Comparative Area Analysis in New York and London

By Elena Tuchina

November 2018

Business Problem

The relocation company that specializes in relocating finance professionals between New York and London needs a tool to help their customers find suitable neighborhoods/districts of the city they are moving to by narrowing down a list of potential areas based on their location preferences in their home city.

This process can be greatly facilitated by customer's knowledge of their home city and its areas. By identifying what area(s) in their home city the new location should resemble, the cluster analysis can narrows down the list of the areas to consider to a few that the customer should further research/visit before making final decision.

Area Data

New York City neighborhood dataset is taken from New York University Spatial Data Repository (https://geo.nyu.edu/catalog/nyu_2451_34572) and contains the boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.

The London district dataset is created by using information from Wikipedia page (https://en.wikipedia.org/wiki/List_of_areas_of_London) and contains the boroughs and the districts that exist in each borough. Only Inner London boroughs (as per London Government Act 1963) are used in this analysis. The latitude/longitude coordinates were determined using Nominatum geolocator from Geopy library.

Venue Data

Foursquare location data is used for information on existing venues in each area of the city. The venues are aggregated by category for further frequency analysis that is served as basis for area comparison and clustering.

The following venue information was used for analysis:

- Venue name
- Venue category
- Venue latitude
- Venue longitude

Methodology – Stage 1 “Area Profile”

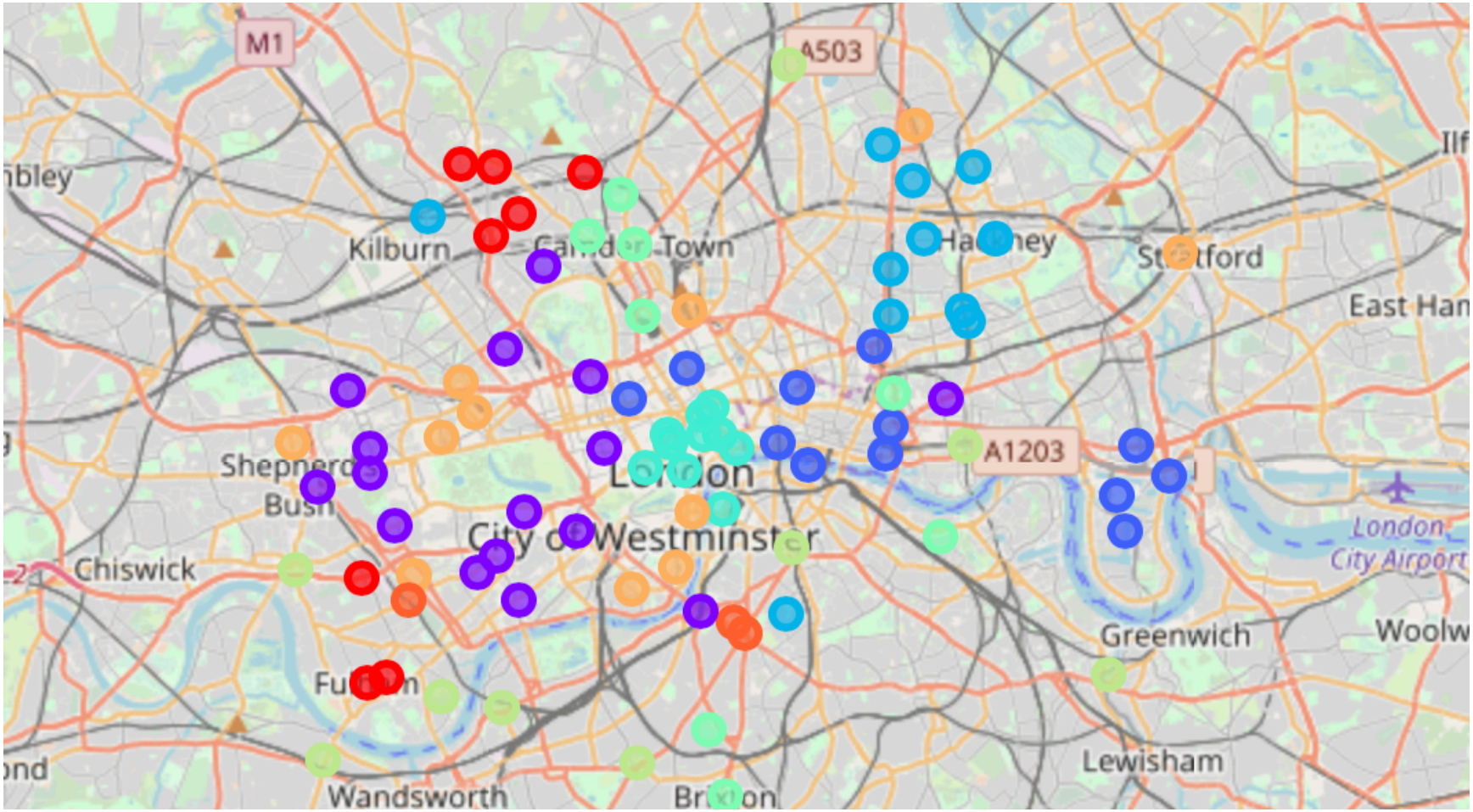
- Venue location data provided by Foursquare is obtained for each area from both New York and London datasets, and then aggregated by venue categories (*Only areas with at least 100 venues reported are retained for further analysis)*
- Aggregated venue data for the chosen New York neighborhood is added to the the aggregated venue dataset for London. This combined dataset is then converted into each area profile by calculating the frequency with which various venue categories appear

