearest Venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius (1 Km) and also corresponding coordinates, venue locations, and names. Now the following data frame is created:

	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Anjanapura	SouthernSuburbs	12.8604	77.5612	State Bank of India ATM	12.861650	77.561634	ATM
1	Anjanapura	SouthernSuburbs	12.8604	77.5612	capital club resorts	12.855979	77.555653	Pool
2	Anjanapura	SouthernSuburbs	12.8604	77.5612	Accenture	12.858980	77.569510	Business Service
3	Anjanapura	SouthernSuburbs	12.8604	77.5612	Axis Bank ATM	12.868820	77.559060	ATM
4	Arekere	SouthernSuburbs	12.8875	77.5970	Decathlon Sports India Pvt Ltd	12.887513	77.597712	Sporting Goods Shop

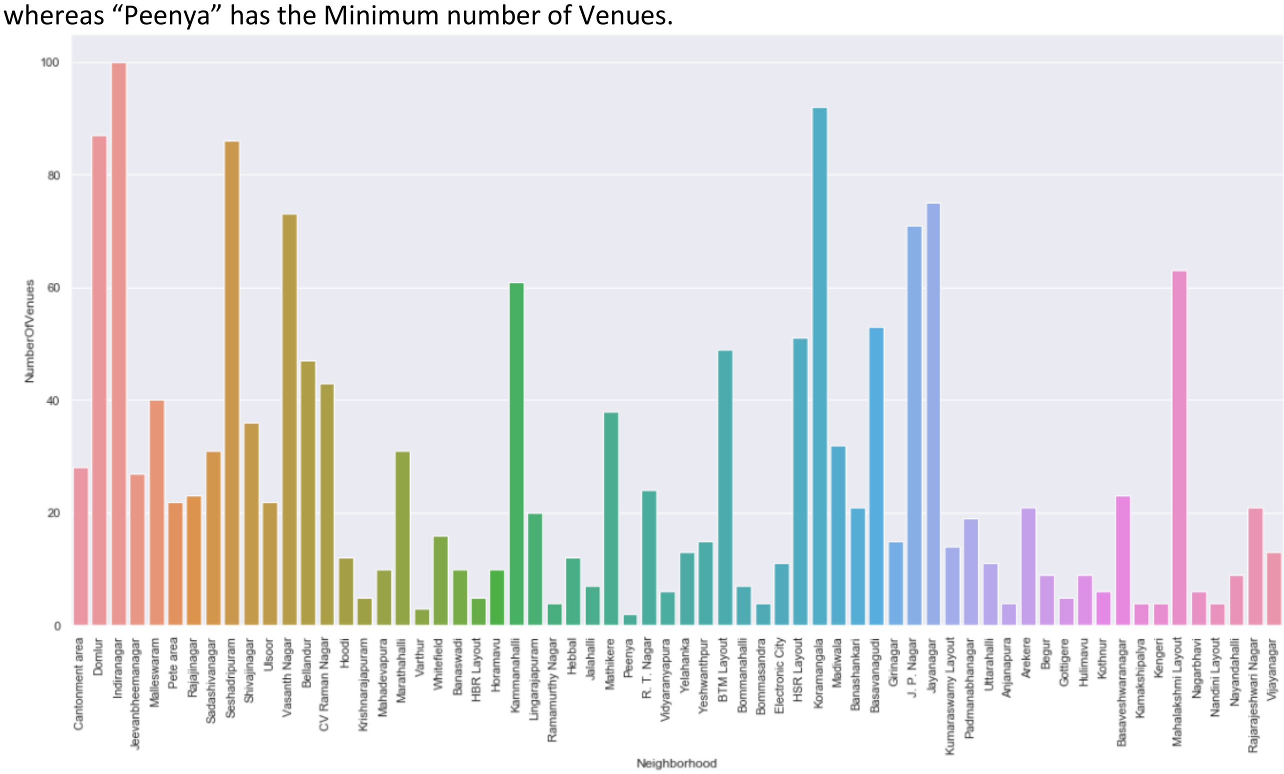
III. Methodology

1. Exploratory Data Analysis

After collecting the information about the venues in each neighborhood of each Borough, the data is grouped and studied to draw some conclusions. This exploring the dataset is important as it gives us initial insights and helps us get a partial idea of the answers that we are looking to find out from the data.

