SCC 460 Data Science Fundamentals Group Project
Project H: Digging Grant Making Data

Students Names and Library Card Numbers:

- Aggarwal, Ravisha *r.aggarwal3@lancaster.ac.uk* 35377464
- Baines, Harry *h.baines3@lancaster.ac.uk* 35315878
- Bellamy, *Emma e.bellamy@lancaster.ac.uk* 35412113
- Hu, Bogong *b.hu5@lancaster.ac.uk* 35323476
- Rushton-Woods, Elizabeth *b.rushton-woods@lancaster.ac.uk* 35136001
- Wang, Zheng *z.wang48@lancaster.ac.uk* 35246404

Overview of the work carried out by each group member:

Aggarwal, Ravisha

- Responsible for creating the Data visualizations for research questions in Power BI app.
- Worked on research and analysis for research question 1 and 3.
- Wrote report sections: Results – Research Question 1 and 3, Potential bias and Validity violations related to this analysis.

Baines, Harry

- Responsible for researching and developing the classifier to categorise grants into a predefined set of labels.
- Contributed to managing grant labels following grant classification and data preprocessing tasks.
- Wrote report sections: Data analysis and data labelling.

Bellamy, Emma

- Task Leader, Meeting Coordinator and Facilitator; kept group focused on goals and set agendas
- Group spokesperson & communicator with 360 giving, prepared & led meetings with 360 giving
- Information Gatherer; researching previous data analysis carried out on grant making data
- Main presenter; wrote dialogue for slides presented
- Wrote report sections: Motivation, Research questions and objectives, Research strategy and methods, Examining validity issues and potential biases

Hu, Bogong

- Meeting note taker: taking notes for some meetings for further review.
- Building the basic structure and revising the presentation slides.
- Wrote report sections: Introduction (Description and Background).

Rushton-Woods, Elizabeth

- Pre-processing the data, and creating different forms of the data for use in the classifiers and in the analysis.
- Manually redefined themes and relabelled grants so that they better described the data.
- Latex: Brought the document together, created and edited, filling in missing gaps.
- Wrote report sections: Pre-processing and R Pre-Processing, assisted with Data Analysis Methods, Bias and Validity and Conclusion

Wang, Zheng

- Helped Ravisha with visualization.
- Work on Research Question 2: Where are the local causes situated?
- Wrote report section: Results – Research Queation 2: Where are the local causes situated? And Potential Bias and Threats to Validity.

I. INTRODUCTION

360Giving is a charity who supports grant-making bodies from the UK to publish grant data about who, where and what they fund in a uniform and consistent way. But "what" they fund or the categorized types of funds that are distributed is a knowledge gap for most funders. In order to have a better understanding of the data in 360Giving's GrantNav database, we chose The Co-operative Group as our research target, from the 115 registered funding bodies. This organization is one of the world's largest consumer co-operatives, owned by its members who decide what they would like to support from 4,000 local causes located throughout the UK.

360Giving's vision is for grant-making in the UK to be more informed, effective & strategic [1]. Our project brief was very broad "to build a better picture of the funding landscape", therefore to gain a better understanding of the data in the 360Giving database 'Grant Nav'. It contains over 348,000 grants, 198,000 recipients, and 115 funders which range from private foundations, to The National Lottery, and individual charities themselves. Some fields, such as grant title and its description are compulsorily completed. However a major challenge for 360Giving is to improve the data quality overall as most of the 79 features are sparsely populated or descriptive, especially fields related to geographical data. The database includes £26bn worth of grants, with the main purpose of our data project to increase the understanding of grant-making organizations, enabling them to make better informed decisions and design sector-changing strategic plans [1].

Previous research work carried out by 360Giving had focused on whether grant-makers fund the same organisations; visualisations include a funding playground of how many recipients have received funding from two or more funders [2] and a network map showing recipients of 3 or more funders [3]. In addition 360Insights tool gives organisations the ability to view a dataset by amount, date, recipient type, region and more. One of the main gaps that we saw, and therefore our main motivation, was the lack of knowledge on the nature of what is being funded.

