Leveraging location and competitive datasets to evaluate business opportunities in Central London

Capstone Report for IBM Data Science Professional Certification

Written by Thomas Reimer

August 2020

1 Introduction

When planning to open a new business, it is increasingly difficult to identify candidate locations in urban locations, due to the saturation of markets. However, taking a data driven approach is increasingly common for a business to get the edge in this competitive landscape.

This report will explore the opportunities for opening a new cafe in a vibrant area within central London. In order to make this decision effectively, the analysis will factor in information about the competitive landscape and leverage various data science techniques to inform the final recommendations.

While this analysis will focus on the launching of a new cafe, a similar methodology could also be applied across a range of business opportunities, thus making this report relevant to all entrepreneurs looking to launch a consumer facing business in the city centre!

2 Data

As highlighted previously, the main focus of this analysis will be to identify candidate locations for a new cafe, by combining location and competition data - detailed below.

2.1 Sources

1. Location Data
    a. To identify candidate locations, we will explore multiple approaches: either using the UK Postcode Locations or by defining our own grid of potential locations
2. Competitive Data
    a. To provide insight on the competitive landscape, the **Foursquare API will be used, which provides detailed information about the existing venues in the city**

2.2 Data Preparation

As mentioned previously, the first investigation will compare the merits of an arbitrarily defined grid of potential locations, with those of the UK postcode areas.

Furthermore, when looking at the Foursquare data on different venues, it will be important to focus on those which are complementary and competitive, in order to inform the analysis. Therefore, when planning for the new cafe, the following definitions will be used:

- Complementary: bars, restaurants
- Competitive: cafes

Also, in order to find a vibrant location in the city centre, it will be important to factor in the centrality of candidate locations, and whether they are more likely to be an entertainment destination or a quiet residential area.

3 Methodology

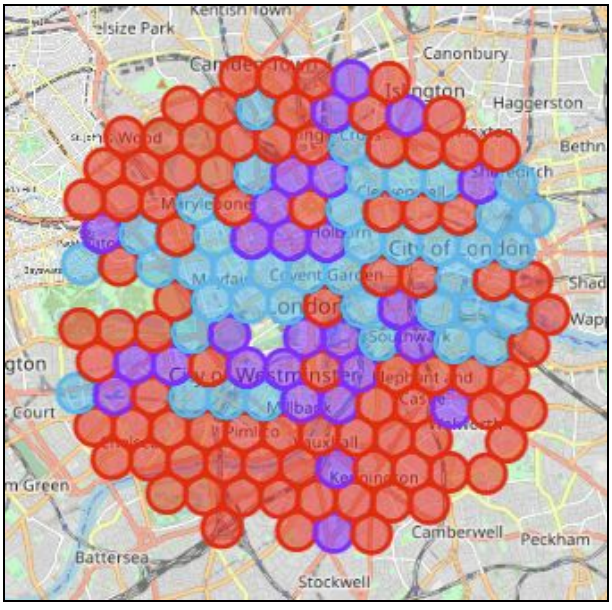
The investigation was approached in a number of steps:

1. A data frame was created with candidate locations, by comparing postcode data with a defined grid of locations - with the latter being used for the analysis
2. The Foursquare API was then used to provide information about the local venues, for each of the candidate locations, with this information a few analyses were conducted:
    a. Calculated the number venues in the area
    b. Cleaned the venue categories, to provide a manageable number of relevant categories for analysis
    c. Understanding the distribution of locations along some key dimensions
    d. Classifying locations, using a k-means algorithm (including the decision as to the best value of k - i.e. how many clusters to create)
    e. Leveraging the clusters to efficiently select out a long list of potential locations for the cafe
    f. Building and leveraging a location score (based on competing and complementary venues) to enable the selection of a short list

