Lancaster University Future Places Centre: Technical Reports

The Future Places Centre (the FPC) is researching how people understand the places they live in – their workplaces and homes, their towns and the countryside. Key to such understandings are various forms of 'data' - much of that computer-derived but much of it also produced by cultural practices: in stories about places written down, thought about and endlessly reimagined. 'Data' labels many types of evidence and various ways of seeing that continuously evolves.

The FPC's technical reports and publications present some of its findings on this subject: data and place.

For further information about any particular report or publication, please contact the author or the FPC - fpc@lancaster.ac.uk

The FPC is funded by the EPSRC under grant number: EP/TO22574/1

Technical report: What's the question?: Using commercial data in community organizations

Published: FPC technical report, June 2024

Socio-demographic data can be linked to geography, to the places in which people live. This data is primarily gathered to enable government policy, and for commercial organisations offering retail services. This paper explores whether this data can be used for long term research seeking to create cultural change.

What's the question?: Using commercial data in community organizations

Paul Alcock, Jan Hollinshead, Richard Harper
School of Computing and Communications, *Lancaster University*
{p.alcock1,j.hollinshead1,r.harper}@lancaster.ac.uk

***Abstract*—This paper illustrates a methodology for leveraging commercial socio-demographic and lifestyle data to inform charitable endeavours, focusing on a community food project's objectives in enhancing engagement and targeting individuals in need. The dataset comprises 816 rows across 59 consumer 'profiles' designed to address consumer facing, retail business needs. Through iterative refinement and data analysis techniques, the study navigates the complexities of using this data to define abstract concepts and deriving actionable insights that apply to the charitable organisation. These goals are distinct from those of retail organisations. Despite encountered limitations, the study demonstrates the potential of doing so, explaining how commercial data can be repurposed to optimize resource allocation and enhanced community engagement, fundamental needs that the community project in question concerned itself with. Implications for the other community or charity cases are remarked upon.**

I. Introduction

In today's data-driven landscape, commercial enterprises are increasingly leveraging data to pursue a market advantage, optimize operations and enhance strategic decision-making. Organizations are increasingly using third-party datasets to do so, using these to unlock insights about populations to inform their business strategy. Socio-demographic data are especially useful for consumer facing, retail enterprises. However, in a world where the attention of these consumers is increasingly expensive and competed for, charitable organizations are also turning to the same data to see if insights from that data can guide their own impact strategies. The use of data for these needs is relatively under-explored.

In this paper, we detail a methodology for harnessing commercial data for a charitable organization, specifically addressing the goals outlined by a community food project that entail seeking to enhance engagement on certain behaviours, identify individuals in need in relation to those behaviours, and evaluate the feasibility of different behavioural impact strategies. The dataset obtained from a private organization offers a comprehensive view, surpassing common open-source datasets, such as those provided by the Office for National Statistics (ONS), by amalgamating multiple data sources

A. Problem Statement

Closing Loops is a collaborative community project aiming to cultivate a regenerative food economy in the Lancaster District. Some of the project's goals are focused on promoting a healthy lifestyle through high quality, sustainably produced food from the local area. In supporting their aims, the project is interested in optimizing its strategic efforts to increase general engagement with their events and glean any actionable insights which may inform event organizing. We obtained a data set from a commercial data consultancy organization called CACI Limited, which categorizes geographic areas based on various socio-economic and lifestyle factors. This dataset contained 59 consumer profiles, which were applied to each postcode in Lancashire.

II. Methodology

A. Requirements Gathering

After an initial discussion of goals with multiple teams within the community project, a few overarching themes emerged. All teams expressed a desire to identify and map geographic areas which were likely to engage with the project's resources and events. Additionally, the project was keen to identify populations who would benefit most from their work.

An initial step involved deciphering abstract phrases, such as, 'likely to engage' and 'likely to benefit'. A number of discussions were necessary to precisely define these terms for the project's context and find representations of this within the data. Ultimately, it was determined that a sufficient definition of 'likely to engage', incorporating a holistic weighting of interest in the project's goals, against lifestyle factors which may limit an ability to participate. Several fields within the dataset were found which contained relevant interest scores for this interpretation, covering cooking, organic food, environmental concern etc. Figure 1 shows an extract of these fields in the 'interest & hobbies' category from the dataset.

