Codeclan: Final Project Documentation

Scottish Household Survey: Access to green space and its effects on neighbourhoods and communities

Jonathan Parker
October 2021

Context

Business intelligence and data-driven decision making

What insights can the business/organisation gain from your analysis and how will your analysis help the business/organisation make better decisions?

The analysis highlights the demographic groups and geographic areas (at local authority level) that have the worst access to green or blue space, and the potential effects this has on how people perceive and rate their local neighbourhoods and communities.

This analysis will help the business determine where and how to focus their efforts on improving access to green space for those who need it most.

Domain knowledge and the business context

Briefly describe the business/organisation and where your analysis fits within its aims/activity.

This project is based on the Scottish Household Survey which is conducted by the Scottish Government every year and is designed to provide reliable and up-to-date information on the composition, characteristics, attitudes and behaviours of private households and individuals in Scotland, at both the national and local level.

It covers a wide range of topics to allow links to be made between different policy areas, including:

- Access to green space;
- Community belonging; and
- Neighbourhood rating

The analysis in this project will assist the Scottish Government in their assessment of current access to green space based on the most recent data (2019). It also gives an insight to potential links between access to green space and community/neighbourhood ratings, at both demographic and geographic levels.

Data

Internal and external data sources

Briefly describe your data sources and which were internal (provided by the organisation) and external (not sourced from the organisation's data)

This project used three data sources provided by the organisation:

- Distance to Green or Blue Space (percentage of adults living within various walking distances to their nearest green or blue space
- Neighbourhood rating (how adults in Scotland rate their local neighbourhood as a place to live)

- Community belonging (how strongly adults in Scotland feel they belong to their immediate neighbourhood)

All use data captured as part of the Scottish Household Survey and are from https://statistics.gov.scot/

Types of data

What kind of data did you work with? E.g. categorical and numerical data and their sub-types.

All three data sets use a mix of categorical, numeric (all values are percentages) and time series data (year that the data was captured).

The project also includes a spatial analysis using polygon data.

Data formats

What format did your data come in? E.g. all downloaded flat files (CSV) or any data from APIs, scraping etc.

The three data sets provided by the organisation were .csv files.

The shapefile data used to create the spatial analysis was constructed from four files (.dbf, .prj, .shp, and .shx) sourced from https://spatialdata.gov.scot

Data quality and bias

Briefly describe the quality of the data and whether you have any reasons to suggest the data is biased e.g. only data from a specific demographic even though a broader demographic would be of interest to the organisation.

All data used in this project was sourced from either statistics.gov.scot or spatialdata.gov.scot and is therefore of very high quality.

The data was captured via the Scottish Household Survey which is based on a random sample of people living in private residences, leaving very little chance of bias.

Ethics

Ethical issues in data sourcing and extraction

Do you have any ethical concerns regarding the sourcing and extraction of your data?

The Scottish Household Survey is voluntary and based on a random sample. Taking part in the survey is not mandatory and all data is anonymised.

No concerns regarding the sourcing and extraction.

Ethical implications of business requirements

Are there any ethical implications of the business requirements?

The Scottish Government use the data captured in the Scottish Household Survey to provide reliable and up-to-date information on people and areas in Scotland with a view to improving public services and many other policy areas besides.

No ethical implications of the business requirements.

Analysis

Stages in the data analysis process

What were the main stages in your data analysis process?

- Selecting a brief and sourcing appropriate data
- Downloading and exploring data (including reading the latest annual report published alongside the survey to acquire more domain knowledge)
- Cleaning, wrangling and joining datasets to enable analysis
- Plotting data and using visualisations to interpret results
- Writing presentation based on findings and delivering recommendations to business

Tools for data analysis

What were the main tools you used for your analysis?

All cleaning, wrangling and analysis for this project was conducted in R using R Studio.

Trello was also used to identify deliverables and track their progress during the course of the project.

Descriptive, diagnostic, predictive and prescriptive analysis

Please report under which of the below categories your analysis falls and why (can be more than one)

Descriptive Analytics tells you what happened in the past.
Diagnostic Analytics helps you understand why something happened in the past.
Predictive Analytics predicts what is most likely to happen in the future.
Prescriptive Analytics recommends actions you can take to affect those outcomes.

This analysis can be categorised as descriptive analytics as it is based on survey data which is taken once a year. It provides insights on that annual snapshot of data which captures Scottish people's responses to questions on their local area.