Bubble Tea Venues in relation to Higher Education Institutions in London

Sarah Jam

June 2020

1. Introduction

1.1 Backround

Bubble tea, alternatively known as Boba, is a popular choice of drink among students and young adults [1][2]. It is most recognized by the tapioca balls in milk tea or various other kinds of drinks. In metropolitan cities, bubble tea shops have been populating the streets to take advantage of this trend. London has a high student population, due to the many higher education institutions, and high tourism, which leads to a perfect environment for this trending drink. Place with high tourism have already been populated with bubble tea shops, but the areas near student populations have not been explored.

1.2 Business Problem

The objective of this project is to analyze and suggest the best locations for a new bubble tea shops in London, United Kingdom. As students and young adults are the primary consumers of bubble tea, the focus will be around higher education institutions in London.

1.3 Audience

This research will be useful for any individuals who are interested in opening a bubble tea shop in London, specifically targeting the student population. It may also benefit individual looking to understand factors in the consumption rates of bubble tea by students in London.

1.4 Data

The following will be used:

- Excel document of higher education intuitions and their number of students, taken from Higher Education Statistics Agency (HESA), to determine which intuitions to analyze based on student population. [3]
- CSV file of higher education intuitions' addresses, taken from Office for Students (OfS), to be converted into latitude and longitude coordinates [4]
- Bubble tea/Boba shop's data retrieved from the Foursquare API.

2. Methodology

2.1 Data Cleaning

Using the Panda's library, the CSV file retrieved from the Office for Students (OfS) was converted into a data frame. and all information beside address and legal name were cleaned from the data. Intuitions which are not in London were dropped from this data frame. The Excel document from HESA was converted into a data frame as well. All information beside intuition name and total number of students were cleaned. Then, intuitions which had less then 1000 students were dropped from this data frame. This is so institutions are reasonable large enough to be considered a primary customer to these shops. These data frames were then merged and any location which then lack an address or total of students were then dropped from the compiled data. Address were then converted in to coordinates using the Geopy Library via Nominatim. If coordinates were unable to be retrieved for an address, that location was dropped.

2.2 Data Processing

A map of London was created with the intuitions to be examined through the Folium library to give an idea of the distribution of these intuitions. Request were then sent to the Foursquare API using the final list of intuitions. Two sets of requests were sent out with different queries: 'Bubble Tea' and 'Boba'. This was to make sure that all relevant shops were covered. Shops that were not specialty bubble tea shops but sold bubble tea were included. Information retrieved from Foursquare were name of venue, distance from intuitions, full address and type of shop. Duplicates were removed accordingly. A count was then retrieved of the number of bubble tea venues within an 800-meter radius, which is about a 10-minute walk, of the intuitions.

k-means clustering was performed to group intuitions. K-means clustering, identifies k number of centroids, and then distributes every data point to the nearest cluster, keeping the centroids as small as possible. It is a popular unsupervised machine learning algorithm and suitable for this project. Intuitions were grouped into 3 clustered based on the number of bubble tea venues nearby.

Additional line plotting of the number of venues verses the number of students was done to see if there was a correlation between the two aspects.

3. Results

The k-means clustering show categorized the intuitions as follows:

- Cluster 0 (Red): intuitions with 0-1 bubble tea venues
- Cluster 1 (Green): intuitions with 2-3 bubble tea venues
- Cluster 2 (Purple): intuitions with 4+ bubble tea venues

It is visualized on the map below (Figure 1). The line plot below is the correlation of number of venues and number of students (Figure 2).