We can achieve our project goal of building a better picture of the funding landscape step by step. Firstly, we can build an overview of what the members of The Co-operative Group value most by categorizing the data into 12 themes and sum the grants of each theme. Then we build an overview of grant amounts on different themes in different locations and investigate the correlation between them by calculate the frequency and the total amount of grants. Finally, we sum the amounts awarded to each theme in different areas. and visualize the major awarded themes in different regions or in different constituent countries.

Therefore the research questions are based around increasing the knowledge on grant themes:

- What themes of grant types are awarded by the Co-operative Group?
- Where are the local causes situated? Is there correlation between themes and locations?
- Is there any geographical variation in the amount awarded to recipients based on theme?

It is therefore a relational type of study, looking at the relationship between themes and geographical information and the relationship between themes and amount of funds allocated in different regions.

In an interview with our project sponsor, the need to understand more fully the areas organisations are funding in was clearly expressed, as every funder has different aims. We also understood that there is a limited data literacy amongst some of the publishers and funders. Therefore using skilled data analysis techniques will allow the potential users of 360Giving, funders and recipients to gain new insights. Our objective was to be able to contribute something to the pre-existing knowledge that could help grant-makers with their decision-making. In addition, our project sponsor was not just interested in the end product, which themes of grants are allocated, but how we achieved these results and whether the process could be generalised to other grant-makers.

II. METHODOLOGY

A. *Research Strategy*

The strategy that we chose to answer the research questions had an element of design and creation as our approach including developing a new method of categorising grants that had not been seen during our research review. With a vast quantity of data to analyse, it was evident that there was not enough time to explore the entirety of the database, therefore our main strategy was building a case study on one of the bigger grant-makers in order to be able to obtain a rich insight into how it allocates funds.

In investigating a funder in depth in a case study, we could look in a more holistic way at the different ways in which labelling themes of grants could enrich the knowledge of grant-making in an organisation and extend our research questions to cover wider topics. This descriptive case study will lead to a more detailed analysis on the themes of grants; how themes may be interconnected or inter-related with geography and the amount of funds that are allocated to beneficiaries [4]. Criteria of the funder for the case study was based on our overall aim of understanding the grant making in the UK and finding appropriate grant-makers where the grants are not specific to one type, such as Sports England or Department for Transport. Taking into account the time available for the project, data needed to be fairly clean and relatively free of missing data, so many grant-makers who had large-scale incompleteness in their geographical information on beneficiaries were disregarded. It seems that there is not a typical instance to choose as a case study that can stand as representative of the whole class, as funders have such varying strategic aims.

After carefully consideration, we chose The Co-operative Group as our case study for a variety of reasons. It is one of the top 5 biggest funders, with very varied beneficiaries, a broad aim of supporting local causes and it is also one of the funders with better data quality in geographical fields. Moreover, The Co-operative Group holds an unique opportunity as the money raised comes from Co-operative Group members, and they decide which causes to support. This ignites questions related to what happens when individuals are allowed to choose their own causes, do they chose causes based on their beliefs, their location or those causes that are located closest to them in their local community. Our research may provide evidence that develops a greater understanding of people in the UK, their personal motivations for giving, and their views about the world around them. In analysing the types of themes that are popular with these members of the public, or analysing the types of themes that are not being supported, this will enrich the knowledge about grant-making and the outcome may be that other organisations may redefine where to concentrate their efforts. In addition, by gathering evidence about how grants are allocated by Co-operative Group members, the public, gives us a better idea of which issues are most important to individuals in the UK.

Another outcome of our research is the new technique for analysing grant making by theme; if this could be adopted by 360Giving , they could generalise it to all funders and add it to their current tools. As grant-makers analyse their distribution of funds, they may be able to look at their data in a new way, as it will give them an opportunity to study grants using our labels in a way that has not been studied before and find whether their aims for grant allocation have been met.

