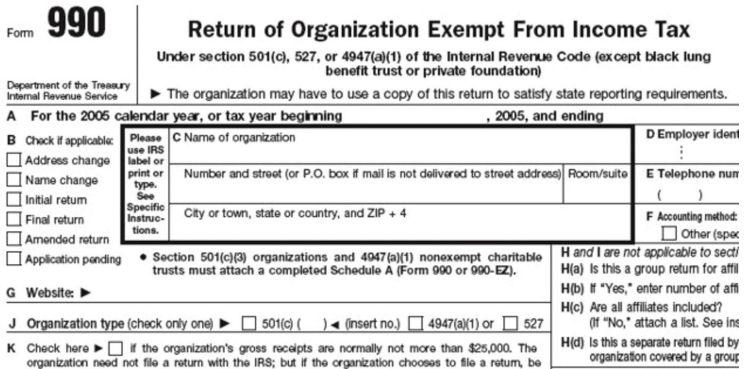
# Predicting Not-for-Profit Executive Compensation

A data science project using classification to model executive salaries as reported on IRS form 990



Form **990**

**Return of Organization Exempt From Income Tax**

Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except black lung benefit trust or private foundation)

Department of the Treasury
Internal Revenue Service

► The organization may have to use a copy of this return to satisfy state reporting requirements.

A For the 2005 calendar year, or tax year beginning , 2005, and ending

B Check if applicable:
- Address change
- Name change
- Initial return
- Final return
- Amended return
- Application pending

Please use IRS label or print or type. See Specific Instruc-tions.

C Name of organization

Number and street (or P.O. box if mail is not delivered to street address) | Room/suite

City or town, state or country, and ZIP + 4

D Employer ident

E Telephone num

( )

F Accounting method:
- Other (spec

• Section 501(c)(3) organizations and 4947(a)(1) nonexempt charitable trusts must attach a completed Schedule A (Form 990 or 990-EZ).

H and I are not applicable to secti

H(a) Is this a group return for affil

H(b) If "Yes," enter number of affi

H(c) Are all affiliates included? (If "No," attach a list. See ins

H(d) Is this a separate return filed by organization covered by a group

G Website: ►

J Organization type (check only one) ► ☐ 501(c) ( ) ◄ (insert no.) ☐ 4947(a)(1) or ☐ 527

K Check here ► ☐ if the organization's gross receipts are normally not more than $25,000. The organization need not file a return with the IRS; but if the organization chooses to file a return, be

# Scope & Interest in This Topic

1.5M not-for-profit organizations in the US advocate for causes that are highly sympathetic. These organizations easily pull at the heartstrings of donors. But how is that money being spent?

My goal with this project was to design a tool that could be used to promote not-for-profit accountability and fair compensation.

Charity Navigator has used a numbers-based rating system to score over 9,000 not-for-profits. A model that can identify outliers could be used to flag specific organizations that are misusing donations.

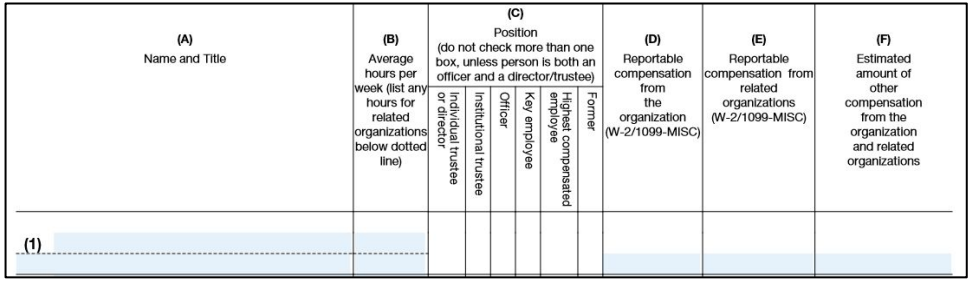
Additionally, the Board of Directors of a not-for-profit could use this tool to determine a fair salary based on the resources and nuances of its not-for-profit.


# Data Source

First made available by the IRS in 2016 and hosted by both Amazon Web Services (xml format) and Open990 (csv format) an organization dedicated to not-for-profit transparency.

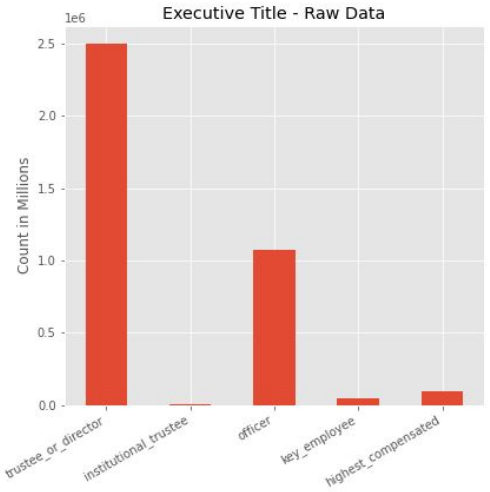
- **Compensation** for 4M executives
  - Time span between Jan/2017 - Aug/2019
  - Limited to full time -> 130K executive payments
- **Governance** data for 1.4M not-for-profits
  - Time span between 2010 - 2018
  - Limited to 2015 - 2017 -> 270K unique organizations

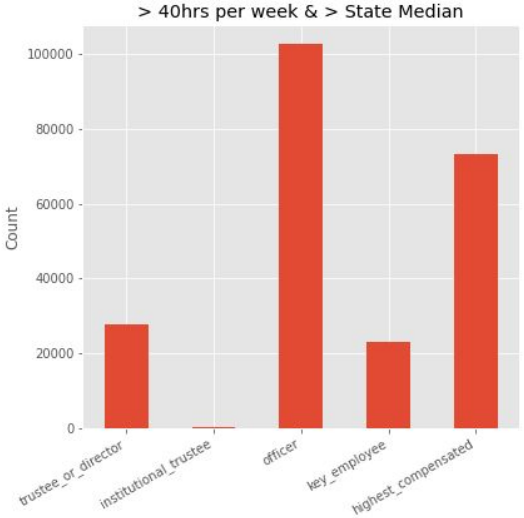
# Data Assumptions

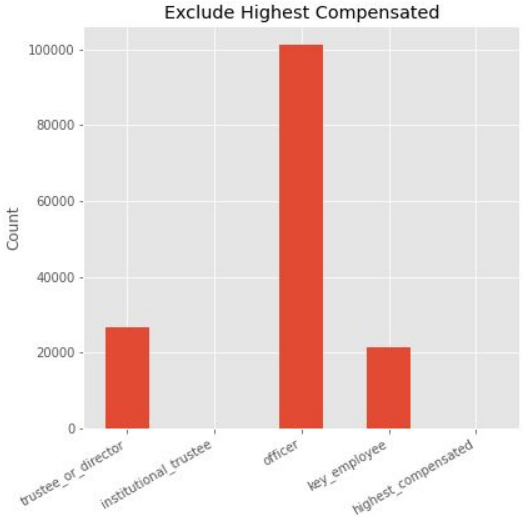


| (A) Name and Title | (B) Average hours per week (list any hours for related organizations below dotted line) | (C) Position (do not check more than one box, unless person is both an officer and a director/trustee) | | | | | | (D) Reportable compensation from the organization (W-2/1099-MISC) | (E) Reportable compensation from related organizations (W-2/1099-MISC) | (F) Estimated amount of other compensation from the organization and related organizations |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Individual trustee or director | Institutional trustee | Officer | Key employee | Highest compensated employee | Former | | | |
| (1) | | | | | | | | | | |

