**Personal questions**

1. **~~What is ALL in the survey? Why there’s a big chunk of ALLs?~~**

**~~\*My guess is that all the ALL values are more or less the sum of all the values in the 3 tables. But still doesn’t make sense… I’m thinking on just dropping all the ALL values and do analysis on the remaining values.~~**

**~~OR… ALL represents all the data collected and then this data has been filtered or grouped afterwards into relevant data. So the values different from ALL were still part of ALL. But only the relevant data has been gathered. So I can discard ALL.~~**

1. **~~Measurement?~~**
2. **~~Value?~~**
3. **~~Does value correspond to community belonging in community\_belonging?~~**
4. **~~Simd\_quintiles… what does the 20% or 80% represent?~~**
5. **~~Issue encountered while trying to calculate council areas and vicinity to green spaces. The data is not proportionate to the population the council area corresponds to. E.g. South Lankarshire seems to have the most surveys and therefore is also the city with the closest and furthest access to green spaces.~~**

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?

#### **Domain knowledge and the business context**

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

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)

#### **Types of data**

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

#### **Data formats**

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

#### **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.

Ethics

#### **Ethical issues in data sourcing and extraction**

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

#### **Ethical implications of business requirements**

Are there any ethical implications of the business requirements?

## **Analysis**

#### **Stages in the data analysis process**

What were the main stages in your data analysis process?

#### **Tools for data analysis**

What were the main tools you used for your analysis?

#### **Descriptive, diagnostic, predictive and prescriptive analysis**

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

**BUSINESS QUESTIONS**

General insight on how people feel like in Scotland about their communities. Relationship between distance to outdoor space, and neighbourhood ratings.

(compare how people feel about their communities in extreme values of council areas)

* Are there certain groups that have local access to green space?
* Are there groups that are lacking access?
* What is the big difference in how far people have to walk to access their green space?
* Are there differences between rural and urban areas?
* How do people in neighbourhoods with good access to green space differ from those who have no good access? Are there differences 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?

**Reproducibility**

* Create appropriate structure with raw data, analysis etc…

**Comment your code**

**FINAL PRESENTATION**

* Talk through your approach to the business question
* Discuss any challenges encountered and how you addressed these
* Present a story of your final insights which compels your client to action
* If your solution would require implementation, discuss how this could be done
* Any enhancements you would pick up if you’d had more time
* Any lessons for the future

**TIMELINE**

**1/10 – 12-10**

1/12:

* Explore and define task
* Finish wrangling and cleaning

**Abstract**

The Scottish Household Survey (SHS), which began in 1999, is a continuous survey based on a sample of the general population in private residences in Scotland. It is financed by the Scottish Government. The aim of the survey is to provide representative information about the composition, characteristics and behaviours of Scottish households, both nationally and at a more local level. The survey covers a wide range of topics to allow links to be made between different policy areas, with a particular focus on information to aid policy decisions on transport and social inclusion.

From 1999 up to 2011 the survey followed a fairly consistent survey design. From 2012 onwards, the survey was substantially redesigned to include elements of the Scottish House Condition Survey (SHCS) including the follow-up Physical Survey component. More details can be found in the survey documentation.

Neighbourhood rating:

* Every year the Scottish Household Survey (SHS) asks adults to rate their neighbourhood as a

place to live. The data is reported here is representative of the adult population.

Green\_spaces:

* The question asking about walking distance to nearest greenspace informs the National Indicator on Access to Green and Blue Spaces. However before 2019, even though the list of spaces included riversides and beaches, the question text used the term "green or open space". In 2019 for one half of the sample this was changed slightly to "green, blue or open spaces" to check if including the word "blue" made any difference to the response given to the question about walking distance or the biennial follow-up questions asking about satisfaction with and frequency of use of nearest green or blue space. No statistically significant difference was found.

Community\_belonging: