Analysis of London Local Districts and their development in the coming years

By Riccardo Napolitano, May 31st 2020

The UK left the EU on the 31st of January 2020, 3 and a half years after the referendum.

However, until the end of the transition period (planned for the 31st of December 2020), the UK will effectively remain in the EU's customs union and single market and continues to obey EU rules.

Over the past few years, there has been a lot of analysis to try to estimate the impact of Brexit on London, and how the Capital will react to a reduction of net migration and business investments.

In this environment, it becomes significantly important to examine and understand the city's development and growth, identifying early opportunities in underdeveloped districts.

Using the Trend-based population projections for London published by Greater London Authority (GLA) Demography and venue types from Foursquare API, this project will provide:

1. Classification of districts as highly developed, downtown and less/underdeveloped
2. Footprint of future London expansion
3. Exploration of underdeveloped districts based on extraordinary population growth
4. Identification of business opportunity in urbanizing districts

In order to use the data from GLA Demography, an excel file was imported into a Jupyter Notebook where, using Python, it was cleaned and aggregated with venue information obtained using Foursquare's API. Finally, different Python packages were used for preparing visualizations and applying machine-learning concepts such as clustering.

Data

The following data sources have been used to identify the problem and quantify the results:

1. Population Projections https://data.london.gov.uk/dataset/trend-based-population-projections
2. Foursquare Developers Access to venue data: https://foursquare.com/

Methodology

The methodology includes:

- Data retrieval, exploration and wrangling.
- Performing K-means clustering algorithm to segment districts.
- Visualizing population projections and district segments.
- Understand growth pattern and urban shift.

Data retrieval, exploration and wrangling

1. The population dataset, an excel file, is provided by GLA Demography and aggregates population projections for different geographical divisions from 2011 to 2050. Spatial geographies used in this dataset are:

   - London Government Statistical Service Codes (gss_code) of the 33 Local District Authorities
   - London Government Statistical Service Names (gss_name) of the 33 Local District Authorities

   Data of Population Projection from 2011- 2050 was retrieved in a Pandas DataFrame and grouped by gss_name (renamed as Area).

   From this data a new column was created containing % of population growth from 2011 to 2050.

   Latitude and Longitude of each district was retrieved using Geocoder from GeoPy Library in Python.

	Area	Latitude	Longitude
0	Barking and Dagenham, London, United Kingdom	51.554117	0.150504
1	Barnet, London, United Kingdom	51.653090	-0.200226
2	Bexley, London, United Kingdom	51.441679	0.150488
3	Brent, London, United Kingdom	51.563826	-0.275760
4	Bromley, London, United Kingdom	51.402805	0.014814

2. Foursquare API was used to explore types of venues and their frequencies in each district. This data was issued to classify districts based on their urban development.
   Foursquare identifies 10 top-level categories, with some sub-categories that have not been considered in this project.

Arts & Entertainment	College & University	Events	Food	Nightlife Spot
Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport

The results were analyzed using a Box Plot.
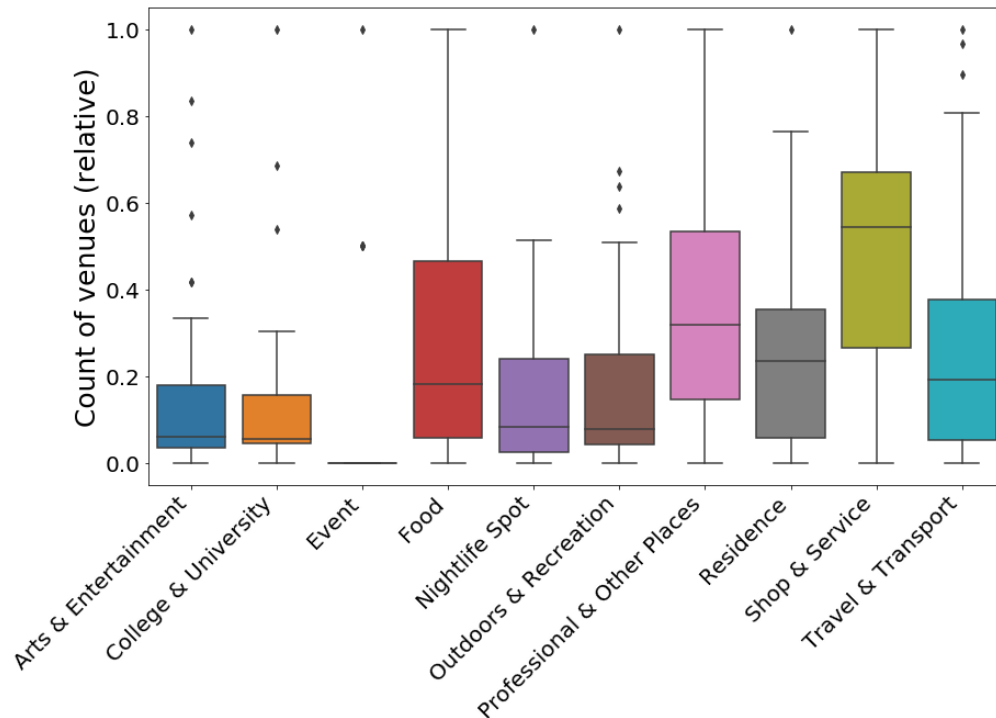
Some outliers for "*Art & Entertainment*" and "*Outdoors & Recreation*". Not enough data for "*Event*", so this category was dropped.