- Identifying full time executives vs. part time trustees or directors (>40hrs per week & > median state income)
- Exclude executives with salaries over 1M
- Final number of executives compensated - 130K



# Primary Features

---

## Compensation:

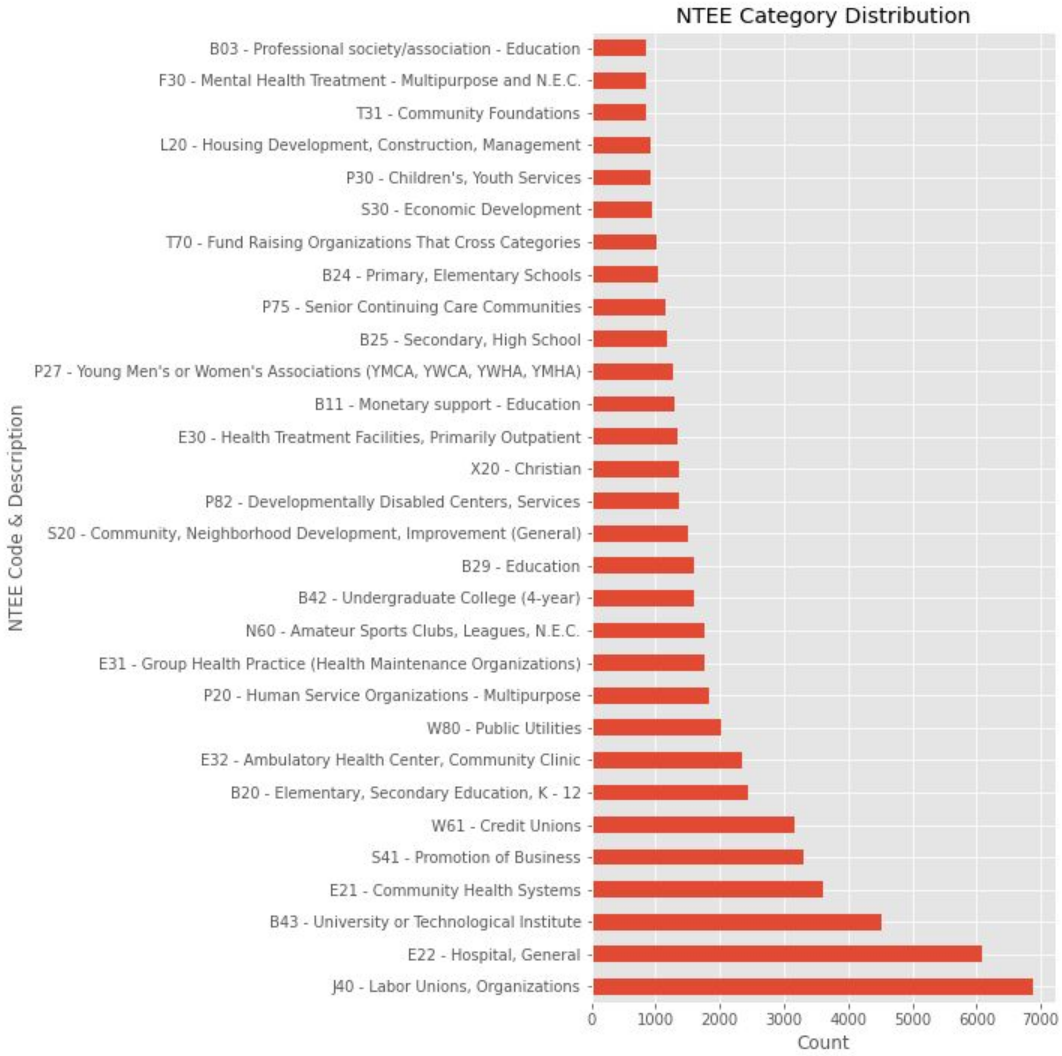
- EIN - organization unique identifier
- Subsection - 501(c) type
- NTEE Code - categories or purpose for the organization
- State
- Formation year
- Assets
- Liabilities
- Expenses
- Revenue

## Governance:

- EIN - organization unique identifier
- Voting Member Count - board of directors
- Employee Total Count
- Volunteer Total Count
- Salaries Expense
- Governance controls(Yes/No):
  - Conflict of interest policy?
  - Whistleblower policy?
  - CEO Compensation review?
  - Document Retention Policy?
  - Delegation of Management Policy?
  - etc.

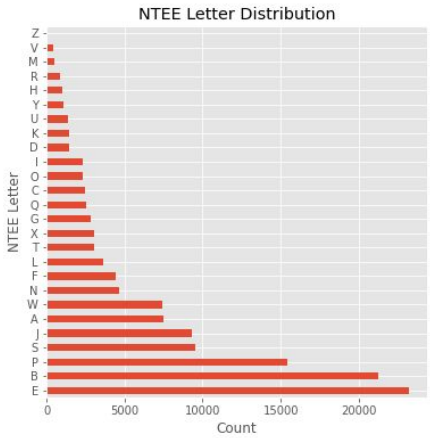
# Data Exploration - NTEE Designation

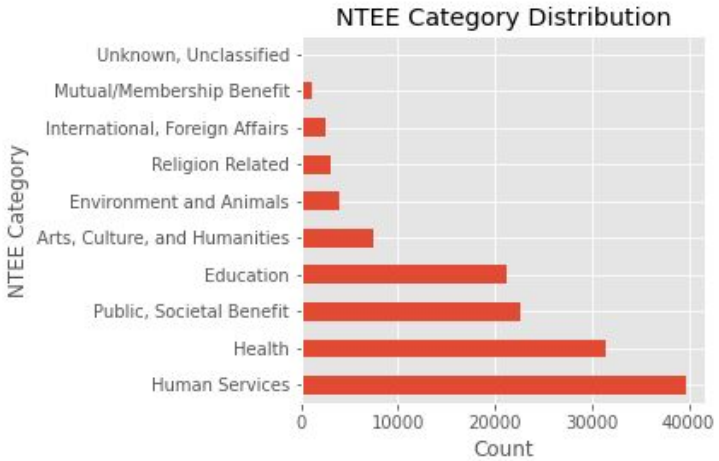
- NTEE (National Taxonomy of Exempt Entities) codes represent the mission or purpose of the not-for-profit (over 600 unique codes)
- The taxonomy was created in the 1980s.