Initially we planned the project at a high-level, we brainstormed methods of being able to extract themes from the data available to us in the 360 giving database. We used the 360 giving database as our main source of data that pre-existed, then we generated new qualitative data, the grants themes. The GrantNav database does not have predefined themes in a dropdown box nor a field to give a one-word label to categorize the grant. In order to be able to create new knowledge on the types of grants, we needed to define a set of labels to allocate the grants to.

We considered a multitude of different approaches to obtain labels for the themes, including identifying the most frequent words or phrases that emerge from the data, or creating regular expressions in R, that would allow us to find and match words or strings in the body of text fields, e.g. if the word "University" appears in either the grant title or description, it is most likely to be categorised as an educational grant. In order to identify the categorisation of the themes, we looked the major groups in The International Classification of Non-profit Organizations but not all of the groups seemed relevant, for example; Housing, Law, Advocacy and Politics, Philanthropic Intermediaries, and International Groups were not relevant for our data. Then at a more-detailed level, a machine learning algorithm was developed to build the theme labels.

We started the data analysis without pre-conceived ideas in mind. After initially analysing the labels, we found that a large portion of grants had a community label. So we gathered more data on how this large category could be broken down into smaller and more meaningful chunks. We found that the process needed to be iterative as we moved forwards with new labels and backwards as adjustments were made to the algorithms.

Finally we evaluated our results to see whether our research questions had been answered. The main criteria that we used was accuracy, precision, recall and AUC.

B. Data Collection and Pre-Processing

All the data from 360Giving is freely available on their website, however there is a large amount of data, with over 300000 data points. As we only focus on grants made by the Co-Operative group, which consists of only 12021 data points, we first must remove the unnecessary grants. This is simple, as we create a new data frame by creating a conditional statement that the identifier must be the Co-operative Group.

In the original dataset, there are over 50 features of each data point, including funder, recipient and postcode. Many of these features are missing for the Co-operative Group data, such as various location identifiers, which were likely used for other funders or recipients. We therefore eliminated the empty and duplicate features from our dataset, resulting in a new dataset with 12021 data points and only 15 features.

Finally, for out analysis we wish to be able to plot information graphically. The dataset does include the postcode for each recipient, however we need to transform this into latitude and longitude, so the visual tool is able to understand the information. Using the R package PostcodesioR [5], we can convert the postcodes into latitude, longitude and country. A new feature of the data was created for each data point to display this information.

C. Data Analysis Methods

Following data-preprocessing, 12021 Co-operative Group grants were available to analyse. However, these grants were not already categorised into a set of pre-defined themes. Therefore in order to answer our first research question regarding themes of grants awarded by the Co-operative Group, we had to explore the potential approaches we could adopt in order to classify grants into categories.

Before classification the first step was to decide which features of the data were a good representation of the grants under consideration. Concatenating the title, description and recipient name columns for each grant produced textual descriptions summarising each grant.

Many machine learning approaches already exist to solve this problem. Topic modelling is a common unsupervised technique (i.e. no training is required), and Topic classification enables a more supervised approach, in which we train a model capable of predicting categories for unseen data. However, we felt implementing a solution from scratch would take a considerable amount of time and wouldn't necessarily produce accurate results for the further analysis we would undertake. Therefore, we decided to utilise the content classification feature of the Google Cloud Natural Language API to categorise the grants [6].

Using the Cloud Shell in the Google Cloud Platform, we execute a Python script which calls the *classify_text* method for all the grants, which returned the categories and the confidence levels for the categories the grants were classified into. Following classification, we exported the results to a Google BigQuery database, enabling a CSV file to be exported for further analysis.

Following classification, we discovered 4078 out of the total 12021 grants were unclassified. We discovered this was mainly due to ambiguous or vague textual descriptions. Given this comprised 34% of the entire dataset we were dealing with, we decided to utilise another technology known as Google AutoML [7]. This method is a supervised form of machine learning - providing a labelled dataset of textual items enables a classification model to be trained and tested on new textual data. Using our previously labelled grants as a training dataset (7943 items), we trained a model to predict the labels for the unclassified grants.