Performing K-means clustering algorithm to segment local districts

The following conclusions have been made for the different number of clusters:

• 2 clusters showed only the uptown/downtown divide

• 3 clusters added clustering within the downtown

• 4 clusters added clustering within uptown and downtown

• 5 and more clusters were difficult to interpret

The "Elbow Method" confirmed that the optimal number of clusters for this project was 4.
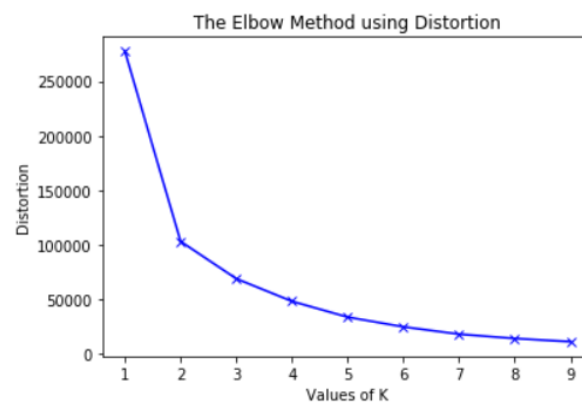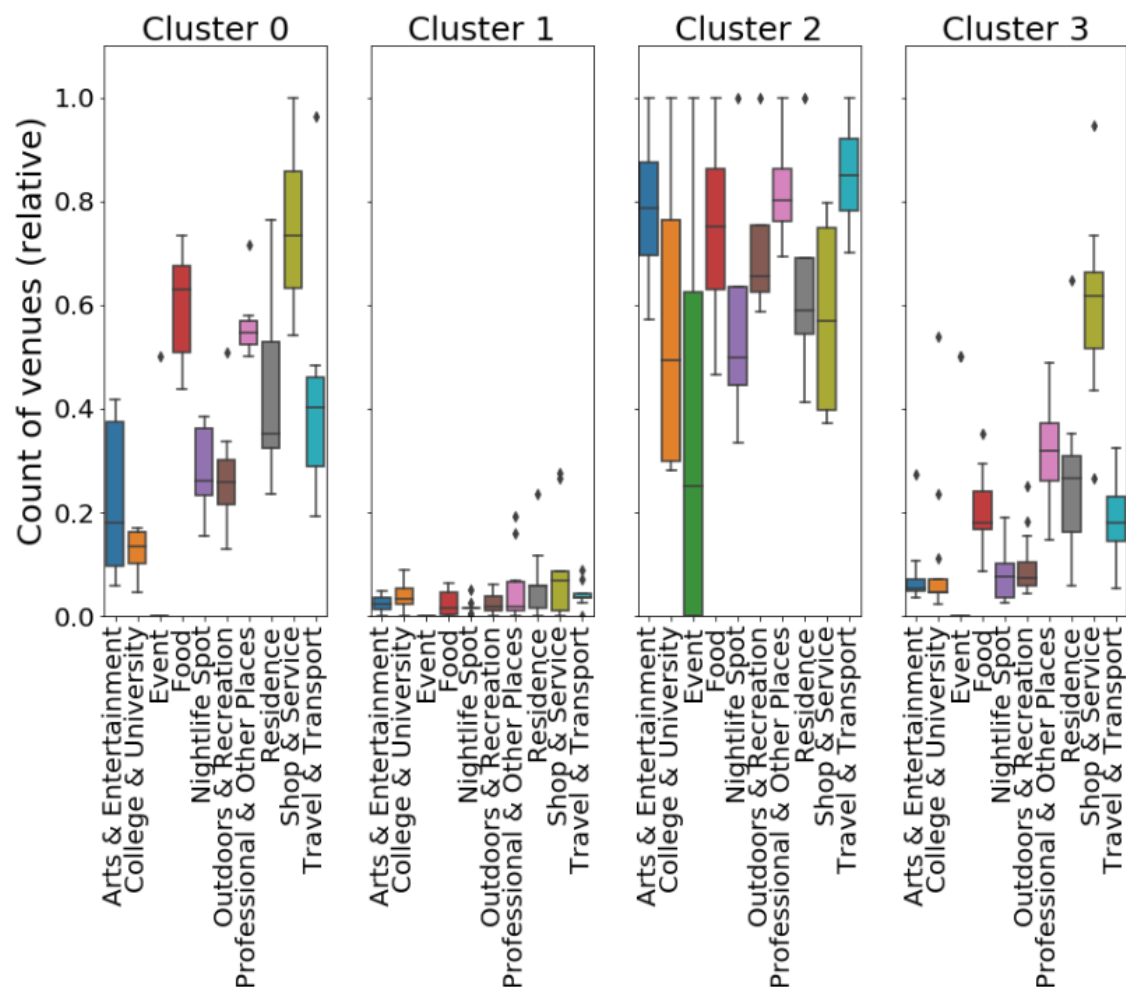
Some interpretations of the clusters, based on the Box Plot above:

