Data Analysis

Beyond Imagination Life Survey
ImaginationLancaster

May 2023 (v2)

imagination LANCASTER | Lancaster University | CATAPULT Connected Places | UKRI Research England

Background

Lancaster University, Lancaster City Council and Blackburn with Darwen Council have come together to initiate the Beyond Imagination Life Survey. Commissioned by Lancaster University and undertaken by BMG Research the survey of just under 3,000 residents aims to gather a representative picture of views, attitudes and experiences, which will in turn inform strategy, activities and research opportunities for ImaginationLancaster, the local authorities and other local partners.

Lancaster City Council and Blackburn with Darwen Council priorities have formed a key part of the survey development throughout, resulting in a survey structure which combines the needs of different stakeholders, based on four themes:

- Health and Wellbeing
- Wealth and Opportunity
- Sustainability, Transport and Travel
- Connected Communities and Services

Bringing the data to life

Connected Places Catapult were commissioned to develop an interactive data dashboard solution to explore and visualise the Life Survey data, and to enable future interactive data analysis, research and visualisation. To this end, a series of case studies have been produced by Connected Places Catapult based on research themes developed with the project partners in Beyond Imagination workshops.

Introduction

This is Data Analysis Report for the Beyond Imagination Life Survey created by Connected Places Catapult in collaboration with ImaginationLancaster, Blackburn with Darwen Borough Council and Lancaster City Council. This report begins with an overview and discussion of the Life Survey data and format. Then key statistics are presented including geographical distribution, demographic variables and socioeconomic variables. Next the important issue of data segmentation is discussed. This is followed by a presentation of the data anomalies that were encountered and a report on data missingness. The report is concluded with an outlook and recommendations for future work.

Overview of the Life Survey Data

This section presents an overview of the Life Survey, and covers the structure of the questions, answers and the survey methods.

Questions

The 2943 respondents were asked 75 questions. Many of the questions were multipart, so in total there were over 300 interconnected variables in the data to explore. These were organised into 344 question-sub question combinations for the dashboard.

Theme	From	To	Count
Age and Gender	2	3	2
Social integration	4	9	6
Local priorities	10	19	10

Housing	20	23	4
Sustainability	24	30	7
Physical and mental health	31	44	14
Digital	45	56	12
Demographics	57	76	20

Table 1: Organisation of survey questions broken down by theme. "From" and "To" refer to question numbers.

Answers

For each question, the Life Survey gave the respondents a set of answers or categories to choose from for each question. Therefore, the data is inherently categorical, in contrast to numerical. These answers could be as simple Yes-No, or they could be more complicated categories, such as the type of home broadband. However, some sets of categorical answers have an inherent ordering to them, such as the age categories or household income levels. In these cases, the data is ordinal. Naturally, it is easier to see trends and relationships in analysis involving these variables due to their inherent order.

Survey Type

The survey was carried out by two different methods: (1) Computer-Assisted Personal Interviews (CAPI) and (2) Computer-Assisted Web Interviews (CAWI). For clarity CAPI were face-to-face interviews, whereas CAWI were online interviews. The later were necessary due to the restrictions of the Covid-19 pandemic. In total there were 1504 CAPI and 1439 CAWI.

Key Statistics of the Life Survey Data

This section presents key statistics of the survey data covering the geographical distribution of responses, as well as demographic and socioeconomic variables.

Geographical Distribution

The Life Survey interviewed 2,943 respondents. 1,644 were in Lancaster and 1,299 were in Blackburn with Darwen. The distribution at the ward level is shown in Table 2 for Lancaster and Table 3 for Blackburn with Darwen. These tables show that for less than 100 respondents were interviewed per ward for almost all wards, which can make it hard to get a place-based understanding of other questions in the Life Survey. 'WARD' and 'AREA' fields of the Life Survey data provide data on the Ward and Local Authority.

Ward	Count	Ward	Count	Ward	Count
Bare	88	Heysham North	65	Scotforth West	95
Bolton & Slyne	109	Heysham South	87	Silverdale	25
Bulk	91	John O'Gaunt	94	Skerton East	68
Carnforth & Millhead	67	Kellet	22	Skerton West	69
Castle	52	Lower Lune Valley	51	Torrisholme	63
Ellel	58	Marsh	79	University & Scotforth Rural	13
Halton-with-Aughton	29	Overton	25	Upper Lune Valley	23
Harbour	92	Poulton	61	Warton	18
Heysham Central	54	Scotforth East	64	Westgate	82

Table 2: Distribution of respondents in Lancaster by ward.