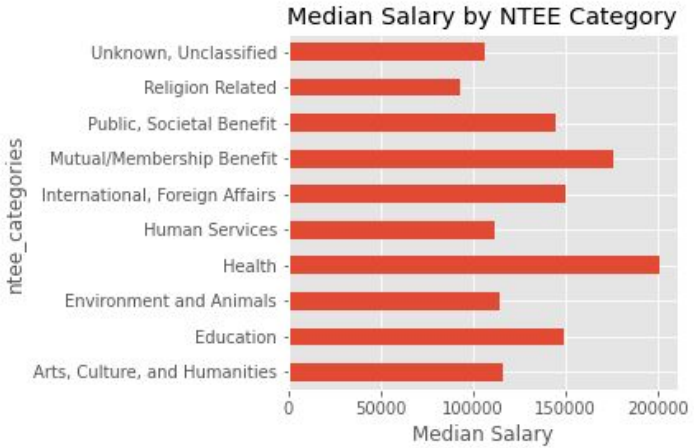


# Feature Engineering - NTEE Designation

- Extracted letter designation.  After modeling this level of granularity wasn't helpful.
- As a result, I wrote a function to consolidate the letters into 10 broad categories



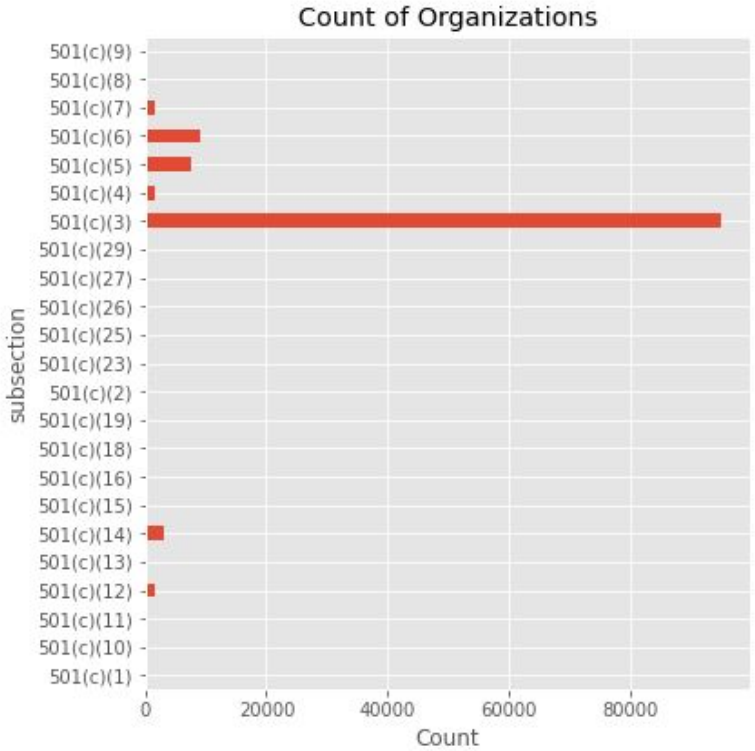



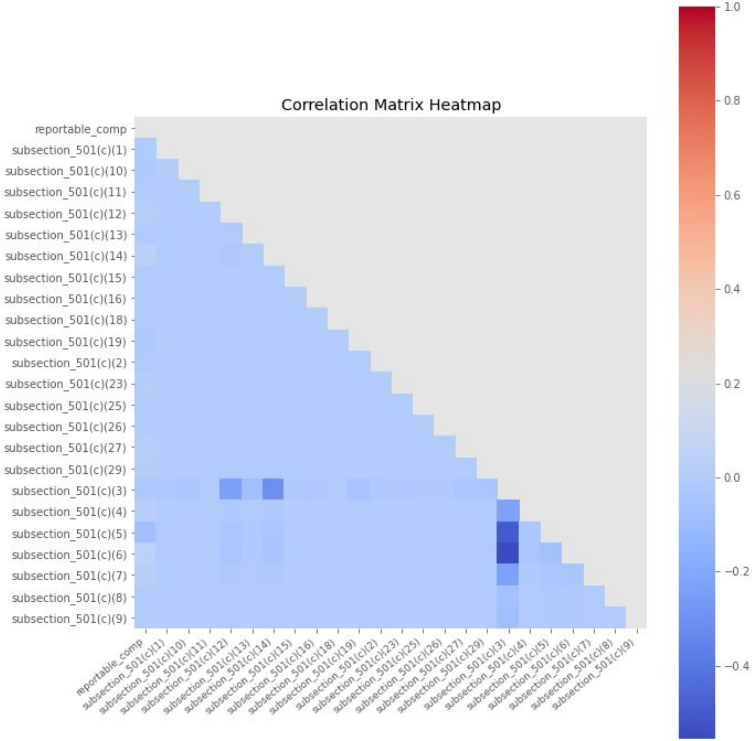
# Data Exploration - 501(c) type heavily imbalanced

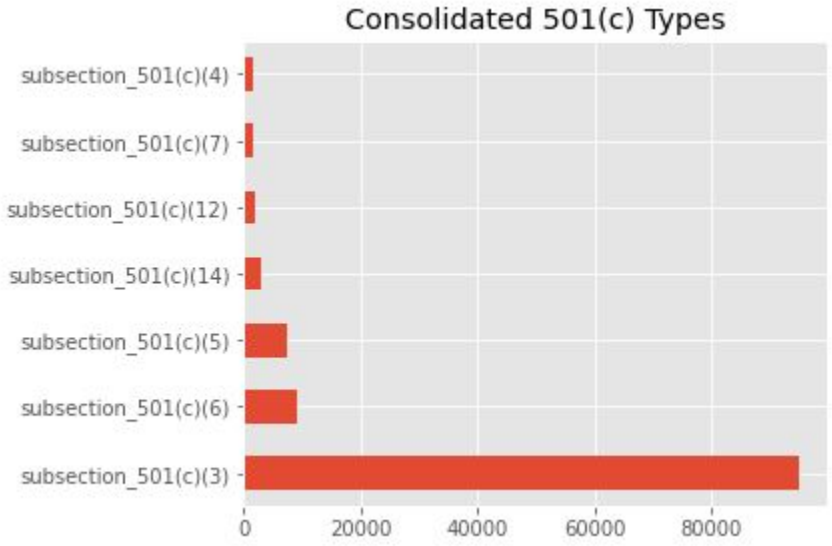
Many 501(c) types can be highly specific and sometimes obscure:

