

COURSERA CAPSTONE PROJECT REPORT

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

Business Problem

- The objective of this capstone project is to analyze and select the best location in the city of Austin, Texas to open a new Indian restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where can I open an Indian restaurant in Austin, Texas.

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.

➤ Sources of data:

- Gathered data from the website named Zip Atlas. <http://zipatlas.com/us/tx/austin/zip-code-comparison/percentage-indian-population.htm>
- Foursquare API for venue data

Data in panda dataframe

	#	Zip Code	Latitude	Longitude	City	Population	%Indians(Asian)	National Rank
0	1	78705	30.293474	-97.738268	Austin, Texas	26825	0.0463	#162
1	2	78728	30.453764	-97.686695	Austin, Texas	17298	0.0425	#202
2	3	78726	30.430488	-97.842530	Austin, Texas	6480	0.0347	#292
3	4	78717	30.488309	-97.764829	Austin, Texas	8148	0.0271	#453
4	5	78759	30.400789	-97.755969	Austin, Texas	40547	0.0254	#499

Methodology

- Collect data from <http://zipatlas.com/us/tx/austin/zip-code-comparison/percentage-indian-population.htm> Clean and process the data into the dataframe.
- Using Foursquare API, find top 100 venues within the radius of 1500 meters for the zip codes.
- Use K-means cluster analysis to cluster the zip codes and examine the data.
- Compare the cluster analysis results with demographic data which contains the percentage of Indian population.

