Project Report

**Capstone Project - The Battle of Neighborhoods**

By Anel Mukasheva

May 2020

Table of Contents:

1. Introduction:
   1. Background;
   2. Problem;
   3. Target audience;
2. Data:
   1. Data description;
   2. Data preparation;
3. Methodology;
4. Results
5. Discussion
6. Conclusion

**Introduction**

**Background**: Starting a new business or expansion of an existing business is always challenging for newcomers and even for experienced entrepreneurs since it is always associated with a broad spectrum of risks. In order to minimize the possibility of business collapse, a person needs to consider various aspects of business management and try to predict possible effects of decisions made. Those aspects cover a large range of issues one of which is the choice of location for opening a new facility or new branch of an existing facility inasmuch as the choice of location can significantly influence the ability of a business to market itself. The aim of this study is to investigate the major criteria which can influence the choice of the location for business and to help businessmen to find the most appropriate location in accordance with the need of their business.

**Problem**: A person aims to open a new medium-size casual-class Asian cuisine restaurant in London, however, he does not know where exactly. This study will examine 32 London’s boroughs with respect to demographic indexes such as borough population and average income. Another aspect lying under the scope of the study is evaluation of the level of competition in different boroughs in the sphere of restaurant business.

**Target audience**: Businessmen, restaurateurs, investors and other people interested in the systematic approach to the choice of location.

**Data**

**Data description:**

There are three major data source that were used during investigation process:

* “Borough data” – provides information on population estimates of each borough[[1]](#footnote-1).

*This data was used in order to analyze possible extension of customer base.*

* "HMRC Survey of Personal Incomes" – provides information on average income in each borough[[2]](#footnote-2).

*This data was used for forecasting customers’ capacity to pay.*

* Foursquare.com – provides information about venues located in the particular geographical location.

*This data was used for evaluating the number of competitors and for finding borough with the lowest number of Asian cuisine restaurants.*

**Data preparation:**

1. Data on population was loaded from web page using Beautiful Soup package and stored in pandas data frame: population\_df;
2. Data on average income per borough was loaded from excel file using pandas read\_excel and stored in data frame: income\_df;
3. For population\_df columns containing irrelevant data were dropped and only columns with boroughs’ names and population for 2018 were left:

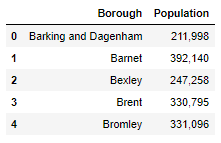


Image 1. Section of population\_df after data cleaning

1. Income\_df contained irrelevant columns that we dropped and only columns with boroughs’ names and mean income for 2017-2018 were left:

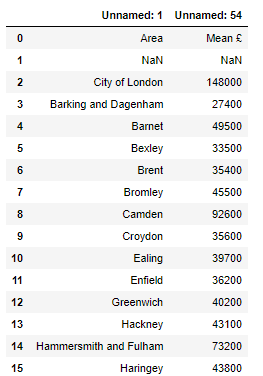


Image 2. Section of income\_df after dropping irrelevant columns

1. Income\_df also contained rows with no data or data related to areas outside of London that were dropped and only rows related to London boroughs were left:

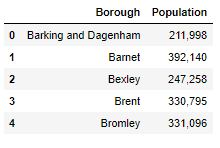


Image 3. Section of income\_df after dropping irrelevant rows

1. After cleaning two data frames separately, they were merged together in accordance with boroughs’ names:

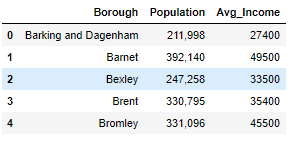


Image 4. Section of borough\_df after merging population\_df and income\_df

1. Type of data in Population and Avg\_Income columns were changed from Object type to Int type.
2. Information on geospatial data for each borough was added using GeoPy’s geolocator:

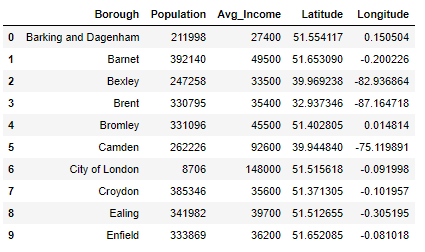


Image 5. Section of borough\_df after changing data type of columns and adding geospatial information

**Methodology**

1. Map of London with boroughs was created in order to check whether there are mistakes in geospatial data obtained during data preparation step.

