

Demographic Catchment Area Profiles using Fuss- Free, Low-cost Approach Solutions

[https://github.com/wpoates1/
storeareacusters](https://github.com/wpoates1/storeareacusters)

Bill Oates

Cloud Capability Lead

ProjectiveGroup

bill.oates@projectivegroup.com



Aims of the session

Explore cloud-based tooling for geospatial analysis



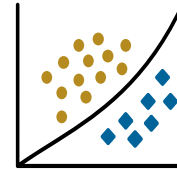
Specific use case for the use of GCP, but can be generalised to other platforms

UK Census Datasets

census2021

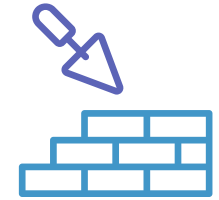
Open data sources, including recent bulk data releases

Basic Unsupervised ML techniques



Clustering, Principal Component Analysis

Foundations



Jumping-off point for future work

Introductions

Bill Oates

Ben Harland

Ben Lambiase

Chris Murchison

Inspiration - Area Classification

Inspiration and baseline methodology

<https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications>

Areas classified based on 100+ census variables

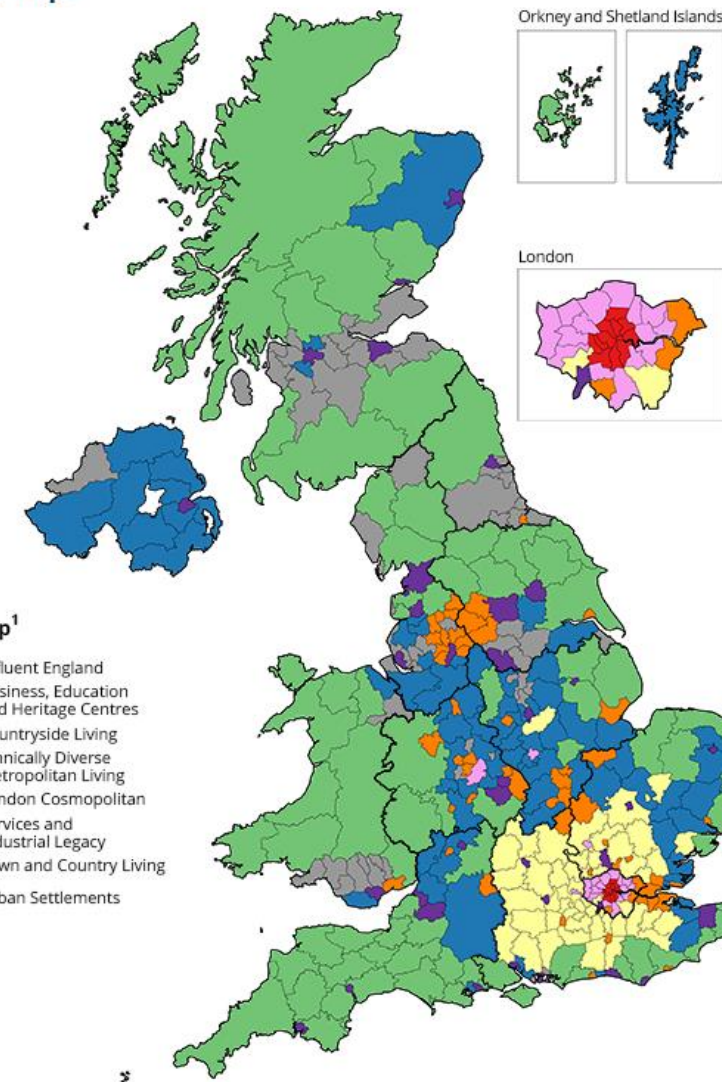
Local authorities, workplace zones and smaller areas (Output Areas)

Group like-areas together – find similarities

Useful for understanding performance characteristics at national scale

2011 Area Classification for Local Authorities: Supergroups

UK



Source: Office for National Statistics licensed under the Open Government Licence v3.0.

Contains OS data © Crown copyright 2017
Contains LPS Intellectual Property © Crown copyright and database right (2017). This information is licensed under the terms of the Open Government Licence (www.nationalarchives.gov.uk/doc/open-government-licence/version/3).