Results

```
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]
```

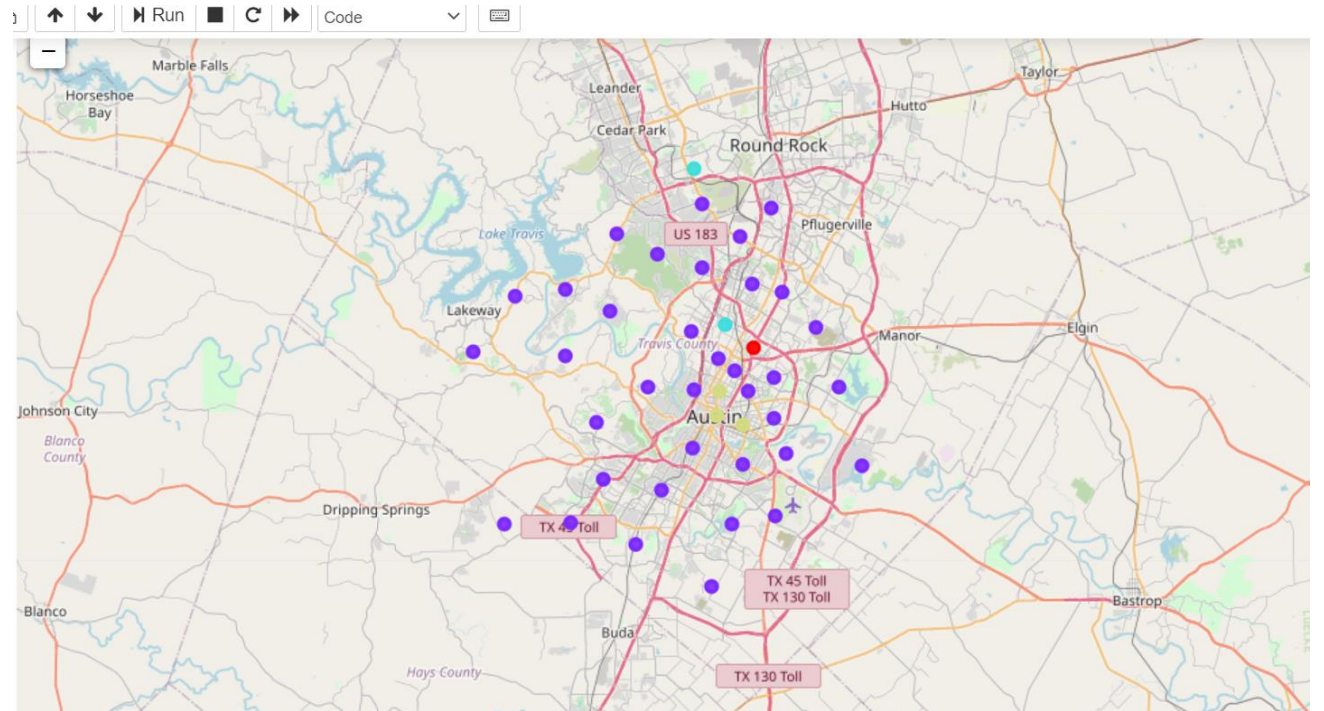
```
In [179]: ▶ print('Cluster 0 has an average frequency of Indian Restaurants of', c0["Indian Restaurant"].mean())
          print('Cluster 1 has an average frequency of Indian Restaurants of', c1["Indian Restaurant"].mean())
          print('Cluster 2 has an average frequency of Indian Restaurants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

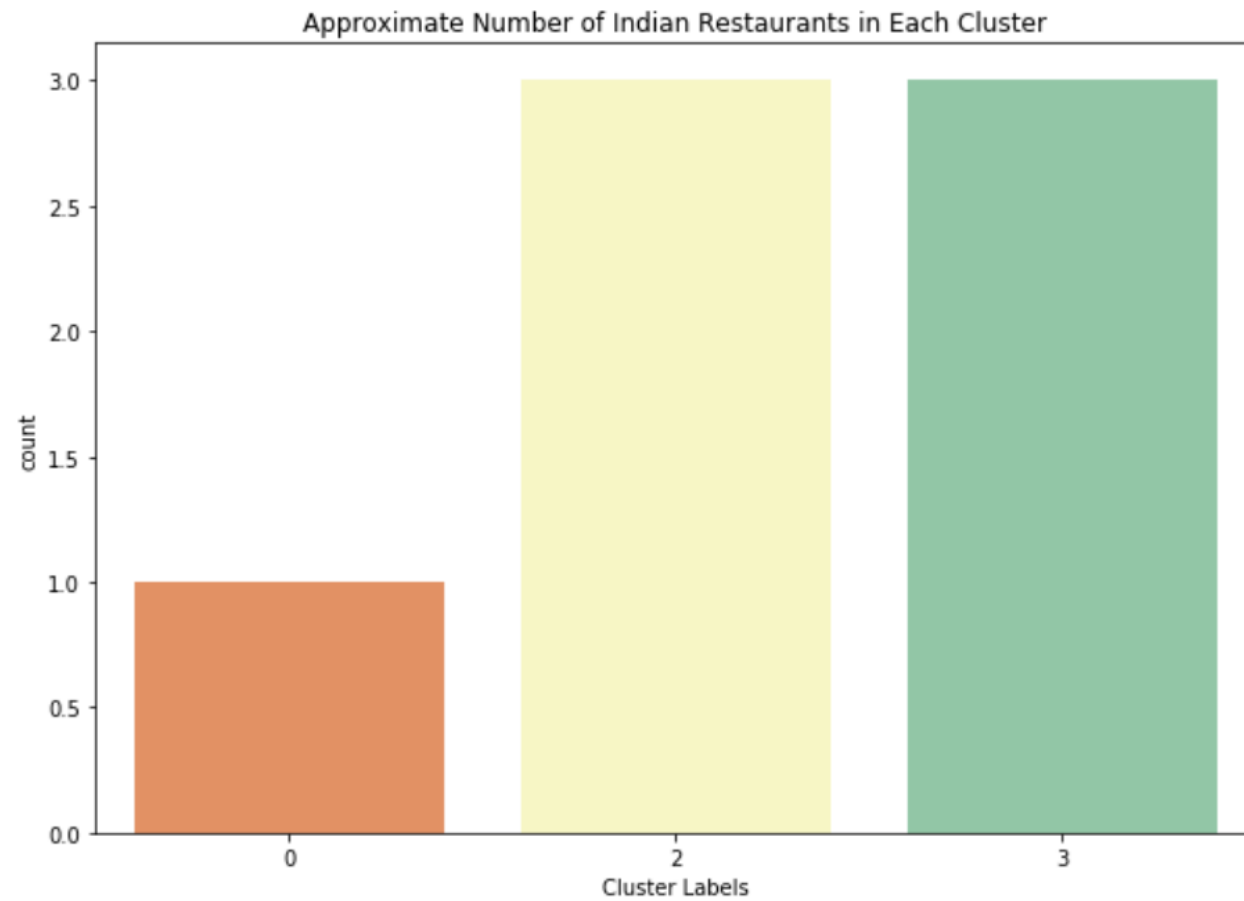
After looking at each cluster, we can say that cluster2 has the highest frequency whereas cluster1 has the lowest.