Ward	Count	Ward	Count	Ward	Count
Audley & Queen's Park	73	Darwen East	89	Mill Hill & Moorgate	82
Bastwell & Daisyfield	59	Darwen South	85	Roe Lee	73
Billinge & Beardwood	71	Darwen West	81	Shear Brow & Corporation Park	53
Blackburn Central	68	Ewood	88	Wensley Fold	78
Blackburn South & Lower Darwen	100	Little Harwood & Whitebirk	64	West Pennine	60
Blackburn South East	92	Livesey with Pleasington	83		

Table 3: Distribution of respondents in Blackburn with Darwen by ward.

Demographic Variables

To allow for demographic analysis, the survey captured age (Question 2 – Q2), gender (Q3), ethnicity (Q62) and religion (Q75). The demographic variables are discussed below.

Age & Gender

Only adults (aged 18 and over) were interviewed for the Life Survey. Figure 1a shows that the majority of respondents were in the 35-74 age group, and more under 35s were surveyed than over 75s. Figure 1b shows that significantly more Females were surveyed than Males, and few were surveyed from the self-described gender group. Figure 1c shows that a higher relative proportion of males were surveyed in the 18-24 and over 65 age groups.

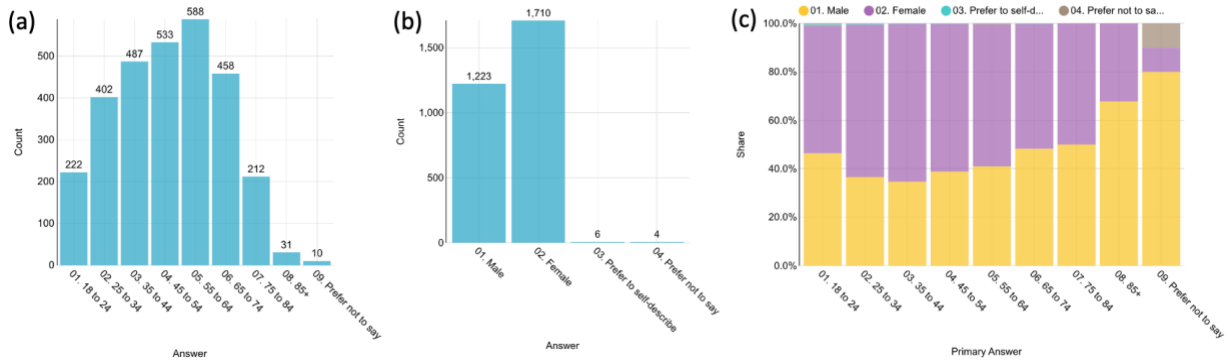

Figure 1: (a) Distribution by age. (c) Distribution by gender. (c) Distribution of gender (legend) by age group. Full details of the truncated items in the legend are: 03. Prefer to self-describe; 04. Prefer not to say.

Ethnicity

Figure 2 shows that the majority of respondents identified as English, Welsh, Scottish, Northern Irish or British (2,398 - 81.5%). The other groups that were well represented in the survey were Pakistani (171 - 5.8%) and Indian (167 – 5.7%).

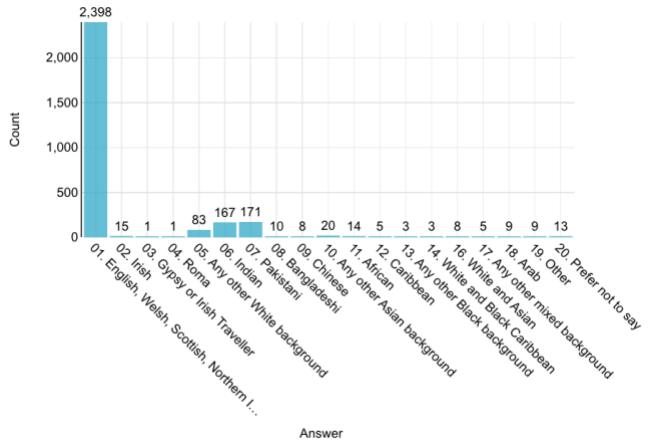

Figure 2: Distribution of ethnicity. Full details of the truncated answers are: 01. English, Welsh, Scottish, Northern Irish or British.

