Snappy title goes here

Charities, funding, and Twitter

|  |  |  |
| --- | --- | --- |
| Natalie Polack University of Stirling npp00002@students.stir.ac.uk | Tom Wallace University of Stirling tom.wallace@stir.ac.uk | Vikki Richardson University of Stirling vikkirichardson@hotmail.co.uk |

ABSTRACT

T

KEYWORDS

A

1 INTRODUCTION

This research paper explores factors that may predict charity use of Twitter. Twitter is a social networking platform that is commonly used by charities but remains understudied academically (Obar, Zube, & Lampe, 2012). Charities have many uses for Twitter, including advertising charitable campaigns, interacting with their users or supporters, and even communicating directly with other charitable organizations. Twitter has a relatively low cost of use, only requiring staff time. This may present an opportunity for smaller charities, but larger charities are likely to have the capacity to dedicate staff to the activity and so are still, likely, at an advantage. Our study attempts to predict various aspects of charity Twitter usage (such as overall use, popularity, and networking activity) based on publically available data on charities features (such as funding source, overall income, who the charity helps, and number of staff). This study is important because there is a lack of published research on the factors which relate to successful use of social media by charities. Twitter has huge potential for charitable organisations and knowing which factors contribute to success or failure on the platform could help implement measures that lead to more of the former and less of the latter.

The next part of this paper, section 2, describes pervious literature and research in this area that feeds into a brief summation the research questions and data management tasks. Sections 3 and 4 then detail the data processing and analysis methods used before the results are discussed in section 5. The paper concludes with a summary of the implications of the findings in section 6.

2.2 Research questions

**RQ1**: Is funding source related to use of Twitter?

**RQ2**: Are charities which seek to help the public more popular on Twitter?

**RQ3**: Does number of staff, rather than size, determine a charity’s active use of Twitter?

2.3  Data sources and management tasks

Given the research questions above, three data sources were required to undertake this research. To obtain data on the outcome, Twitter use by charities, we downloaded data directly through the Twitter API (2018). For data on charity features, we scraped the UK Charity Commission website (2018) which holds a record for every registered charity in England and Wales (Scotland and Northern Ireland are covered by separate regulators). Given these two data sources, we could sample an arbitrary number of charities, right up to the full population of ~168,000 that are registered with the regulator. However, the Charity Commission website does not record detailed financial information for charities which is key to RQ1. Therefore, we make use of data collected by Alcock and Mohan of the Third Sector Research Centre (2017). This dataset is based on surveying a stratified random sample of registered charities which contains detailed financial information, such as funding source. As this dataset is the smallest, it forms our sample and we then add to this data from the Charity Commission and Twitter. This gives us a sample of 12,150. Each of these data sources needed to be interacted with in different ways and presented different data management challenges which are detailed in section 3.

Because the Alcock and Mohan data provided the sample, it was sourced and managed first. The Charity Commission and Twitter data were then collected in parallel. The data sets were then recombined to form a final analysis dataset as shown in the code tree below.

[Code tree diagram]

2 SCENARIO

2.1 Literature and related work

Twitter use by charities is common. The Lloyds Digital Index found that 44% of the, more than one hundred thousand, charities they surveyed in 2016 were using social media in some form (Lloyds Bank, 2016). Of these platforms, Twitter is by far the most used platform for charitable organisations (Guo & Saxton, 2014). But what do charities use Twitter for? A common use case is broadcasting one-to-many messages to followers to convey information or solicit donations (Phethean, Tiropanis, & Harris, 2015; Waters & Jamal, 2011). Another use of Twitter is networking activity between charities, either because they share a common purpose or for support. Infrastructure organisations, which are charities established to help support other charities, can use Twitter to share links to information, resources, and training (Dayson & Sanderson, 2014).

Overall, the literature, which has been published so far, seems to concur that social media use helps make a charity successful (Lloyds Bank, 2016; McCabe & Phillimore, 2012). Therefore, it is important to understand what factors make a charity an active user of social media and this has largely been neglected by the literature thus far.

3  DATA PROCESSING & ANALYSIS METHODS

3.1 The Alcock and Mohan data

The data set chosen as a starting point for this project was from the UK Data Archive and it contains detailed financial information for charities in England and Wales collected via a survey. The data is longitudinal and contains information from the financial years 2006/07 up to 2013/14. For this project, it was decided to use data from one financial year. The year selected was 2011/12, as the information for the more recent years was less complete.

The ‘CharityCharacteristics.csv’ file was imported into Python and a pandas data frame was created. Then df.reset\_index followed by df.set\_index was used to set a two level index using Charity Number (‘ccnum’) and Financial Year (‘financial\_year’). Next df.drop was used to drop a list of all the financial years apart from 2011/12. The two level index was no longer needed at this point, so the index was reset to Charity Number and the pandas data frame was saved as a JSON object: ‘charity\_oneyear.json’.