Following evaluation of the trained classifier we obtained a precision of 83% and a recall of 79%. Precision gives a ratio of how many grants were correctly predicted into a given category. Recall gives a ratio of how many actual labels were predicted correctly. From this evaluation we see the trained classifier performed reasonably well.

True label \ Predicted label	Arts and Entertainment	Gardening and Renovation	Health	Hobbies and Leisure	Jobs and Education	People and Society	Pets and Animals	STEM	Sports	Support Groups	Travel and Tourism
Arts and Entertainment	75%	2%	-	4%	2%	12%	-	-	6%	-	-
Gardening and Renovation	-	59%	-	4%	4%	33%	-	-	-	-	-
Health	-	-	88%	1%	1%	10%	-	-	-	-	-
Hobbies and Leisure	-	1%	1%	77%	2%	17%	1%	-	-	-	-
Jobs and Education	2%	-	-	1%	87%	10%	-	-	-	-	-
People and Society	2%	1%	2%	3%	3%	86%	0%	-	2%	-	-
Pets and Animals	-	-	-	57%	-	43%	-	-	-	-	-
STEM	-	-	-	-	50%	50%	-	-	-	-	-
Sports	-	-	2%	-	-	20%	-	-	78%	-	-
Support Groups	-	-	33%	-	-	33%	-	-	-	33%	-
Travel and Tourism	12%	-	-	12%	-	53%	-	-	-	-	24%

Fig. 1. Confusion matrix after the Google Natural Language API.

A large percent of the data was classified in 'People and Society', which resulted in analysis and visuals, and was quite an ambiguous theme. As a result, we split this category up into the smaller theme 'Community', and manually reclassified some of the data points into other categories such as Hobbies and Leisure using key words in the grant description. The Scouting and Guiding theme was also created as a large part of the People and Society theme comprised of these types of grants.

III. Results

A. Research Question 1: What themes of grants are awarded by The Co-operative Group?

After training our classifier, we categorised the data into a total of 12 themes where the Co-operative Group grants are being awarded between 2017 and 2019. The total value of the Co-operative Group grants was £56.3million, with the majority of these grants belonging to the Hobbies and Leisure (19% of total) theme, followed by Jobs and Education (18% of total), Community (16% of total) and Health (11% of total). Hobbies and leisure occupies the largest percentage of grants awarded and includes grants such as volunteering activities for school children, self-development workshops, youth cafés, womens self-help workshops and festive events. Jobs and Education includes grants for summer schools, educational trips, tuition for disadvantaged, career counselling workshops and other activities related to upgrading skills for higher studies and job hunting. The community theme includes grants for foodbanks for underprivileged areas, building community centres and cleaning neighbourhoods.

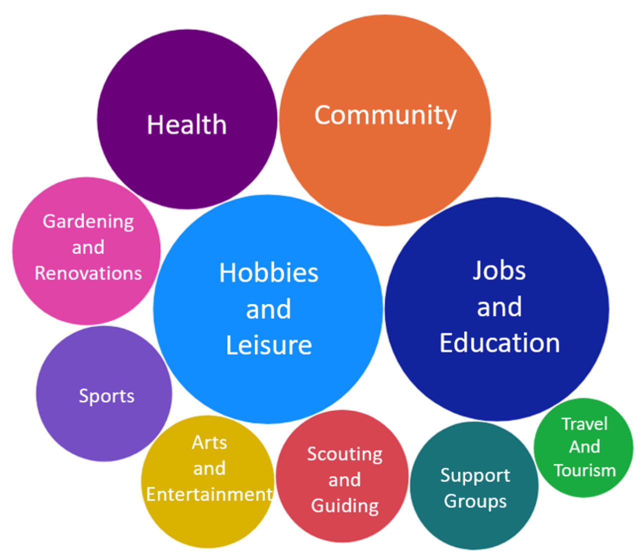

Fig. 2. Comparison of the different themes by number of grants in each theme.