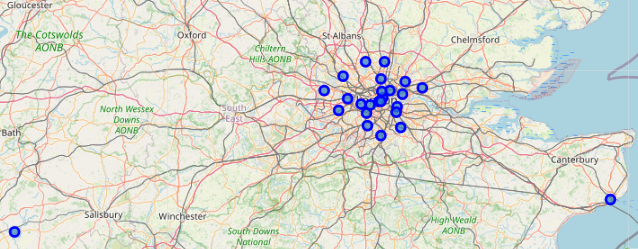


Image 6. Map of London with blue markers indicating 32 boroughs in borough\_df

1. As it can be seen from the map on Image 6, there are two points located too far from others. Those two points were checked and fixed

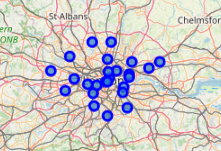


Image 7. Fixed map of London with blue markers indicating 32 boroughs in borough\_df

1. It was decided to implement multi-stage selection process, however, in order to implement this procedure it was essential to choose more crucial criterion between population and average income. Since the aim is to start casual-class medium-size restaurant, we are more interested in high foot traffic and thus, higher value of borough population has higher priority rather than higher average income per borough.

Hence, analysis algorithm includes 3 steps:

1. Find top 15 most populated boroughs in London;
2. Find top 5 richest boroughs among 15 most populated ones from point 1;
3. Find top 100 popular venues for each borough from point 2;
4. Find borough where the number of Asian cuisine restaurants is the lowest;
5. Find top 15 most populated boroughs in London:

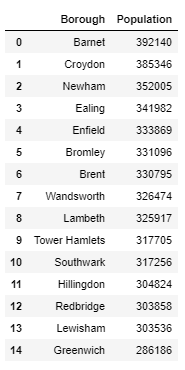


Image 8. Top 15 most populated boroughs in London

1. Find top 5 richest boroughs among 15 most populated ones:

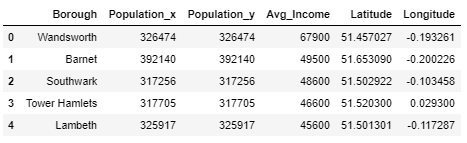


Image 9. Top 5 richest boroughs among 15 most populated boroughs in London

1. Find top 100 popular venues for each borough:

For each of 5 rich-populated boroughs 100 venues were loaded using Foursquare API.



Image 10. Section of London\_venues dataframe showing 100 venues for each of rich-populated borough

1. Find borough where the number of Asian cuisine restaurants is the lowest:

From London-venues data frame only venues containing word Restaurant were filtered:



Image 11. Section of London\_restaurants data frame showing restaurants in 5 rich-populated boroughs

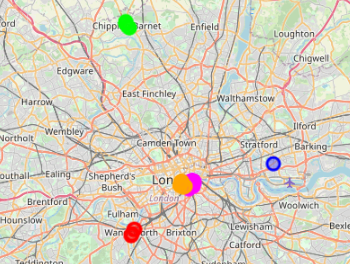
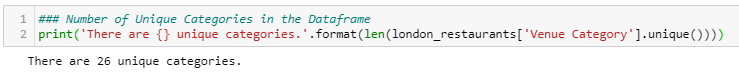


Image 12. Restaurants from London\_restaurants dataframe indicated on the map

As it was found there are 26 unique categories of restaurants in 5 boroughs selected:



Those 26 categories:

Italian Restaurant 7

Restaurant 7

Chinese Restaurant 4

Korean Restaurant 3

Asian Restaurant 3

Ramen Restaurant 2

Modern European Restaurant 2

Thai Restaurant 2

English Restaurant 2

Turkish Restaurant 2

Fast Food Restaurant 2

Indian Restaurant 2

Vietnamese Restaurant 2

Mexican Restaurant 1

Dim Sum Restaurant 1

Latin American Restaurant 1

Spanish Restaurant 1

Vegetarian / Vegan Restaurant 1

Israeli Restaurant 1

Tapas Restaurant 1

Argentinian Restaurant 1

Sushi Restaurant 1

Eastern European Restaurant 1

Japanese Restaurant 1

French Restaurant 1

Portuguese Restaurant 1

As it can be seen, Asian, Chinese, Korean, ramen, sushi and Japanese restaurants are counted separately. All those restaurants were accumulated into one dataframe:



Image 13. All Asian cuisine restaurants from 5 boroughs accumulated into one dataframe

After all Asian restaurants in all 5 boroughs were accumulated into one dataframe, number of restaurants in each borough was found:

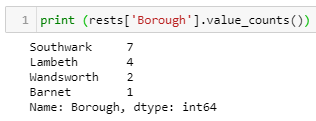


Image 14. Number of Asian restaurants in each borough

**Results**

After investigation of 32 London boroughs it turned out that the most appropriate place for starting new casual-class medium-size Asian cuisine restaurant is in Barnet borough since it has high population, high average income and low number of competitors.