In the ‘CharityCharacteristics.csv’ and the ‘charity\_oneyear.json’ files, the data for the source of funding is split into a large number of variables and the names used are not particularly informative. A new variable was created in the data frame and the addition operator was used to combine the source of funding data from different columns:

* **Government funding** ('government\_funding')   
  – combination of ig100, ig110, ig121, ig125, ig161, ig162, ig163 and ig180

Two new variables were created to assess the proportion of each charity’s funding that came from Government Funding and General Public Funding:

* **Government Funding Proportion** (‘Prop\_government\_funding’)   
  = Government funding / total income (‘itotal’)
* **Public Funding Proportion** ('Prop\_general\_public\_funding')  
  = ig600 / total income (‘itotal’)

The data types of the variables used in these calculations were checked to ensure they were compatible using df.dtypes and they were found to be integers (‘int64’). Lastly, the variable for each charity’s total income (‘itotal’) was renamed as **'Income2011-2012'** to distinguish it from up-to-date information collected via a web scrape and the variable for funds generated from the general public (‘ig600’) was renamed as **'Funds\_general\_public'**.

After creation and renaming of the variables of interest, the filter () function was applied to the data frame to create a new smaller data frame with data for 12 150 charities and only 6 variables:

* **'ccnum'** – the Charity Number which was kept as the index
* **'Government\_funding'** – total funding from the Government
* **'Prop\_government\_funding'** – proportion of funding coming from the Government
* **'Prop\_general\_public\_funding'** – proportion of funding coming from the general public
* **'Income2011-2012'** – Income from 2011 - 2012
* **'Funds\_general\_public'** – funds generated from the general public

3.2 The Charity Commission data

It was desirable to join the charity commission data to the UKDA data for two reasons. Firstly, it contains updated income data and secondly, it contains some data not found in the UKDA data.

3.3 The Twitter data

The

3.4 Combined data

The

4  ANALYSIS METHODS

4.1 Univariate methods

The

Histograms, summaries, one-way tables

4.2 Bivariate methods

The

Correlations, chi2

4.3 Multivariate methods

The

OLS

Logit

5 RESULTS & DISCUSSION

5.1 Question 1

The

5.2 Question 2

The

5.3 Question 3

## Summary statistics of the variables

In order to determine whether the number of staff employed by a charity, rather than the size of the charity, plays a role in a charity’s successful use of Twitter, three variables were selected from the final data set: ‘Final\_analysis\_file.json’.

* **‘Staff’** - Number of staff employed by the charity, measured as full-time equivalent (FTE)
* **‘Twitter following’** - Number of people / organisations the charity follows on Twitter is used as a measure of active use of Twitter, as some charities may choose not to Tweet very often
* **‘Income2018’** – Total income from the web scrape carried out in 2018

The df.describe() command was used to produce statistics for these three variables and the results are shown in Table 1. The number of NaN present for all variables in the data frame was counted using df.isnull().sum(): Staff: 1455; Twitter following: 7696; and Income 2018: 2692 but Python automatically excludes these when calculating summary statistics.

**Table 1: Summary statistics for Staff, Twitter following and Income2018**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Staff** | **Twitter following** | **Income2018** |
| **Data type** | float64 | float64 | float64 |
| **Count** | 10695 | 4454 | 9458 |
| **Mean** | 14.730 | 1000.057 | 1351087 |
| **Standard deviation** | 23.887 | 885.686 | 1655516 |
| **Minimum** | 0 | 0 | 0 |
| **25%** | 0 | 289 | 138250 |
| **50%** | 0 | 758.500 | 706700 |
| **75%** | 22 | 1505.500 | 1900000 |
| **Maximum** | 102 | 3874 | 7400000 |

## Correlation between Staff and the number ‘Following’ on Twitter

The correlation between the number of Staff and the number ‘Following’ on Twitter was calculated using df['Staff'].corr(df['Twitter following'] and a weak positive correlation of 0.07 was measured.

## Linear Regression of ‘Staff’ predicting ‘Twitter Following’ controlling for size

A linear regression was carried out (y = m1x1 + m2x2) using ‘Staff’ and controlling for size, using ‘Income2018’ as the independent variables: x1 and x2, and setting ‘Twitter following’ as the dependent variable: y. The statmodels.api module was used and the results are shown in Table 2.

**Table 2: Results of the Linear Regression Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dependent:**  **Twitter following** | **Coef.** | **Std error** | **P>|t|** |
| **Staff** | -0.992 | 0.756 | 0.189 |
| **Income 2018** | 0.00008 | 0.00001 | 0.000 |
| **Constant** | 867.481 | 21.347 | 0.000 |

**R-squared = 0.017 ; Prob = 0.0000000000007;   
AIC = 55160; BIC = 55170**

The R-squared value takes into account the number of independent variables used in the model and in this case it is 0.017. The constant coefficient is the intercept with the y axis and is 867.4814. The standard error indicates the accuracy of the constant, ‘Staff’ and ‘Income2018 coefficients with a low value indicating a higher accuracy, and the confidence interval shows the range that the coefficients will probably fall within.

The p-value of 0.189 for ‘Staff’ indicates that changes in this variable are not associated with changes in ‘Twitter following’ – the y variable. However, the p-value of 0 for ‘Income2018’ is statistically significant, so the null hypothesis that size of a charity does not affect the number of people / organisations that the charity follows on Twitter can be rejected. Note: Size of the charity is measured by the Income in this case.

6 CONCLUSIONS

6.1 Summary

The

6.2 Future work

The

REFERENCES