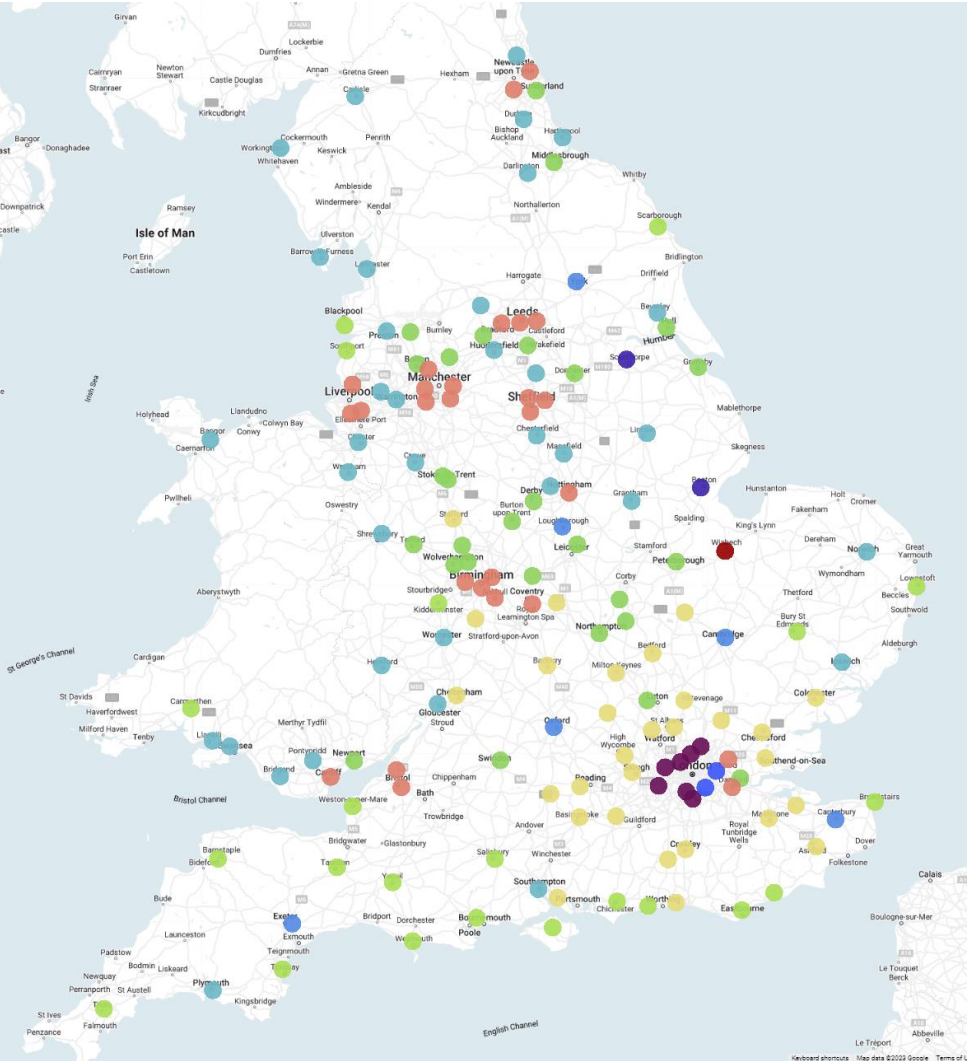
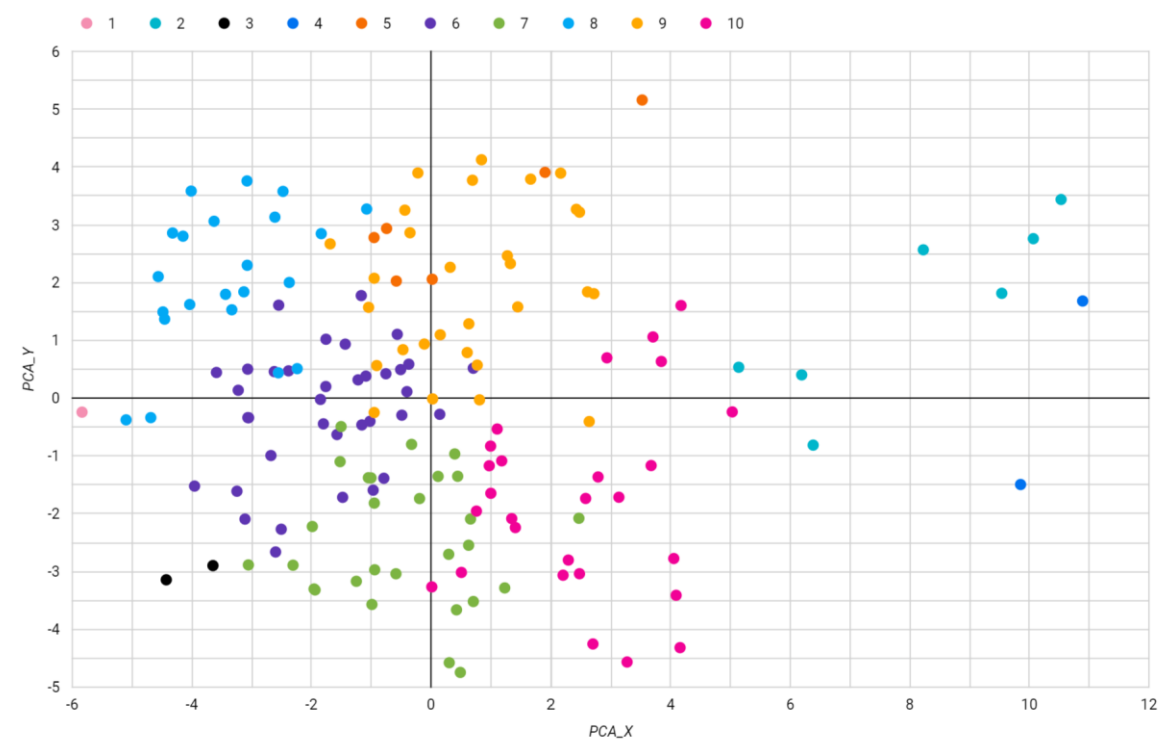
¹ The superscript (1) indicates that these are the corrected and revised supergroup clusters.

