**Topic: Where to Open a Coffee Venue in London**

Annemarie Donohue

**1 Introduction**

**1.1 Background**

London is the capital of England, and the United Kingdom. Here, culture flourishes from the English people. Buckingham Palace, Tower of London and Westminster Abbey are some of famous attractions. Along with culture comes the food and drink associated with it. With the UK known for tea drinkers, there was actually a 20% fall of tea consumption in the 1970-80s (*Ferdman*). Research explains a probable cause for the switch would be folks not wanting to spend time brewing the tea when a caffeine alternative, coffee, can be made in seconds from a pre-portioned coffee pod. Coffee sales from local shops have increased by 700% in 2019 (*Crowe*). 51% of those orders are lattes, which require milk frothers. This means that these drinks are easier purchased than made at home, hence the drastic increase in coffee shop revenue. From organic ground beans to different preferences of milk, some want to sit down and chat while others want to run in and out of the shop. Most café aesthetics are laid back and calming, which serves to those who want to sit down. The coffee shop relies on catering to those running late and need a boost of caffeine. With the increase in coffee demand, there is a potential for high net revenue in coffee locations.

When opening a venue, an owner must ensure that there is a need for their products. The ideology can be compared to a bodega in New York City. Known for selling basic convenience needs, owners fill the demand of locals by providing everyday products in close proximity to their apartments. With high demand, these bodegas sell necessities at high rates. A new owner must think in a similar manner. Even if your product is outstanding, there must be a demand from the consumer or else the business will suffer.

Taking into account that London (pandemic aside) is a bustling city, the coffee shop will need to account for other similar venues in the area, and what their specific shop provides that is unique. Along with individuality, a lot of coffee shops are known for having great business due to their location. Not having to cross the street in the wrong direction when running late is ideal. However, having too many locations can also be an issue (*Purdy*). With the example of Starbucks, there are so many locations that the company is actually hurting itself, with the competition of it’s own stores. The article explains how there needs to be the right amount of locations selling the product to meet the demands of the consumer, while also being coinvent. This is why the study of where to open the next coffee location in London will take into consideration what locations are already in the surrounding area. The coffee shop will have to be able to fill a demand, and be located close to stores that are commonly associated with coffee locations.

This project also includes the idea that stores are correlated with other stores in the area. A common example would be food establishments at shopping malls. Some may go to a restaurant for lunch and later pop into the mall, or a customer is hungry after a few hours of walking around with heavy purchases. A shopping mall that is lacking well prepared, vegan food, could benefit from an owner basing their menu around veganism. Both the bodega, Starbucks and mall examples will be infiltrated into the project. When looking to open a venue in London, analysis of which boroughs are lacking the product must be found, along with a high quantity of venues that often are associated with the new owner’s products. Profitability can be found by finding which boroughs lack the venues the owner wants to open, and has many venues normally correlated with the new venue.

**1.2 Problem**

This project will focus on finding quantities of venue categories per borough, and finding venues highly correlated with coffee venues. I included both coffee shop and café in this project so the owner can determine which would be better for them. Since there is a fine line between the two, I assumed that a caffeine fanatic owner would be interested in both. The goal will be to find boroughs that are suitable for the new coffee venue in London.

**2 Data**

The data must show the boroughs of London, and the venue categories associated with those boroughs. This will be used for regression analysis to find correlation between most common venues. One data set was chosen from Wikipedia, focusing on the geography of London. The other data set includes latitude and longitude to use as a foreign key when merging to the Wikipedia data set, but includes counts of venues and their categories as well.

