|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Business question | Sub question(s) | Relevant dataset(s)? | 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? | Green spaces | 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? | Green spaces | 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? | Green spaces | 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) | Green spaces | Grouped dataset  Grouped bar chart |  |  |  |
|  | Are there differences in access by the household type? | Green spaces | 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? | Green spaces | 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 other ethnic groups. There is evidence of a racial disparity here. |  |
|  | Are there any differences by local authority? | Green spaces | 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, we can compare them as a small case study to illustrate differences between local authorities. |  |
|  | Are there any differences between rural and urban areas? | Green spaces | Grouped dataset  Grouped bar chart | On the graph, people living in rural areas look to have better access to green spaces within a 5 min walk from their residence – 77.8% vs 65.3% | This difference was significant at the 99.99% level (p < 0.01) |  |
| How has access to green space changed over time? | What is the trend over time in access across all of Scotland? | Green spaces | Line graph showing overall access over time for all Scotland | This is holding very steady in general and shows that average access across Scotland hasn’t changed very much. | It's probably not very important to include the overall graph |  |
|  | What is the trend over time in access for different demographic groups? | Green spaces | Line graphs showing the mean percentage for different distances to green space over time faceted by demographic grouping variables that were significant in Q1 (SIMD, ethnicity, age, household type) | -Access to green space for the 20% most deprived is improving over time  -Access to green space for other ethnic groups is getting worse over time  -Access to green space for people 65+ and pensioner households has gotten a bit worse | It would be good to include these in the demographics to show whether access is improving or not for people who had significantly less access |  |
|  | What is the trend over time in access for different geographical areas? | Green spaces | Line graphs showing the mean percentage for different distances to green space over time faceted by geographic grouping variables that were significant in Q1 (urban/rural, LA comparison) | -Rural access is showing an overall downward trend in access to green spaces while urban access has shown very slight improvement.  -West Dunbartonshire access looks quite unstable while access in East Lothian hasn’t changed very much | Not sure how important these are, the council comparison is not really that interesting here. It could be good to include the urban-rural graph just for consistency with the demographic groups comparisons |  |
| 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? | Neighbourhood rating | Box plots  Independent samples diff in means test | The test was significant showing that people who have better access to green spaces (less than 10 min walk) are more likely to rate their neighbourhood as “Very good” or “Fairly good” |  |  |
|  | Are there differences in how they rate their communities? | Community belonging | Box plots  Independent samples diff in means test | The test was significant showing that people who have better access to green spaces (less than 10 min walk) are more likely to rate their neighbourhood as “Very good” or “Fairly good” |  |  |
| Is there any way to predict which households would have higher ratings? | Can a statistical model accurately predict neighbourhood ratings using other variables as predictors? | Aggregate dataset | Random forest - AUC  Logistic regression - AUC | The AUC for the random forest model on test data was 0.774.  The AUC for the logistic regression model on test data was 0.815 | The logistic regression model was best at predicting neighbourhood ratings |  |
|  | Which variables are important for predicting neighbourhood rating? | Aggregate dataset | Random forest – importance  Logistic regression – statistical significance |  |  |  |

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 |

Aggregate dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Definition | Values | Missing data |
| Year | year of the measurement | 2012 - 2019 | None |
| 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 | None |
| Neighbourhood rating | 4 point qualitative scale for participants to rate their neighbourhood | * very poor, * fairly poor, * fairly good, * very good * no opinion (spontaneous) | None |
| Distance to nearest greenspace | 5 categories for how long it takes to walk to the nearest greenspace (not including private gardens) | * A 5 minute walk or less * More than a 30 minute walk away * Within a 21-30 minute walk * Within a 6-10 minute walk * Within an 11-20 minute walk * Don't know | None |
| Satisfaction with nearest greenspace | 5 point qualitative scale for participants to rate their satisfaction with their nearest greenspace | * Fairly dissatisfied * Fairly satisfied * Neither satisfied nor dissatisfied * No opinion * Very satisfied * Very dissatisfied | 9988 missing  Could impute with no opinion? |
| Age | Participant’s age – three categories | * 16-34 years * 35-64 years * 65 years and over | None |
| 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 | None |
| Economic status | Employment status | * Full Time Employment * Part Time Employment * Training * Other * Retired * Self Employed | 1 missing  Can drop |
| Household size | Number of people in the person’s household | * 1 - 10 | * None |
| Highest education level | The person’s highest completed education level | * Degree, Professional qualification (Above SVQ Level 4) * HNC/HND or equivalent (SVQ Level 4) * Standard grade or equiv (SVQ level 1 or 2). * Higher, A level or equivalent (SVQ Level 3) * Other qualification | * 10920 missing |
| Nearest green space use | How frequently the participant visits their nearest greenspace | * Several times a month * Once a week * Once a month * Several times a week * Every day * Not at all * Less often * Don't know | * 16766 missing   Could impute with don’t know? |
| Volunteering last twelve months | Whether or not the person/ people have volunteered in the last 12 months | * Yes * No | * 6778 missing * Could create a new category called “don’t know” or “no information” |
| N persons | The number of people with the observed ratings and characteristics in the other columns | * 1 - 94 | None |