## 2) Data

### a. What data will be used

For this analysis data from the ONS and Foursquare will be used.

ONS Data
The ONS publishes every 10 years since 1981 the results of the UK population census. The data available provides a view on the population demographics, the spending habits, the working population and much more. It is used by companies and public bodies to perform economic forecasts. The ONS also made available a postcode mapping table linking the census data to geographic areas. Hence it is possible to combine the census data with geographic data to identify the population demographics for defined locations.
For this analysis several datasets from the ONS will be used:
- UK Postcode Directory
- UK weekly expenditures by household
- UK population by OAC (see the UK postcode directory overview for definitions)
- UK LSOA shape files

This data will be used in conjunction with the Foursquare API which will provide the point of interests (venues) in Birmingham.

### b. Source of the data

#### i. UK Postcode Directory

**Source**
The data is available at this location: https://geoportal.statistics.gov.uk/datasets/ons-postcode-directory-february-2019
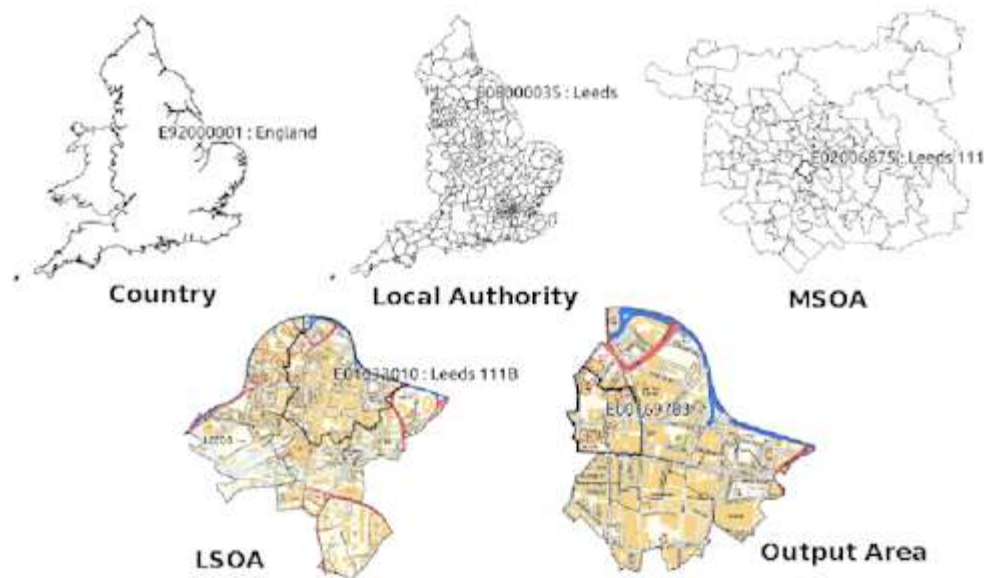
**Use**
This data will be used to map the population of Birmingham by socio economic groups and spending but also to map the working population. The original dataset is at Output Area level which is the most granular data point for the census geography. This level is too small therefore the data will be aggregated at Lower Super Output Areas (see below).

**Data overview**
It provides a mapping of the UK postcodes to various geographic areas. For the census data, specific geographic boundaries were created. These areas are called Census Output Areas. The most granular level is the Output Area (OA) and the highest is the Country.
Each postcode belongs to an Output Area. For this analysis we will use the LSOA level (Lower Super Output Area.

Overview of the Census boundaries:

The dataset is made of over 30 fields. Here is a snapshot of the main fields:

| pcds | oa11 | lsoa11 | msoa11 | oac11 | lat | long |
|------|------|--------|--------|-------|-----|------|
| B1 1AA | E00175658 | E01033625 | E02006899 | 2B2 | 52.47666 | -1.90354 |

- pcds is the postcode
- oa11 is the Output Area code
- lsoa11 is the Lower Super Output Area code
- msoa11 is the Middle Super Output Area code

## ii. Mid-2017 Population Estimates (West Midlands)

**Source**
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationesti
mates/datasets/censusoutputareaestimatesinthewestmidlandsregionofengland

**Use**
The data will be use create the age groups for the population of Birmingham.

**Data overview**
The data provide the population estimate (2017) at oa11 level and by age (from 0 to 90+ by increments
of 1 year).

## iii. UK weekly spending by household

**Source**
https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expendit
ure/datasets/averageweeklyhouseholdexpenditurebyoutputareaclassificationoacgroupuktablea52

**Use**
This data will provide an estimate of the spending in restaurants and hotels in an area by calculating
the average weekly spending by person in an output area. The ONS only provides spending for
restaurants and hotels together hence this will be use as a proxy for restaurant spending.

**Data overview**

Provides the weekly household expenditure on good and services in the UK by area and by output area category group. The file provides the average number of persons per household.

| OAC group | OAC code | Weekly Restaurants and Hotels expenditure by household (in £) |
|---|---|---|
| Farming Communities | 1A | 51.50 |
| Rural Tenants | 1B | 51.80 |
| Ageing Rural Dwellers | 1C | 39.10 |
| Students Around Campus | 2A | 71.70 |
| Inner-City Students | 2B | 62.30 |
| Comfortable Cosmopolitans | 2C | 45.50 |

Each Output Area is assigned an OAC group which broadly describe the socio-economic groups (OAC Group).

### iv. UK working population (West Midlands)

**Source**
The dataset is available on the NOMIS website which is part of the ONS.
https://www.nomisweb.co.uk/query/construct/submit.asp?menuopt=201&subcomp=

**Use**
The data will provide a view on the working population of Birmingham by LSOA level. The working population is people who are working in the area but are not necessarily living there.

**Data overview**
All usual people aged 16 to 74 in employment in the area the week before the census.

### v. UK lsoa11 shape files
**Source**
https://geoportal.statistics.gov.uk/datasets/lower-layer-super-output-areas-december-2011-full-clipped-boundaries-in-england-and-wales

**Use**
This file contains the geographic shapes of the LSOA in the UK.

### vi. Foursquare API
The data will be use to get the venues in Birmingham.