COURSERA CAPSTONE PROJECT REPORT

BATTLE OF NEIGHBORHOODS- FINDING A LOCATION TO OPEN AN INDIAN RESTAURANT IN AUSTIN,TX

INTRODUCTION

Austin is the capital city of Texas in United States and is the fourth largest city in Texas and fourteenth largest city in United States. It has been one of the fastest growing large cities in the United States since 2010. The city consistently appears as one of the best cities to live in and best city to move and retire and is also considered as the most livable city based on its amenities, crime rates, cost of living, housing, and other factors. Austin, Texas offers a unique diversity of lifestyles within its boundaries, making the city an option for people in every walk of life. It is also known for its outstanding food and great live music venues.

Being a fast-growing city with diversified ethnicities, Austin can be a great place to start a business. Austin is better known for amazing Mexican food, Italian restaurants and in most recent years' sushi bars. But when it comes to Indian food, it may not be the first thing that comes to our mind when we think of ethnic food. Having lived in Austin for a few years, there are not many authentic Indian restaurants in Austin and would love to see more. To open an Indian restaurant, it requires some serious considerations, one of them is the location which affects the success or failure of the business.

The objective of this project is to analyze and select the best location in the city of Austin to open a new Indian restaurant.

TARGET AUDIENCE AND DEMOGRAPHIC FACTS

The target audience would be the people who are interested in opening or growing an Indian restaurant. The results obtained or the data obtained through cluster analysis would be useful for entrepreneurs who would want to know more about a particular location which is best suited for them to open a restaurant.

DATA

To solve the problem, we will need the following data:

- Austin data containing the zip codes.
- Latitudes and Longitudes of those neighborhoods to plot the map and get the venues.
- Venue data particularly the data related to restaurants.

The Austin data is extracted from http://zipatlas.com/us/tx/austin/zip-code-comparison/percentage-indian-population.htm . It is from the website named Zip Atlas, where it is a structured collection of zip codes, area codes, city and state demographic, social and economic profiles. The data extracted from this website consists of Austin zip codes, location and demographics.

After getting the required data, we will use Foursquare API to find the top 100 venues within a radius of 1500 meters. After finding out the top venues, we will use the cluster analysis method and compare it with the demographic data to find out the best location.

METHODOLOGY

The data is extracted from the website http://zipatlas.com/us/tx/austin/zip-code-comparison/percentage-indian-population.htm and this data is then cleaned and processed into a dataframe. After getting the required data Folium is used to map the zip codes along with its coordinates which are latitudes and longitudes. Folium is a python library used for visualizing the geospatial data.

Next, we will have to find the top 100 venues and this can be done using Foursquare API. This provides location-based experiences with diverse information about venues, users, photos, and check-ins. We only need top 100 venues within a radius of 1500 meters for each zip code. After getting this data we need to find Indian restaurants returned by Foursquare API.

To analyze the data, we first use One Hot Encoding. To run machine learning algorithms on the data, we need numerical data about the existence of the venues. One hot encoding helps us do that by creating new (binary) columns to indicate the presence of each possible value from the original data. This means that each venue in each zip code will be labeled as 1 in their correct category.

After this, we group the dataframe by the zip codes to get the mean of the frequency of occurrence of each venue category.

Next, we run the cluster analysis to cluster the zip codes and then examine these clusters and find out the average frequency of Indian restaurants in each cluster. Finally, we will compare the average frequency of Indian restaurants and demographics of each zip code which contains the percentage of Indian population. This will help us to find a location where we can open an Indian restaurant.

DISCUSSION AND RESULTS

Based on our cluster analysis result, we found that cluster2 has highest frequency of Indian restaurants when compared to the rest of the clusters and cluster1 has the lowest. However, we can see that cluster3 has low frequency and high number of Indian restaurants and this maybe because of higher number of zip codes in that cluster. Cluster1 has highest number of zip codes but there are no Indian restaurants. To make a decision on which zip code would be better to open a restaurant, we looked at the demographics of Indian population in those zip codes. When we look at the demographics of top nine zip codes with highest to lowest Indian population, I found that most of the zip codes belong to cluster1. Which means that cluster1 has no Indian restaurants in the areas where the population of Indians are high. The most Indian populated zip code belongs to cluster0. Sometimes the reason for the zip codes having Indian population but no Indian restaurant is because the overall population for those zip codes is low. But there are some zip codes which have high overall population and high Indian population but no restaurants for example zip codes 78759,78758. Therefore, we can take these zip codes into consideration to open an Indian restaurant because they are highly populated as well as they have high Indian population.

