

Analyzing the Neighborhoods in Jaipur

for Starting a Restaurant

Applied Data Science Capstone Project

By: Samiksha Gupta

Date: May 23, 2021

Table of Contents

1. Introduction.....	3
2. Data Collection.....	4
2.1 Neighborhoods Data.....	4
2.2 Geographical Coordinates.....	4
2.3 Venue Data.....	5
3. Methodology.....	6
3.1 Feature Extraction.....	7
4. Model Tranning.....	10
4.1 Unsupervised Learning.....	10
5. Result.....	12
6. Discussion.....	15
7. Conclusion.....	16
8. Final Comments.....	17

1. Introduction

Jaipur is the capital and the largest city of the Indian state of Rajasthan. As of 2011, the city had a population of 3.1 million, making it the tenth most populous city in the country. Jaipur is also known as the *Pink City*, due to the dominant color scheme of its buildings. It is located 268 km (167 miles) from the national capital New Delhi.

Jaipur is a popular tourist destination in India and forms a part of the west Golden Triangle tourist circuit along with Delhi and Agra. It also serves as a gateway to other tourist destinations in Rajasthan such as Jodhpur, Jaisalmer, Udaipur, Kota, and Mount Abu. Jaipur is located 616 km from Shimla.

Personally, I have been brought up in Jaipur and have loved the city from the bottom of my heart. It is one of the major hubs of the world and is extremely diverse with people from various ethnicities residing here. The multi-cultural nature of the city of Jaipur has brought along with it numerous cuisines from all over the world. The people of India generally love food and I personally love to try different cuisines and experience different flavors. Thus, the aim of this project is to study the neighborhoods in Jaipur to determine possible locations for opening a restaurant. This project can be useful for business owners and entrepreneurs who are looking to invest in a restaurant in Jaipur. The main objective of this project is to carefully analyze appropriate data and find recommendations for the stakeholders.

2. Data Collection

The data required for this project is collected from various resources. The data and their source is described below.

2.1 Neighborhoods Data:

The neighborhood or location data is scrapped from website <https://www.zricks.com/Locality/Jaipur> and parsed using html parser using BeautifulSoup. The neighborhood column of the pandas dataframe is created by this location data.

2.2 Geographical Coordinates:

The geographical coordinates of the neighborhoods are obtained from GeoPy library in Python. The address or location is provided in the geocode function in the library and it returns the latitude and longitude of the provided address. We pass the locations that we scrapped from <https://www.zricks.com/Locality/Jaipur> website and we get latitude and longitude for every address. If any location in geopy library returns none value that we return 'Nan' for it. This data is relevant for plotting the map of Jaipur using the Folium library in python.

```
In [4]:  def get_lat_lng(address):
         geolocator = Nominatim(user_agent='foursquare_api')
         location = geolocator.geocode(address)
         if location == None:
             latitude = 'Nan'
             longitude = 'Nan'
         else:
             latitude = location.latitude
             longitude = location.longitude
         return(latitude,longitude)
```

Fig. 1 Function that convert the address into latitude and longitude

2.3 Venues Data:

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighborhoods in Jaipur and is used to study the popular venues of different neighborhoods.

```
In [12]: ► locations_df.head(10)|
```

```
Out[12]:
```

	neighborhood	latitude	longitude
0	22 Godam	26.902774	75.7927199
1	Achrol	27.1332143	75.9566333
2	Adarsh Nagar	26.8986698	75.8163567
3	Ajairajpura	Nan	Nan
4	Ajmeri Gate	26.91722115	75.81681843102842
5	Ambabari	26.9446462	75.7783245
6	Amer	26.9888269	75.8558916
7	Amrapali Circle	26.9115079	75.7438297
8	Anand Lok	Nan	Nan
9	Anand Nagar	26.9154576	75.8189817

Fig. 2 The Scrapped and structured data

3. Methodology:

Using folium, a map was plotted to show how the different neighborhoods are spread all over the Jaipur.

