**Capstone Project**

**The Battle of Neighborhoods**

*By: Gargi Misra*

A picture containing sky, outdoor, water, city

Description automatically generated

**Table of Contents**

1. Introduction/Business Problem
2. Data Description
3. Methodology
4. Results
5. Conclusion
6. **Introduction/Business Problem**

If you are a first-time visitor to London, it can seem like an overwhelming place because there are so many popular places to visit, such as the iconic clock tower called ‘Big Ben’, London Bridge, London Eye, Westminster Abbey, Buckingham Palace, Windsor Castle, museums, and the list goes on. Although 11.3 million visitors are expected in 2021, which is nowhere near the 21 million visitors in 2019, it is still a very large number for a city with a population of approximately 9 million. Besides so many tourist attractions, what kinds of restaurants are available in London? This project will help answer this question using Foursquare API, which is an independent location data platform that provides diverse detailed information about venues and their locations.

1. **Data Description**

**For this project, geographical data, such as boroughs (neighborhoods), coordinates (latitude and longitude) will be needed to display these locations on maps.**  
**Wikipedia is a good source to retrieve London’s list of boroughs and geopy library will be used to get coordinates for those boroughs. Additionally, information regarding venues in each borough will be obtained from Foursquare API, which is an independent location data platform that provides diverse detailed information about venues and their locations.**

**Data Source for London’s boroughs:**

<https://en.wikipedia.org/wiki/List_of_London_boroughs>

1. **Methodology**

* Exploration Data Analysis
  + Find the names of London’s boroughs, area and population by web-scraping
  + Clean the data so that irrelevant, incomplete or null values are eliminated
  + Get the geographical coordinates (latitude and longitude) of all the boroughs
* Data Visualization
  + Use Foursquare API to retrieve information about the different types of venues, categories and their count
  + Obtain information about different types of restaurants
  + Visualize the types of restaurants by their count using Seaborn library
* Cluster Boroughs of London
  + Cluster boroughs by calculating the frequency of restaurant types using K-means algorithm from sklearn clustering library
  + Visualize these clusters using Folium, a map rendering library

1. **Results**

London has 32 boroughs and Barnet has the highest population of 395,896 within an area of 33.49 square miles and Bromley has the highest area of 57.97 square miles with 332,336 people in 2019. Barnet is more densely populated.

Table

Description automatically generated

Table

Description automatically generated

The Folium library plots a map where we can see the all the boroughs in London.

Map

Description automatically generated

Fast Food and Italian Restaurants have the highest number of restaurants compared to other restaurants followed by Restaurant category, which is not very descriptive in the food it offers. Indian Restaurant has the fourth highest frequency of restaurants in London.

Chart, bar chart

Description automatically generated

When I checked for the Top 10 Most Common Venues for each Borough, there was a variety of restaurants depending on the borough. For instance, if you were in Barnet, the most common restaurant would be Fast Food, whereas, if you were in Croydon, you would find more Indian restaurants listed there. These are the first five boroughs:

Table

Description automatically generated

In order to find clusters of similar restaurants, we used kmeans algorithm, and the Elbow Method showed the best k-value, which is the number of clusters. For this project, the best number of clusters was 4. Once again, the folium map showed the 4 clusters of 10 most common restaurant types in London.

Map

Description automatically generated

Cluster 1:

Table

Description automatically generated

Cluster 2:

Table

Description automatically generated

Table

Description automatically generated

Cluster 3:



Cluster 4:



A picture containing chart

Description automatically generated

1. **Conclusion**

According to the results, Cluster 2 has the highest number of boroughs with the most variety of cuisine types. Kingston upon Thames has the most variety in the restaurants (27) followed by Islington (26), whereas, Redbridge and Richmond upon Thames has the least number of venue categories in restaurants. The cluster map shows a concentration of purple dots that are in the heart of London, which is Cluster 2.

Cluster 1 had missing values for restaurants for boroughs: Barking and Dagenham, Bexley, Brent, Greenwich, Newham, Sutton, Tower Hamlets, and Waltham Forest which is an anomaly. Further research would be needed to figure out why there were no listings of restaurants in these boroughs, however, it is out of the scope of this project.