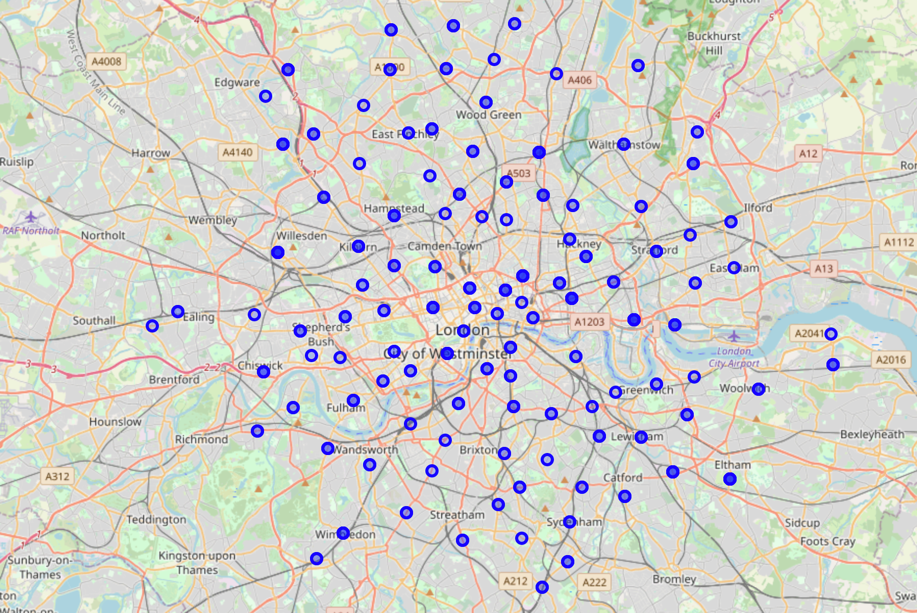
**3 Methodology**

**3.1 Data Collection**

Two main forms of data were used for this analysis. The first being from Wikipedia, listing details of London, such as location, London borough, post town, postcode district, dial code and OS grid relief. This data will be used for borough coordinates in London. Foursquare focused on venues within the boroughs. This data set included Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue and Venue Category. Note that there are 32 official boroughs in London, but the neighborhoods category includes some groupings of boroughs for relevance. An example would be the neighborhood Bexley, Greenwich, which includes both Bexley and Greenwich boroughs.

**3.2 Data Cleaning**

The data from Wikipedia was collected specifically for the latitude and longitude associated with the borough. This data set extracted, transformed and loaded details into a new data frame. Data cleaning included renaming and only keeping the columns borough, town and post code. The post code was for geography, while the borough and town were qualitative naming. The final size was 308 rows with 3 columns. To create a map of London, geocode was used to associated the latitude and longitude coordinates with certain boroughs. The coordinates for London are (-0.13, 51.51). Coordinates of boroughs were then integrated into the data set, with columns consisting of borough, town, post code, latitude and longitude. This was used to map the boroughs throughout London, using folium, as shown below.

****

The second part of cleaning used Foursquare to locate venues and their coordinates. The nearby venues were loaded into a data frame after using the client ID, client secret, access token and version from my Foursquare account. The final columns loaded into the data set were Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue and Venue Category. These variables were chosen in order to properly merge with the first data set.

In order to have an understanding of the venues in London, the two datasets were merged. The Wikipedia data set used the primary key Borough, which correlates to the Foursquare foreign key Neighborhood. This was congruent with Latitude and Neighborhood Latitude, and Longitude and Neighborhood Longitude keys as well. Neighborhood refers to boroughs, but also common boroughs or areas close to each other that are commonly associated, hence the new terminology. The   
final columns in the merged data set were Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue and Venue Category.