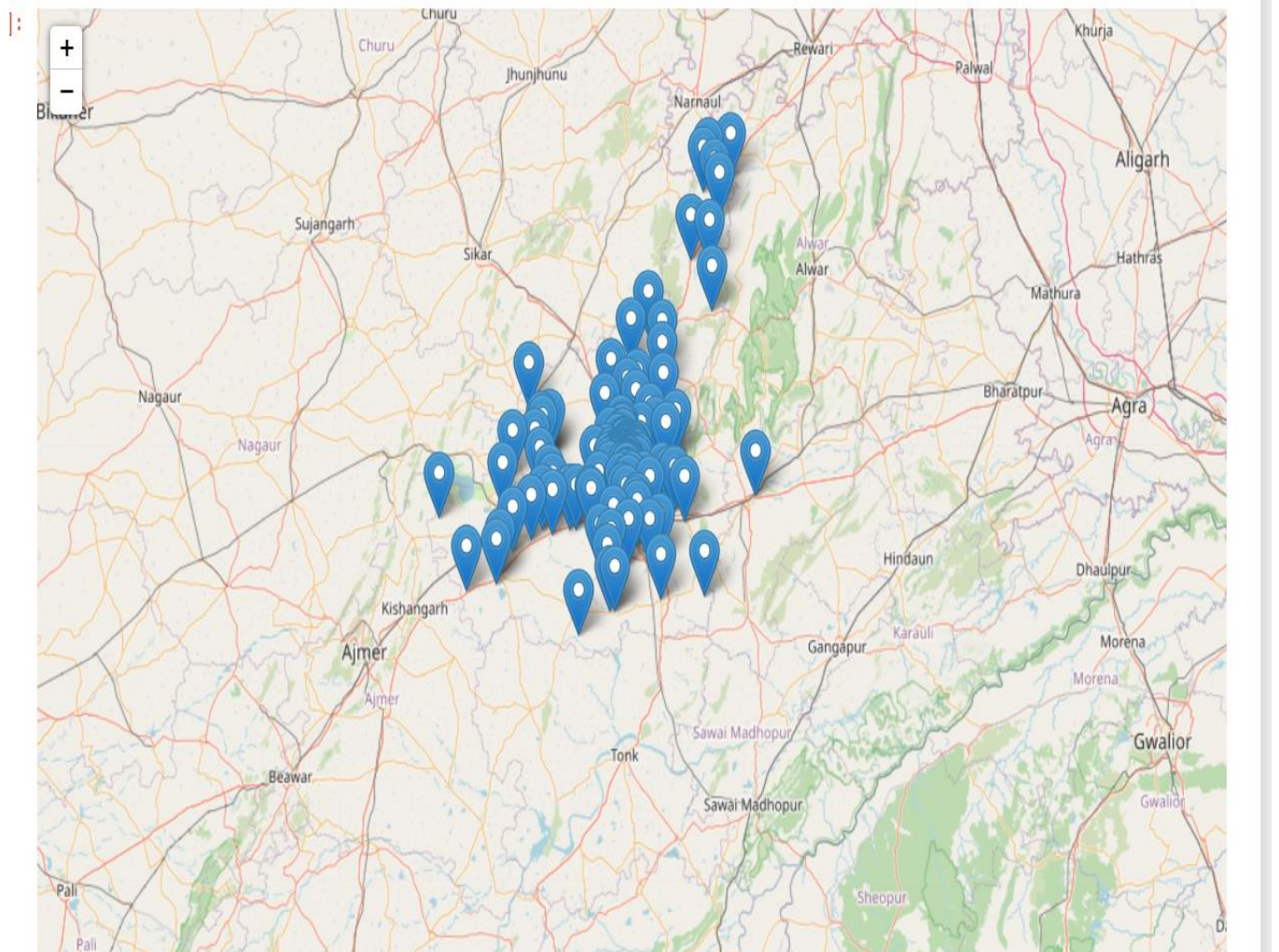


Fig. 3 Map of Jaipur and its neighborhood locations.

3.1 Feature Extraction:

Feature extraction was carried out to obtain features from the Foursquare API data which was used for building the unsupervised learning model. In order to achieve this, the “Venue Category” Column had to be converted to some form of numeric value to be used for building the model. This was achieved by one-hot-encoding method which takes all the unique categories and creates a column for each category. Then, if a neighborhood venue belonged to that category, it would get a value of 1 for that row in that row in that specific category column and if a neighborhood venue does not belong to the particular category, the value would be 0. This process was repeated for all the venues in all neighborhoods and the result was a sparse matrix containing the neighborhood name and all unique category columns with either 1 or 0 based on whether the neighborhood venue belonged to that category or not.

