

# Data

To identify the problem and quantify the results the following data have been used:

1. Population Projection <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 neighborhoods
- Visualizing population projections and neighborhood segments
- Understand growth pattern and urban shift

## Data retrieval, exploration and wrangling

1. The population dataset, an excel file, is provided by GLA 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 is retrieved in a Pandas DataFrame and grouped by gss-name (renamed as Area).

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

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

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