• Cluster 0 (blue) has moderate scores with shops and services being the most popular. These are developed residential suburbs.

• Cluster 1 (green) has low frequencies for all venue categories. These appear to be underdeveloped districts.

• Cluster 2 (red) has consistently high frequencies for all venue categories. This is the most diversely developed part of city.

• Cluster 3 (magenta) has high frequencies but with less residential places and more professional places. These are the developed professional or industrial suburbs.

Plotting latitude and longitude confirmed that the most developed area of the city is concentrated in the most central zone of the city, and the outer areas are relatively less developed.
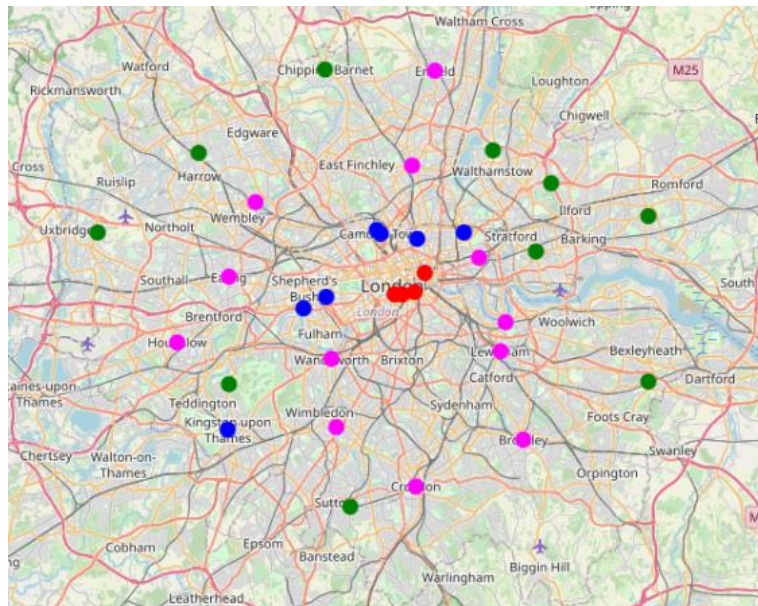
## Visualizing population projections data and districts segments

