**Abstract**

This research project provides a better understanding of the impact of mission statements of tax-exempt organizations on fundraising, specifically in the United States, Australia, the United Kingdom, and Canada. We utilize the Python Requests and BeautifulSoup libraries to web scrape mission statements, names, revenue, and other relevant data from the IRS, GuideStar, and Charity Navigator. After generating the data, we use techniques drawing from Natural Language Processing (NLP) and statistics to identify patterns in mission statements across different categories of tax-exempt organizations like religious charities and medical nonprofits. NLP libraries used include Natural Language Toolkit (NLTK), Regular Expressions (RegEx), and SpaCy. In full, this project provides insight into the impact of mission statement language on fundraising efforts. All project code, data, and documentation can be found at github.com/Hariaksha/NLP-Charity-Research.

**Motivation**

The nonprofit sector largely consists of small, community-based organizations working with meager resources. For this reason, nonprofits rely on resources and research to operate more effectively. Prior researchers have begun to understand the impact of different linguistic structures and patterns on economic and organizational performance. This project furthers these efforts by addressing a research gap surrounding empathy and compassion in mission statements. We evaluate the impact of "self VS other" linguistic structures on revenues of tax-exempt organizations. This research will provide a better understanding of the impact of linguistic features and can be applied to increase charitable donations and elevate the impact of a wide array of nonprofits, thereby touching the lives of countless individuals.

**Background**

*Natural Language Processing*

* Natural Language Processing (NLP) is the branch of Artificial Intelligence that enables computers to understand, manipulate, and generate human language.
* This technology is vital for many applications involving unstructured and textual data.

*Nonprofits in America*

* Most nonprofits are small and community-based, and 92% operate with less than $1 million a year.
* There are 1,859,826 tax-exempt organizations in the country.

*Language and Impact*

* Prior research indicates that mission statement language correlates with economic and organizational performance.

**Hypotheses**

1. Mission statement language will affect revenue of tax-exempt organizations.
2. Mission statements with language identifying as more donor-serving will correlate with higher revenue.
3. The effects of mission statement language on revenue will be less powerful for tax-exempt organizations with higher asset codes.
4. The effects of mission statement language on revenue will differ slightly among categories of tax-exempt organizations, such as education versus health nonprofit.

**Methodology**

*Data Collection*

The research dataset consists of many tax-exempt organizations in the United States. The most important data for each organization are the mission statements and the revenue. We obtain the EINs, addresses, NTEE codes, and financial information from the [Internal Revenue Service (IRS)](https://www.irs.gov/charities-non-profits/exempt-organizations-business-master-file-extract-eo-bmf), which has information about all tax-exempt organizations in the United States organized in CSV files. The mission statements, organization names, and URLs were scraped from [GuideStar](https://www.guidestar.org/) and [Charity Navigator](https://www.charitynavigator.org/). The web scraper used to collect these data is titled 'scraper.py'. This script uses the csv, requests, BeautifulSoup, time, and openpyxl Python libraries.

Each mission statement was preprocessed and tokenized using spaCy's 'en\_core\_web\_lg' English NLP model.

*Preliminary Data Analysis*

Before running our main analysis of labeling mission statement as more donor-serving or society-serving, we use the Quantitative Analysis of Textual Data R package (quanteda) to measure the linguistic richness, [linguistic readability](https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests), and mean sentence length of each mission statement. We then find the relationship between each of these values and the revenue of each charity.

*Division of Data*

Before we begin our main analysis of the data, we organize our data by sectors and by assets. This allows us to see if overall linguistic trends and revenue relationships differ among sectors and among nonprofits with varying amounts of assets. The labeled NTEE codes and asset codes are pictured below.

|  |  |
| --- | --- |
| **NTEE Code** | **Sector** |
| A | Arts, Culture, and Humanities |
| B | Education |
| C | Environment |
| D | Animal-Related |
| E | Healthcare |
| F | Mental Health & Crisis Intervention |
| G | Voluntary Health Associations & Medical Disciplines |
| H | Medical Research |
| I | Crime & Legal-Related |
| J | Employment |
| K | Food, Agriculture and Nutrition |
| L | Housing and Shelter |
| M | Public Safety, Disaster Preparedness and Relief |
| N | Recreation and Sports |
| O | Youth Development |
| P | Human Services |
| Q | International, Foreign Affairs and National Security |
| R | Civil Rights, Social Action & Advocacy |
| S | Community Improvement & Capacity Building |
| T | Philanthropy, Voluntarism and Grantmaking Foundations |
| U | Science and Technology |
| V | Social Science |
| W | Public and Societal Benefit |
| X | Religion-Related |
| Y | Mutual and Membership Benefit |
| Z | Unknown |

|  |  |
| --- | --- |
| **Asset Code** | **Description ($)** |
| 0 | 0 |
| 1 | 1 to 9,999 |
| 2 | 10,000 to 24,999 |
| 3 | 25,000 to 99,999 |
| 4 | 100,000 to 499,999 |
| 5 | 500,000 to 999,999 |
| 6 | 1,000,000 to 4,999,999 |
| 7 | 5,000,000 to 9,999,999 |
| 8 | 10,000,000 to 49,999,999 |
| 9 | 50,000,000 to greater |

*Relating Mission Statement Language to Revenue*

We use NLP and statistics to determine the relationship between mission statement language and revenue. The specific linguistic structures that we distinguish between is donor-serving and society-serving language. Examples of varying mission statements are found below.

Appeal to self (donor-serving, how does donating benefit you?):

* To provide tools and for members to run their businesses efficiently and effectively, work safely, and be an advocate for the members.
* Render financial aid to members and their families in addition to support of Catholic various charities.
* The purpose of the Alliance for Performance Excellence is to educate its members in performance excellence and support their provision of performance excellence services and education to their clients.

Appeal to society (community-serving, how does donating benefit the world?):

* The Society of St Vincent de Paul (SVP) Charity is a Christian voluntary organization working with poor and disadvantaged people. Inspired by our principal founder, Frederic Ozanam, and our patron, St Vincent de Paul, we seek to respond to the call every Christian receives to bring the love of Christ to those in need: "I was hungry and you gave me food" (Matthew 25). No work of charity is foreign to the society.
* Our mission is to provide support, information and an atmosphere where children who are fighting or have survived Retinoblastoma can socialize with other similarly affected children.
* "Helping people live in harmony with their environment. We offer Horticulture Therapy outreach programs and publish and sell books that meet our mission statement."

After we find a measurement for the language of each mission statement, we will create a linear model correlating that variable with the revenue of each charity. This allows us to see if donor-serving or society-serving mission statements are more effective to increase charitable giving.

**Future Steps**

The next steps are:

* Generate data from the United Kingdom, Australia, and Canada
* Create computational linguistic measurement of "self VS other" language and assign respective value to each mission statement.
* Analyze the data and note important patterns.
* Train ML model on training data to estimate revenue of tax-exempt organization given mission statement and asset code.
* Test ML on testing data.

**Conclusion**