The remaining themes are Gardening and Renovation, Sports, Arts and Entertainment, Scouting and Guiding, Support Groups and Travel and Tourism. Gardening is a popular pastime in the UK, and these grants support services such as community gardens. Renovation ranges from fixing roofs, flooring, ceiling and upgrading equipment in various places such as community centres and churches. Sports grants support services such as supporting children with disabilities and special needs access sports and sport facilities, health and yoga workshops for the elderly, and constructing community gyms. A large portion of the grants are awarded to Scouts and Guides around the country. These grants help fund trips for the groups and help them volunteering and carrying out community services.

B. Research Question 2: Where are the local causes situated? Is there correlation between themes and locations?

For this research question, we analyse the data by both the numbers of grants and the total amount of grants.

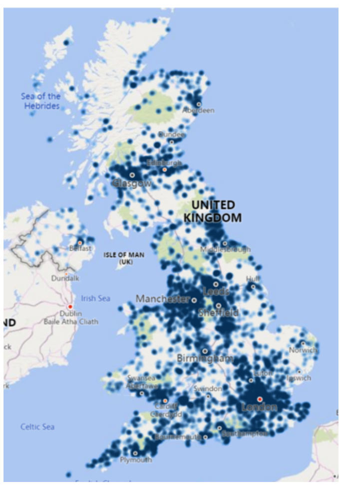

Fig. 3.

In figure 3, we use the heat map to show the frequency and the total amount of grants awarded by The Co-operative Group over past 2 years in different places in UK. Each blue point in the map indicates a grant. A darker point denotes higher amount awarded, lighter one denotes the grant with lower amount.

Figure 3 provides evidence that larger UK cities have a higher concentration of grants, for example London and Manchester are big cities in UK, and they both have a large concentration of grants compared to other cities. This may be due to the fact that these cities host many charity's headquarters, and so the recipient of the grant is 'based' in these larger cities.

This may also be due to the fact that larger cities come with a greater population size, typically with more deprived areas and thus community initiatives that the Co-operative Group typically funds.

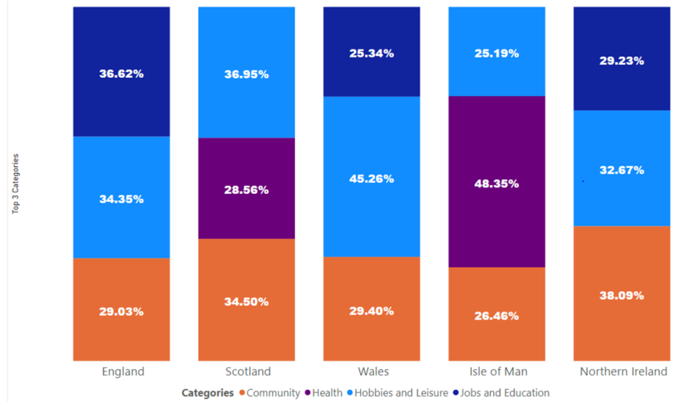

Fig. 4. Top three themes in each country.

Figure 4 shows how themes of grants are different over different countries of UK. As we can see in the chart, grants to Northern Ireland are mostly categorized by Community with 38.09% of over grants. Wales have the most grants related with Hobbies and Leisure (48.35%). Health related grants are dominated in Isle of Man with 48.35% of grants are labeled as Heath. Grants categorized by both Community and Hobbies and Leisure have 36.95% and 34.5% of total grants respectively in Soctland. England has a much balanced category of grants with 36.62% of Jobs, 34.35% of Hobbies and Leisure and 29.03% of Community.

These 5 countries can be classified into 2 categories: one category containing Scotland and Isle of Man where Health related grants are one of the major grants, and the other category containing England, Wales and Northern Ireland, who all have grants of Hobbies and Leisure as their one of major grants instead of Heath related grants. This can be explained by the fact that different countries may have different dominate industry.

In Soctland and Isle of Man, Health related industries may have a dominant power. In England, Wales and Northern Ireland, recreation related industry may be more prosperous that other countries.