Today's specific objectives

Area Classification for Retail Store data

- Create a Project in Google Cloud to contain the work
- Download Output Area level datasets from ONS and from NOMIS websites
- Upload the data to the cloud project
- Create BigQuery tables with the raw data
- Re-shape (PIVOT) the data to row and column format
- Merge the data into a single big file
- Spatially create catchment areas for each of the 160 Stores in our dataset
- Using the population variables from each catchment
 - Use Kmeans clustering to group together similar store areas
 - Apply Principal Component Analysis to test and visualise the clustering output
- Plot the stores on a map!

KMEANS CLUSTERS - 10 Clusters, KMEANS++ method
Clustering projected onto 2-dimensional PCA space



Key Techniques Used – Spatial SQL

■ ST_GEOPOINT

- Used to create the Point Spatial Data Object (GEOMETRY data type) from latitude and longitude

Create geom column SQL

```
1  -- ADDS geom COLUMN AND CREATES THE SPATIAL POINT DATA
2  -- FOR THE STORES TABLE
3
4  ALTER TABLE `_PROJECT_._DATASET_.stores`
5      ADD COLUMN IF NOT EXISTS geom GEOGRAPHY;
6
7  UPDATE `_PROJECT_._DATASET_.stores`
8      SET geom = ST_GEOPOINT(longitude,latitude) WHERE true;
```

■ ST_DISTANCEWITHIN

- Selects all the Census Output Areas within a given distance (10KM) of each of the 160 stores

