501c3 Job Title Standardization

This document outlines three interrelated products created to standardize the titles of individuals from Nonprofits from 501c3 organizations. Data for this project came from the IRS’s public release of e-filed nonprofit tax return data on Amazon Web Services. The three products, when ran sequentially, clean standardize and categorize 85.9% of all of the titles of people listed on the nonprofit 990s, Schedule J. This covers the years 2010-2015. The three products are: Clean and Standardized the Title, Thesaurus of Title Categories, and Titles Sorted into Categories.

Clean and Standardized the Titles

The first data product, Clean and Standardized Titles, outlines the process by which we cleaned and standardized the titles of people listed on schedule J. First, we capitalized every letter in the Titles. Then we added a space to the beginning and end of each title, to make it easier for the code to differentiate between complete words. Next, we removed any punctuation or numbers for the titles. As well as stop words (i.e., of, and, for, etc.). Finally, we corrected common misspellings and removed words for months of the year. This process left us with a much more standardized set of Titles.

----------------------------------------------------------

Title Cleaned Title

----------------------------- ----------------------------

VP OF OPERATIONS VICE PRESIDENT OPERATIONS

PRESIDENT/DIREC PRESIDENT DIRECTOR

CFO CFO

CFO CFO

PRESIDENT PRESIDENT

ACADEMY LEADER ACADEMY LEADER

DIRECTOR OF MARKETING DIRECTOR MARKETING

EXECUTIVE DIRECTOR EXECUTIVE DIRECTOR

VP Development & Alumni Rel VICE PRESIDENT DEVELOPMENT

ALUMNI REL

STORE MANAGER STORE MANAGER

Thesaurus of Title Categories

Once the titles were standardized and cleaned, we found the 1000 most common titles; these accounted for 75% of the dataset. We used these titles to create 17 categories. We used using the ASU taxonomy created by the Lodestar Center Compensation Survey, as a starting poing to create the categories[[1]](#footnote-1). Though, we added several categories and dropped the category volunteer management (None of the people listed on the Schedule J’s fit into that category. We sorted the 1000 most common titles into these categories. It’s important to note that these categories were not mutually exclusive, as an individual could hold multiple positions. The Categories are listed below; the bolded categories are ones we added in addition to the Lodestar taxonomy:

*Title Categories*

Executives

Financial Management

Fundraising and Development

Public Relations

Human Resources

Information Technology

Program Management

Legal and Public Policy

Facilities and Properties

Administrative Support

Health and Medical Management

Health and Human Services

Training and Education

**Treasurer**

**Secretary**

**Department Head**

**Generic Management**

**Academic management**

**Professional**

**Medical Professional**

**Academic Professional**

**Other**

**Trustee**

Titles Sorted into Categories

The final product takes the cleaned and standardized titles and sorts them into the categories above. As mentioned before, these categories were not mutually exclusive so that someone with the title’s CEO/CFO, would be placed into both the Executive and Financial Management categories. In addition to sorting the titles according to a 1-1 match, we also used the thesaurus of categories to generate common phrases associated with the category. These common phrases were used to categorize titles that were outside of the 1000 most common titles. For instance, the Information Technology category used the abbreviation’s IT, CIO, CDO, as well as the words “Information” and “Technology” in order to classify anyone who had those titles as part of the Information Technology category. Finally, these categories were used to create abstract titles.

1. <https://lodestar.asu.edu/sites/default/files/asu_lodestar_center_compensation_survey_job_positions.pdf>. [↑](#footnote-ref-1)