This dataframe was taken grouped by the neighborhood name and the average value was taken for all categories. The result is shown

in figure which shows only top 10 rows.

Out[100]:

	ATM	Antique Shop	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Astrologer	Athletics & Sports	Auto Dealership	...	Tea Room	Temple	Train Station	Travel Agency
Neighborhood															
22 Godam	0.000	0.0	0.034483	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0
Adarsh Nagar	0.000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.043478	0.0	0.0	0.0
Ajmeri Gate	0.000	0.0	0.000000	0.000000	0.0	0.064516	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0
Ambabari	0.000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0
Amer	0.000	0.0	0.000000	0.083333	0.0	0.083333	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0
Amrapali Circle	0.000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0
Anand Nagar	0.000	0.0	0.000000	0.000000	0.0	0.080000	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0
Anand Puri	0.000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.058824	0.0	0.0	0.0
Anita Colony	0.000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0
Arjun Nagar	0.125	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0

10 rows x 16 columns

Fig. 4 Neighborhoods and average value of each category

A dataframe was also created which contained the top 10 most common venues of all neighborhoods. Though this is not a part of Feature Extraction, it is important to provide a glimpse into what this dataframe looks like as it will be used later to combine the results from the unsupervised learning model.

Out[66]:

	Neighborhood	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 Most Common Venue	10th Most Common Venue
0	22 Godam	Hotel	Café	Shopping Mall	Indian Restaurant	Fast Food Restaurant	Department Store	Dessert Shop	Pizza Place	Lounge	Resort
1	Adarsh Nagar	Hotel	Café	Shopping Mall	Women's Store	Department Store	Pizza Place	Park	Indian Sweet Shop	Indian Restaurant	Lounge
2	Ajmeri Gate	Indian Restaurant	Hotel	Historic Site	Arts & Crafts Store	Coffee Shop	Café	Market	Zoo	Juice Bar	Dessert Shop
3	Ambabari	Hotel	Shopping Mall	Multiplex	Department Store	Coffee Shop	Clothing Store	Park	Asian Restaurant	Dessert Shop	Diner
4	Amer	Palace	Historic Site	Bed & Breakfast	Arts & Crafts Store	Café	Restaurant	Hotel	Indian Restaurant	Art Gallery	Well
...
110	Transport Nagar	Shopping Mall	Zoo	Convenience Store	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store
111	Udyog Nagar	ATM	Indian Restaurant	Diner	Food Truck	Dessert Shop	Dhaba	Dog Run	Donut Shop	Electronics Store	Fabric Shop
112	Vaishali Nagar	Pizza Place	Coffee Shop	Multiplex	Athletics & Sports	BBQ Joint	Sandwich Place	Fast Food Restaurant	Food	Dessert Shop	Dhaba
113	Vidhyadhar Nagar	Hotel	Pizza Place	Arts & Crafts Store	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop
114	Vishwakarma Industrial Area	ATM	Indian Restaurant	Diner	Business Service	Food Service	Dessert Shop	Dhaba	Dog Run	Donut Shop	Electronics Store

115 rows x 11 columns

Fig. 5 Neighborhoods and their most common venues

4. Model Training

Now we can use KMeans clustering method to cluster the neighborhoods.

First we need to determine how many clusters to use. This will be done using the Silhouette Score.

We will define a function to plot the Silhouette Score that will be calculated using different number of clusters.

4.1 Unsupervised Learning:

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes. A cluster refers to a collection of data points aggregated together because of certain similarities. You'll define a target number k , which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the center of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The '*means*' in the K-means refers to averaging of the data; that is, finding the centroid.