Figure 1

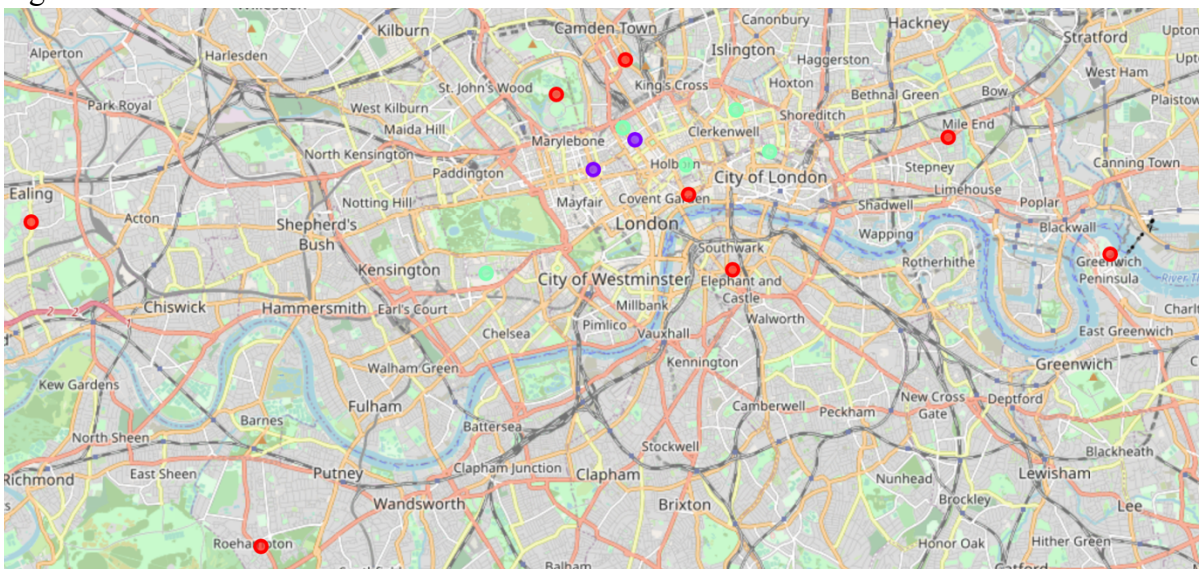
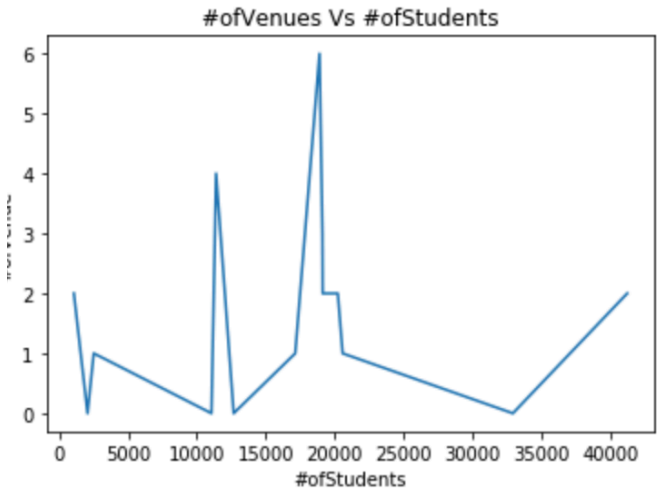


Figure 2



4. Discussion

4.1 Observations

It should be noted that intuitions with a greater number of bubble tea venues were concentrated in Central London. As one moves further out the number of venues decrease. Cluster 0 reflects areas of great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing bubble tea venues. Cluster 2 reflects area of higher competition and less opportunity. This research suggest that it is better to open a new bubble tea venue in areas near cluster 0. Additionally, the research demonstrated that there is currently no correlation between the number of students at an institution and the number of bubble tea venues nearby. Which supports an previous statement that the potential of bubble tea venues near higher education intuitions has not been greatly explored.

4.2 Limitations and Further research

It must be noted that this research is examines one factor, the number of bubble tea venues. Other factors, such as general population size, location of all an intuition's buildings, and public transportation, surely influence the best location for a new bubble tea venue. At this time, such information is not readily available and would require more research on those topics in general. This project was also restricted by the fact it could only use free and available resources, that may not be completely up to date or completed. It must be noted that the OfS did state that they were continuing to update their data. Also, the number of students from HESA were based on intuitions that choose to self-report their numbers. There may be intuitions that have a student population greater than 1000, but did not choose to report to HESA. Thus, the sample size may be smaller.

5. Conclusion

In this project, the relationship between higher education intuitions and the number of bubble tea venues was analyzed. Intuitions were clustered through k-mean based on the number of bubble tea venues nearby. This project suggest that a new bubble tea venue will find success in areas near cluster 0 intuitions. These finding will help relevant parties to understand the opportunities and potential areas of establishing a bubble tea venue near a higher education intuition. For those researching, factors in the consumption of bubble tea by student, this research may provide insight in the number of shops that are near students.

Sources

[1] https://beveragebusinessworld.com/latest-news/why-is-bubble-tea-taking-the-uk-by-storm-25-02-2019/

[2] https://www.socialstandards.com/blog/boba-bubble-tea-consumer-trends#:~:text=Boba%20is%20highly%20dependent%20on,the%2090th%20percentile%20for%20boba.

[3] https://www.hesa.ac.uk/data-and-analysis/students/where-study

[4] https://www.officeforstudents.org.uk/advice-and-guidance/the-register/the-ofs-register/