C. Research Question 3: Is there any geographical variation in the amount awarded to recipients based on theme?

Fig. 5. Major recipients geographically plotted based on theme and amount awarded.

Figure 5 explains the geographical variation of themes for largest recipients of grants by the Co-operative Group. The size of the bubble for different cities denotes the amount awarded to an organization and the colour of the bubble denotes the theme, given in the legend. The recipient organizations who receive multiple grants do not necessarily receive grants in the same theme. For example, The British Red Cross receives grants for both Hobbies and Leisure as well as Support Groups. But it receives major grants for Hobbies and Leisure (0.14 Million pounds in 2017), so we have shown the British Red Cross belonging to Hobbies and Leisure theme here. On the contrary, some organization's grants belong solely to one theme. For example, Alzheimer Scotland and North Aberdeenshire Services receive grants for Health theme only.

Since the British Red Cross was the largest recipient in 2017 (£256K) and its main headquarters are located in London, we observe that there is a large bubble around London belonging to Hobbies and Leisure theme. Similarly, The Co-op Foundation has been the largest recipient of the grants in the year 2018 (£238K) with their main headquarters in the Greater Manchester Area, explaining the large bubble around Manchester belonging to the Community theme. These were the default local causes in each of the 2 years, suggesting why there was a substantial funding here. In contrast, other organizations like Girlguiding have different locations across the UK, and so their grants are spread across different geographical regions. In total receive more funding but they are allocated into many smaller amounts.

It can be seen that Scotland has 2 major recipient organizations - Alzheimer Scotland – North Aberdeenshire Services (Health) and Arbroath Lifeboat Station (Community). This is also concluded from the analysis of our second research question that Health and Community are among top 3 major themes for grant receiving in Scotland.

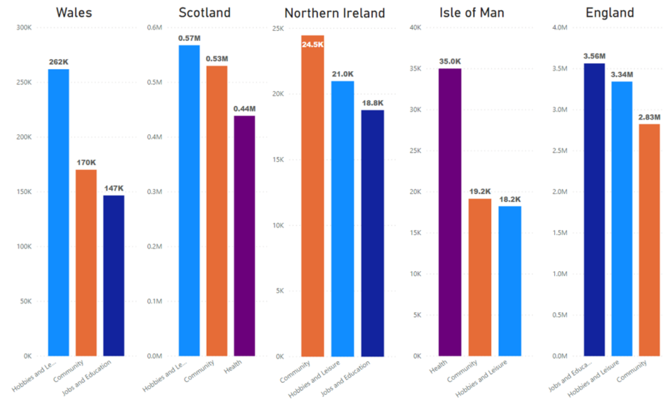

Fig. 6. Amount awarded in the top three themes of each country.

Figure 6 depicts that Scotland and Wales had majority of their grants being awarded to Hobbies and Leisure theme which is approximately 0.5million and 262k respectively. On the contrary, Isle of Man had significantly lower preference for grant giving to Hobbies and Leisure and majority of its grants were awarded to Healthcare, amounting to 35K. Although this amount is much lower as compared to Scotland and Wales, but accounting for the relative sizes of the countries, the number of recipients ae also lower in Isle of Man. England received grants valued 3.5M alone for Jobs and Education, which is the highest among all the countries. Thus we can observe an inclination on investing towards Jobs and Education for England, community for Northern Ireland, Hobbies and Leisure for Wales and Scotland, and healthcare for Isle of Man.

D. Bias and Validity

There are potential biases in the allocation of funding by The Co-operative Group. To be accepted as a new local cause, the project must benefit the local community and the organisation cannot be run for private profit [8]. In previous years, recipients that were registered charities, found it is easier to apply for funding, however the criteria for applications changes on a yearly basis depending on the Co-operative Group's strategic direction. The other major bias is that fact that The Co-operative Group's biggest pay-outs are to the charities that they sponsor in that given year. The British Red Cross and The Co-operative Group Foundation charity are the beneficiaries that received by far the largest pay-outs, in 2017 and 2018 respectively, as they were the default choice for a member did not specifically choose a local cause.

