**The Battle of Restaurants in London**

**By Valeriya Kovalenko, 2020**

1. **Introduction** **and** **Business problem**

This project will compare London restaurant categories based on Foursquare data. There are a large number of restaurants in London, and therefore it is very difficult to understand where to go for a newly arrived person. Based on Foursquare user reviews, I will group these restaurants so that Foursquare can determine where to go.

1. **Data** **and** **Data Acquisition Approach**

In this project, I will use the Foursquare API to get the necessary data on the location of restaurants in London: (Venue Name, Venue ID, Venue Location, Venue Category, Count of Likes). You need to get the latitude and longitude coordinates of London, and also use the Foursquare API to get a list of all places in London.

1. **Methodology** **section**

In this study, the coordinates of London were obtained and venues from the site were requested. Created a dataframe with the name, id, category and coordinates. Excluded attractions that do not meet the conditions of restaurants.

Then, the number of likes for each restaurant was requested and summarized from the site.

Categorical variables for restaurants have been created to group them by the number of likes. User preferences are also taken into account by type of cuisine. This is necessary in order to divide the restaurants into groups for easy search.

In order to divide the restaurants into groups, the data collected from Foursquare was used. Next, a k-means clustering algorithm was created that groups restaurants into clusters so that people who want to eat in Moscow can easily see which restaurants are better, which cuisine is available and where in Moscow they can eat.

1. **Results section**

As a result of data processing and research in London, a relatively small number of venues (100) were found, of which only 22 are restaurants. After collecting data on the number of likes and grouping restaurants by type of cuisine, a cluster analysis was performed. Clustering results are shown in Figure 1.

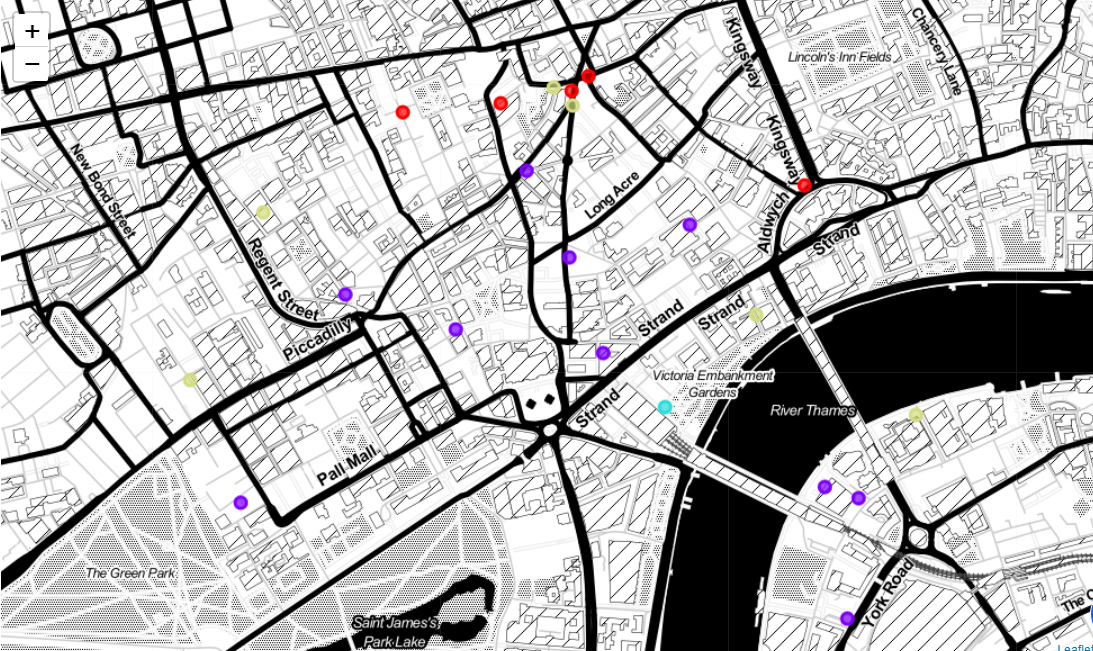


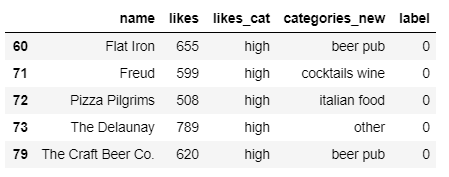
Figure 1. Clustering London restaurants.

Based on the number of **likes** and category of restaurant, restaurants are divided into 4 clusters:

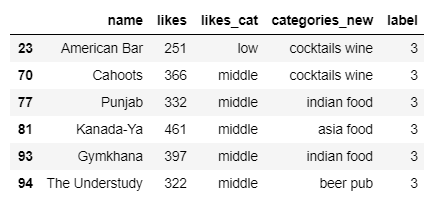
1. Very high marks, only one bar:



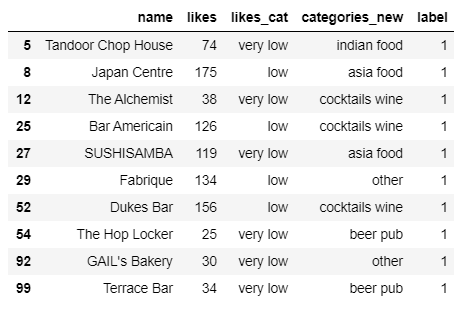
1. High marks, mostly pubs and Italian food:



1. Average ratings Asian and cocktail bars:



1. Low and very low, mainly Asian and cocktail bars:



1. **Discussion section**

Quite interesting observations can be made from the results of data processing. The number of restaurants with low rating is greater than with high (Figure 2):

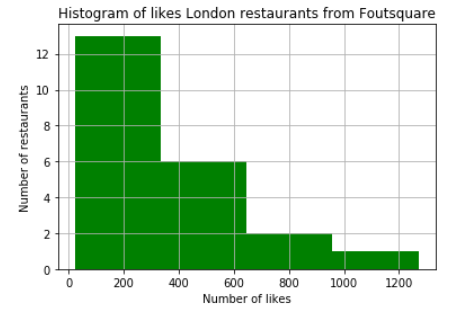


Figure 2. Preference distribution histogram

The distribution of estimates differs from the normal one, which most likely indicates a lack of sampling, most likely Foutsquare is not as popular in London as in North America. Also done box plot data (Figure 3):



Figure 3. Preference distribution histogram

The maximum number of likes is 1270, the minimum is 25. At the same time, the median value is less than the average - 286 to 340, which indicates that most restaurants are shifting towards a low number of likes.

The advantage of clustering from the usual classification is also quite obvious - classification by percentiles identified 4 groups, in which several restaurants received high marks. Thanks to clustering, we were able to see that one bar is especially loved by the residents of London, and it forms a separate cluster. What increases the value of research in relation to classification.

1. **Conclusion section**

Gathering data on London restaurants from Foutsquare revealed the undisputed leader among restaurants - this is Gordon's Wine Bar. Obviously, people who come to London should visit this place most beloved by local establishments. Residents of London are obviously more attracted to Pubs and Italian restaurants than to Asian-Indian cuisine.