**3.4 Model Building**

**3.4.1 Model Building: Descriptive Analytics**

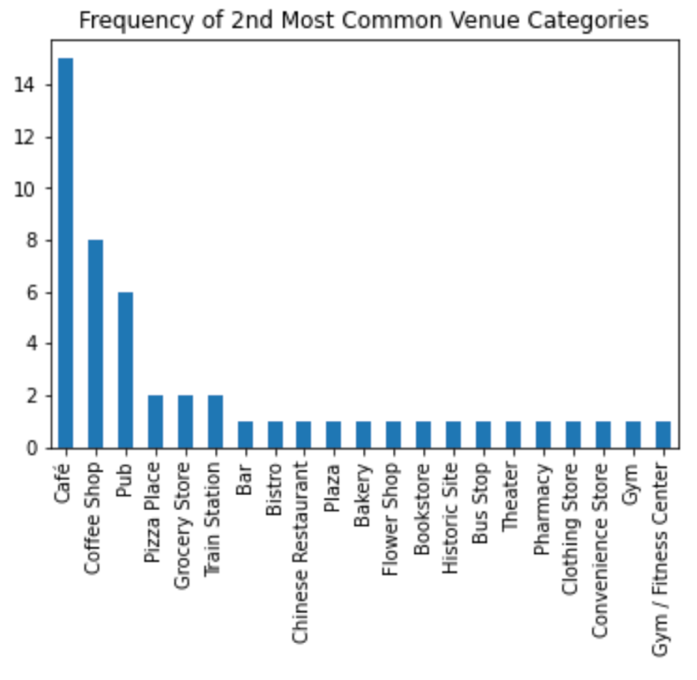
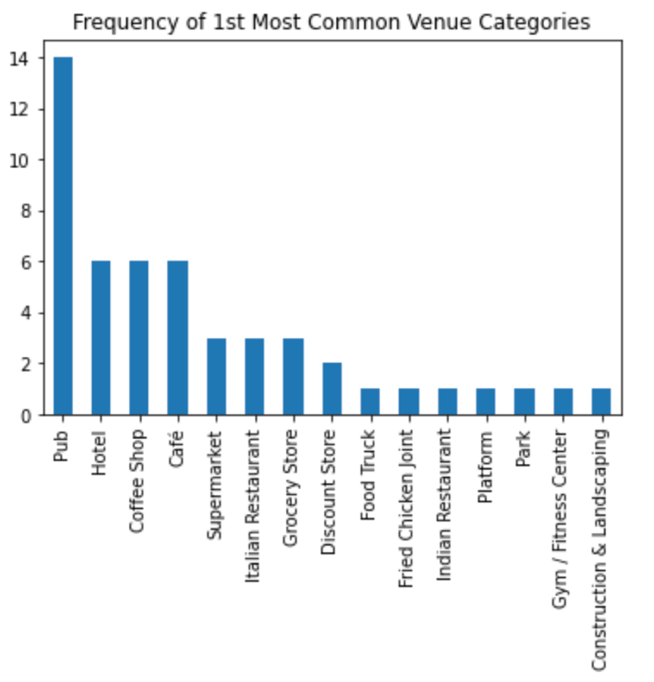
The goal was for the relationship between areas and venue categories to be analyzed, establishing the type and count of venues in each neighborhood. I sorted and counted venues for all of London. I arranged a new data frame where the top 10 venue categories for each of the 49 neighborhoods would be shown. I then counted the venue categories that were the ‘1st most common venue category’ for all of London. I also included ‘2nd most common venue category’ for comparison. The total amount of venues in that neighborhood were included to put the number of venue categories into perspective. This descriptive model was produced to understand which venue categories were most popular per neighborhood in London.

**3.4.2 Model Building: Predictive Modeling**

In order to understand the association of venues, k-means clustering model was used. This is a supervised machine learning model with a goal of clustering based on similar measurements, determined by distance functions. This modeling can be used for classification and/or regression. I started by merging the original data from Wikipedia and Foursquare to get the borough, town, post code, latitude, longitude, cluster label, and 1-10 most common values. To start building the model, I created 5 clusters of venue categories, grouping by neighborhood. I grouped by neighborhood and not by venue category because the goal was to find which neighborhood to put the location in, therefore trying to find details based on the location. Each cluster was located by cluster label in order to be individually analyzed. Along with the k-means model, I also integrated the quantitative counts of venue categories per cluster through a pivot table. This analysis provides addition information to the association of venues through k-means clustering. This clusters will indicate the association portion of the project. Seeing which venue categories are correlated with coffee locations will provide insight on where to open a new shop.