- 501(c)(6) - Business Leagues, Chambers of Commerce, Real Estate Boards, etc.  The NFL is included within this subsection.
- 501(c)(5) - Labor and Agricultural Organizations
- 501(c)(14) - Credit Unions
- 501(c)(3) - religious, charitable, scientific, literary, or educational purposes, for testing for public safety, to foster national or international amateur sports competition, for the prevention of cruelty to children, women, or animals.
- 501(c)(23) - Any association organized before 1880 for the purpose of providing insurance and other benefits to veterans or their dependents

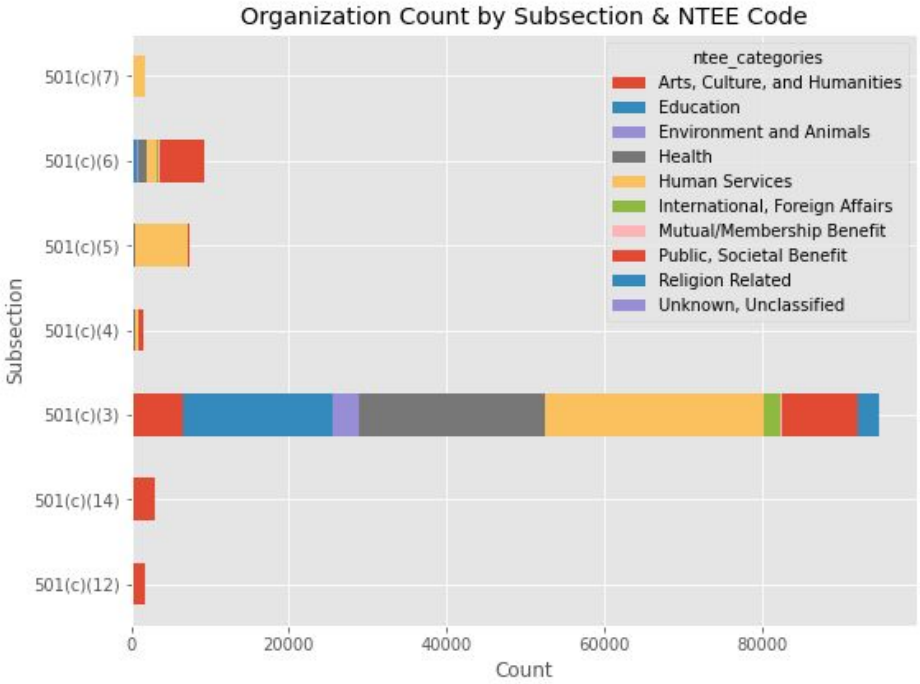


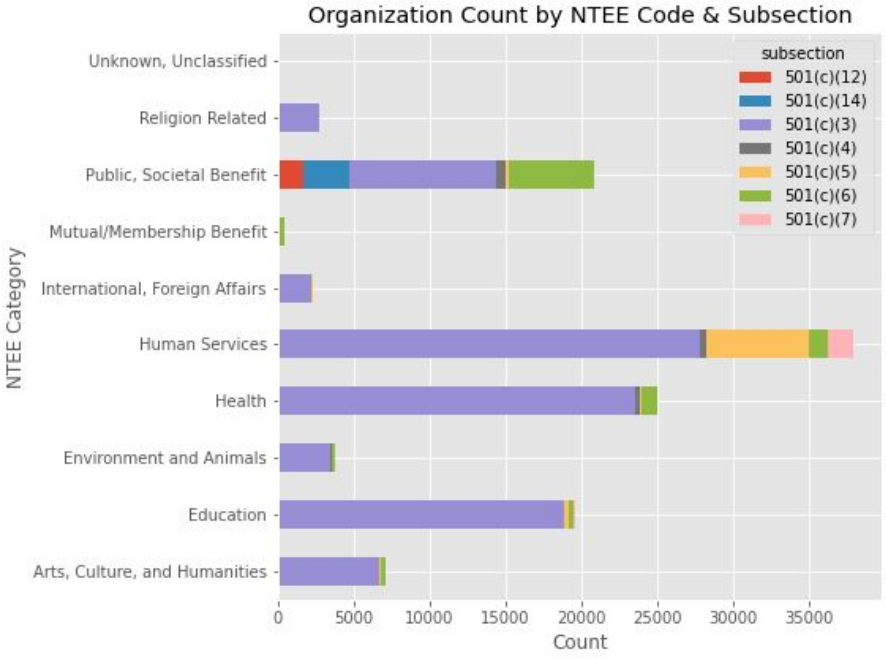
# Feature Engineering - Consolidating 501(c) Subsections