Religion

Figure 3 shows that the majority of respondents either followed no religion, were Christian or were Muslim. A relatively high proportion did not disclose their religion. There were also other religions not accounted for by the survey that made up more of the responses than the religions with a small representation.

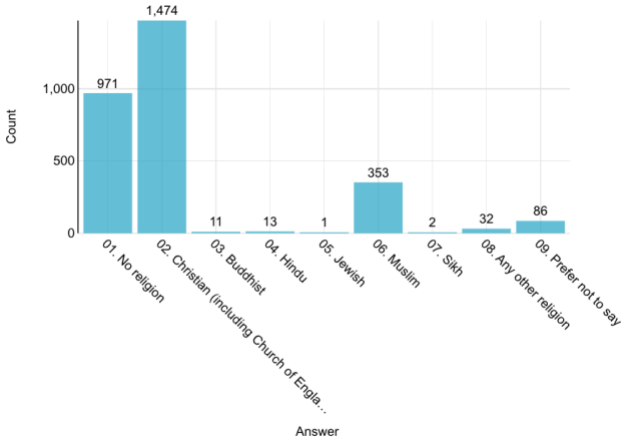

Figure 3: Distribution of religion. Full details of the truncated answers are: 02. Christian (including Church of England, Catholic, Protestant and all other Christian denominations)

Ethnicity and Religion

Further analysis found that the English, Welsh, Scottish, Northern Irish or British ethnic group was distributed with 38.6% no religion and 57% Christian. The Indian ethnic group was distributed with 87.4% Muslim, 4.8% Hindu and 3.6% Christian. The Pakistani ethnic group was distributed with 96.5% Muslim. The any other religion group was comprised of English, Welsh, Scottish, Northern Irish or British and “any other White background” ethnic groups.

Socioeconomic Status Variables

To allow for socioeconomic analysis, the survey captured education (Q69), employment status (Q70), occupational sector (Q73), household income (Q67) and home ownership (Q63). The socioeconomic variables are discussed below.

Education

Figure 4 shows that the majority of respondents had attained an educational level between GCSE (or equivalent) and Bachelor's Degree (or equivalent). 10.7% of respondents had no qualifications.

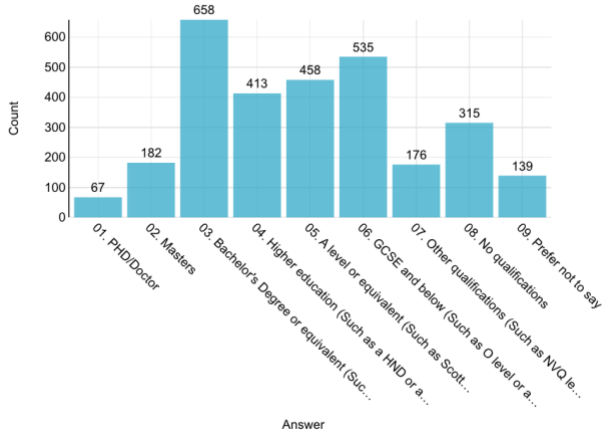

Figure 4: Distribution of highest level of educational qualification. Full details of the truncated answers are: 03. Bachelor's Degree or equivalent (Such as a NVQ level 5); 04. Higher education (Such as a HND or a NVQ level 4); 05. A level or equivalent (Such as Scottish Highers or NVQ level 3); 06. GCSE and below (Such as O level or an RSA Diploma); 07. Other qualifications (Such as NVQ level 1)

Employment Status

Figure 5 shows that the majority of respondents are either in full time (34.5%) or part time (14.8%) work, retired (25.8%) or a homemaker (4.9%). There was also 6.8% who were not in work due to ill health or disability, and a further 3% that were out of work for less or more than 6 months.

