|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Business question | Sub question(s) | Relevant dataset(s) and variables? | What visualisation(s) or tests will answer the question? | What did the test/chart tell us? | Is it important? Why? | What evidence can explain/support this finding? |
| What differences in access to green spaces are evident in the data?   * Who has more or less access? (demographics/ individual characteristics) * Which parts of the country have more or less access? (geographical) | Are there any differences in access between age groups? | Age  Distance to green space  Percent | Grouped dataset  Grouped bar chart | The only difference that stands out is between people over 65 and the other two groups | The difference was significant in the Kruskal-Wallis test, meaning that, on average, significantly fewer people over 65 have access to a green space within 5 minutes walk from their place of residence when compared with the other age groups. |  |
|  | Are there any differences in access between sexes? | Gender  Distance to green space  Percent | Grouped dataset  Grouped bar chart | There aren’t any differences between age groups that are standing out in the grouped bar charts | Not important, no major differences. | N/A |
|  | Are there any differences in access between the most and least privileged SIMD quintiles? | Simd\_quintile  Distance to green space  Percent | Grouped dataset  Grouped bar chart | It looks like there is a difference between the most deprived 20% and the least deprived 80%. On average 59% of people in SIMD 1 can access a green space in a 5 minute walk or less while 67% of people in the least deprived 80% can access a green space in 5 min or less. | Important. There is a statistically significant difference between the most and least privileged SIMD quintiles, showing that significantly more people in the least deprived 80% can access a green space within 5 min of their residence than people in the most deprived 20% |  |
|  | Are there any differences in access between type of tenure? (e.g. more social rentals could be available in urban areas) | Type\_of\_tenure  Distance to green space  Percent | Grouped dataset  Grouped bar chart |  |  |  |
|  | Are there differences in access by the household type? | Household\_type  Distance to green space  Percent | Grouped dataset  Grouped bar chart | The difference that stands out in the bar plots is between pensioners and the other household types | In the ANOVA, the difference between pensioners and the other groups was significant, showing that pensioners have less access to green spaces |  |
|  | Are there differences in access by ethnicity? | Ethnicity  Distance to green space  Percent | Grouped dataset  Grouped bar chart | There is a noticeable difference in access between white people and people classified as “other” ethnicity. White people have significantly better access to green spaces | The two sample difference in means test was significant, showing that significantly more white people can access a green space within 5 min walk from their residence when compared with people from “other” ethnicities. |  |
|  | Are there any differences by local authority? | Ca\_name  Distance to green space  Percent | Grouped dataset  Grouped bar chart | There are definitely some differences, but too many local authorities to do testing properly. The best approach might be to pick out two Las that are very different and compare them. | East Lothian and West Dunbartonshire are very different |  |
|  | Are there any differences between rural and urban areas? | Urban\_rural  Distance to green space  Percent | Grouped dataset  Grouped bar chart |  |  |  |
| How do people in neighbourhoods with good access to green space differ from those who have no good access? | Are there differences in how they rate their neighbourhoods? |  |  |  |  |  |
|  | Are there differences in how they rate their communities? |  |  |  |  |  |
| Is there any way to predict which households would have higher ratings? |  |  |  |  |  |  |

About the datasets:

Neighbourhood\_rating

This dataset focuses on the neighbourhood rating and allows you to breakdown these ratings in different ways. We can look at the data over time (years), geographically (LA, urban/rural classifier), and broken down by participant characteristics (SIMD, sex, type of household, type of tenure). We can also see this rating in relation to the walking distance to a green space (this relates to another dataset – green\_space). This implies that they expect some relationship between neighbourhood rating and walking distance to a green space. There are no missing values

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Definition | Values | Possible actions |
| Feature\_code | Code that identifies the local authority area | All local authority areas and a code for all of Scotland | Join LA authority names for a clearer breakdown |
| Date\_code | year of the measurement | 2013 – 2019 | Nothing |
| Measurement | Type of measurement | * percent, * upper 95% confidence interval * lower 95% confidence interval | We can look at how wide the confidence intervals are by plotting all three lines. We have quite a lot of data, so I would not expect them to be super wide since our sample is pretty big. |
| Units | only one distinct value, letting us know that we are measuring the percent of adults giving each rating |  | We can probably drop this |
| Value | the percent of adults assigning a particular rating |  |  |
| Neighbourhood\_rating | 4 point qualitative scale for participants to rate their neighbourhood | * very poor, * fairly poor, * fairly good, * very good * no opinion (spontaneous) | Transform to ordered factor  Can transform to binary poor/good scale for predictive model |
| Gender | Technically the participant’s gender since the survey says that participants can define themselves in another way than male/female, but this info was not recorded in the dataset, so closer to biological sex. | * Male * Female * All |  |
| Urban\_rural\_classification | 2-fold urban rural classification | * urban, * rural * all |  |
| Simd\_quintiles | Scottish index of multiple deprivation. Note that this is not an actual quintile breakdown, there are only 3 distinct values | * 20% most deprived (SIMD 1), * 80% least deprived (SIMD 5) and * all. |  |
| Type\_of\_tenure | Whether the person owns or rents the property | * all, * owned mortgage/loan, * owned outright, * social rented, * private rented, * other |  |
| Household\_type | Breakdown by type of household | * all, * adults, * with children, * pensioners |  |
| Ethnicity | breakdown by ethnicity | * all, * white, * other |  |
| Walking\_distance\_to\_nearest\_greenspace | The amount of time it takes to walk to the nearest greenspace (for example, a park, countryside, wood, play area, canal path, riverside or beach – not including private gardens) | * all, * don’t know (spontaneous only), * less than 10 minutes, * more than 10 minutes |  |

