Introduction / Business Problem

Each year, many mid-career professionals relocating from London seek smaller cities. Reasons include
● More accessible price of housing
● Desire to raise small children in a safer, cleaner environment
● Desire to change pace (e.g. from a hectic banking career to owning a small local business) ...and so on

Norwich is an interesting choice of location, as it is only 2 hours away from London by train, and provides access to a majestic countryside and the scenic seasides of Yarmouth and Cromer.

In this study, we explore the neighborhoods of Norwich, attempting to cluster the neighbourhoods in the city based on the types of venues present. The output is a small number of categories, with their defining characteristics (in terms of mix of venues) which can be presented to prospective home-buyers to inform their purchasing decisions.

The key stakeholders are mid-career professionals relocating from London, as well as property agents and property sellers based in Norwich itself

Data

1. Definition of neighborhood boundaries and centroids.
● The UK Office for National Statistics maintains sets of boundaries used for different purposes, including Census, Electoral, and Local Authority boundaries.
● To zoom in at the right level, we use Lower Layer Super Output Areas (LSOA) as the basis for boundaries. LSOAs are a geospatial statistical unit used in England and Wales to facilitate thereporting of small area statistics. They are created and maintained by the ONS.
● They have a minimum population of 1000 with a mean size of 1,500.

2. Addition of data on local venues: Foursquare
● Foursquare provides the leading source of search-anddiscovery data on what types of venues exist in a given area, as well as additional information such as usage, ratings, etc.
● We use the locations data from Foursquare to map the types of venues which exist in the vicinity ofeach neighborhood.

Methodology: ONS location boundaries and centroids

1. Raw Data
● API call to ONS portal to extract GeoJSON file of Norwich LSOAs
● ONS data in the form of geoJSON polygons, displayed here in Folium map

2. Processed data
● GeoPandas library used to compute centroids, and overlaid onto Folium map

Methodology: Foursquare data

1. Most common venue categories
● Pubs are highest frequency, (not surprising, given the city's national reputation for a large number of independent breweries).
● Others: grocery stores, supermarkets and a large number of restaurants.
● Slight inconsistency with Foursquare's categories; some Restaurants are marked by type of cuisine, others are simply marked "Restaurant".
● Overall conclusion: OK for our purposes

2. Least common venue categories
● Some instantly make sense (e.g. given a small city, two bus stations seems about appropriate).
● Others slightly misleading. e.g. only one Shoe Store in Norwich - However we know it is common for such to be inside a Shopping Mall
● Rare case ("Buffet") whose significance is not apparent at all.
● Overall conclusion: OK for our purposes

Selection of k: Silhouette analysis

● Silhouette is a technique which provides a succinct graphical representation of how well each object in a cluster has been classified
● The silhouette value (or score) per object effectively measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation).
● Value ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. A negative value indicates that the object is more closely matched to another cluster than it is to its current cluster
● 4 k values tested (4,5,6,7); k= 6 is chosen since it gives the fewest number of negative kscore values

Results & Analysis
We are able to produce a map of the different clusters, and can observe from the map the distribution of the 6 clusters. We note that 2 are non-contiguous (ie they are spread over more than one geographical region. These are clusters 0, 2 and 5.
Cluster 0 is good for the professional who wants to be outside the city centre, and enjoy access to open spaces and parks, but would still like to have the convenience of easy shopping, restaurant and cafe options.
Cluster 1 is for the professional who wants to stay relatively close to the city and enjoy the convenience of multiple cafes and bars
Cluster 2 is for the professional who loves DIY, and doesnt mind being away from the social centre of the city. She can still reach the odd restaurant or bar, but she has better access to Furniture and Home supply stores
Clusters 3, 4, and 5 are more remote, and would suit the professional who is far less interested in the social scene, and rather places a premium on peace and quiet.