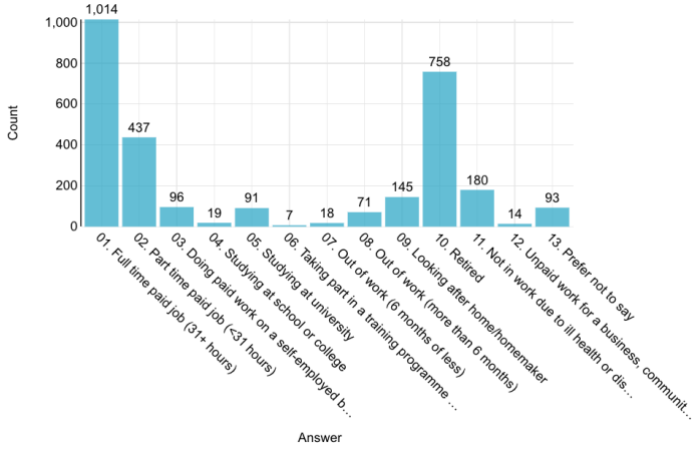

Figure 5: Distribution of employment status. Full details of the truncated answers are: 03. Doing paid work on a self-employed basis or within your own business; 06. Taking part in a training programme e.g. traineeship or apprenticeship; 11. Not in work due to ill health or disability; 12. Unpaid work for a business, community or voluntary organisation.

Occupational Sector

Figure 6 shows that the majority of the 1,547 respondents that were in employment worked in health or social work activities (21.4%) or education (17.3%) with a broad distribution in other sectors.

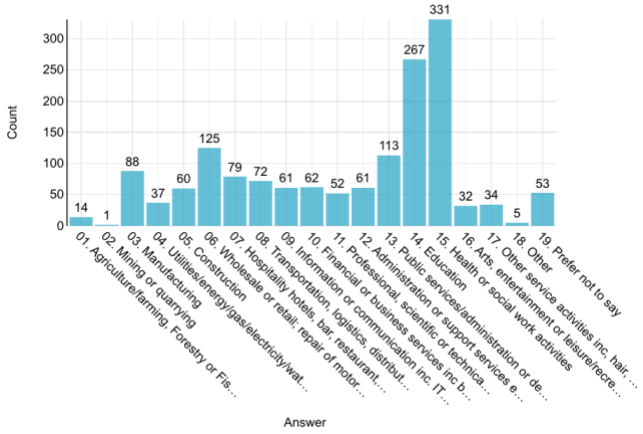

Figure 6: Distribution of occupational sector for those in employment. Full details of the truncated answers are: 01. Agriculture/farming, Forestry or Fishing; 04. Utilities / energy / gas / electricity / water supply / sewerage or waste management; 06. Wholesale or retail; repair of motor vehicles and motorcycles; 07. Hospitality hotels, bar, restaurant, catering; 08. Transportation, logistics, distribution, or storage; 09. Information or communication inc. IT, technology; 10. Financial or business services including banking, insurance, estate agents; 11. Professional, scientific or technical activities e.g. laboratories, vets; 12. Administration or support services e.g. recruitment, call centre, cleaning; 13. Public services / administration or defence, including local & national government; 16. Arts, entertainment or leisure / recreation, inc. theatre, museums; 17. Other service activities including hair, beauty, personal services, repair services or membership organisations.

Household Income

Figure 7 shows that there is a broad distribution of household income centred around the £20-25k income bracket. There was a high number of respondents that preferred to not disclose their household income.

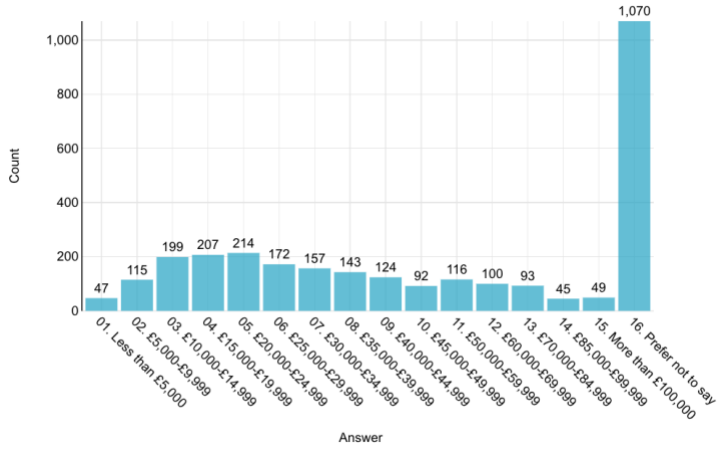

Figure 7: Distribution of household income.

Home ownership

Home ownership can be used as a proxy for wealth. Figure 6 shows that the majority are homeowners with 41.7% owning their accommodation outright and a further 30.1 % owning with a mortgage or loan. The other major group are renters that make up 23.4% of the respondents.