A common criticism of case studies is that it only relates to the case under study, and the Co-operative Group is maybe not representative of all grant-makers in that it's a cooperative, not a private foundation nor charity. Indeed The Co-operative Group is a fairly unique case as Co-operative Group members choose the causes, rather than the funding organisation itself. The degree to which conclusions in our study are valid for other funders will need to be considered; firstly are all the labels relevant to that organisation or do the theme types need to be modified or extended. Secondly, if the funder allocates in a generalist way with a broad ranging variety of beneficiaries like The Co-operative Group, we can map out the degree of proximal similarity [9] to see if conclusions can be generalized to others.

Many of the themes are closely related to each other, and so there is a potential bias in how the classifier has predicted the theme for each grants, given the classifier was trained on an unknown dataset. Hobbies and Leisure can be closely related to community theme, and so there may be bias as the classifier may have incorrectly predicted leisure activities and classified them as Community. However, we hope to have minimised this bias as we have attempted to manually relabel some of the grants. This however may introduce researcher bias.

Our findings violate external validity as these observations are specific to Co-operative Group in United Kingdom, we cannot generalize the same grants giving trend for other organization in UK or outside.

IV. Conclusion

With The Co-operative Group frequently changing its application criteria for local causes and only two years of historical data currently being provided in GrantNav, it is difficult to investigate the themes over time and to predict the themes that individuals will support in the future as the application criteria changes. Therefore in further work, we could generalise to other times and reduce threats to validity following the release of The Co-operative Group local cause pay-outs in 2019, which would give an addition year of data and allow more forward-looking analysis to be performed. Alternatively we could select an alternative funder to see whether our findings can be generalised.

In terms of grant labelling we envisage future work could involve training a supervised machine learning classifier on a larger portion or even the entire GrantNav database in an attempt to improve classification accuracy. Developing a bespoke classifier from scratch is also a potential area to explore, given we utilised the Google Natural Language API to categorise the grants. We also believe the set of labels we produced could be further subdivided into separate labels in order to better understand the distribution of grants across a broader range of themes.

We have seen that Hobbies and Leisure, Jobs ans Education, and Community are most granted themes for years 2017-2018 together. Of the total value of 56.3 million, these 3 themes alone receives 43% of the total grants the are being awarded. This is also supported by the fact that the top 2 grants recipients for year 2017-2018 – The British Red cross and Co-op foundation typically recieve Hobbies and leisure and Community respectively. However at a microscopic level, the preference of themes for grants changes considerably. For Wales and Scotland, Hobbies and Leisure were top most awarded theme whereas for Isle of Man, healthcare received the maximum grants. Thus we can conclude that there is significant geographical variation in the amounts awarded for each theme for different countries.

References

[1] 360Giving, "360Giving Strategy," 2019. [Online]. Available: https://www.threesixtygiving.org/strategy/

[2] D. Kane, "Funding Playground - who funds with who in the UK?" 2018. [Online]. Available: https://www.threesixtygiving.org/2018/11/27/funding-playground-who-funds-with-who-in-the-uk/

[3] ——, "360 Giving 2017 grants network," 2019. [Online]. Available: https://public.flourish.studio/visualisation/155446/

[4] B. J. Oates, *Researching information systems and computing*. London: SAGE Publications, 2006.

[5] E. Walczak, "PostcodeioR." [Online]. Available: https://cran.r-project.org/web/packages/PostcodesioR/index.html

[6] Google, "Google Cloud Natural Language API." [Online]. Available: https://cloud.google.com/natural-language/docs/

[7] ——, "Google AutoML." [Online]. Available: https://cloud.google.com/automl/

[8] C.-O. Group, "Co-op Local Community Fund," 2018. [Online]. Available: https://causes.coop.co.uk/

[9] W. M. K. Trochim, *Research methods knowledge base*, 3rd ed. Mason, Ohio.: Atomic Dog/Cengage Learning, 2008.