

Project Title: The Beverage Battle of the Neighborhoods in London, UK

INTRODUCITON:

Coffee, tea, and alcohol have been an integral part of British people's lives. The British consume 60 billion cups of tea per year, according to the Tea and Infusions Organisation. Regarding coffee, a study, named Project Café 2017 UK, conducted by the Allegra World Coffee Portal, suggested that coffee is a booming business as the total UK coffee shop market grew 6% in 2016 with an estimated 22, 845 outlets (see <https://www.morningadvertiser.co.uk/Article/2017/03/24/Coffee-shops-to-become-new-local-outnumbering-pubs-by-2030>). More pubs and bars in the UK recently started selling coffee and tea in addition to alcoholic drinks in order to provide customers a wide diversity of choices. London is the top one commercial city and the most populated place in the UK. As such, it is not surprising to learn that this city has the largest number of coffee shops, cafés, pubs, and bars.

Given these observations, there is a great need for coffee shops, cafés, pubs, and bars to have stable and quality supplies of beverages or/ and their ingredients.

Problem:

The city is facing a problem: The beverage supplies are not reliable and less ideal. Many coffee shops, cafés, pubs, and bars in London heavily rely on suppliers located outside of the city and therefore suffer unstable and delayed supplies. This problem, however, opens a window of opportunity for the beverage supply market. A stakeholder is interested in setting up a company to supply beverages and the ingredients to the coffee shops, cafés, pubs, and bars in the city of London, but would need to investigate the most ideal location for the company.

DATA DESCRIPTIONS:

Data needed for this project is from 'the List of areas of London' on Wikipedia (https://en.wikipedia.org/wiki/List_of_areas_of_London).

The dataset contains 533 areas in London, UK; each of which contains information of 'Location', 'Borough', 'Post town', 'PostalCode', 'Dial code', 'OS grid ref'. After data downloading, the column name, 'Location' was replaced with 'Neighborhood' to make it easier to understand. Moreover, the column, 'Dial code', was removed as it will not be used later.

To obtain the information of latitude and longitude of these areas, 'OS grid ref' was converted to latitude and longitude with this link (<https://gridreferencefinder.com/batchConvert/batchConvert.php>). A new data frame was subsequently created to store the converted information, and merged with the first one.

METHODOLOGY:

First of all, the python packages needed for data cleaning, visualization, and analysis, including the BeautifulSoup, wikipedia, pandas, numpy, geopy, folium, scipy, and sklearn, were installed and imported to the Jupyter Notebook.

After clearing the data, I only focused on data in the post town, London, to ensure that the business location derived later would be in the main London area. Next, I

garnered information about the top 5 most common venues in each neighborhood of the main London area using the Foursquare API.

The exploratory analysis aimed to offer an understanding of customers' favorite neighborhoods for coffee, tea, and alcohol. To this end, I applied k-means clustering with k=5 to segmenting the neighborhoods based on information of the 5 most common venues in the respective neighborhoods.

Judging by the output of the top 3 most common venues, it could be apparent to decide which cluster had the highest frequency of the top 3 most common venues being either pubs, bars, coffee shops, or cafés. Nevertheless, I conducted a chi-squared test using the stats function of 'scipy' to examine if this conclusion was statistically supported.

For the cluster with the highest frequency of all of the top 3 most common venues being one of the target categories (i.e., coffee shops, cafés, pubs, and bars), it was further analyzed. Within this cluster, the locations of the neighborhoods with all of the top 3 most common venues being the target categories would be presented in the map using the folium package.

To narrow down the choices of neighborhoods for the company location, I selected neighborhoods showing all of the top 4 common venues being the aforementioned target categories. I also made use of the 'search for venues' function in the Foursquare with the search terms, 'coffee', 'tea', 'pub', and 'bar', to identify places where coffee, tea, and alcohol were sold in the identified neighborhoods as the potential customers for this new beverage and ingredient supply company.