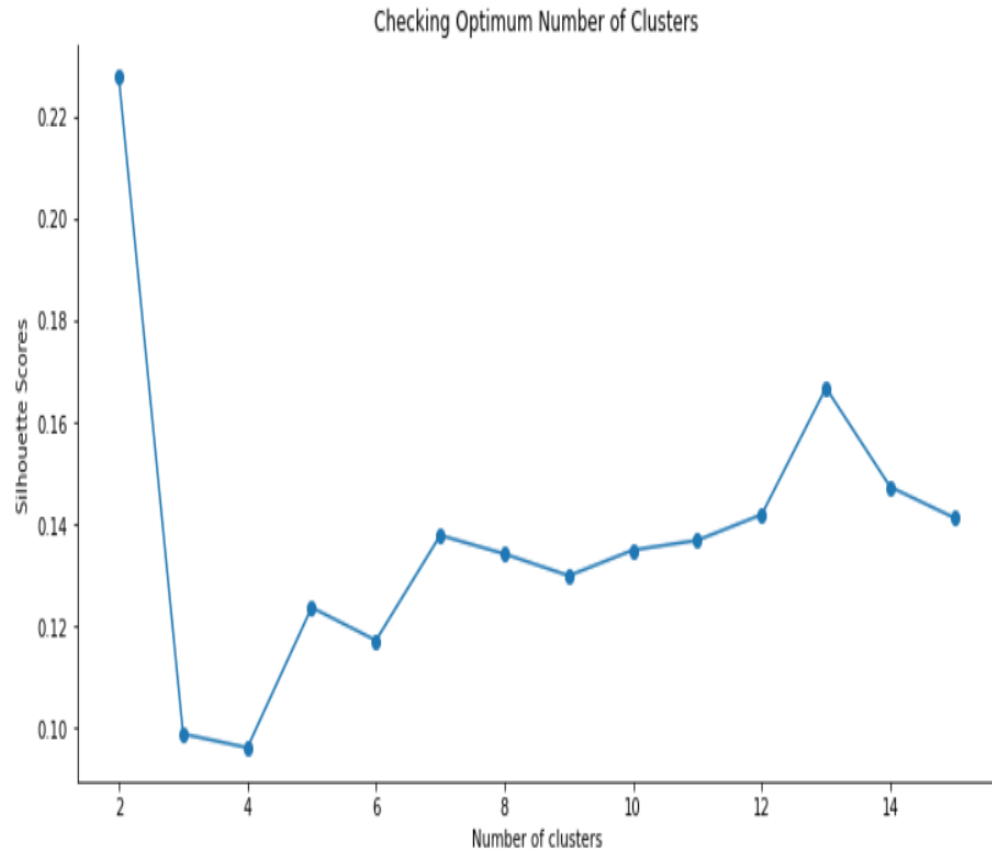


Fig. 6 Graph for number of clusters and model accuracy.

We can see that the silhouette scores are not very high even as we increase the number of clusters. This means that the inter-cluster distance between different clusters is not very high over the range of k -values. However, we will try to cluster our data as best as we can. For this, we will use 2 clusters for our clustering model since it provides the highest silhouette score as seen above.

5. Results

The K means model generate clusters of dataset in two categories and label them as 0 and 1.

They can be visualized by plotting them onto a map using Folium library.