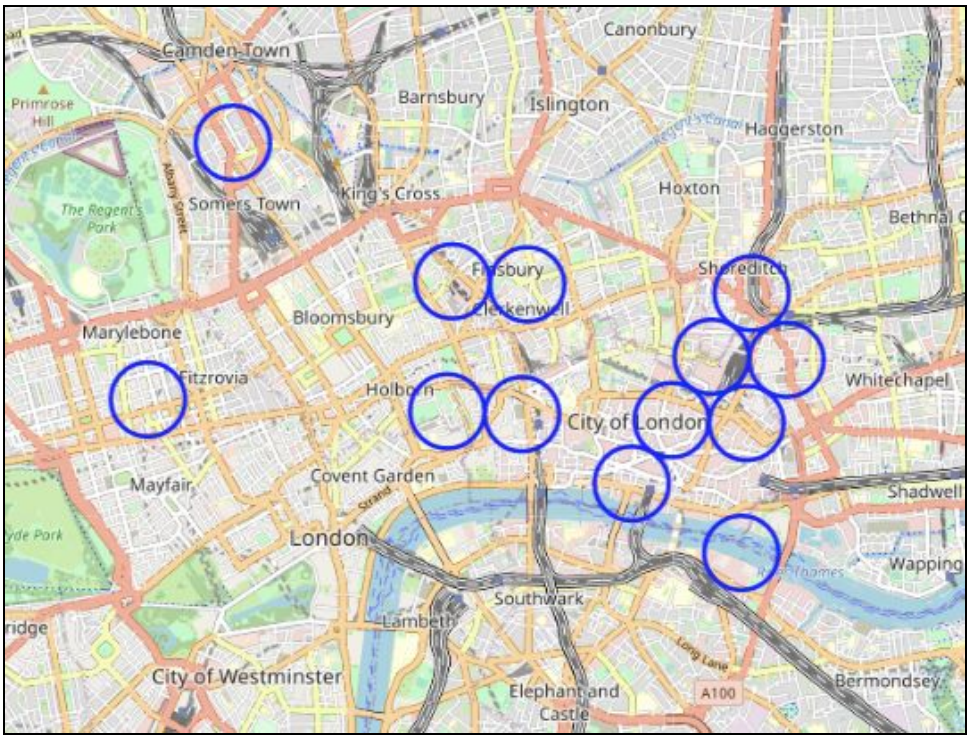
Please refer to jupyter notebook for full details and process:
https://github.com/tomreimer16/data-science-capstone/blob/master/capstone-london-analysis.ipynb

4 Results and Discussion

In summary, it was possible to define a grid of 162 candidate locations (600m apart from one another) and leverage the above methodology to filter down to the best possible locations. A cluster of 41 locations was identified, which met the criteria of being central, vibrant, and with relatively fewer existing cafes. These are shown in blue, in the below visual:



Furthermore, this long list was reduced to a 13 location short list, by leveraging the location score which was built. This allowed for a quantification of the location by trading off competing venues and complementary venues. The final locations for consideration are shown below:



Please refer to jupyter notebook for full results:
https://github.com/tomreimer16/data-science-capstone/blob/master/capstone-london-analysis.ipynb

5 Conclusion

To conclude, we have thus far identified locations that are suitable, based on analysis of Foursquare data on the local venues, and accounting for the following criteria:

- Central
- Less competed
- More vibrant, and with complementary venues

The next steps will be as follows:

- Factor in available properties for the cafe, and overlay these with the existing recommendations
- Conduct the same analytical process with the actual candidate locations, and leverage the results to inform the final choice

Areas for potential, further investigation, could be:

- Build additional datasets into the analysis, such as:

    - Public transport data as a proxy for footfall
    - Any publicly available data on spending, to inform the spending power of the locality
- Also, looking to connect with an additonal API, such as Rightmove or Zoopla, to pull in available properties, would provide an additional level of automation

Appendix

References

UK Postcode Data - https://www.freemaptools.com/download-uk-postcode-lat-lng.htm

Links

Analysis in jupyter notebook:
https://github.com/tomreimer16/data-science-capstone/blob/master/capstone-london-analysis.ipynb

Presentation available here:

https://github.com/tomreimer16/data-science-capstone/blob/master/CapstonePresentation.pdf