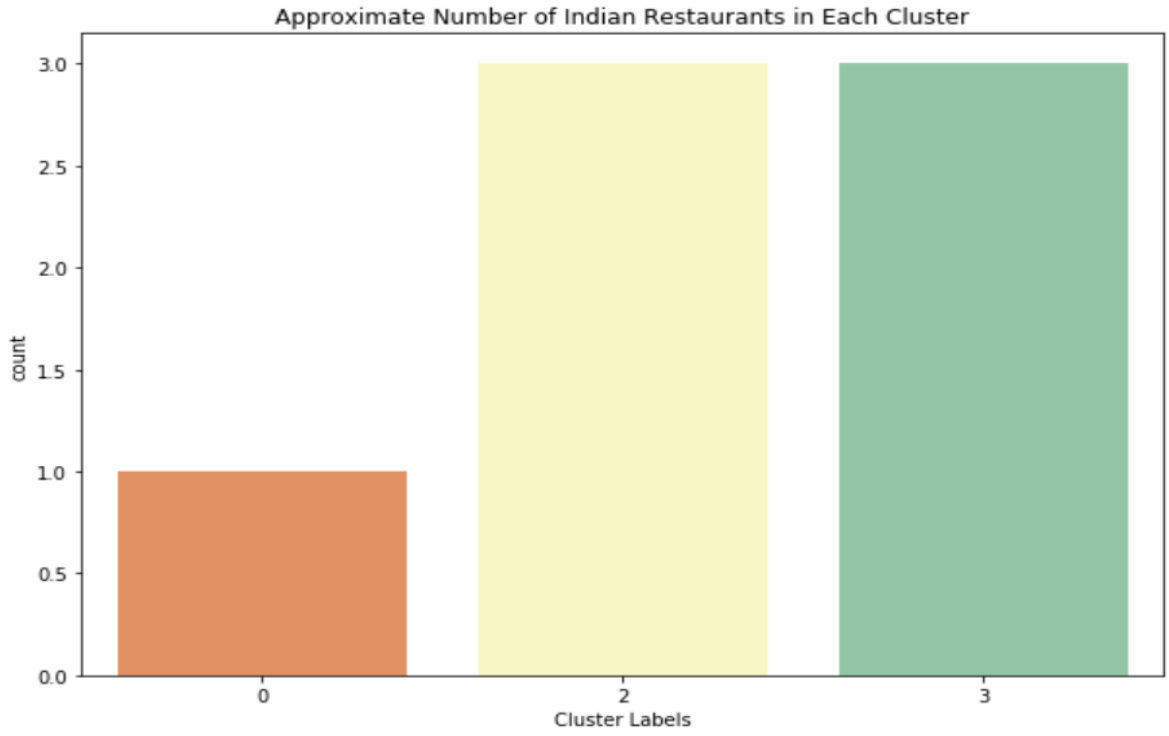
Fig: Average frequency of Indian restaurant for each cluster

In [178]:
```
c0 = clusters.loc[clusters['Cluster Labels'] == 0]
c1 = clusters.loc[clusters['Cluster Labels'] == 1]
c2 = clusters.loc[clusters['Cluster Labels'] == 2]
c3 = clusters.loc[clusters['Cluster Labels'] == 3]
```

In [179]:
```
print('Cluster 0 has an average frequency of Indian Restaurants of', c0["Indian Restaurant"].mean())
print('Cluster 1 has an average frequency of Indian Restaurants of', c1["Indian Restaurant"].mean())
print('Cluster 2 has an average frequency of Indian Restaurants of', c2["Indian Restaurant"].mean())
print('Cluster 3 has an average frequency of Indian Restaurants of', c3["Indian Restaurant"].mean())
```

```
Cluster 0 has an average frequency of Indian Restaurants of 0.013888888888888888
Cluster 1 has an average frequency of Indian Restaurants of 0.0
Cluster 2 has an average frequency of Indian Restaurants of 0.025561706284597852
Cluster 3 has an average frequency of Indian Restaurants of 0.01
```

After looking at each cluster, we can say that cluster2 has the highest frequency whereas cluster1 has the lowest.



