Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

Predicting suitable location for a new Indian restaurant
Girish
May 20, 2020

1.1 Background
Indian restaurant or Dhaba or Punjabi dhaba is a roadside restaurant in the Indian subcontinent. They are on highways, generally serve local cuisine, and also serve as truck stops. They are most commonly found next to petrol stations, and most are open 24 hours a day. Since many truck drivers in the Indian subcontinent are of Punjabi descent, and Punjabi food and music are popular throughout the Indian subcontinent, the word dhaba has come to represent any restaurant that serves Punjabi food, especially the heavily spiced and fried Punjabi fare preferred by many truck drivers.

Dhaba roadside eateries are a common feature on Punjab's national and state highways. Earlier frequented only by truck drivers, today eating at a dhaba—urban or roadside—is a trend. Thus, Punjabi dhaba has become a part of the culture of the Punjabi people.

1.2 Problem
Client is keen to know a suitable location to open a new Indian restaurant as part of their business expansion in NewYork.

1.3 Interest
Obviously, our client and Others who are interested in Indian food cuisine would also be interested.

Data

Data acquisition and cleaning

2.1 Data sources

GeoSpace data
Data source : https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm Description : By using this geo space data we will get the New york Borough boundaries that will help us visualize choropleth map.

-New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
Data source : https://cocl.us/new_york_dataset
Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.

-Indian resturants in each neighborhood of new york city. Data source : Fousquare API To achieve this, we would be using the Foursquare location data explore New York in particular to the restaurants and food habits of the people staying in the neighborhood and analyze the performance of existing restaurants via their ratings.

The idea is to chart out potential locations which has Indian traits. We would consider ratings of restaurants to bring insights of how good those restaurants are. The results will be show-cased on map for better understanding.

Based on definition of our problem, factors that will influence our decission are:

- number of existing restaurants in the neighborhood (Indian restaurant)

Additional or auxillary data sources. Following data sources will be needed to extract/generate the required information:
Yelp_url="https://www.yelp.com/search?find_desc=Punjabi+Restaurant&find_loc=New+York%2C+NY+11418"
Wiki_url="https://en.wikipedia.org/wiki/Indian_Americans"
The idea is to chart out potential locations which has Punjabi traits. We would consider ratings of restaurants to bring insights of how good those restaurants are. The results will be show-cased on map for better understanding.

Based on definition of our problem, factors that will influence our decission are:

* number of existing restaurants in the neighborhood (Punjabi restaurant)

* distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

Yelp_url="https://www.yelp.com/search?find_desc=Punjabi+Restaurant&find_loc=New+York%2C+NY+11418"

Wiki_url="https://en.wikipedia.org/wiki/Indian_Americans"

Methodology

In this project we will direct our efforts on detecting areas of New York that have low restaurant density, particularly those with low number of Punjabi restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required data: location and type (category) of every restaurant within 6km from Ozone Park. We have also identified Punjabi restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of New York - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Punjabi restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Panjabi restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Results and Discussion

Our analysis shows that although there is are lots of restaurants in New York with high density near the Airport Area. As we move away from the Airport area, we could see a gradual decrease in the number of Indian Restaurants.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Alexanderplatz) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Italian restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Purpose of this analysis was to only provide info on areas close to New York and for Punjabi Restaurants using FourSqaure API. Although we had problems fetching the required information from Foursqaure with regard to "Punjabi" we manually scraped data related to New York and from websites for those areas, reasons which would make them suitable for a new restaurant regardless of competition. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify potential areas in New York to host a new Punjabi Restaurant away from density distribution.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.