**4 Results**

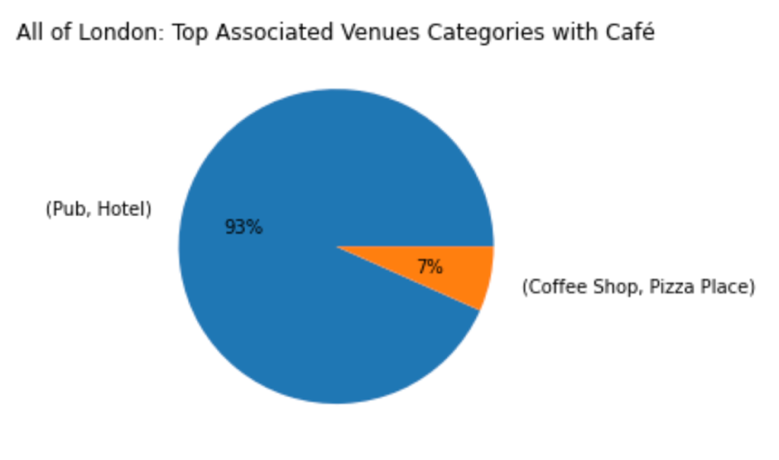
The top 10 venues per neighborhood showed which areas had a high amount of product. Considering that there is only so much demand, a new store owner must consider if the demand is already being met when many venues already exist in that category. In the 49 neighborhoods, ‘Coffee Shop’ and ‘Café’ come in second and third (third tied with hotel) for the most common venue in a neighborhood. That means that most of the demand is met, but there are still boroughs that could use coffee. Graph 1 and 2 below shows a bar chart of 1st and 2nd most common venue categories in London. Graph 1 includes the most common venue categories of pub, hotel, café and coffee shop, and Graph 2 shows café and coffee shop. There is already a direct indication that these venues categories are associated and popular throughout London.

****

***Graph 1*** ***Graph 2***

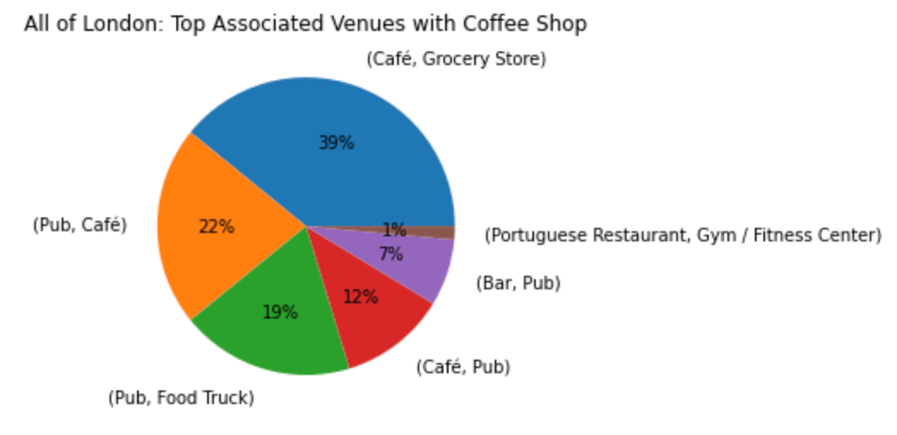
The k-means model was able to show associations between different venue categories. Considering that coffee shop and café are very similar, I was specifically looking for the results for both categories. The most relevant cluster is 1, showing many correlations between venues with coffee, which will be further analyzed later on (Graph F). Cluster 2 showed correlation with grocery store and park, cluster 3 with supermarket and historical site, cluster 4 with Chinese restaurant and pharmacy, and cluster 5 with bus stop and bakery.