**Discussion**

Among all 32 boroughs of London the most populated boroughs are Barnet, Croydon and Newham

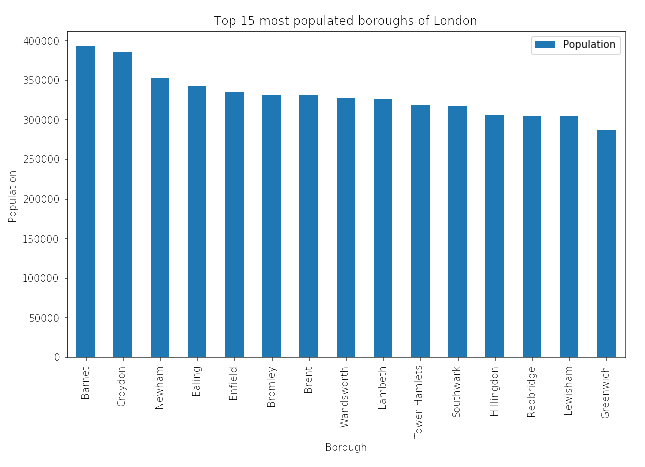


Image 15. 15 most populated boroughs of London

Among 15 most populated boroughs in London the richest boroughs are Wandsworth, Barnet and Southwark.

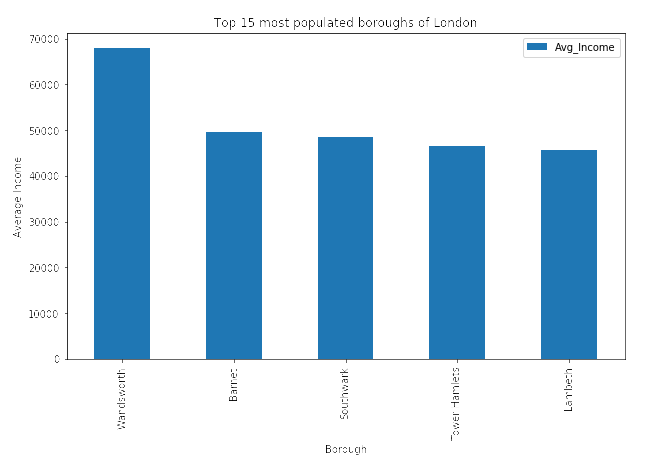


Image 16. 5 boroughs that have the highest population – average income proportion

These three boroughs have the highest population – average income ratio, and thus, in this study they are considered to be the most financially profitable boroughs.

Under the scope of current study, the aim was to find the most appropriate location for casual-class restaurant, that is why emphasis was placed on higher foot traffic and on higher borough population respectively. However, if the goal was to start premium-class restaurant the emphasis of the study would change to higher average income per borough and steps of multi-stage selection would be Average Income -> Population -> Competitors instead.

**Ways of study improvement:**

* Since boroughs area was not investigated under the scope of the study, there is a possibility that selected boroughs are not only highly populated but also have large land area. In order to increase the precision of the study, population density needs to be taken into consideration instead of population value itself.
* Using Foursquare data composition of each boroughs could be investigated, and number of office building could be taken into account.
* Crime rate per borough also could be considered since low number of competitors could be caused by higher number of crimes happened in the area of investigation.
* There are plenty of other additional variables that could increase the precision and usefulness of the study.

**Conclusion**

In this study I evaluated 32 London boroughs with perspective of starting a new restaurant using borough population, average income and number of competitors in each of them. It was revealed that top 5 most populated and richest boroughs are Wandsworth, Barnet, Southwark, Tower Hamlets and Lambeth. After evaluating number of Asian cuisine restaurants, it was revealed that the most appropriate borough for starting restaurant is Barnet due to low competition level. However, study results can be improved by adding additional variables such as population density, crime rate and etc.

1. <https://www.citypopulation.de/en/uk/greaterlondon/> [↑](#footnote-ref-1)
2. <https://data.london.gov.uk/dataset/average-income-tax-payers-borough> [↑](#footnote-ref-2)