# Data Exploration - 501(c) type & NTEE Category Interaction



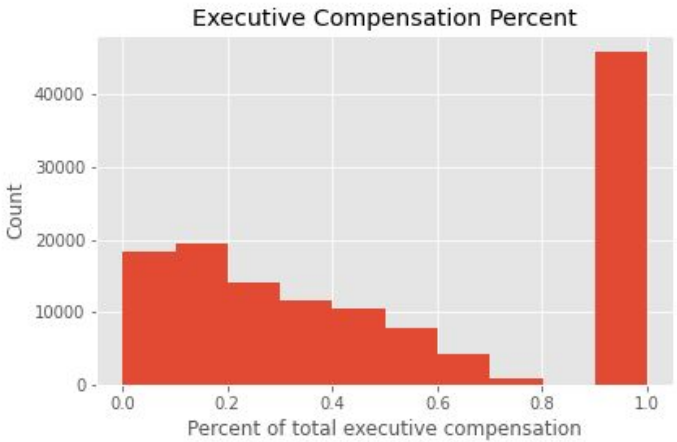



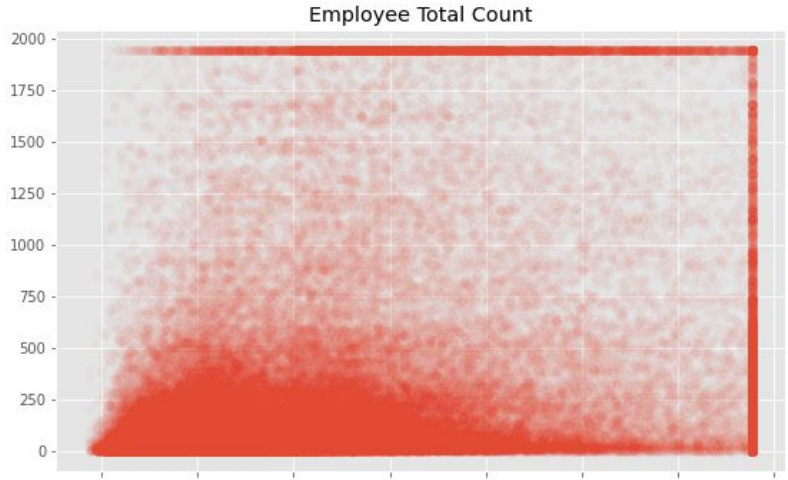
# Data Exploration: Continuous Features

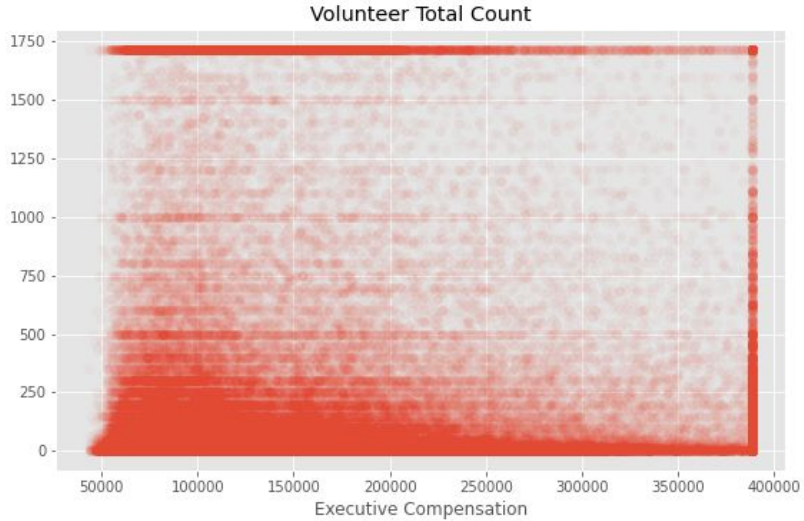
---

Almost half of all executives are the only executive at the not-for-profit.

Lower executive compensation is associated with more volunteers.