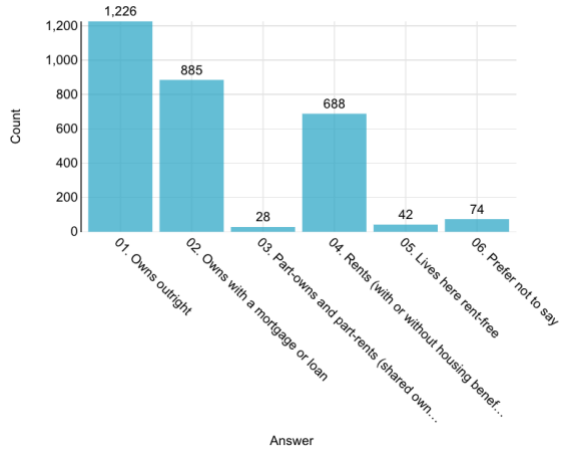

Figure 8: Distribution of home ownership. Full details of the truncated answers are: 03. Part-owns and part-rents (shared ownership); 04. Rents (with or without housing benefit).

Potential Issues of Data Segmentation

It is important to be cautious when drawing conclusions from small sample sizes, which is the case for the Life Survey data. This is particularly so when the data is segmented by more than one variable from age, ward, gender, etc. For this reason, the Exploratory Dashboard introduced The Rule of 100 in attempt to encourage users to think statistically and cautiously about their data discoveries.

The Rule of 100 is a common statistical rule of thumb stating that "In general, the size of a sample group should (ideally) be at least 100 respondents". A sample of size 100 is the typical number needed for common statistics such as the percentage share and mean to have a margin of error of approximately ±10%. Note that 100 is not exact, sometimes more samples will be needed, sometimes less will be needed. Naturally, any advanced user or researcher using the data will perform their own statistical analysis outside of the Exploratory Dashboard.

The following example shows how this rule might be broken when the data is segmented into small groups using multiple slices. A user might want to understand the views of young males in their ward of interest. Here there are 3 slices: age, gender and ward. It is apparent from Tables 1 and 2 that there is unlikely to be more than 100 respondents in the ward of interest. The second segmentation by gender will roughly split this group into two (male and female – Figure 1 shows that the other groups are negligible). The number of male respondents in the ward of interest will be approximately 50. The final segmentation by age will slice the group of 50 into 8 groups (neglecting those that preferred not to say), and so there is likely to be less than 10 respondents in the young (18-24) male group in the ward of interested. It is straightforward to see that the views of 10 people in a group cannot represent all voices and opinions of that group. Any strong effect that is being seen from this small group would be unlikely to survive if 10 different young males from the ward were interviewed.

The main way to overcome this segmentation issue would be to expand the scale of the survey to increase the sample size. In practice there will be high financial and time costs to doing so, because halving the margin of error typically requires four times the number of samples.

Those interested in further details of the statistics of sample size in surveys should consult The RCSI Sample Size Handbook by RM Conroy or The Survey Research Handbook by P Alreck and R Settle.

Data Anomalies

This section presents findings on anomalies in the Life Survey data. First an analysis of missing responses in the data is presented. This is followed by further challenges in the data that were discovered in researching the case studies.

Missing data

The most prevalent data issues with the survey arose from missing responses. Options were provided to not answer a question using "prefer not to say" and "don't know", and respondents could also not answer the question, which was recorded as "refused to self-complete". These missing answers can weaken the statistical validity of any analysis when a large subset of the group do not provide a meaningful response. If a question is multi-part then missing a single subquestion prevents an overall score being calculated. However, it is important to note that for some questions these "missing" answers were meaningful, and also that respondents reserve the right to not disclose information.

Table 4 shows that questions around health and wellbeing, socioeconomic status, and feelings towards the council received the most missing responses. The high number of missing responses for these types of questions are to be expected. Health, wellbeing and socioeconomic questions are personal, potentially sensitive and there could be privacy concerns that this personal information could be used to identify individuals. The same is true for council related questions, but this could also be a further indication of a lack of understanding of the council's role and functions amongst the public.

