Applied Data Science Capstone - Final Project

Table of Content

1. BUSINESS PROBLEM

- According to www.food.gov.uk, there are more than 14,000 restaurants in London and about 9 million people. That is why opening a new restaurant there can be an extremely challenging task. According to several surveys, up to 40% of such start-ups fail in the very first year. Let's suppose, an investor has enough time and money, as well as a passion to open the best eating spot in London. What type of restaurant would it be? What would be the best location for it?

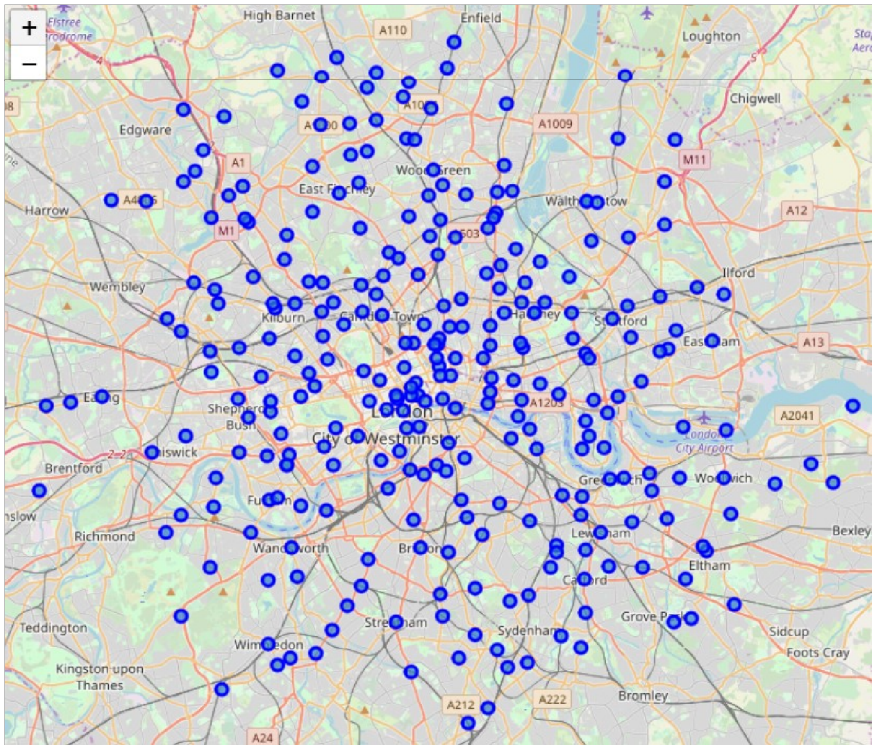
- Target audience: investors, entrepreneurs, and chefs interested in opening a restaurant in London, who may require objective advice regarding what type of restaurant would be most successful and where exactly it should be opened.

2. Methodology and Data

- Step 1. Using a table on https://en.wikipedia.org/wiki/List_of_areas_of_London, collect information about London boroughs and locations, excluding records whose "Post Town" is not London.
- Step 2. Use the Geopy and Folium libraries to get the coordinates of all locations and map geospatial data on a London map.
- Step 3. Using Foursquare API, collect the top 100 restaurants and their categories for each location within a 500 meter radius.
- Step 4. Group collected restaurants by location and by taking the mean of the frequency of occurrence of each type, preparing them for clustering.
- Step 5. Cluster restaurants by k-means algorithm and analyze the top 10 most common restaurants in each cluster.
- Step 6. Visualize clusters on the map, thus showing the best locations for opening the chosen restaurant.

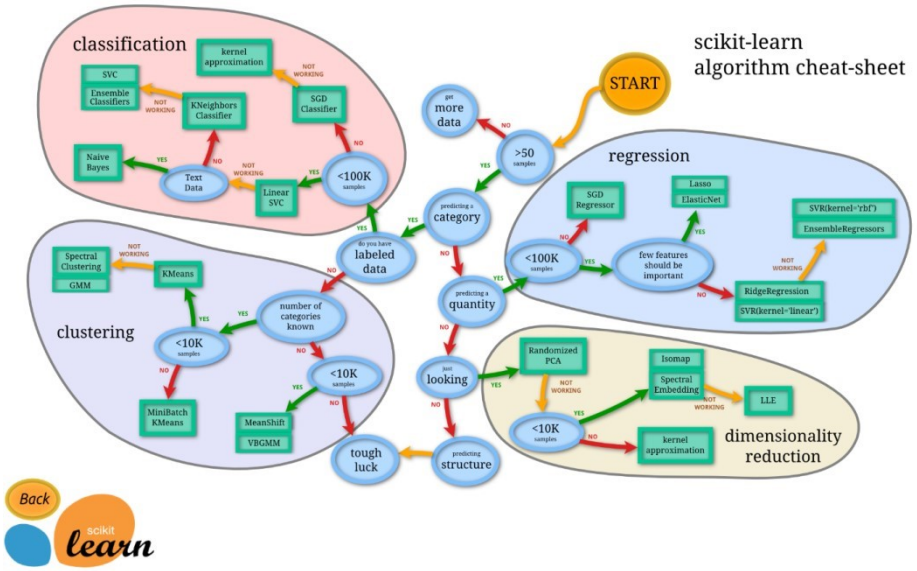
3. Exploring London Neighborhoods

- Using Wikipedia we collected all of London's neighborhoods.
- Using geospatial libraries we added geographical coordinates.
- With some cleaning and wrangling, we obtained 288 locations in London.



4. Exploring London Restaurants

- We utilized the Foursquare API to get the top 100 restaurants in each neighborhood.
- Thus we obtained 6357 restaurants of 131 individual types.
- We applied one-hot encoding and grouped them by taking the mean of the frequency of occurrence of each type.
- Then we clustered restaurants using the k-means algorithm based on their types similarity. The k-means is an unsupervised machine learning algorithm for clustering unlabeled data.