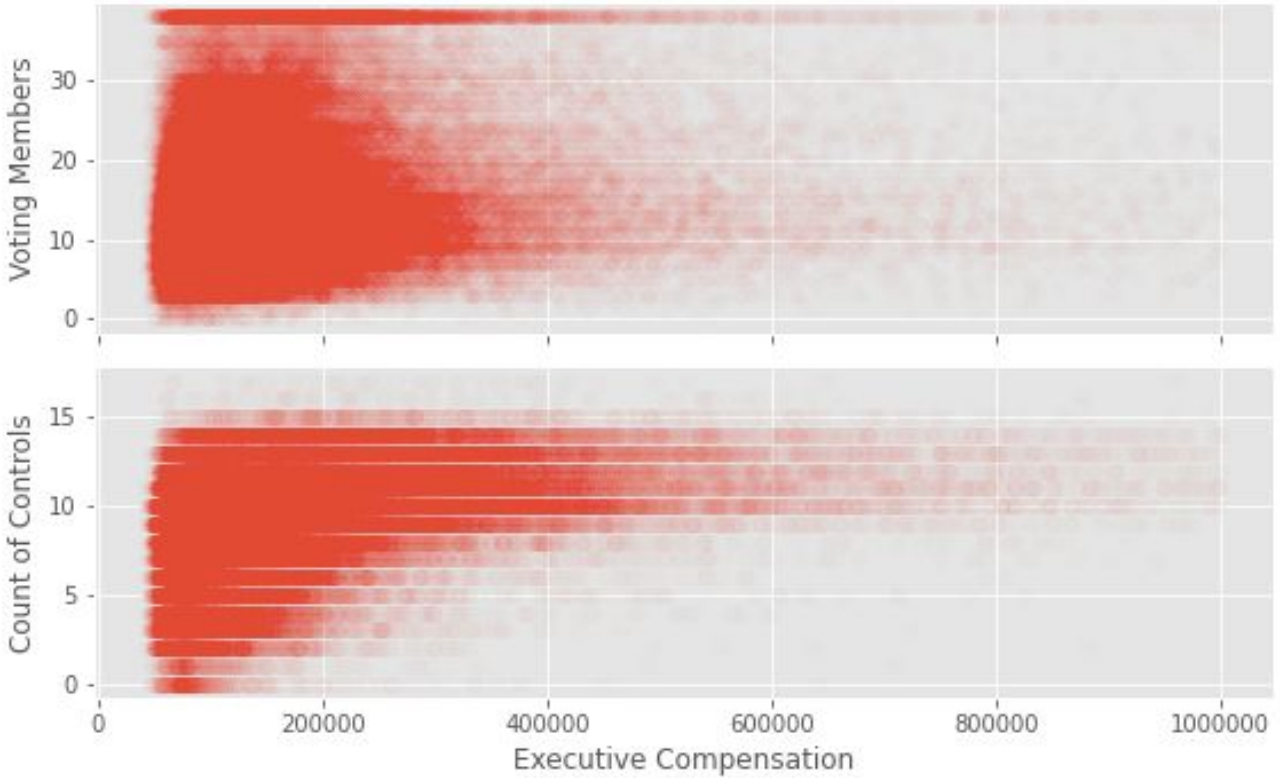



# Data Exploration:  Governance Controls

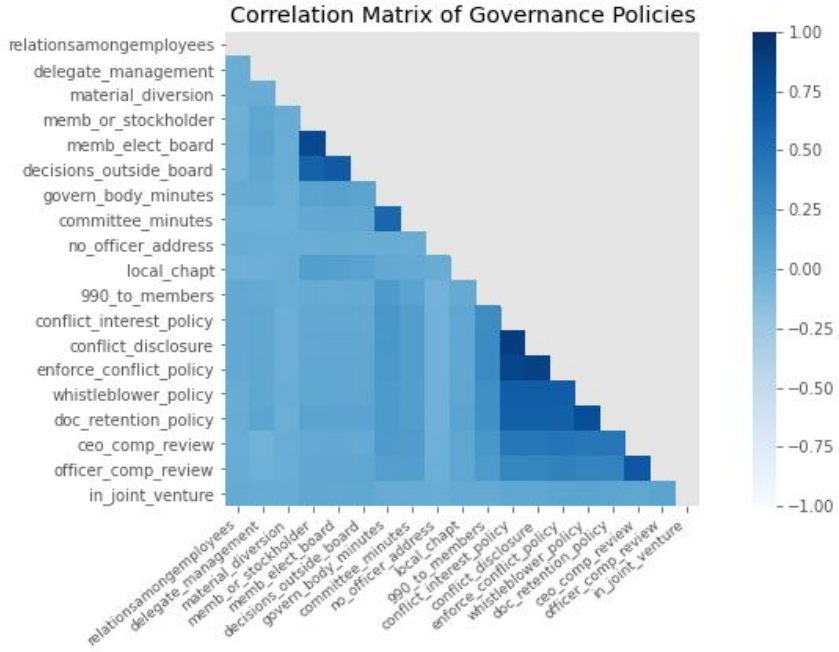
---

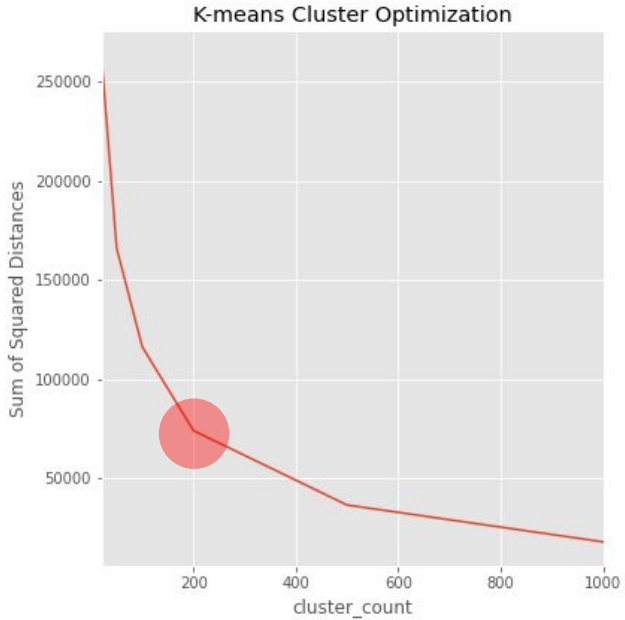
Executive compensation increases as governance controls increase.



# Feature Engineering: Governance Feature Clusters

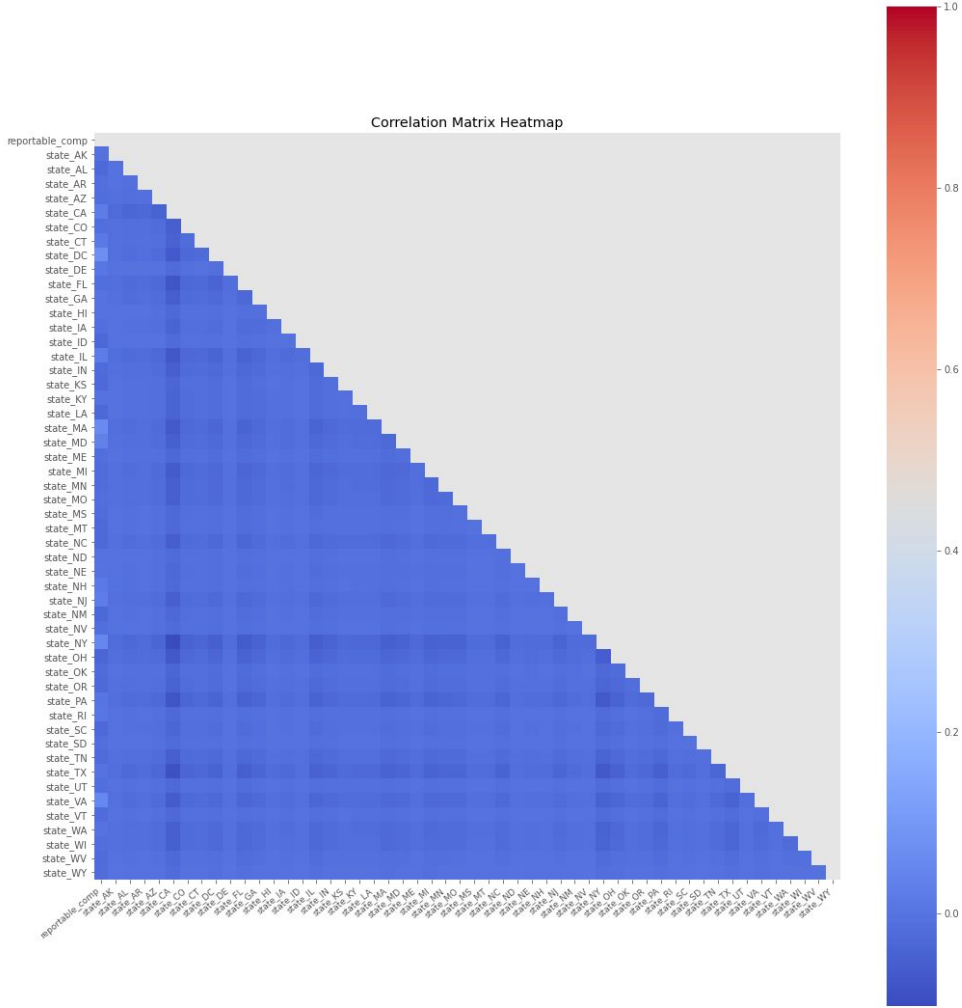
Clustering governance data into 200 clusters using K-means



# Feature Engineering: Not-for-Profit Location

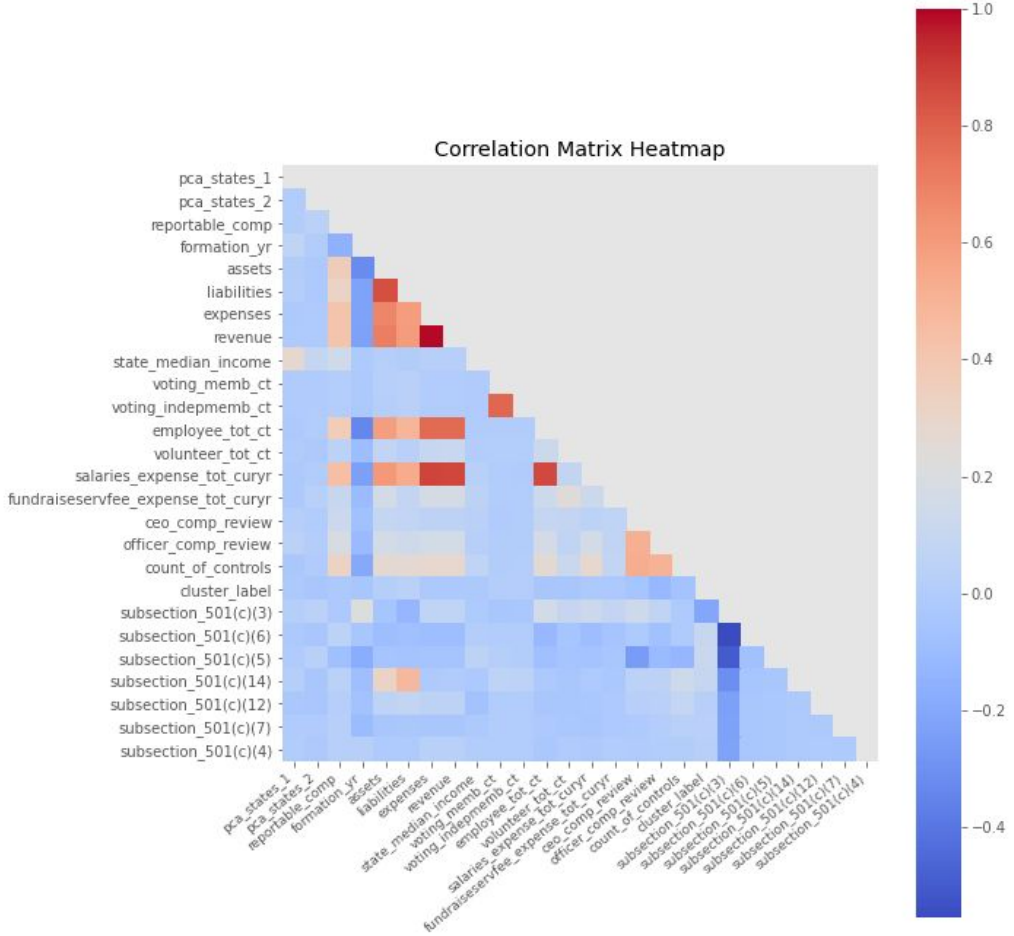
- Top correlated States consolidated into PCA features.
- Obtained median salaries by State dataset, joined that as a feature to add additional location information.