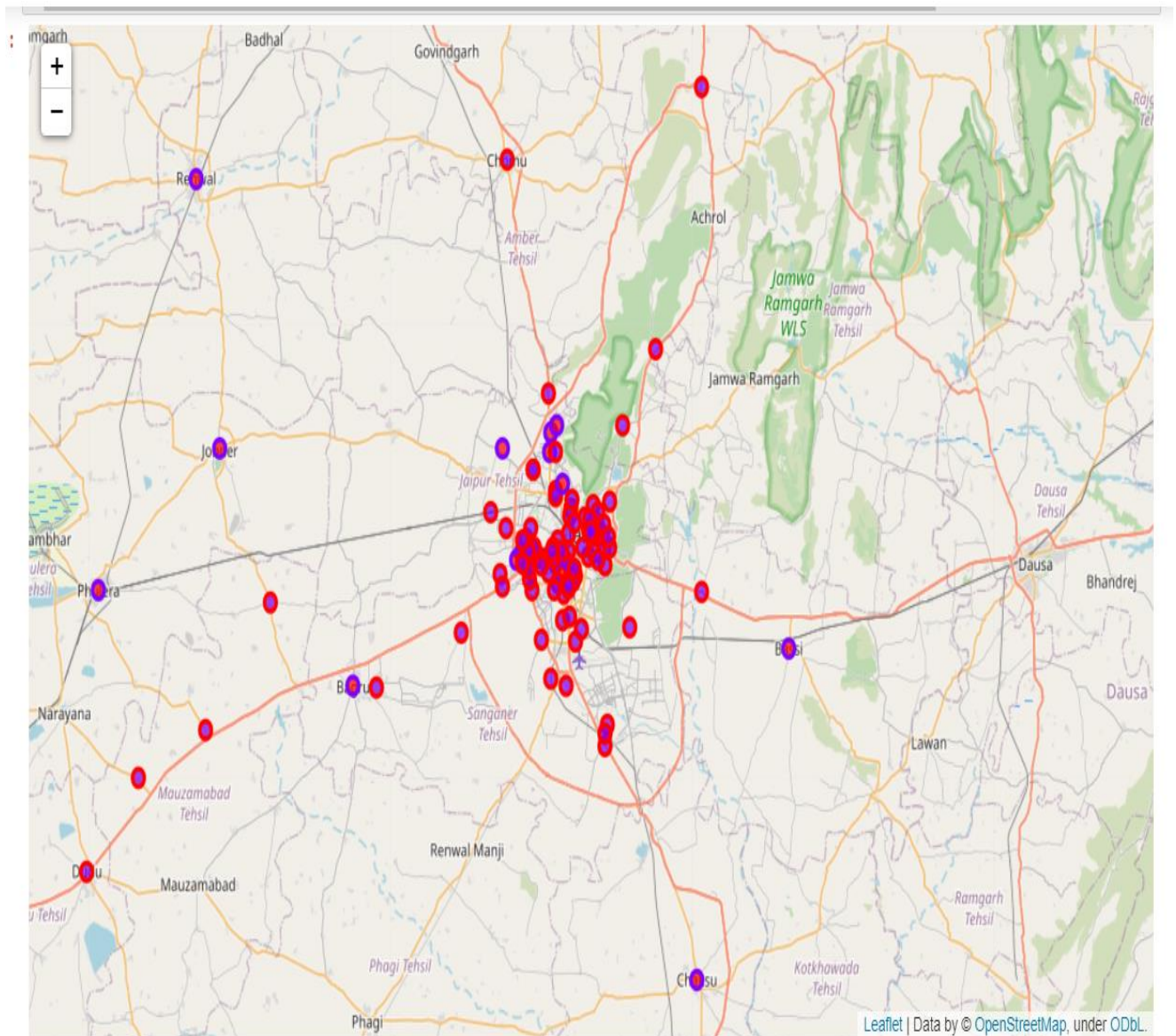


Fig. 7 Map having two clusters of the neighborhood data set

Table for cluster one having label 0 is showing below.

];

	neighborhood	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 Most Common Venue	10th Most Common Venue
12	Bagru	ATM	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store	Fabric Shop
19	Bassi	ATM	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store	Fabric Shop
25	Chaksu	ATM	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store	Fabric Shop
28	Chitrakoot	ATM	Athletics & Sports	Lawyer	Vegetarian / Vegan Restaurant	BBQ Joint	Multicuisine Indian Restaurant	Business Service	Fast Food Restaurant	Farm	Fabric Shop
50	Jobner	ATM	Platform	Asian Restaurant	Food Service	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store
57	Karni Vihar	ATM	Multicuisine Indian Restaurant	Athletics & Sports	Business Service	Health & Beauty Service	Lawyer	IT Services	Vegetarian / Vegan Restaurant	Train Station	Fast Food Restaurant
58	Kartarpura	ATM	Arts & Crafts Store	Café	Shopping Mall	Astrologer	Fried Chicken Joint	Historic Site	Food	Dhaba	Diner
59	Kotputli	ATM	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store	Fabric Shop
75	Nehru Nagar	ATM	Business Service	Electronics Store	Food Service	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Fabric Shop
79	Niwaru	ATM	Clothing Store	Fast Food Restaurant	Food Service	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store
81	Phulera	ATM	Train Station	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store
88	Renwal	ATM	Bus Station	Food Service	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store	Fabric Shop
99	Shiv Nagar	ATM	Snack Place	Indian Restaurant	Motorcycle Shop	Hotel	Art Museum	Food Truck	Diner	Dog Run	Donut Shop
111	Udyog Nagar	ATM	Indian Restaurant	Diner	Food Truck	Dessert Shop	Dhaba	Dog Run	Donut Shop	Electronics Store	Fabric Shop