Deeper investigation included visualization for better understanding, which is under the Results section of the code. The pie chart for Graph A is sorted so that the most common venue category is café, and the 2nd and 3rd are listed next to the percentages. Pie charts were chosen because every borough or neighborhood had a different number of venues, so percentages would establish a ratio that would be comparable for all areas. These percentages indicate the probability of the sequence for that neighborhood. First, I wanted to see the top categories for all of London associated with café. Graph A depicts pub and hotel at 93%. This shows an extremely high correlation between cafés, pubs and hotels. Coffee shops and pizza places were correlated, but only 7%.



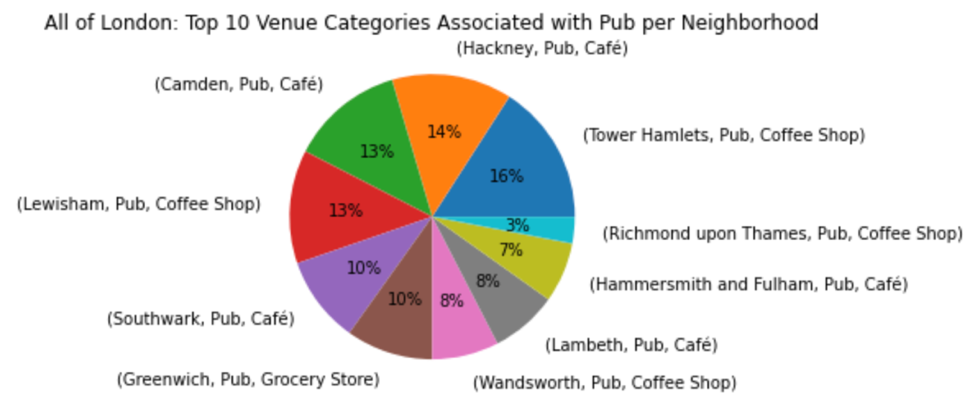
***Graph A***

Coffee shops for all of London is set up the same way as the pie chart above, except coffee shop is listed as the 1st most common venue in Graph B. The top correlation was café and grocery store at 39%. Pub and café or café and pub made a combined percentage of 34%. This once again concludes that high quantities of pubs has a direct relationship with coffee location quantities.



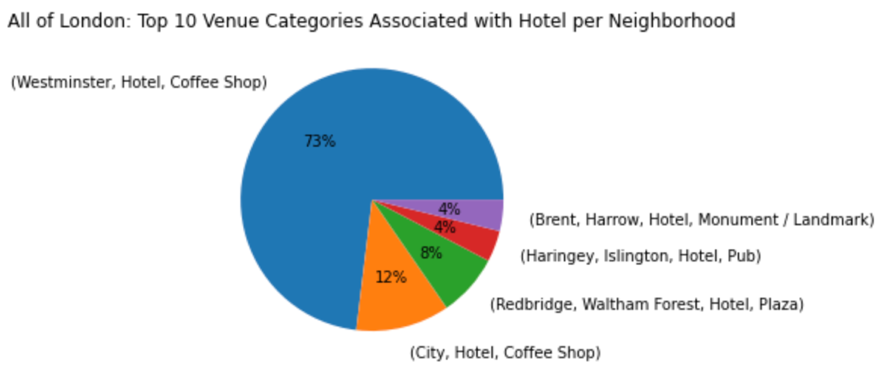
***Graph B***

Next, pubs were analyzed as being the top venue category for all of London. Graph C shows the borough, 1st and 2nd most common venue. Pub and either café or coffee shop fills 90%, with pub and grocery store being the other 10. The top for pub then coffee shop are Tower Hamlets (16%) and Lewisham (13%). Pub then café show Hackney (14%) and Camden (13%). This graph indicates that Tower Hamlets would be the best place for a new coffee shop, and Camden for a new café.