From this, GIS maps were produced which gave a granular view of interest in specific topics, as they related to postcode location. Determining an 'ability to participate' however, was more challenging to define, as no fields directly addressed this concept, requiring an interpretation of available fields. Although some major concepts which may contribute to an 'ability to participate' were discussed, there remained a gap between those which seemed central to the concept ('ability to participate') and the fields in the data.

Time poverty seemed especially relevant, for example. Time poverty refers to an individual feeling constantly pressed for time, juggling numerous responsibilities and commitments in daily life. This concept is challenging to score, however, in part due to representing a mix of psychological perspective (a person's feelings) and sociological facts about those persons (income, family size, etc). Besides, it is difficult to find a

1	Theme	Variable
2	Interests & Hobbies	Charity / Voluntary Work
3	Interests & Hobbies	Cinema
4	Interests & Hobbies	Cookery
5	Interests & Hobbies	Eating Out
6	Interests & Hobbies	Exercise / Sports
7	Interests & Hobbies	Gardening
8	Interests & Hobbies	Playing Golf
9	Interests & Hobbies	Healthy Eating
10	Interests & Hobbies	Hiking / Walking
11	Interests & Hobbies	Organic Foods
12	Interests & Hobbies	Pets
13	Interests & Hobbies	Reading Books
14	Interests & Hobbies	TV
15	Interests & Hobbies	Vegetarian Products

Fig. 1.

strong consensus in the literature on what these contributing factors would be for any chosen population. However, an initial review of prior research identified household composition and having young children, specifically, as a major contributor to both perceived time poverty and its sociological determinants.

An issue, shared with both abstract indicators, was joining together the rows to create a single composite indicator to simplify our analysis. A number of potential methodologies were explored, but it was quickly determined that we couldn't develop a detailed enough picture to be sure that we were weighing certain factors accurately. Ultimately, we opted for a simplified approach, averaging all the fields we had identified together, without weighing factors individually.

B. Analysis and Refinement

An issue shared with abstract indicators for all socio-demographic features was joining together the rows in the data sets to create a single composite indicator to simplify analysis around such things as time poverty. At this stage, there was no way to sanity check the weighting of particular factors, and so composite indicators were created from a mean average of all fields we judged relevant, without any opinion on the significance of specific fields. Once we had gathered what we thought were all the relevant data for further processing, a number of data analysis techniques were used to extrapolate useful information.

For example, a Pearson correlation was performed to understand and explore any trends between fields in the dataset. This analysis showed that, the interests and hobbies relevant to the community food project were all highly correlated, giving confidence that multiple teams within the project could make use of a single 'general interest' score to locate people with shared interest, and therefore optimize their project's marketing activities. Furthermore, the correlation study gave some support for our understanding of time poverty, showing significantly lower correlations for relevant interests and having children, with a particular emphasis on households with

	A	B
1		Cookery Interest
2	Children aged (0-5) 0	0.99
3	Children aged (0-5) 1	0.9
4	Children aged (0-5) 2+	0.41
5	Children at home : 0	0.99
6	Children at home : 1	0.99
7	Children at home : 2	0.97
8	Children at home : 3+	0.72
9	Single Parent	0.42
10	Couple No Children	0.99
11	Couple with Children	0.99
12	£0 - £20,000	0.95
13	£20,000 - £40,000	0.99
14	£40,000 - £60,000	0.99
15	£60,000 - £80,000	0.98
16	£80,000 - £100,000	0.83
17	£100000+	0.57
18	Unemployed	0.34
19	Job seeker's allowance	-0.02
20	Universal Credit	0.05

Fig. 2.

more than 2 children under 5. Figure 2 shows an extract from the correlation study conducted.

