

Categorisation of Neighborhoods in the Norwich urban area

Anthony Maina

Data sources

Analysis will be based primarily on two datasets:

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 will 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 the reporting 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.

Data on LSOA boundaries is published free of charge as either pdf or Feature Layer on the ONS website.

In this study, we will use API calls to retrieve the geospatial boundary data. We will collect this data as a **GeoJSON data**, and manipulate it to derive insights

The screenshot shows the Open Geography Portal website. The header includes the Office for National Statistics logo and navigation links. The main content area features a search bar and a list of boundary categories. A table at the bottom right lists specific products.

Product Name	Product Type	Release Date
2011 Boundaries	PDF	15/11/2019
2011 Boundaries	Shapefile	23/10/2019

geoportal.statistics.gov.uk/datasets/lower-layer-super-output-areas-december-2011-full-clipped-boundaries-in-england-and-wales/data?where=UPPER(LSOA11NM)%20like%20%27%25NORWICH%25%27

Office for National Statistics

ONS Linked Data Portal | ONS Release Calendar | ONS Methodology | Media | About ONS

Open Geography Portal | Boundaries | Documents | Lookups | Maps | Names and Codes | Postcodes | Products | All Data

Lower Layer Super Output Areas (December 2011) Full Clipped Boundaries in England and Wales

Last updated 7 months ago | 34,753 Records

Please zoom in to see the filter applied to the map

Overview | Data | API Explorer

4/17/2019 | Feature Layer | Custom License

Download | APIs

Showing 1 to 10 of 83 | LSOA11NM: norwich

objectid	LSOA11CD	LSOA11NM	LSOA11NMW	st_area(shape)
26103	E01026791	Norwich 005A	Norwich 005A	205149.2048165
26104	E01026792	Norwich 005B	Norwich 005B	356110.0370575
26105	E01026793	Norwich 005C	Norwich 005C	292194.535432
26106	E01026794	Norwich 005D	Norwich 005D	756144.2250725
26107	E01026796	Norwich 005F	Norwich 005F	372798.9630095
26108	E01026797	Norwich 001A	Norwich 001A	3636730.700981

Download | APIs

Full Dataset

OGC WMS

<https://ons-inspire.esriuk.com/arcgis>

OGC WFS

<https://ons-inspire.esriuk.com/arcgis>

GeoService

<https://ons-inspire.esriuk.com/arcgis>

GeoJSON

<https://opendata.arcgis.com/datasets>

Filtered Dataset

OGC WMS

<https://ons-inspire.esriuk.com/arcgis>

OGC WFS

<https://ons-inspire.esriuk.com/arcgis>

GeoService

<https://ons-inspire.esriuk.com/arcgis>

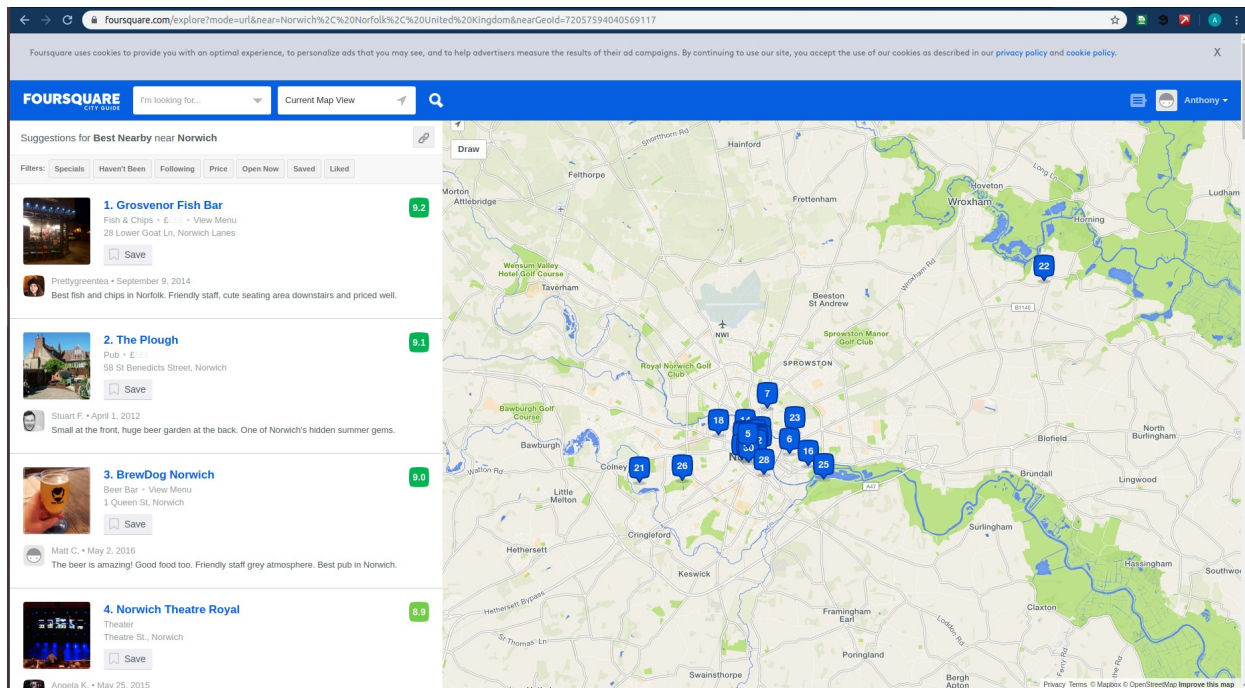
GeoJSON

<https://opendata.arcgis.com/datasets>

2. Addition of data on local venues: Foursquare

Foursquare provides the leading source of search-and-discovery data on what types of venues exist in a given area, as well as additional information such as usage, ratings, etc.

We will use the locations data from Foursquare to map the types of venues which exist in the vicinity of each neighborhood.



3. How the data will be used

The data will be used as follows:

1. Extract and refine boundaries data
 - a) Extract the data via an API call and inspect the boundaries using a visualisation package (e.g. Folium)
 - b) Once happy with the data, and any outliers or anomalies have been removed, compute the centroids which will form the “centrepoin” of each neighborhoods
2. Add Foursquare Data
 - a) For each neighborhood, extract the list of venues within a given radius (e.g. 1km); the radius will be defined based on the closeness of the neighborhoods)
 - b) Use “one-hot encoding” to convert the list of venue categories per neighbourhood into a Feature Set
 - c) Normalise the featureset by grouping the values by mean
 - d) Use **k-clustering** to define clusters based on the similarity of their features (ie mix of venue categories)
3. Perform checks and sensitivity analyses
 - a) Check the k-cluster score for different values of k, (ie the elbow method)
 - b) Check the similarity of cluster members to each other (Silhouette method)
 - c) Tune the model based on changing k as well as the radius
4. Produce results and discussion