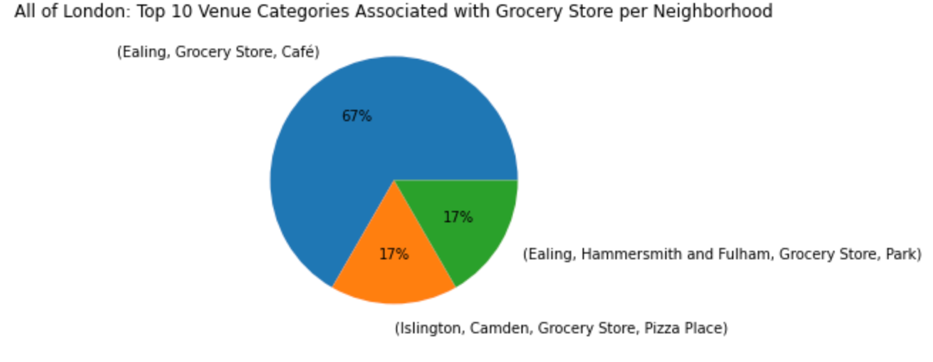


***Graph C***

Since hotel and grocery story were also correlated, their pie charts were included for additional information. After seeing consistent association pubs, these graphs were niche. In Graph D, Westminster showed a high correlation between hotel as the 1st common venue and coffee shop as 2nd common venue at 73%. Graph E showed Ealing with a high association with grocery store and café at 67%.

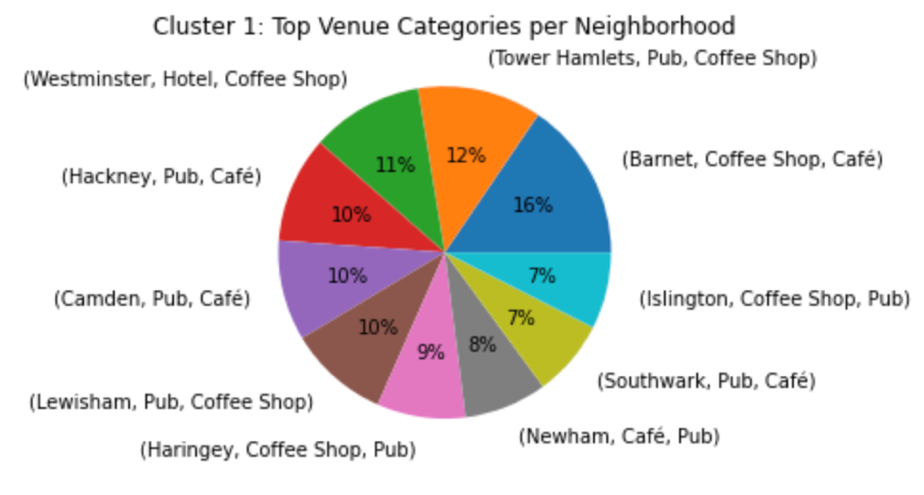


***Graph D***



***Graph E***

Coming back to the k-means analyzation, cluster 1 was the most useful to coffee of all clusters. Many of the venue categories included pub, coffee shop or café, so a pie chart, Graph F, was created to visualize the top venue categories per borough from cluster 1. As shown below, the top 1st and 2nd most common venue categories all range from 7-16%, so all are within a low 9% range. The highest correlation would be in Barney with coffee shop then café at 16%, then Tower Hamlets with pub and coffee shop at 12%. Because these percentages are so close, this visualizes the high correlation between coffee shop, café and pub between the top 10 neighborhoods .



***Graph F***

Considering that the goal is to find the most probable location for highest net revenue, the borough with the highest correlation with coffee will have a direct relationship with net revenue. The results also take into consideration that having a coffee location as the top venue category means the demand is mostly likely met, so the analysis takes this into consideration, focusing coffee as the second most popular. As stated previously, cafés are associated with pubs and hotels, and coffee shops with cafés, grocery stores and pubs. For cafés, Hackney and Camden show signs of prosperity. In Graph F, both of these boroughs are tied for 4th at 10% each with pub then café. In Graph C, they are almost tied 2nd with 14% and 13%. There is also a niche correlation in Graph E where Ealing has a 67% correlation of grocery store and café. For coffee shop, Westminster would be the option. Hotel and coffee shop are 3rd highest (11%) in Graph F, and by far the highest (73%) for venues associated with hotel in Graph D. Tower Hamlets is also a great option a coffee shop. In Graph F, Pub and coffee shop have the 2nd highest percent correlation (12%), and highest percentage (16%) in Graph C.