To determine any obvious groupings of the 59 consumer profiles, a k-means clustering technique was employed. k-means iteratively partitions a dataset into K clusters, where K represents a predetermined number of clusters. As the 'likelihood of engagement' was taken to mean a holistic weighting of interests in the project's goals, against limitations for participation, k was given a value of 4; representing all possible combinations of high or low interest, and high or low ability to participate. Quickly, this method was abandoned as the variations in interest scores were small which could either be reflecting a true lack of variance in the population, or the true variance may be obscured by the presentation of the data. Since the data provides pre-processed, clustered groups, instead of raw, unprocessed data, it is probable that groupings are made based on commercially relevant, economic factors, which may disproportionately obscure the variation of factors like 'interest in cooking'.

Ultimately, although we had some idea of the relevant factors, the project's members did not feel comfortable in making any assumptions around the importance of specific data fields, and therefore it was a challenge to further refine the data.

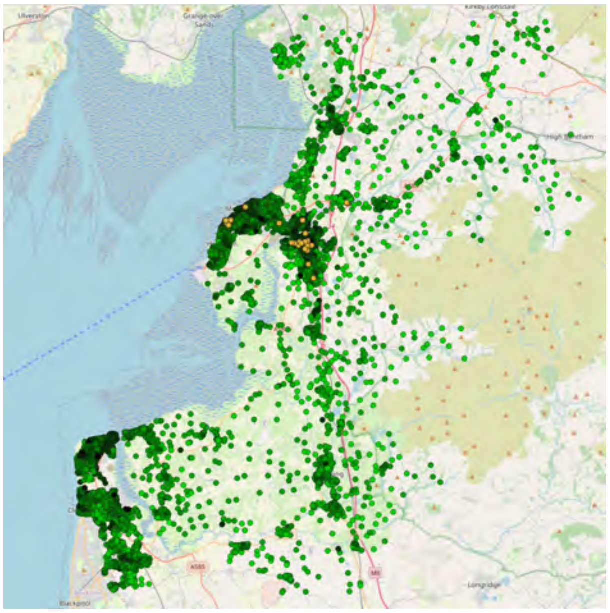

Fig. 3. Composite Interest w/ Churches

III. Results

Despite some initial issues in defining abstract terms, then, we managed to generate several useful outcomes from our analysis which can be actioned by the Closing Loops team.

A. Market Size

A core aim of one of the teams within the charity was the founding of a 'food hub' within the Lancaster district, promoting the sale of local, organic produce. The data revealed that, based on an existing interest in organic food, the potential market size for this venture would be 6,553 people in the Lancaster district, around 4.4% of the population.

B. Event Locations

In addition to market size approximations, the data was able to inform a location scouting process for cooking lessons that the project was aiming to host. A number of venues had previously been selected by the team in question as possible locations for cooking lessons, many of which were places of worship with access to kitchens. Figure 3 shows a visualization of our composite interest score by postcode, identifying populations with both interest and lack of limiting lifestyle factors. Additionally, pictured in yellow are the initially considered lesson locations.

IV. Conclusion

While the methodology yielded valuable insights, certain limitations persisted.

Firstly, it was a challenge to extract meaningful information from the data in line with the project's requirements, as any concepts which were not directly reflected by the data required further research and discussion with Closing Loops teams to extrapolate relevant information.

Secondly, as the data had been processed to create consumer profiles for a retail enterprise focus, some of the insights relevant to the community project may have been obscured. As the interests of enterprise focus largely on economic factors, features of most interest to a community project like Closing Loops may have not been prioritized when creating each consumer profile. In effect, this would move less economic-relevant fields in each profile closer to the UK average, disproportionately obscuring the variance in those fields between postcodes. Nonetheless, the data was able to identify trends between consumer profiles, which ultimately supported the creation of GIS visualizations; enabling the teams to optimize their marketing efforts and target desirable postcodes.

In summary, the methodology outlined in this paper illustrates an iterative process of translating abstract concepts embedded in commercial socio-demographic data into actionable insights for a social community endeavour. Through repurposing data, community projects like Closing Loops can optimize their efforts in identifying populations that share values and are ultimately more likely to engage with their functions and events. Additionally, such data can be used to identify populations that community or charitable projects deem desirable to target.

Acknowledgement: This research was funded through the EPSRC Future Places Centre at Lancaster University. The authors gratefully acknowledge the support of the Closing Loops team.