Methodology – Stage 2 “Segmentation”

- Areas are segmented by applying k-means clustering method to area profiles
- This relatively simple model is fast and sufficient for the complexity of the task at hand, and will allow for adding further dimensions to area profiles in the future
- The chosen number of clusters is 10 (based on the optimal performance of the model)

Results – Area Clusters

The analysis shows clusters of London districts on the map, with each cluster drawn in a different color:



Results – List of Areas

The result is a list of London districts, most similar to the chosen neighborhood in New York:

```
Enter neighborhood name from the list above:  Upper West Side
You have chosen to perform analysis for:  Upper West Side
```

Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Clus Lab
Aldwych	Theater	Hotel	Coffee Shop	Dessert Shop	Restaurant	History Museum	Pub	French Restaurant	Clothing Store	Cocktail Bar	
Charing Cross	Theater	Hotel	Ice Cream Shop	Plaza	Wine Bar	Lounge	Dessert Shop	Bookstore	Steakhouse	Bakery	
Chinatown	Theater	Hotel	Ice Cream Shop	Steakhouse	Plaza	Bakery	Lounge	Bookstore	Wine Bar	Seafood Restaurant	
Covent Garden	Theater	Hotel	Ice Cream Shop	Wine Bar	Coffee Shop	Steakhouse	French Restaurant	Pub	Bookstore	Beer Bar	
Holborn	Coffee Shop	Theater	Hotel	Pub	Dessert Shop	Indian Restaurant	History Museum	Restaurant	Plaza	Cupcake Shop	
Lambeth	Hotel	Theater	Pub	Coffee Shop	Event Space	Bookstore	Beer Bar	Plaza	Sandwich Place	Korean Restaurant	
Soho	Theater	Ice Cream Shop	Hotel	Bakery	Steakhouse	Pizza Place	Plaza	Lounge	Coffee Shop	Bookstore	
St James's	Hotel	Boutique	Clothing Store	Indian Restaurant	Theater	Dessert Shop	Steakhouse	Plaza	Hotel Bar	French Restaurant	
St Giles	Theater	Coffee Shop	Hotel	Ice Cream Shop	Bookstore	French Restaurant	Dessert Shop	History Museum	Steakhouse	Cocktail Bar	
Temple	Theater	Hotel	Coffee Shop	Pub	Dessert Shop	Restaurant	Cocktail Bar	History Museum	Building	Park	

Results – Discussion

- The city areas returned by the model represents a narrowed down list that most closely match customer's preferences.
- This list can now be used to perform in-depth research by the customer and relocation agency to take into account additional criteria and housing availability, followed by customer visit to the area before making final decision.

Conclusion

- The application of the analysis by the relocation company would result in allowing finance professionals to relocate quickly and successfully while staying focused on their professional transition, and with reduced cost for business.

- Further steps to improve the model could include incorporation of additional information that enhance the description of areas such as property data, educational (schools, universities, etc) and medical facilities, parks and other recreational centers, crime and safety records, ethnic composition etc.