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

Completed by Thomas Reimer for IBM Data Science Professional Certification

Background & Context

A young entrepreneur has been lucky enough to secure funding for their new business, a cafe!

They are based in London, where it is difficult to find an ideal location due to the saturation of the market.

However, they have enlisted support to take a data driven approach, to get the edge in this competitive landscape.

This following slides will outline the work conducted to find the location for their new cafe - aiming for a vibrant area within central London.

The analysis will factor in information about the competitive landscape and leverage various data science techniques to inform the final recommendations.

Data

Sources:

- Location Data - the final approach defined a grid of potential locations
- Market Data - the Foursquare API provided detailed information about existing venues

Data Preparation

- The focus was on complementary and competitive venues, so the venue data was filtered and cleaned according to some rough definitions:
  - Complementary: bars, restaurants
  - Competitive: cafes
- To find a vibrant location in the city centre, the centrality of candidate locations was factored in, and whether they are more likely to be an entertainment destination or a quiet residential area.

Methodology

The investigation was approached in a number of steps:

1. A data frame was created with candidate locations
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

Results & Discussion

Started with a grid of 162 candidate locations (600m apart from one another)

Leveraged the previously outlined methodology to filter down to the best possible locations - within a cluster of 41 locations which were central, vibrant, and with relatively fewer existing cafes. These are shown in blue, in the below visual:
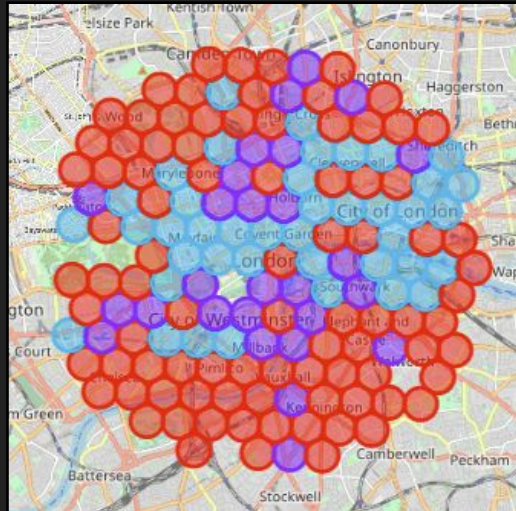
Results & Discussion

The 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

Conclusion

Thus far, initial locations have been identified 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