Coursera Capstone

IBM Applied Data Science Capstone

Opening an Indian Restaurant in New York City, USA

By: Kunal Gupta

August 2020

INTRODUCTION

- Indian cuisine is very popular around the world. Hot curries with lots of chillies and a side of raita to cool down. Dishes are based on rice and often vegetarian or with seafood. Coriander, ginger, cumin, cardamom, saffron and nutmeg flavored flavor makers/spices.
- As the popularity of different international cuisines is increasing worldwide and Indian diaspora being one of the largest among the world. Many international food chains do not want to miss this opportunity to set-up their restaurants.
- As the popularity of different international cuisines is increasing worldwide and Indian diaspora being one of the largest among the world. Many international food chains do not want to miss this opportunity to set-up their restaurants.

BUSINESS PROBLEM

- The objective of this capstone project is to analyze and select the best locations in New York City, USA to open a Indian cuisine restaurant.
- Using Data Science Methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the question:
    - If a successful owner of multiple mid to high-end restaurants decided to open a new Indian restaurant in New York City, where would you recommend they open it?

DATA

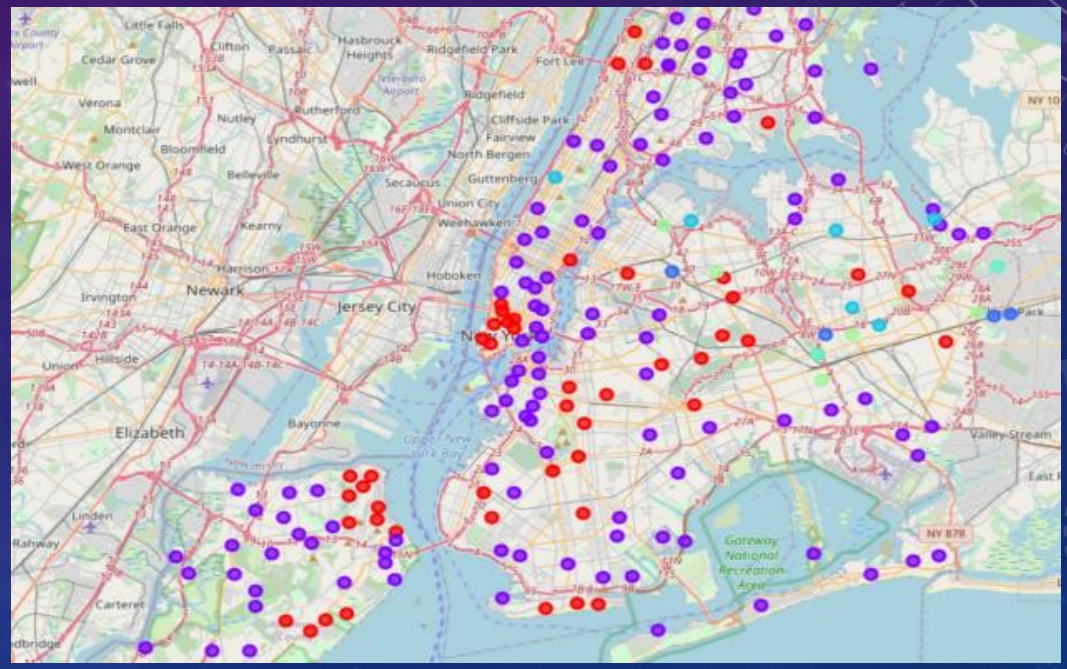
- **Data Required**
  - ✓ List of Neighborhoods in New York City.
  - ✓ Latitude and Longitude coordinates of each neighborhood.
  - ✓ Venue data of each neighborhood.
- Sources of Data
  - ✓ yourneighborhood.co
  - ✓ Geocoder Package for Geographical Coordinates
  - ✓ Foursquare API for Location Data

METHODOLOGY

- Web scraping the webpage for the list of neighborhoods.
- Getting the geographical coordinates using geocoder package.
- Using Foursquare API to get venue data.
- Group data by neighborhoods and taking the mean of the frequency of occurrence of each venue category.
- Analyzing approaches:
    1. Filtering venue category for Indian restaurants and using k-means clustering.
    2. Finding top 10 most popular venues in each neighborhood and performing K-means clustering.
- Visualizing the results of both approaches on map using folium.

RESULTS

- Approach 1 (Indian Restaurant Concentration) : Categorized into 5 clusters
  - ✓ Cluster 0 and 3: These clusters have very less number of Indian restaurants.
  - ✓ Cluster 1: This cluster have no Indian restuarants.
  - ✓ Cluster 2 and 4: These clusters have moderate number of Indian restaurants.
  - ✓ Cluster 5: This cluster have large number of Indian restaurants.



- Approach 2 (Gastronomy of the neighborhood): Clustered the data into 2 clusters
  - ✓ Cluster 0: In this cluster, we can see that there are a lot of gastronomy related venues like coffee shop, pizza place, Thai restaurant, Mexican Restaurants, pub, etc.
  - ✓ Cluster 1: In this cluster, gastronomy is not represented as pizza places and fast food are in top.

DISCUSSION

- Approach 1(Indian Restaurant concentration):
  - ✓ Most of the Indian restaurants are located in the eastern and southern part of the New York City, with the highest number in cluster 5 and moderate number in cluster 2 and 4, very less number in cluster 0 and 3 with no Indian restaurant in cluster 1 which is the central area.
- Approach 2(Gastronomy of the neighborhood):
  - ✓ Most of the gastronomy related venues are present in the northwestern area of the city which are part of cluster 0. While in cluster 1 there are venues which does not represent gastronomy which are present in all other parts of the city.

CONCLUSION

- Approach 1(Indian Restaurant concentration):
  - ✓ we can conclude that Cluster 0 will be great to start an Indian restaurant because of less competition.
- Approach 2(Gastronomy of the neighborhood):
  - ✓ Cluster 0 of the second approach will be the best option to start a highend Indian restaurant.
- Finally we can conclude that the neighborhoods present in the intersection of Cluster 0 of approach 1 and cluster 0 of approach 2 will be the great choice to open an Indian restaurant.

Note – the list of neighborhoods in the clusters in both approaches is different.

THANK YOU

By: Kunal Gupta

August 2020