	Borough	Neighborhood	NumberOfVenues
0	Central	Cantonment area	28
1	Central	Domlur	87
2	Central	Indiranagar	100
3	Central	Jeevanbheemanagar	27
4	Central	Malleswaram	40

Now the above data is plotted to visualize and it was found that “Indranagar” has the Maximum number of Venues, whereas “Peenya” has the Minimum number of Venues.



2. Data Preprocessing

To build our models, two datasets need to be defined.

a) “Venue Preference”

This is the dataset containing the top 10 venues in each of the neighborhoods which will be used at the end of the Use Case to tell the user about what venues he has competition with.

	Neighborhoods	1th Most Common Venue	2th Most Common Venue	3th 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
0	Anjanapura	Venue_capital club resorts	Venue_Accenture	Venue_Axis Bank ATM	Venue_State Bank of India ATM	Venue_ಗೋಪಾಲ್ ಕ್ಯಾಂಟೀನ್ (Gopal Canteen)	Venue_Goli Vadapav	Venue_Gold's Gym	Venue_Gold's gym	Venue_Golds Gym
1	Arekere	Venue_Domino's Pizza	Venue_Cafe Coffee Day	Venue_Reliance Mart	Venue_Baskin-Robbins	Venue_Pizza Hut	Venue_Benison Super Market	Venue_Natural Ice Cream	Venue_Adyar Anand Bhavan (A2B)	Venue_Baker's Stop
2	BTM Layout	Venue_Cafe Coffee Day	Venue_Subway	Venue_Domino's Pizza	Venue_Faaso's	Venue_Cafe Orio	Venue_Narmadha's Hyderabad Biriyani	Venue_Core Fitness	Venue_Marwa Restuarant	Venue_McDonald's
3	Banashankari	Venue_Domino's Pizza	Venue_Corner Adda	Venue_Anna Kuteera	Venue_Croma	Venue_Pizza Hut	Venue_Cafe coffee day	Venue_Caf-Eleven	Venue_Adyar Ananda Bhavan	Venue_Kamakya theatre

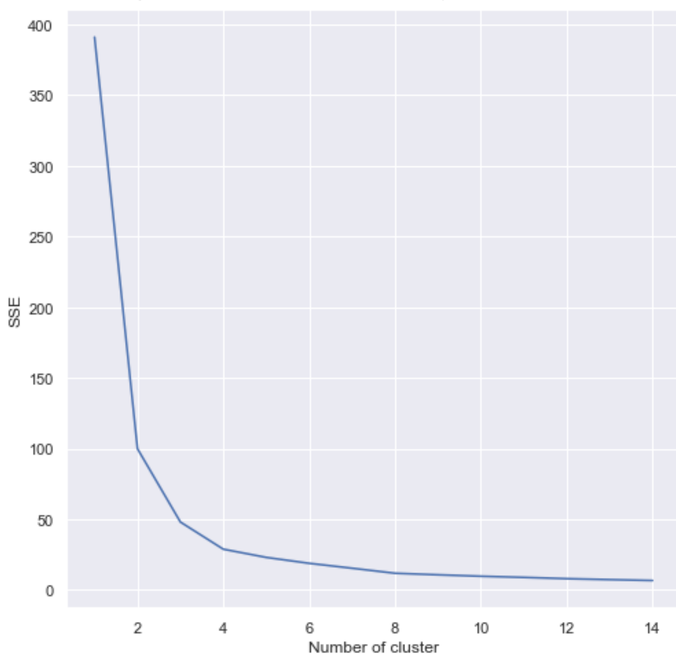
b) “Neighborhood Preference”

This is the dataset that is clustered into different groups based on the similarity and the features of the locations.

	Borough	Neighborhoods	Latitude	Longitude	Normalized_Population	Normalized_Income	Ranking
0	SouthernSuburbs	Anjanapura	12.8604	77.5612	0.955027	0.655989	0.707110
1	SouthernSuburbs	Arekere	12.8875	77.5970	0.128036	0.406048	0.206135
2	Southern	Banashankari	12.9255	77.5468	0.821235	0.880079	0.718645
3	NorthEastern	Banaswadi	13.0120	77.6471	0.637135	0.809771	0.601988
4	Southern	Basavanagudi	12.9421	77.5754	0.425425	0.975031	0.553973

3. Segregating Data into Clusters using K Means Clustering

At first, using the Elbow method, an optimum value of k (i.e., the number of Clusters) was found out to be 4.