Table for cluster one having label 1 is showing below.

	neighborhood	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 Most Common Venue	10th Most Common Venue
0	22 Godam	Hotel	Café	Shopping Mall	Indian Restaurant	Fast Food Restaurant	Department Store	Dessert Shop	Pizza Place	Lounge	Resort
1	Adarsh Nagar	Hotel	Café	Shopping Mall	Women's Store	Department Store	Pizza Place	Park	Indian Sweet Shop	Indian Restaurant	Lounge
2	Ajmeri Gate	Indian Restaurant	Hotel	Historic Site	Arts & Crafts Store	Coffee Shop	Café	Market	Zoo	Juice Bar	Dessert Shop
3	Ambabari	Hotel	Shopping Mall	Multiplex	Department Store	Coffee Shop	Clothing Store	Park	Asian Restaurant	Dessert Shop	Diner
4	Amer	Palace	Historic Site	Bed & Breakfast	Arts & Crafts Store	Café	Restaurant	Hotel	Indian Restaurant	Art Gallery	Well
...
108	Tilak Nagar	Indian Restaurant	Hotel	Arts & Crafts Store	Café	Market	Zoo	Fast Food Restaurant	Coffee Shop	Dessert Shop	Museum
109	Tonk Phatak	Hotel	Train Station	Jewelry Store	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop
110	Transport Nagar	Shopping Mall	Zoo	Convenience Store	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop	Electronics Store
112	Vaishali Nagar	Pizza Place	Coffee Shop	Multiplex	Athletics & Sports	BBQ Joint	Sandwich Place	Fast Food Restaurant	Food	Dessert Shop	Dhaba
113	Vidhyadhar Nagar	Hotel	Pizza Place	Arts & Crafts Store	Food Service	Department Store	Dessert Shop	Dhaba	Diner	Dog Run	Donut Shop

100 rows x 11 columns

6. Discussion

By analyzing the two clusters obtained we can see that one of the clusters are more suited for restaurants and hotels, whereas, other cluster is less suited. Neighborhoods in clusters 2, contain a small percentage of restaurants, hotels, cafe and pubs in their top 10 common venues. These clusters contain a higher degree of other venues like train station, bus station, fish market, gym, performing arts venue and smoke shop, to name a few. Thus, they are not well suited for opening a new restaurant. On the other hand, neighborhoods in clusters 1 contain a much higher degree of restaurants, hotels, multiplex, cafes, bars and other food joints. Thus, the neighborhoods in this cluster would be well suited for opening a new restaurant.

1]:

	neighborhood	latitude	longitude	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 Most Common Venue	10th Most Common Venue
0	22 Godam	26.902774	75.792720	Hotel	Café	Shopping Mall	Indian Restaurant	Fast Food Restaurant	Department Store	Dessert Shop	Pizza Place	Lounge	Resort
1	Adarsh Nagar	26.898670	75.816357	Hotel	Café	Shopping Mall	Women's Store	Department Store	Pizza Place	Park	Indian Sweet Shop	Indian Restaurant	Lounge
2	Ajmeri Gate	26.917221	75.816818	Indian Restaurant	Hotel	Historic Site	Arts & Crafts Store	Coffee Shop	Café	Market	Zoo	Juice Bar	Dessert Shop
3	Ambabari	26.944646	75.778324	Hotel	Shopping Mall	Multiplex	Department Store	Coffee Shop	Clothing Store	Park	Asian Restaurant	Dessert Shop	Diner
4	Amer	26.988827	75.855892	Palace	Historic Site	Bed & Breakfast	Arts & Crafts Store	Café	Restaurant	Hotel	Indian Restaurant	Art Gallery	Well

Fig. 8 Neighborhoods that are suitable for opening a restaurant

7. Conclusion

The neighborhoods in cluster 1 contain a variety of food joints like restaurants, tea rooms, bakery, cafe, steakhouse and pubs and also contain very diverse cuisines like Japanese, Indian, Chinese, Italian and seafood restaurants.

