Coursera IBM Data Science Capstone Project

Opening a new Asian restaurant in Western part, Germany

Prepared by:

Sowjanya

Introduction

For this Capstone project, I am creating a hypothetical scenario for a concept Indian restaurant who wants to explore opening an authentic Indian restaurant in West Germany area. The idea behind this project is that there may not be enough Indian restaurants in Germany and it might present a great opportunity for this entrepreneur who is based in Germany. As Indian food is very similar to other Asian cuisines, this entrepreneur is thinking of opening this restaurant in locations where Asian food is popular (many Asian restaurants in the neighborhood). With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

Business Problem

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Burmese restaurant in west, Germany. By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to the business question: In German, if an entrepreneur wants to open a Indian restaurant, where should they consider opening it?

Target Audience

The entrepreneur who wants to find the location to open authentic Indian restaurant

Data

To solve this problem, I will need below data:

- List of neighborhoods in west, Germany.
- Latitude and Longitude of these neighborhoods.
- Venue data related to Asian restaurants. This will help us find the neighborhoods that are most suitable to open a Indian restaurant.

Extracting the data

- Scrapping of west Germany neighborhoods via Wikipedia
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

Methodology

I got the neighbourhoods data for germany. I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I tried using Geocoder package but it was not working so I used the csv file provided by IBM team to match the coordinates of Toronto neighborhoods. After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

Next, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Here, I made a justification to specifically look for “Asian Restaurants”. so I looked for the restaurants closest to Indian cuisine taste.

Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeriods, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “Asian food”. Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

Results

Clusters

The results from k-means clustering show that we can categorize Western Germany neighborhoods into 3 clusters based on how many Asian restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with little or no Asian restaurants
- Cluster 1: Neighborhoods with no Asian restaurants
- Cluster 2: Neighborhoods with high number of Asian restaurants

.

Recommendations

Observations

Most of Asian restaurants are in Cluster 2 which is around Bad Neuenhr Ahrweiler, Bremerhaven areas and lowest (close to zero) in Cluster 1 areas which are Butzbach areas. Also, there are good opportunities to open near la pizza, italian restuarant as the competition seems to be low. Looking at nearby venues, it seems Cluster 1 might be a good location as there are not a lot of Asian restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Burmese restaurant in these locations with little to no competition. Nonetheless, if the food is authentic, affordable and good taste, I am confident that it will have great following everywhere =)

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

References

Foursquare Developer Documentation: https://developer.foursquare.com/docs