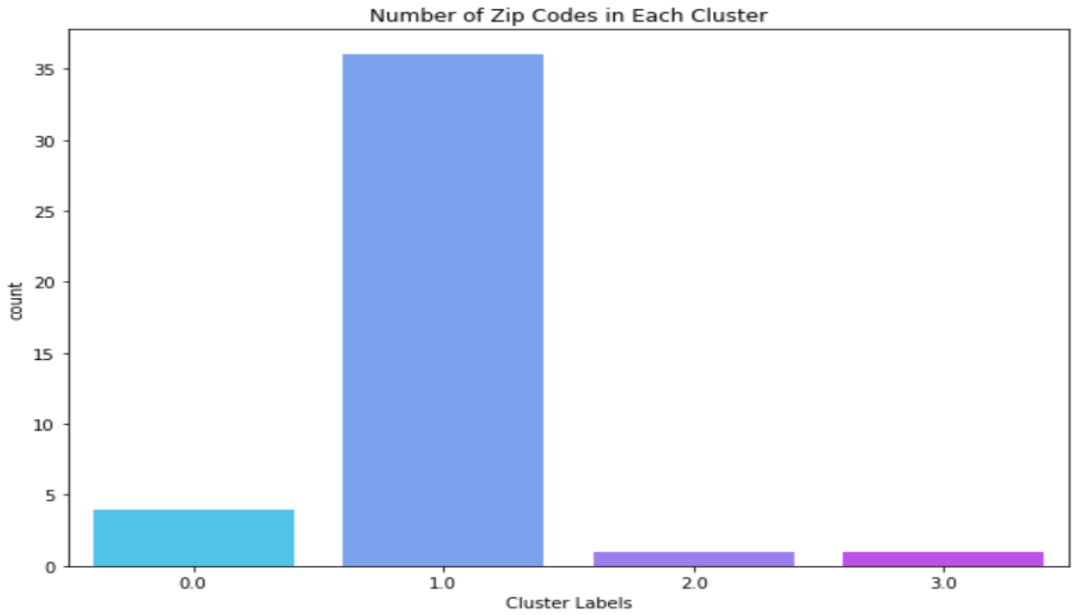


Fig: Demographic data with percentage of Indian population from most to least populated.

	Zip Code	Population	%Indians(Asian)	Latitude	Longitude	City	Cluster Labels	Indian Restaurant
0	78705	26825	0.0463	30.293474	-97.738268	Austin, Texas	3.0	0.010000
1	78728	17298	0.0425	30.453764	-97.686695	Austin, Texas	1.0	0.000000
2	78726	6480	0.0347	30.430488	-97.842530	Austin, Texas	1.0	0.000000
3	78717	8148	0.0271	30.488309	-97.764829	Austin, Texas	2.0	0.027027
4	78759	40547	0.0254	30.400789	-97.755969	Austin, Texas	1.0	0.000000
5	78727	22212	0.0220	30.428300	-97.718143	Austin, Texas	1.0	0.000000
6	78758	42820	0.0212	30.387634	-97.705310	Austin, Texas	1.0	0.000000
7	78746	26023	0.0206	30.296951	-97.811647	Austin, Texas	1.0	0.000000
8	78730	4885	0.0200	30.363632	-97.850355	Austin, Texas	1.0	0.000000
9	78751	14005	0.0187	30.310819	-97.722821	Austin, Texas	1.0	0.000000

LIMITATIONS

After discussing the possibility of opening an Indian restaurant we should also note some limitations to this analysis. To start with the coordinates taken from the website may not be accurate. They are only an approximation for the zip codes which are given along with the coordinates. The limit set to the Foursquare API for 100 venues within the radius of 1500 meters

is also another limitation which has to be considered. This is because there may be other Indian restaurants which were not included in 100 venues and not returned by Foursquare API. Lastly, the data collected for the demographics may also be just an approximation which may have changed. So, before opening a restaurant it is important to know various other factors which lead to success of the restaurant. Many of those factors include consumers, competitors, neighborhoods. For future analysis, considering the neighborhoods where more Indians reside can also give us an insight on where exactly we can open an Indian restaurant which can lead to more success and profits for the restaurant.

CONCLUSION

In this project we have gone through the process of identifying the business problem: where to open an Indian restaurant in Austin, Tx. We further collected data from a website which had information on the zip codes and its coordinates and also information on the demographics on Indian population in those zip codes. We then used Foursquare API to collect information on top 100 venues within the radius of 1500 meters in each zip code. We then used cluster analysis to analyze the data which was grouped into 4 clusters in total. From this analysis, we could find the average frequency of Indian restaurants and later used the demographic data to compare the zip codes with high Indian population and average frequency of Indian restaurants. From the results, we concluded that zip codes 78759, 78758 both have high overall population and high Indian population but there were not many Indian restaurants. I would therefore open an Indian restaurant in those zip codes if given a chance based on the cluster analysis results and demographic data which was provided.