# Feature Engineering: Continuous Features

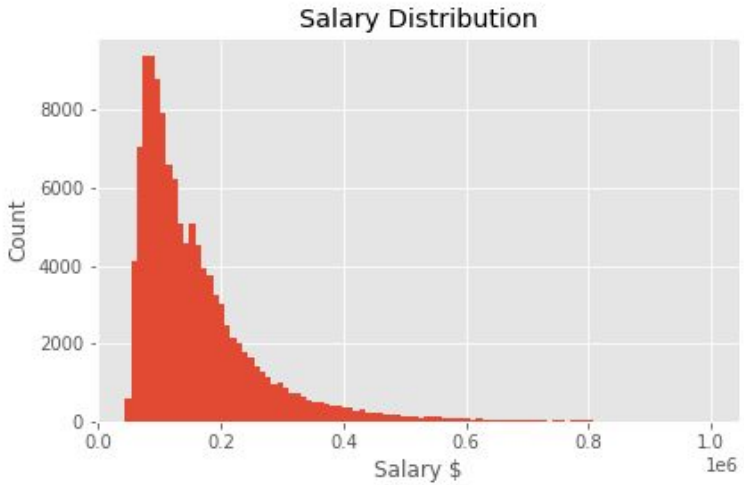
Consolidated top correlated continuous features into 3 principal components.

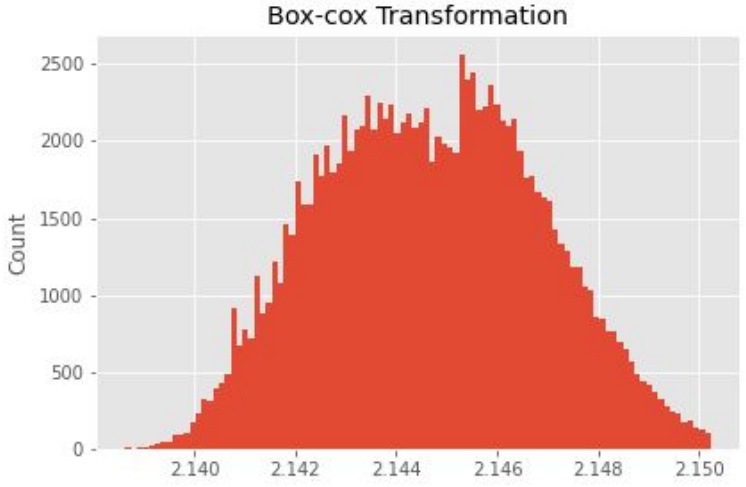
- Assets
- Liabilities
- Revenue
- Salaries expense
- Expenses
- Employee total count



# Target Category Levels

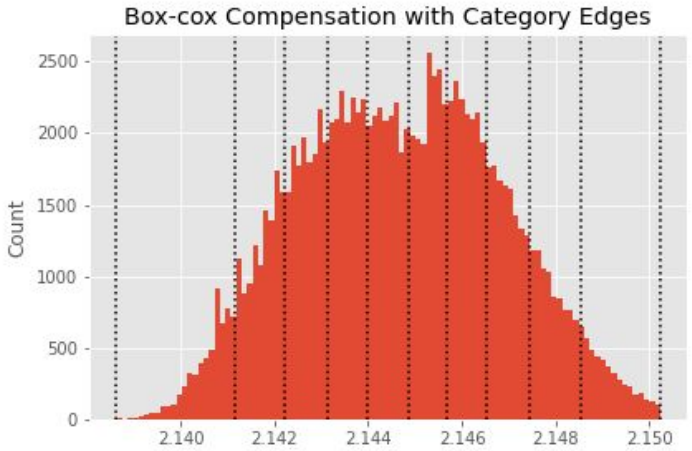
What salary levels should executives be categorized into? Because the salaries aren't normally distributed, I transformed with box-cox.