Districts with population above 300K and growth rate higher than 30% are separated in the DataFrame and selected for the last part of the analysis. These 7 districts were renamed "Developing areas".
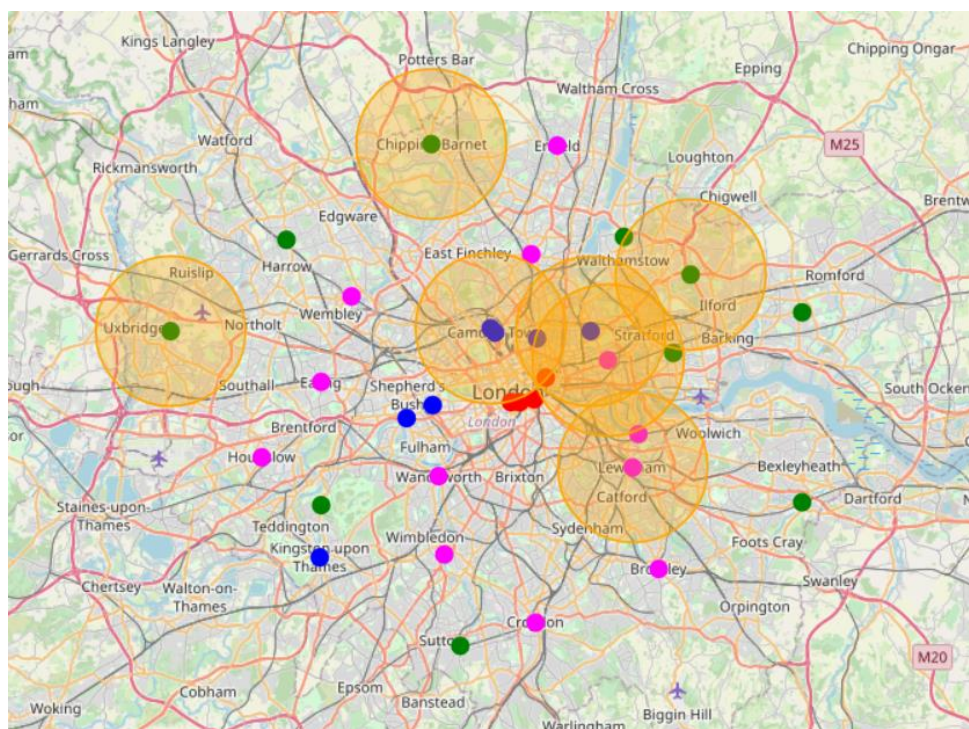
In the map above the developing areas are highlighted by bubbles quantifying the growth in population on top of clustered districts in blue, green, red and magenta.

Visualizing population projections data and districts segments

The stacked bar below highlights the growth of the developing area compared to the rest of the city.
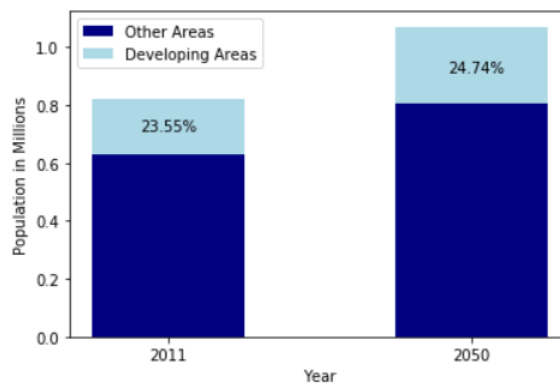
The developing areas have increased their contribution to total population by almost 120bps, from 23.55% to 24.74%.

Results

1. This analysis has highlighted that very few districts of London can be classified as highly developed, based on venues data from Foursquare API. However, the population projection data has showed that some decentralized areas of the city will see a significant growth in the coming years.

2. 7 currently underdeveloped districts, mostly located in the outer area of London, will have a significant growth in population, increasing their contribution to the total population in London by almost 120bps, from 23.55% to 24.74%. This result shows a real shift of population to new areas of the city.

3. Districts like Tower Hamlets and Redbrigde emerge as top growing areas, showing a potential transformation in new centers of interest in London.

4. Predicting these growth patterns in terms of population allows us to assume a consequent opportunity for development of potential businesses and services. A growth of this scale will have significant implication on venues that have been analyzed in this project (like Residence, Shop & Service and Food). However, further analysis will be needed to understand the bigger picture of these developments, including more complex factors (like financial crisis, new migration policies and development of public transportation) that have not been considered in this project due to high uncertainty.

Discussion and Recommendations

The aim of this project was to identify early opportunities in less developed or underdeveloped London districts that are set to grow in the future.

It would be interesting to continue this study to include additional and more complex macro factors that will significantly affect the development of these areas.

Conclusion

Using a combination of datasets from the GLA Demography and Foursquare venue data we were able to analyze, discover and statistically describe districts and population expansion quantitatively to identify some areas of interest.

Given its high-level nature, this project could be used as a first step for a more detailed study using larger datasets and including more complex factors.

More data from GLA Demography about housing development, workforce, population segmentation should be used for true valued quantitative analysis and predictive analytics, which would be most valued by investors and developers to make appropriate investments and to minimize risk.