Question	Description	Count	% Share of total
Q34	Self-reported mobility, self-care, pain/discomfort, anxiety/depression	1460	49.6%
Q67	Household income	1070	36.4%
Q44	Food insecurity	463	15.7%
Q36	Personal wellbeing (ONS4)	459	15.6%
Q35	Mental health and wellbeing (WEMWBS)	447	15.2%
Q14	View that the respondent's council acts on concern of the residents	250	8.5%
Q68	Feelings towards household income	199	6.8%
Q42	Time spent exercising (if exercises)	171	10.7%
Q15	Satisfaction with respondent's council giving opportunity for residents' views	147	5.0%
Q69	Educational Attainment	139	4.7%
Q06	Relationship between those from different backgrounds in the local area	133	4.5%
Q40	Weekly alcohol intake (if drinks alcohol)	131	8.7%

Table 4: The number of missing responses and share of total responses for questions with over 100 missing responses. Missing is defined as record with "prefer not to say", "don't know", or "refused to self-complete". Where these were multi-part questions, the average count and share is used. WEMWBS refers to the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS). ONS4 refers to the Office for National Statistics four measures of personal wellbeing.

Figure 9 shows that there appeared to be two distinct types of respondents that provided missing responses. The chart displays the apparent bimodal (two peaked) distribution of the 2583 respondents that provided at least 1 missing response. The first group containing approximately 80% of these respondents (left peak of the chart) provided 5 missing answers on average, with a standard deviation of 3. The second group (right peak) containing the remaining 20% of these respondents provided 28 missing answers on average, with a standard deviation of 7. This analysis used a two-component Gaussian mixture model.

Further inspection found that those in the second group typically gave missing answers for the multi-part food insecurity and mental health and wellbeing questions. These long multi-part questions explain the discrete jump in missing responses. This could potentially indicate that some respondents are less inclined to answer longer questions.

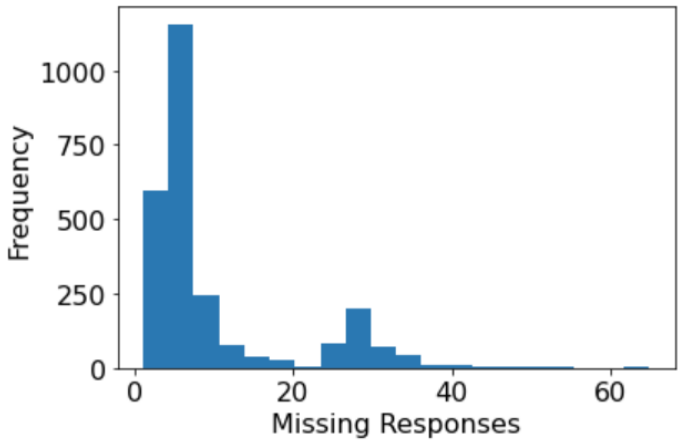

Figure 9: Histogram showing the distribution of the number of missing responses for the 2583 respondents with at least 1 missing response. A missing response is defined as the respondent answering "prefer not to say", "don't know" or did not answer (documented as "refused to self-complete").

Other Anomalies and Challenges in the Data

Here some other anomalies and challenges are discussed. These were encountered when researching the case studies.

Q27 asked respondents what they believed could help them reduce their car or van use for travel. The answer "nothing would help" was the most common. Here this answer is unhelpful, as in principle one of the other improvements is likely to have some impact on car/van use. Alternatively, there could have been a follow up question to better understand why "nothing would help". Additionally, the question failed to disentangle personal versus commercial use of cars and vans.

Q67 asked respondents about their household income. When using this as a socioeconomic variable to segment other questions with, it would be helpful if there was a way of breaking it down to the individual level. Q60 gives the number of adults in the household, but it is not a given that all these adults are contributing to the household income.

It was found in the Poverty case study that different age groups responded to the same question in different ways. The older age groups tended to provide more positive responses when asked about their feelings towards their household income (Q68). This could be interpreted as a generational perception of money/wealth skewing the responses. Yet, this finding is likely to generalise beyond money/wealth. It is conceivable that groups may respond differently to a question and their answers may be on different relative scales. Therefore, care should be taken to understand how groups tend to respond when performing segmentation analysis and drawing strong conclusions from the Life Survey.

Outlook and Future Work

Outlook

This report analysed the data from the Beyond Imagination Life Survey. Key statistics around geographical distribution, demographic variables and socioeconomic variables were presented. The issue of data segmentation was identified, and an example was provided. Data anomalies and missingness were reported on. This report concludes with a list of recommendations for future work to maximise the value of the Life Survey.

