

Computing Lab / Data Warehousing and Business Intelligence

Analytics on the US Nonprofit Sector - *Milestone I*

The nonprofit sector is the fastest growing industry in the United States: the number of registered nonprofit organizations (NPO) is proliferating as well as the economic activity and labor force within it. For example, the number of registered nonprofits increased by 21.5% from 2001 to 2011 and in 2007 the sector had annual revenues of almost \$2 trillion. Additionally, 10% of the entire US labor force worked within the sector; and, if you include the labor of volunteers, for which there were \$62 million in 2007, then the nonprofit sector has a larger workforce than any other industry.¹

Given the quantity as well as variability of this sector, coupled with rudimentary data collection and storage methodologies, our team saw tremendous potential in exploring this field from a data science perspective. Specifically, we hope that our project will yield greater insight into the ways in which NPOs function internally as well as within the larger capitalist economy—the applications of which, we hope, our target users (defined below) will derive business insight.

The Internal Revenue Service (IRS), the government agency responsible for tax collection and tax law enforcement, requires all NPOs to annually file a 990 form to verify that NPOs are in compliance with their federal tax-exempt status. While this information is public, the public is not able to easily analyze the sector as a whole because 1) the IRS just recently digitized their filing system and 2) the architect of the public database, provided by the National Center for Charitable Statistics (NCCS), limits the public to individual NPO queries. Our project addresses the first issue through the help of the NCCS, who just this year digitally converted 990 forms collected from the 2012-2014 filing period, and we seek to address the second issue through the design of our dashboard.

The raw data that we have collected from these 990 forms includes both quantitative and qualitative data in .csv file format. There are approximately 200K records (one per organization) with about 250 columns of data per year. This data includes text (company names, locations), numbers (finances), dates, dictionary sets (state codes) and booleans. Furthermore, we hope to include other sources of data to cross-reference as well as delve deeper into our analysis of the sector within the larger economy. For example, because our data is limited to the 2012-2014 reporting years, we may seek other sources to add historical dimensionality of the sector to our analysis.

Two primary target users of our dashboard guided the development of our analytics. Internally, the executive branch of an NPO could use our dashboard to drive business decisions. For example, executives could compare various financial metrics to their counterparts across the sector. Externally, nonprofit consultancy firms could use this dashboard to both generate leads for marketing to NPOs as well as derive greater insight into the sector. Finally, secondary users of our dashboard include philanthropist and government bodies who seek complementary guidance to their donation and funding decision-making. Specifically, our dashboard will enable users to run queries that answer the following questions:

- Salary Proportionality: how is it related to revenue (both in amount and streams), years of operation, location, number of paid and unpaid employees?
- Revenue Streams: how are revenue streams related to the types and size of nonprofits and/or number of paid and unpaid employees?
- Effectiveness of Fundraising: how effective is fundraising and how might it be dependent on years of operation, location, number of paid and unpaid employees?
- Donating patterns: Are there patterns to giving based on various aforementioned NPO features?
- Additionally, we would like to add some predictive analytics. For example, given the aforementioned analytics, could one who is interested in starting an NPO with various NPO features predict the likelihood of revenue streams, fundraising efficacy and executive salary?

¹ Salamon, Lester M. "America's nonprofit sector: A primer." New York (1992): 1988-1992.