Results

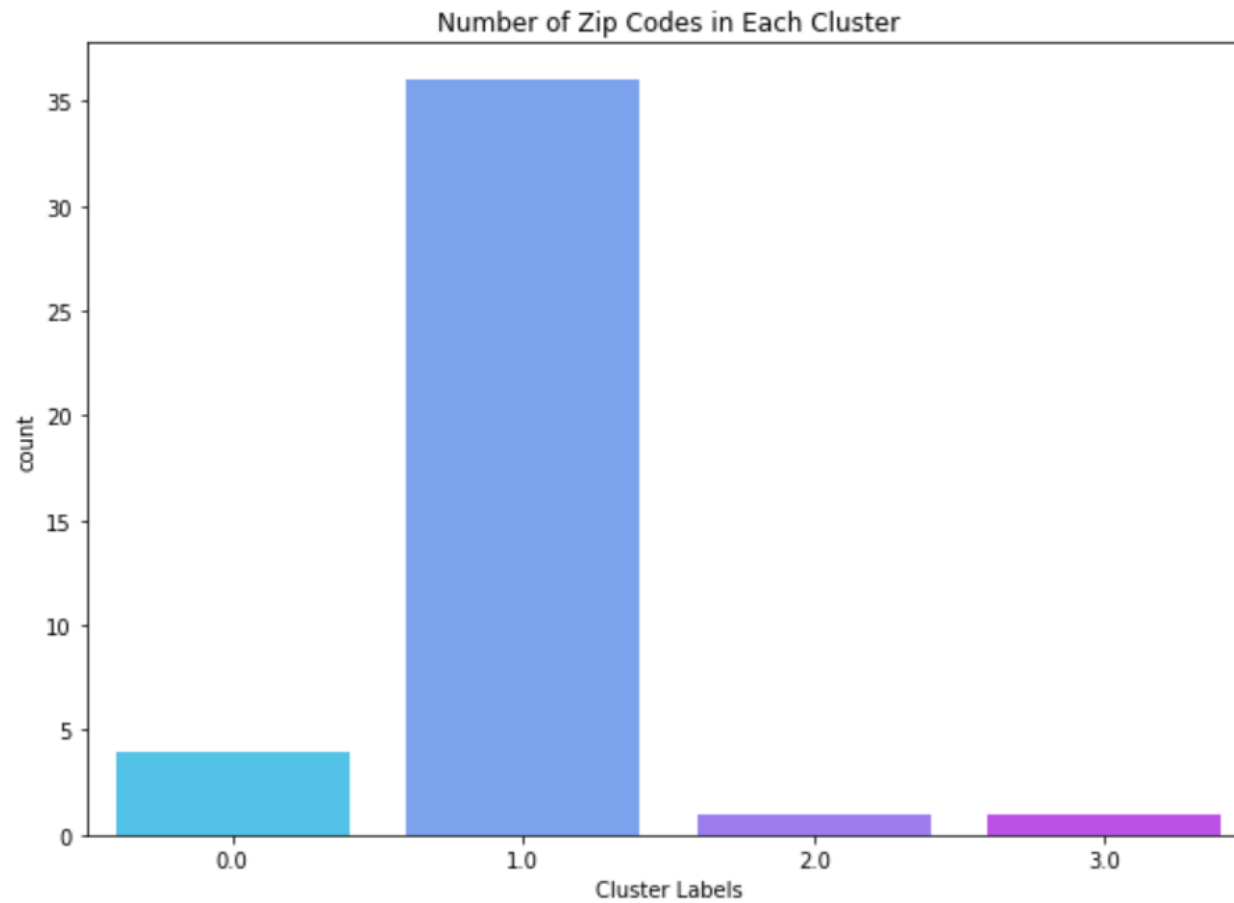
Cluster0-Red, Cluster1-Purple, Cluster2-Light Green, Cluster3-Blue. From the map, we can observe that most of the Indian Restaurants are located in Central part of Austin. Cluster0 and Cluster3 are located in the central part of the city whereas Cluster2 which has the highest frequency of Indian restaurants is located towards north of Austin.



Results



Results



Results

J.

	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

Results and Discussion

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.

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.