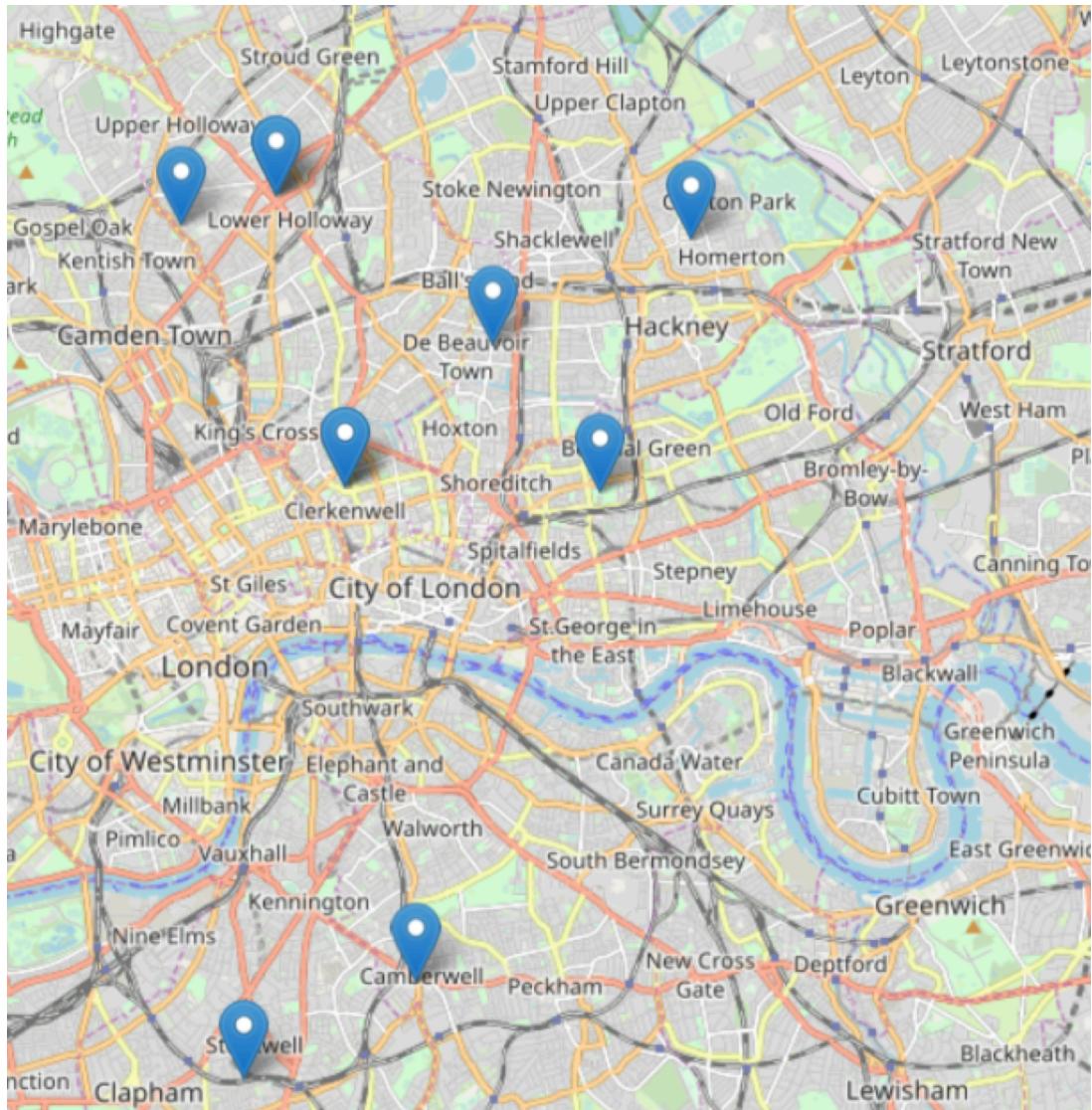
RESULTS:

Among the five clusters, Cluster 1 encompassed the largest number of neighborhoods where the top 3 most common venues were the target categories. This observation was supported by the result of chi-squared test ($\chi^2 = 286.57$, $P < 0.05$).

Within Cluster 1, there were 10 neighborhoods with the top 3 most common venues all being the target categories. However, Clerkenwell and Finsbury shared the same latitude and longitude; similarly, Holloway and Nag's Head shared the same latitude and longitude. Therefore, there were 8 geographically distinct neighborhoods. The exact locations of the 8 neighborhoods are presented in Figure1. To narrow down the neighborhoods, I sought for neighborhoods with all of the top 4 most common venues being the target categories. A single neighborhood, named Stockwell, met such a criterion.

Finally, I conducted a search for coffee shops, tea shops, cafés, pubs, and bars in this neighborhoods as the potential customers for this new beverage and ingredient supply company to reach out. As a result, there were 30 potential customers per search term located in Stockwell.

Figure 1. Distributions of the 8 neighborhoods with the top 3 most common venues being one of the following: coffee shop, café, pub, or bar



DISCUSSION:

To determine the most suitable location for the beverage and ingredient supply company, I utilized the open-access data from the Wikipedia and Foursquare to segment the main London area into 5 clusters based on the top 5 most common venues. Among the 5 clusters, Cluster 1 significantly had more coffee shops, cafés, pubs, and bars as the most common venues. A further search indicated 8 neighborhoods where all of the top 3 most common venues were the target categories (i.e., coffee shops, cafés, pubs, and bars) in London. Among these neighborhoods, Stockwell was considered the most suitable location for the new company as this neighborhood showed all of the top 4 most common venues being the target categories, thereby suggesting the greatest need for beverage and its ingredient supplies.

Although London is a highly populated city and is not short of coffee and tea shops, cafés, pubs, and bars, a steady and timely supply of beverages and the ingredients

is a necessity to these shops, cafés, pubs, and bars. Given this, a better business perspective is not to add more similar services, but to set up a company to supply beverages and other needed ingredients to these places. Following this line of reasoning, it is crucial for the stakeholder to establish the company in the area with the highest demand for beverages and the ingredients. Indeed, according to the analysis, there were approximately 120 potential customers selling coffee, tea, and alcohol in Stockwell.

The strength of this project was the utilization of public data that can be updated easily by users to reflect any instant changes. Yet, the limitation of the project should be mentioned. As data was simply from the Foursquare users, some bias could be introduced to the database. For instance, some customers may have visited other coffee or tea shops, cafés, pubs, and bars more frequently but are not Foursquare users. Hence, their movements cannot be traced on the Foursquare. That being said, the work in this project provides more specific information about the location for the new company at least.

CONCLUSION:

Findings from the current project suggest that Stockwell would be the most suitable neighborhood in the main London area as the location of the new company supplying beverages and the ingredients to coffee and tea shops, cafés, pubs, and bars.