Then the K Means algorithm was applied to group the data into 4 Different Clusters.

	Borough	Neighborhoods	Latitude	Longitude	Normalized_Population	Normalized_Income	Ranking	Cluster
0	SouthernSuburbs	Anjanapura	12.8604	77.5612	0.955027	0.655989	0.707110	0
1	SouthernSuburbs	Arekere	12.8875	77.5970	0.128036	0.406048	0.206135	0
2	Southern	Banashankari	12.9255	77.5468	0.821235	0.880079	0.718645	3
3	NorthEastern	Banaswadi	13.0120	77.6471	0.637135	0.809771	0.601988	1
4	Southern	Basavanagudi	12.9421	77.5754	0.425425	0.975031	0.553973	3

IV. Results

For a given neighborhood (say "Banashankari"), first we will decide which cluster it is the most similar to. Next, all the Neighborhood belonging to that cluster is arranged in descending order according to their Ranking based on Population, Average Income, and Neighborhood.

	Borough	Neighborhoods	Latitude	Longitude	Normalized_Population	Normalized_Income	Ranking	Cluster
0	Southern	Banashankari	12.925500	77.546800	0.821235	0.880079	0.718645	3
1	Southern	Uttarahalli	12.907000	77.552100	0.730264	0.975102	0.706418	3
2	SouthEastern	BTM Layout	12.916600	77.610100	0.847384	0.741745	0.683303	3
3	Southern	Girinagar	12.938600	77.544000	0.793501	0.604236	0.608233	3
4	Southern	Basavanagudi	12.942100	77.575400	0.425425	0.975031	0.553973	3
5	Southern	Kumaraswamy Layout	12.903800	77.561800	1.000000	0.095524	0.533433	3
6	Southern	Padmanabhanagar	12.915600	77.556800	0.564219	0.707752	0.529823	3
7	SouthEastern	Madiwala	12.922600	77.617400	0.915415	0.164316	0.515218	3
8	SouthEastern	HSR Layout	12.908100	77.647600	0.536576	0.681594	0.506846	3
9	SouthEastern	Bommasandra	12.816800	77.698900	0.731856	0.327855	0.480677	3
10	Southern	J. P. Nagar	12.910500	77.585700	0.248431	0.940562	0.453412	3
11	SouthEastern	Electronic City	12.840711	77.676369	0.791153	0.051895	0.413740	3
12	SouthEastern	Bommanahalli	12.898400	77.617900	0.745644	0.103626	0.409091	3
13	SouthEastern	Koramangala	12.927900	77.627100	0.282157	0.586791	0.346455	3
14	Southern	Jayanagar	12.925000	77.593800	0.092088	0.060996	0.067393	3

At last, using the Venue Reference Dataset, Top 10 Venues for the top 3 Neighborhoods is calculated as shown:

	Neighborhoods	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th M
56	Uttarahalli	Venue_Masala	Venue_Domino's Pizza	Venue_Kaggis Bake Shop	Venue_More 4 U	Venue_Stop N Joy	V
2	BTM Layout	Venue_Cafe Coffee Day	Venue_Subway	Venue_Domino's Pizza	Venue_Faaso's	Venue_Cafe Orio	Venue Hyde
15	Girinagar	Venue_Domino's Pizza	Venue_Seetha Circle	Venue_Cuppa	Venue_Aahar Utsav veg	Venue_New Sagar Fast Food	Ver

V. Discussion

In this study, I created the clusters of Neighborhoods in the city of Bangalore based on mainly Income and Population. These two features are non-linearly related to each other, so they must be normalized before proceeding with the model. The non-linear relationship between these features means that as the amount of population increase, it does not necessarily mean that the average income of a neighborhood will also increase. It is true to most of the case but many cases may differ to follow this trend.

VI. Conclusion

The recommendation system is a system that considers factors such as population, income and makes use of Foursquare API to determine nearby values. It is a powerful data-driven model that helps companies make the right decisions before expanding their business. I was successfully able to make a sample recommendation system that recommends the top 3 similar neighborhoods and top 10 venues in each corresponding neighborhoods.