Future Work

Here general themes for future work are discussed. Readers are advised to consult the case study reports for focused future work and extensions to the case study research questions.

Extending the Survey

The issue of sample size was discussed in the earlier section on Potential Issues of Data Segmentation. In general, the main way to overcome this segmentation issue would be to expand the scale of the survey to increase the sample size. In practice there will be high financial and time costs to doing so, because halving the margin of error typically requires four times the number of samples.

An additional benefit of extending the survey would be that the data would gain time dependence. It was found in the Food Insecurity case study that time dependence provided an interesting perspective. This time dependence would allow for a better understanding of how behaviour's and views are changing over the Covid-19 recovery period and as lockdown restrictions lift.

Grouping Wards for Place-Based Analysis

The segmentation by ward puts considerable pressure on statistical validity, as there are typically less than 100 respondents per ward, see Tables 2 and 3. This makes it difficult to draw place-based insights at the ward level. Future work could construct geographically meaningful groups of wards (e.g., covering major settlements and rural areas). These groups will hold a greater number of samples and so will provide more opportunities for statistically reliable findings.

Fusing with Additional Datasets

The Food Insecurity case study demonstrated how the Life Survey data can be combined with and enriched by additional datasets, and how the dashboard solution can enable this fusion. Future work should explore the landscape of datasets that are available. It will be important to consider data licensing and data sharing agreements to legally unlock new data. It is important to note that there can be considerable work involved in fusing disparate datasets.

Documenting the Life Survey

Some questions in the Life Survey have been created or are owned by other bodies and it would be valuable for these to be well documented. Examples include: Q34 ONS4 created by the Office for National Statistics; Q35 The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) ©University of Warwick, 2006, all rights reserved. WEMWBS was developed by the Universities of Warwick, Edinburgh and Leeds in conjunction with NHS Health Scotland; and the Food Insecurity Experience Scale (FIES) developed by the Food and Agriculture Organisation of the United Nations. Each of these examples has useful additional material available online that would be valuable to users when interpreting the data.

Scoring and Indexing the Life Survey

Almost every question in the Life Survey has categorical answers, and a subset of these are ordinal answers. Ordinal variables naturally lend themselves to scoring or indexing. These processes convert variables into numerical scores or indexes (hereafter scores). These can then be combined into an overall score, and this aggregation of multiple related questions can provide a more holistic understanding that is easier to interpret. An example of this is the WEMWBS. Scoring and aggregation enables statistical analysis such as mean, median, mode, variance and more advanced statistics. It also allows for more expressive data and geospatial visualisations, e.g. visualising the average wellbeing score as opposed to visualising the number of respondents that were feeling close to others some of the time. To this end, future work should score and index the Life Survey where possible. Valuable information to help with this will be gained from documenting the Life Survey questions.

Focus Areas

The case study research questions were developed in collaborative workshops. It was found that there was considerable interest on links to mental health and wellbeing. As this is likely to continue to be a strong focus, it could be valuable to analyse and process this part of the dataset further. The creation of metrics will improve analysis and breakdowns with other variables. Alternatively, this focus might indicate that other parts of the survey that received less attention, such as the digital theme, were not as accessible. Here additional documentation of these themes could help to encourage future research questions that make better value of the data.

CAPI versus CAWI

Approximately half of the interviews were face-to-face (CAPI) and half were online (CAWI). Future work could investigate the effects of these interview techniques on the responses.

Cite this

Cruickshank, L., Knight, J., Christou, E. Harper, J., Coupe, G. (2022): Beyond Imagination Life Survey: Topline Reports and Case Studies (Life Survey Data Analysis). Lancaster University.
https://doi.org/10.17635/lancaster/researchdata/551

Appendix

Data Access Statement

Beyond Imagination Life Survey: Topline Reports and Case Studies

Topline Reports of the Life Survey findings and six case studies created in collaboration with Connected Places Catapult are available to download from Lancaster University's Research Directory at: https://doi.org/10.17635/lancaster/researchdata/547

Life Survey: Data Dashboard

A data visualisation dashboard for the Life Survey has been created in collaboration with Connected Places Catapult. Access to the data dashboard is restricted to Lancaster University researchers and officers in Lancaster City Council and Blackburn with Darwen Council. Lancaster University host and manage the data dashboard with access granted on a role-based basis. Conditions for access and a password protected log on to the dashboard can be requested by emailing lifesurvey@lancaster.ac.uk

Life Survey Dataset

Access to the anonymised Life Survey dataset is restricted to researchers within UK universities. The dataset is available on request with an appropriate Data Access Agreement. To request access and receive further information about the dataset and conditions for access please email lifesurvey@lancaster.ac.uk

Local Authority Officer requests to access the dataset

An Information Sharing Agreement between ImaginationLancaster, Lancaster City Council and Blackburn with Darwen Council permitted the sharing of the Life Survey dataset with the local authorities. Access will be granted to officers on a role-based basis.

