Coursera Capstone

IBM Applied Data Science

OPENING A NEW INDIAN RESTAURANT IN LONDON CITY

BY, NIVEDITHA

JUNE 2020

Business Problem

- The objective of this capstone project is to analyze and select the best locations in the London city to open a new restaurant.
- What is the best location for an Indian restaurant in London City?
- In what Neighborhood should I open an Indian restaurant to have the best chance of being successful?

Data

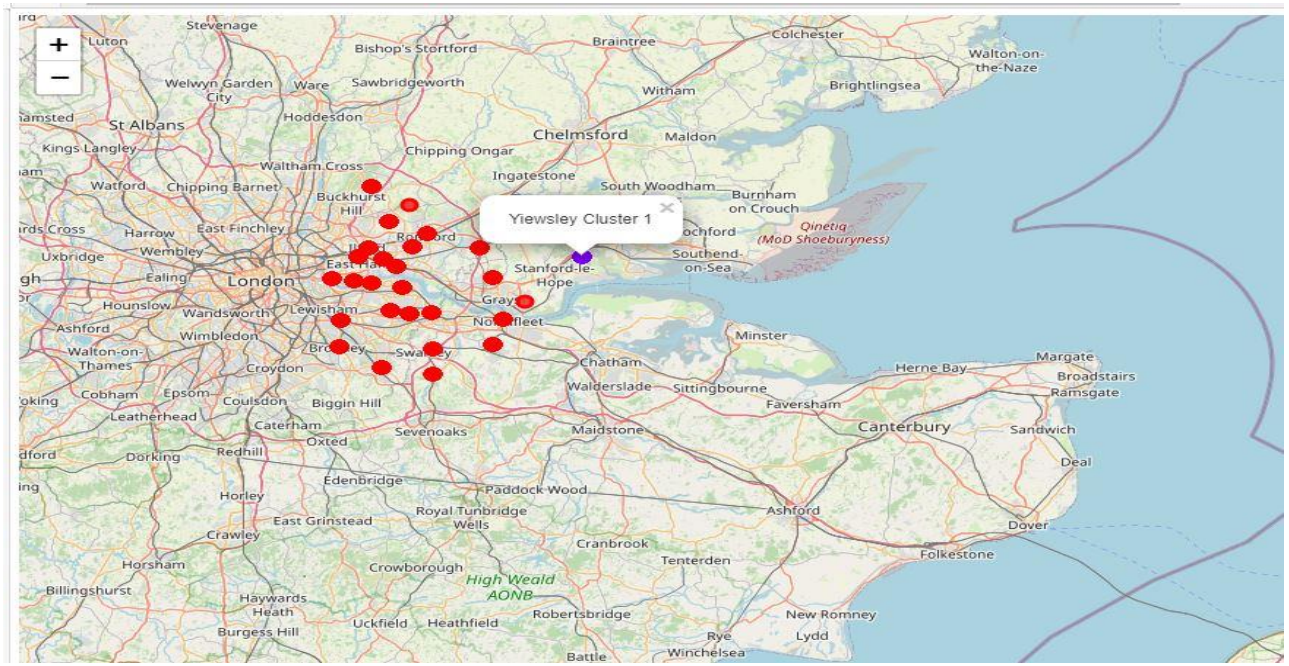
- Data Required
  - List of neighbouthoods in London
  - Latitude and Longitude coordinates of the neighborhoods
  - Venue data related to Indian restaurants.
- Sources of data
  - Wikipedia page of neighbourhoods (https://en.wikipedia.org/wiki/List_of_areas_of_London)
  - A csv file from the dataset with London borough coordinates
  - Foursquare API for venue data

Approach

- Web scraping Wikipedia page of neighborhoods list.
- Get the coordinates from csv file.
- Use Foursquare API to get venue data.
- Group data by neighborhood and taking the mean of the frequency of occurrence of each venue.
- Filter venue for Indian Restaurant.
- Perform clustering to find the best k-means
- Visualize the clusters in the map using Folium

Results

- Categorize the neighborhood into 2 clusters based on the K value
  - Cluster 0 Neighborhood with high number of Indian restaurants
  - Cluster 1 with very dense number of restaurants



Discussions

- Most of the Indian restaurants are concentrated in central London.
- Highest number is cluster 0
- Cluster 1 has very low number in terms no. of Indian restaurants

Recommendations

- Opening a new Indian restaurants in the neighborhoods of cluster 1.
- You can also consider the neighborhoods near to Heathrow airport as there is a major chance in the business

Conclusion

- Answer to the business question: the neighborhoods of cluster 1 is the most preferred locations
- Findings of this project may help stakeholders to capitalize on the opportunities of high potential locations avoiding the crowded areas.
- As a final note, all of the above analyses are depended on the adequacy and accuracy of Four-Square data. A more comprehensive analysis and future work would need to incorporate data from other external databases.

Thank you