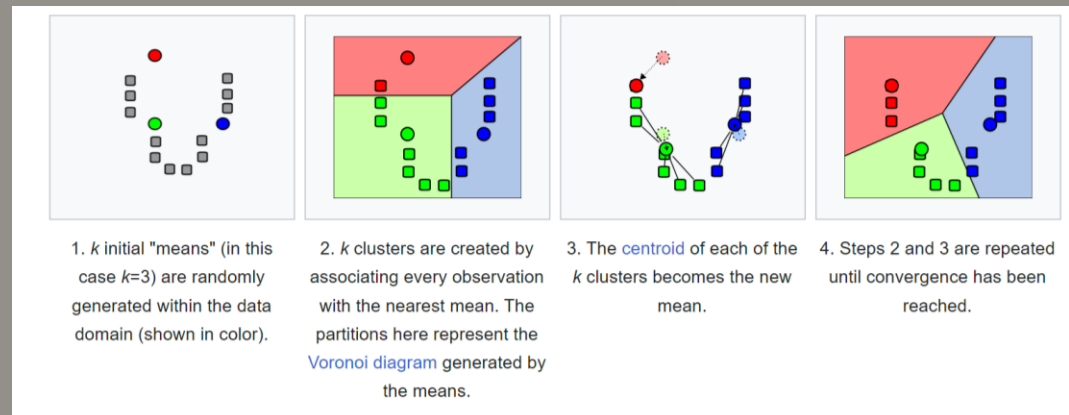
ST_DWITHIN QUERY SQL

```
1  SELECT s.StoreID as Store,
2          SUM(p.TS001Code_1) as Totalpop
3
4  FROM `_PROJECT_._DATASET_.stores` s,
5       `_PROJECT_._DATASET_.PVT_Conflated` p
6
7  WHERE ST_DWITHIN(p.centroid, s.geom, 10000)
8
9  GROUP BY Store
```

Key Techniques Used – AutoML

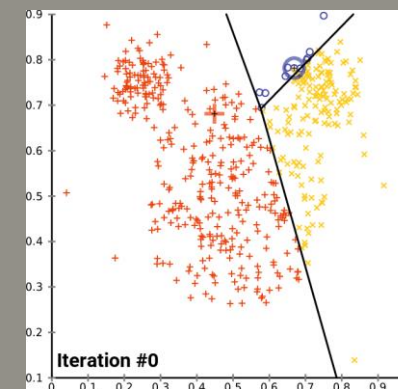
■ KMeans Clustering

- Method to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centre or cluster centroid)



■ Principal Component Analysis

- Enables reduction of the number of dimensions to facilitate plotting



What factors will we be using?

Variable	Summary Method
Output Areas	Count
Household and Non-Household Population	Sum
Non-household Population	Proportion of Total
Density	Persons per Hectare
Aged under 5	Proportion of Total Population
Aged 5-14	Proportion of Total Population
Aged 25-44	Proportion of Total Population
Aged 45-64	Proportion of Total Population
Aged 65-84	Proportion of Total Population
Aged over 85	Proportion of Total Population
Higher and Lower Managerial, Administrative and Professional Occupations	Total persons
Higher and Lower Managerial, Administrative and Professional Occupations	Proportion of Total Population

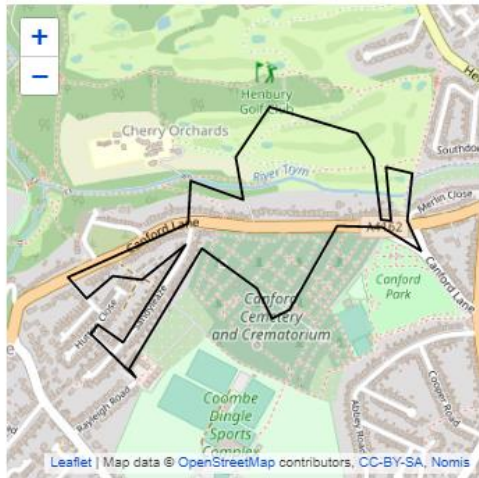
Census Output Areas

E00074345 Output Area Local Area Report

This report covers the characteristics of people and households in E00074345 Output Area in Bristol, City of (GSS code E00074345). Figures are sourced from the 2011 Census key statistics.

There were 269 usual residents as at Census day 2011. Of these, 100% lived in households and 0% lived in communal establishments. The average (mean) age of residents was 47.2 years.

In total there were 115 household spaces. Of these, 114 (99.1%) had at least one usual resident and 1 (0.9%) had no usual residents.



<https://www.nomisweb.co.uk/reports/localarea>

- 188,880 in England and Wales, each one generally with :
 - 40 - 250 households
 - 100 - 625 usually resident persons



<https://geoportal.statistics.gov.uk/datasets/ons::output-areas-dec-2021-boundaries-generalised-clipped-ew-bgc>

Let's get going!!!

———— All the instructions and the SQL code needed for the session is at:

[https://github.com](https://github.com/wpoates1/storeareacusters)
[/wpoates1/](https://github.com/wpoates1/storeareacusters)
[storeareacusters](https://github.com/wpoates1/storeareacusters)