

Opening up a new Indian Restaurant in Denver CO

Mohandas Palatshaha

Jan 2021

Author Note

This project is part of Applied Data Science course deliverable

Abstract

Indian food is one of the most famous dining items in the world and the folks around the world are usually in search of fine Indian Restaurant for enjoying Indian spicy taste.

Our project talks about the business problem of searching for the best place in Denver, CO to open up an Indian restaurant so that it should be a success

Keywords: Machine Learning, Coursera Capstone

Opening up a new Indian Restaurant in Denver CO

The project aims at analyzing the locations in Denver Colorado for opening up a new Indian restaurant. We will be using Foursquare APIs to get the location information and different attributes of various locations to understand the spread of Indian restaurants across the city. We will try to refine this data using machine learning techniques such as clustering and will try to find out the answer for: what would be the best place within Denver to open Indian restaurant so that it would be successful.

Data Gathering

We will be looking for the sub urbs within Denver Co and operating Indian restaurants around the sub urbs within Denver.

Along with this dataset we will be needing the technical details like latitude and longitudes of the gathered restaurants data to visualize it. Also, Indian restaurant venues would be used to cluster the sub urbs to understand the concentration of Indian restaurants across the city so that the least competition spot could be understood

Data Sources

We will be using Wikipedia page to get the sub urbs from Denver CO

(https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Denver) Web scrapping methods would be used to extract this data and convert it into relational dataset. Lat long for the sub urbs would be extracted from Geocoder library of Python. Indian Restaurant venues would be extracted using Foursquare APIs.

Methodologies Used

The project aims to use Web Scrapping technique, data pull using APIs (Foursquare API) and machine learning technique such as K-means clustering to solve for the business problem.

We will be using Folium library to visualize the output obtained and based on this we will try to gather actionable insights

Details of the methodology used is as follows:

For analyzing the best spot for opening up Indian Restaurants in Denver, CO we need the list of the sub urbs in Denver area which is available in Wikipedia at

(https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Denver). The data available here is not a relational data. To convert it into data-frame we have used web scraping technology using BeautifulSoup library from Python. Once we get the data scrapped and converted into data frame we need to add the latitude and longitude for this dataset so that we can easily visualize it. For the purpose we used Geocoder library which helps to pull up lat long for the given dataset.

This dataset now provides us the basic data for starting our analysis further. We are interested in understanding spread of already available Indian Restaurants across Denver CO, to achieve this dataset we will be using Foursquare APIs to get the top venues for the given places. We will be pulling up top 50 venues within 5000 meters of the radius of Denver. For the purpose we have created a free developer account on Foursquare. The dataset then gathered from Foursquare is merged with our original data frame.

The venue data pulled up contains categorical variables which is not much helpful for analysis purpose, to make the analysis meaningful we use onehot encoding on the dataset. As we are interested only in 'Indian Restaurants' we filter out the data for only Indian Restaurant venue.

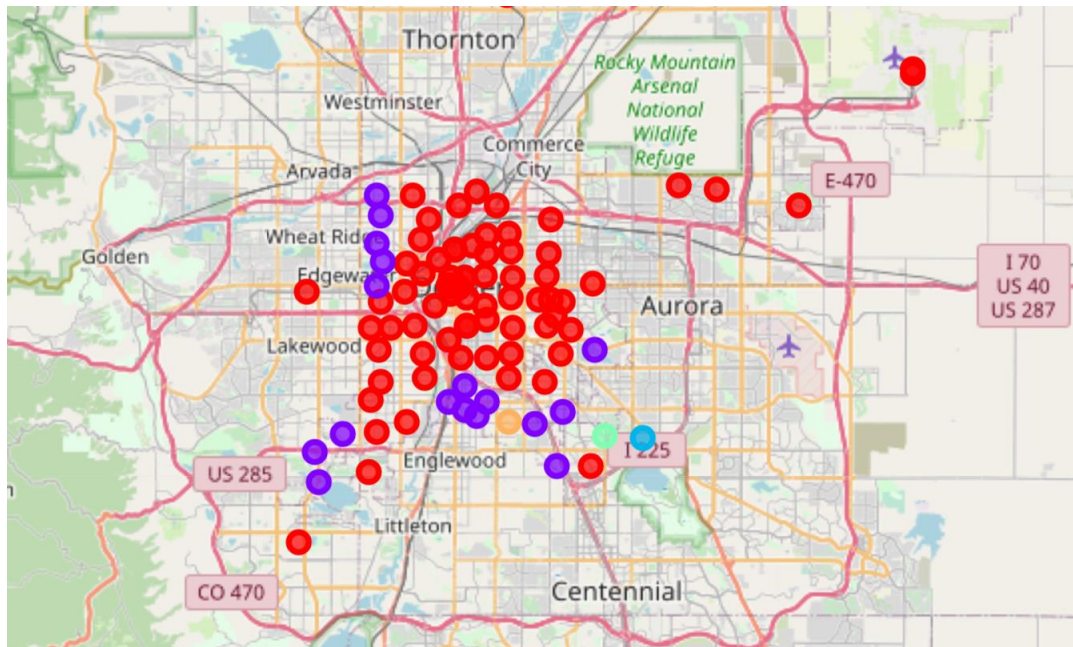
Now our dataset is ready to apply clustering to understand the concentration of the Indian restaurants across the city. We split up neighborhood into 5 clusters based on the frequency of the Indian restaurants in the city, we use K-means clustering to understand the clusters. The output

obtained clearly shows the concentration of the Indian Restaurants across the city which potentially help us answering the business questions.

Based on the results obtained from clustering, we can see following results:

1. Cluster 0: 78 restaurants
2. Cluster 1: 17
3. Cluster 2: 1
4. Cluster 3: 1
5. Cluster 4: 1

As we can observe in the visualization below, most of the Indian restaurants are concentrated around northern and central part of the city (Cluster 0 and Cluster 1) but that is not the case in southern part of the city. The outskirts, near to Englewood, Centennial we hardly see any Indian Restaurant, hence if the new Indian Restaurant would be opened here, we would definitely see lesser competition which could easily be converted into the success of the business.



Way forward

While working on this project, we analyzed only one factor to solve for the business problem i.e. the frequency of the Indian Restaurant at given geography. Though location is one of the most important factor to consider while opening a new business, that is not the only factor, population spread across, favorite tastes of the people living across, ease of transportation to the site are also the deciding factors which should be taken into consideration to take this work further. The results obtained henceforth would be able to derive at more meaningful insights

References

- https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Denver
- <https://developer.foursquare.com/docs>