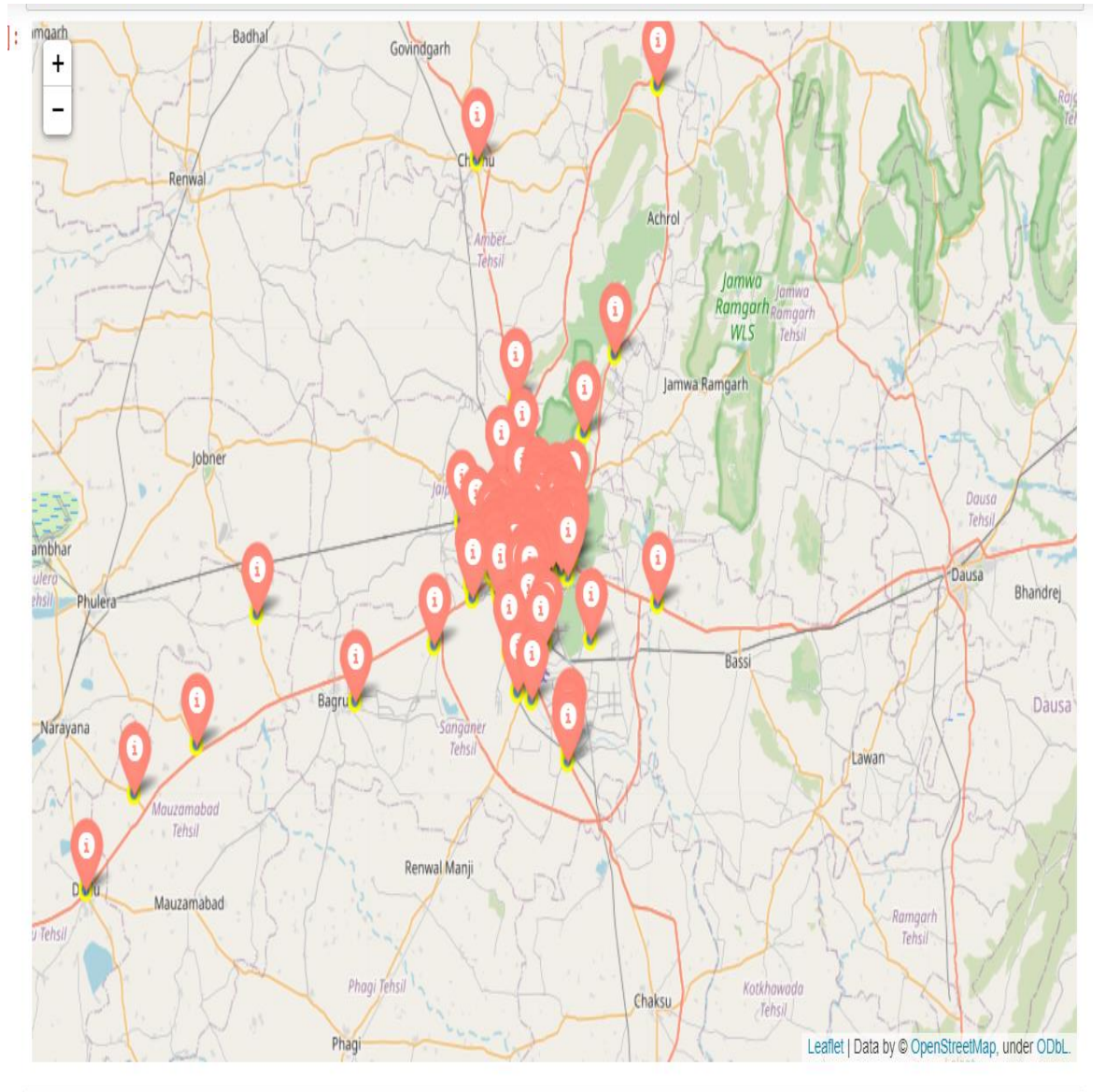


Fig. 8 Locations that are good for opening a restaurant.

8. Final Comments

We have successfully analyzed the neighborhoods in Jaipur, India for determining which would be the best neighborhoods for opening a new restaurant. Based on our analysis, neighborhoods in cluster 1 are recommended as locations for the new restaurant. This has also been plotted in the map above. The stakeholders and investors can further tune this by considering various other factors like transport, legal requirements, and costs associated. These were out of the scope for this project and thus were not considered.