Officers within Lancaster City Council should contact Kirsty Chekansky with dataset access requests.
Email: kchekansky@Lancaster.gov.uk

Officers within Blackburn with Darwen Council should contact Elise Carroll with dataset access requests.
Email: elise.carroll@blackburn.gov.uk

Partners & Funding

ImaginationLancaster

ImaginationLancaster is Lancaster University's interdisciplinary design and architecture research lab. In 2019 ImaginationLancaster was awarded £13.2m for a multi-year project titled Beyond Imagination. Funded by Research England and Lancaster University, Beyond Imagination explores and demonstrates how cutting edge design research can create a healthier, more prosperous and sustainable world.
http://imagination.lancaster.ac.uk

Data Science Institute

Lancaster University's Data Science Institute (DSI) supports interdisciplinary data intensive research across the University, with core themes of: foundations, health, environment, society. It has 155 academic members from 19 depts and works with business, government and third sector partners. Current member activities include work understanding the needs of looked after children, links between the quality of the urban environment and health as well as projects in cyber security.
https://www.lancaster.ac.uk/dsi/

Blackburn with Darwen Council

Blackburn with Darwen is a semi-rural unitary borough located in the south east of Lancashire. It has compact urban areas predominately located around the towns of Blackburn and Darwen, surrounded by countryside. These contrasting areas include some of most and least deprived in England. The current corporate plan has the aim of enabling borough residents to achieve a good quality of life in a vibrant and thriving place, with strong community values, in an inclusive society.
https://blackburn.gov.uk

Lancaster City Council

The Lancaster district includes diverse and attractive city, coast and countryside locales. Lancaster City Council's vision is for the district to thrive as a vibrant regional centre in the north west of England. In December 2021 the council set out its four priorities for 2030, along with strategies for how these can be achieved: A Sustainable District, An Inclusive and Prosperous Local Economy, Healthy and Happy Communities and A Co-operative, Kind and Responsible Council.
https://www.lancaster.gov.uk

Connected Places Catapult

Connected Places Catapult is the UK's Innovation Accelerator for cities, transport and places. We provide impartial 'innovation as a service' for mobility and built environment businesses, infrastructure providers and public institutions to catalyse step-change improvements in the way people live, work and travel. We help develop, implement and commercialise the latest technology and innovation for existing markets, as well as create demand and grow new markets in the UK and globally.
https://cp.catapult.org.uk

BMG Research

Established since 1988, we have more than 30 years' experience of working with our clients to build and deepen our understanding of changing and ever more complex markets, people and society. We work with our clients to fully understand the challenges faced by their organisations, identify priorities for action, and evaluate the impact of change.
https://www.bmgresearch.co.uk/

Research England

We are responsible for funding and engaging with English higher education providers to create and sustain the conditions for a healthy and dynamic research and knowledge exchange system in the higher education sector.
https://www.ukri.org/councils/research-england/

Acknowledgements

ImaginationLancaster would like to thank: Research England, Jez Bebbington at Lancaster City Council and Elise Carol at Blackburn with Darwen Council. Sharon Gowland and Jack Harper at BMG Research. Elliot Christou, William Daubney, Louise Files and Darren Pangbourne at Connected Places Catapult. Professor Jo Knight at Lancaster University's Data Science Institute. Officers of Lancaster City Council, Blackburn with Darwen Council and researchers in ImaginationLancaster who contributed to the design of the Life Survey Questionnaire and Catapult Case studies.

---

imagination.lancaster.ac.uk

ImaginationLancaster

LICA Building

Lancaster University

Lancaster

Lancashire

United Kingdom

LA1 4YW

Email: imagination@lancaster.ac.uk

Twitter: @ImaginationLanc

ISBN: 978-1-86220-386-0