5. Results

Cluster #1

- The most common restaurant: Café.
- The 10th most common restaurant: Eastern European Restaurant.

Cluster #2

- The most common restaurant: Café.
- The 10th most common restaurant: Empanada Restaurant.

Cluster #3

- The most common restaurant: Café.
- The 10th most common restaurant: Eastern European Restaurant.

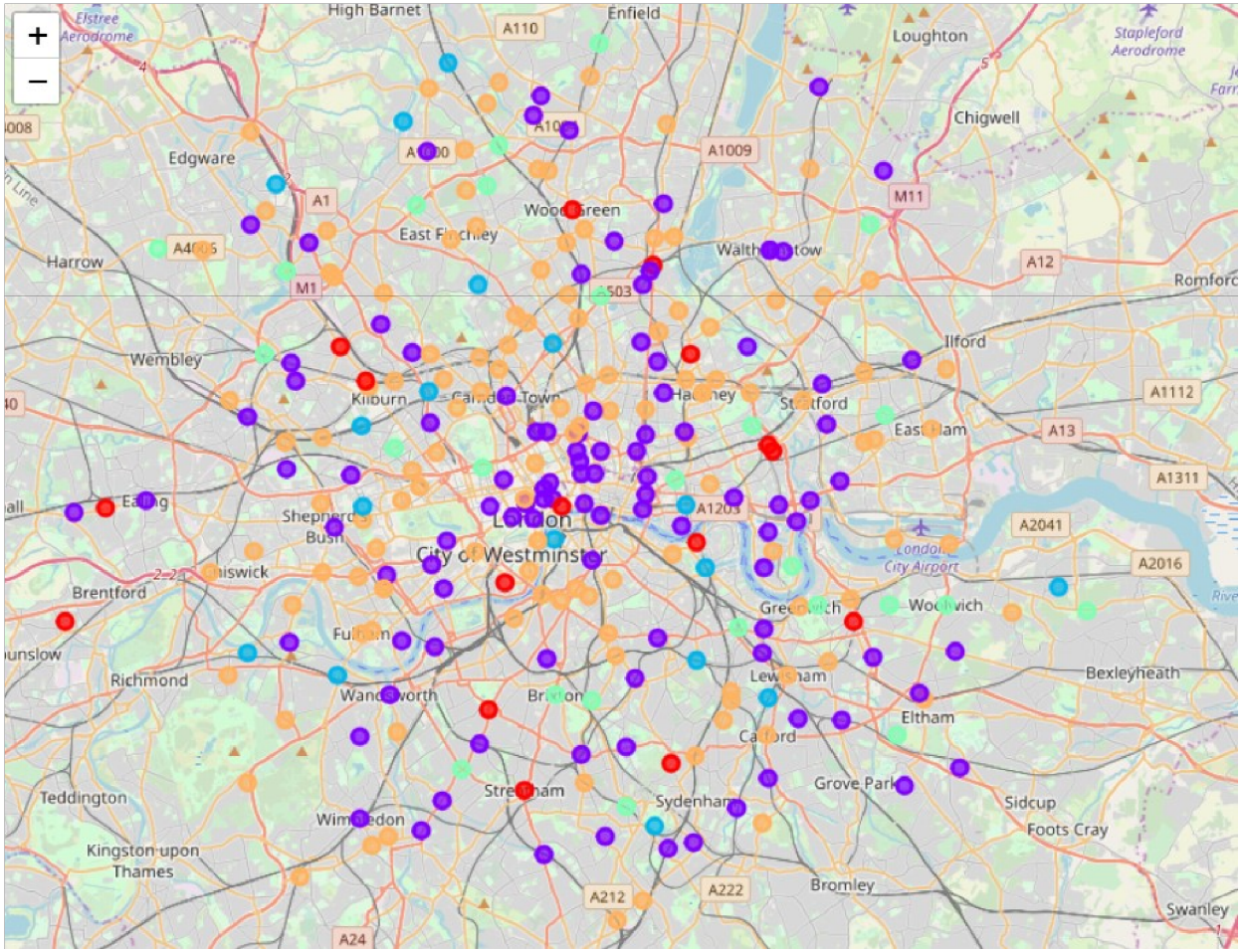
Cluster #4

- The most common restaurant: Indian Restaurant.
- The 10th most common restaurant: Eastern European Restaurant.

Cluster #5

- The most common restaurant: Café.
- The 10th most common restaurant: Eastern European Restaurant.

5. RESULTS - Visualizing Clusters



Cluster 1 - red
Cluster 2 - purple
Cluster 3 - blue
Cluster 4 - green
Cluster 5 - orange

6. Discussion

- Analyzing the most popular restaurants in each cluster, the stakeholder should select the least popular types within the top 10 as a safe choice. In our recommendations we advise selecting from the 10th or 9th positions. This selection is a reasonable balance between being too popular and having no customers.

- Recommendations, based on description of each cluster:

  Cluster 1 Locations: Eastern European or Dumpling Restaurant

  Cluster 2 Locations: Empanada or Ethiopian Restaurant

  Cluster 3 Locations: Eastern European or Ethiopian Restaurant

  Cluster 4 Locations: Eastern European or Dumpling Restaurant

  Cluster 5 Locations: Eastern European or Dumpling Restaurant

- After the type of restaurant is chosen, it is time to select a suitable location, using the final map.

7. Conclusion

- In this report we established a methodology to determine what the most promising type of restaurant is, and where it should be opened.
- This type of analysis can be applied to any city of your choice that has available geospatial information.
- This type of analysis can be applied to any type of venue (shopping, clubs, etc.) that is available in Foursquare database.