**5 Discussion**

The analysis of the top 10 venues per neighborhood was created to show which neighborhoods had the most amount of venues. By seeing the popularity of coffee locations, this was a positive indication that this service is desirable in London. Counting which categories were the most common shows that cafes and coffee shops were popular, therefore profitable in London. Other venues like pub, hotel and grocery store proved to be relevant as well. Between both types of venues analyzed, both can be profitable when looking for the right location.

Graph C and F show all the neighborhoods or boroughs that would be reasonable places for coffee. There is correlation with other venues that already have a high occupancy, predicting revenue and growth for the new shop. Graph D and E are more niche, showing how coffee and hotels in Westminster (19 total) and grocery stores and cafés in Ealing (4 total) have high association. Westminster would be profitable for a coffee shop, and considering café is 3rd most common, a café as well. With a strong correlation and high occupancy, there is a high probability that a coffee location would do well. Although there are variables that go into selecting a location, these correlations are a key takeaway. Finding the niche correlations will ensure revenue greater than pub and coffee, which tends to be popular across a lot of London.

From a marketing perspective, a product has to fill a demand that isn’t already being filled. The owner must ask themselves what sets them apart from other coffee vendors in the area. This can include having a unique ambiance (décor, music, etc), perks for those working (multiple outlets, charging docs) or discounts for students. There are other factors that go into owning a coffee shop, like having organic locally sourced coffee beans leading to a higher price coffee but better quality, or loyalist card to get the 10th drink free. Another factor would be rent. Everyone has different financial capabilities, especially when first opening a location. Having a financial plan to make sure that the rent is manageable is essential, and also understanding that most new businesses do not produce revenue the first year. Finally, there is an importance to realize that tea is much more popular in the UK than coffee. The owner should consider which areas are international, as Americans drink more coffee than tea. Perhaps the menu should also have a moderate length tea section in order to provide for both the natives and international peoples.

**6 Conclusion**

London is bustling city where each borough has it’s own flavor. When opening any venue, there is an importance to get to know the demands of the community in order to ensure that you are providing a product where there need. Coffee shops and cafés are not like other venues because convenience is a large component, not just quality and price. By finding venue categories that are correlated with having a large quantity of coffee locations within the borough, this project was able to analyze the correlations between coffee locations and other venue categories. With the highest correlations being with pubs, hotels and grocery stores, some boroughs that would be profitable include Westminster, Ealing, Hackney and Camden. These analyzations will enable the owner to understand the area in order to product high revenue in their new coffee location.

|  |  |  |
| --- | --- | --- |
|  |  |  |

**7 References**

Crowe, Eamonn. “London Sees 700% Surge in Independent Coffee Shops since 2010.” *SquareMeal*, SquareMeal, 6 Feb. 2020, www.squaremeal.co.uk/restaurants/news/rise-in-independent-coffee-shops-london\_9631#:~:text=New%20statistics%20have%20revealed%20that,to%20now%20more%20than%20400.

Ferdman, Roberto. “The Slow Death of the Most British Thing There Is.” *The Washington Post*, WP Company, 29 Apr. 2019, www.washingtonpost.com/news/wonk/wp/2016/05/04/why-the-british-are-drinking-coffee-instead-of-tea/.

Purdy, Chase. “That Joke about a Starbucks on Every Corner? It's Actually True and Hurting the Company's Sales.” *Yahoo! Finance*, Yahoo!, finance.yahoo.com/news/joke-starbucks-every-corner-actually-113027036.html.