Community belonging dataset

This dataset focuses on the community belonging rating and allows you to breakdown these ratings in different ways. We can look at the data over time (years), geographically (LA, urban/rural classifier), and broken down by participant characteristics (SIMD, sex, type of household, type of tenure). We can also see this rating in relation to the walking distance to a green space (this relates to another dataset – green\_space). This implies that they expect some relationship between community belonging and walking distance to a green space. There are no missing values

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Definition | Values |  |
| Feature\_code | Code that identifies the local authority area | All local authority areas and a code for all of Scotland |  |
| Date\_code | year of the measurement | 2013 – 2019 |  |
| Measurement | Type of measurement | * percent, * upper 95% confidence interval * lower 95% confidence interval |  |
| Units | only one distinct value, letting us know that we are measuring the percent of adults giving each rating |  |  |
| Value | the percent of adults assigning a particular rating |  |  |
| Community\_belonging | 4 point qualitative scale for participants to rate their sense of belonging in their neighbourhood (\*defined as the street they live on an the streets nearby (urban) and the local area (rural)) | * Not at all strongly * Not very strongly * Fairly strongly * Very strongly * Don’t know (spontaneous) | Transform to ordered factor  Can transform to binary poor/good scale for predictive model |
| Gender | Technically the participant’s gender since the survey says that participants can define themselves in another way than male/female, but this info was not recorded in the dataset, so closer to biological sex. | * Male * Female * All |  |
| Urban\_rural\_classification | 2-fold urban rural classification | * urban, * rural * all |  |
| Simd\_quintiles | Scottish index of multiple deprivation. Note that this is not an actual quintile breakdown, there are only 3 distinct values | * 20% most deprived (SIMD 1), * 80% least deprived (SIMD 5) and * all. |  |
| Type\_of\_tenure | Whether the person owns or rents the property | * all, * owned mortgage/loan, * owned outright, * social rented, * private rented, * other |  |
| Household\_type | Breakdown by type of household | * all, * adults, * with children, * pensioners |  |
| Ethnicity | breakdown by ethnicity | * all, * white, * other |  |
| Walking\_distance\_to\_nearest\_greenspace | The amount of time it takes to walk to the nearest greenspace (for example, a park, countryside, wood, play area, canal path, riverside or beach – not including private gardens) | * all, * don’t know (spontaneous only), * less than 10 minutes, * more than 10 minutes |  |

Green spaces dataset

This dataset focuses on the distance to a green space, but unlike the other datasets, this contains all the options for distance to a green space (the others only have a binary split between participants that reported a green space within a 10 minute walk and those that said it took longer). It also allows you to breakdown these ratings in different ways. We can look at the data over time (years), geographically (LA, urban/rural classifier), and broken down by participant characteristics (SIMD, age, sex, type of household, type of tenure). This contains an additional demographic variable (age). There are no missing values

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Values |
| Feature\_code | Code that identifies the local authority area | All local authority areas and a code for all of Scotland |
| Date\_code | year of the measurement | 2013 – 2019 |
| Measurement | Type of measurement | * percent, * upper 95% confidence interval * lower 95% confidence interval |
| Units | only one distinct value, letting us know that we are measuring the percent of adults giving each rating |  |
| Value | the percent of adults that reported each of the options for distance to a green space |  |
| Distance\_to\_nearest\_green\_or\_blue\_space | 3 options for the walking distance to the nearest green or blue space, which is defined as a park, countryside, wood, play area, canal path, riverside or beach – not including private gardens | * A 5 minute walk or less * Within a 6-10 minute walk * An 11 minute walk or more * Don’t know (spontaneous only) |
| Age | Age ranges for participants | * All * 16-34 years * 35-64 years * 65 years and over |
| Gender | Technically the participant’s gender since the survey says that participants can define themselves in another way than male/female, but this info was not recorded in the dataset, so closer to biological sex. | * Male * Female * All |
| Urban\_rural\_classification | 2-fold urban rural classification | * urban, * rural * all |
| Simd\_quintiles | Scottish index of multiple deprivation. Note that this is not an actual quintile breakdown, there are only 3 distinct values | * 20% most deprived (SIMD 1), * 80% least deprived (SIMD 5) and * all. |
| Type\_of\_tenure | Whether the person owns or rents the property | * all, * owned mortgage/loan, * owned outright, * social rented, * private rented, * other |
| Household\_type | Breakdown by type of household | * all, * adults, * with children, * pensioners |
| Ethnicity | breakdown by ethnicity | * all, * white, * other |