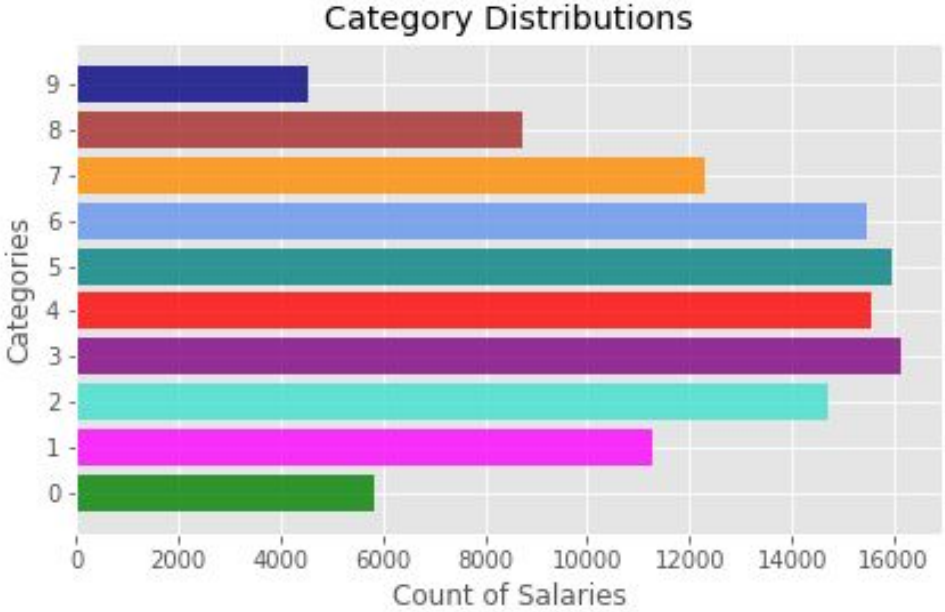


# Target Category Levels

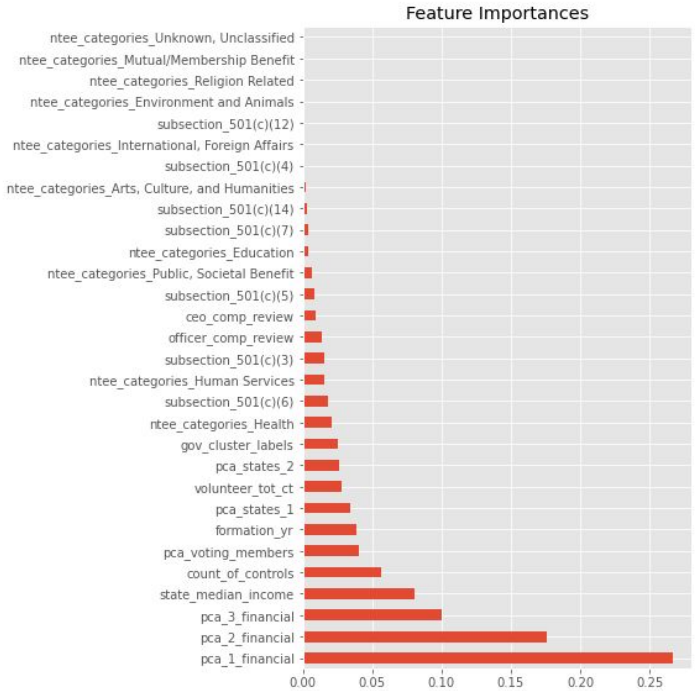
Found breakpoints using [Jenks](#) method with bins set to 10.

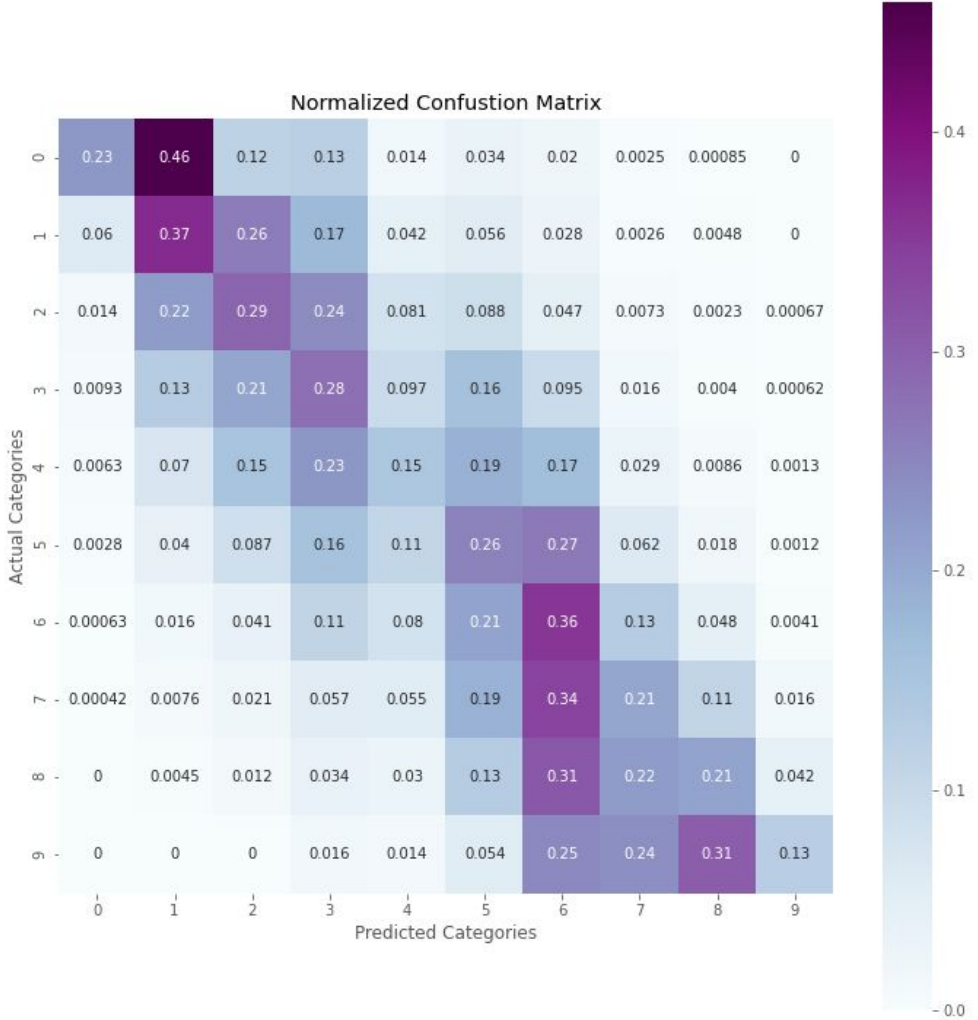



# Model 1:
# Random Forest Classifier

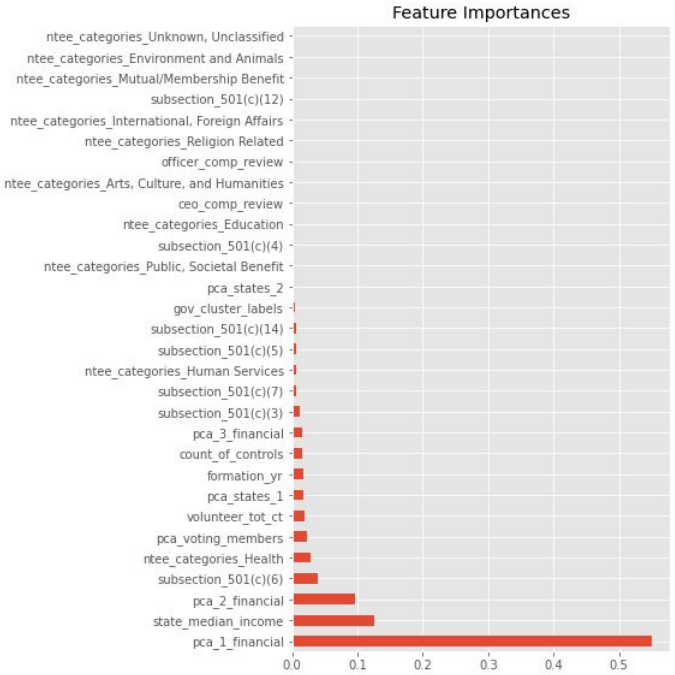
Train, Test Score: **.35, .26**

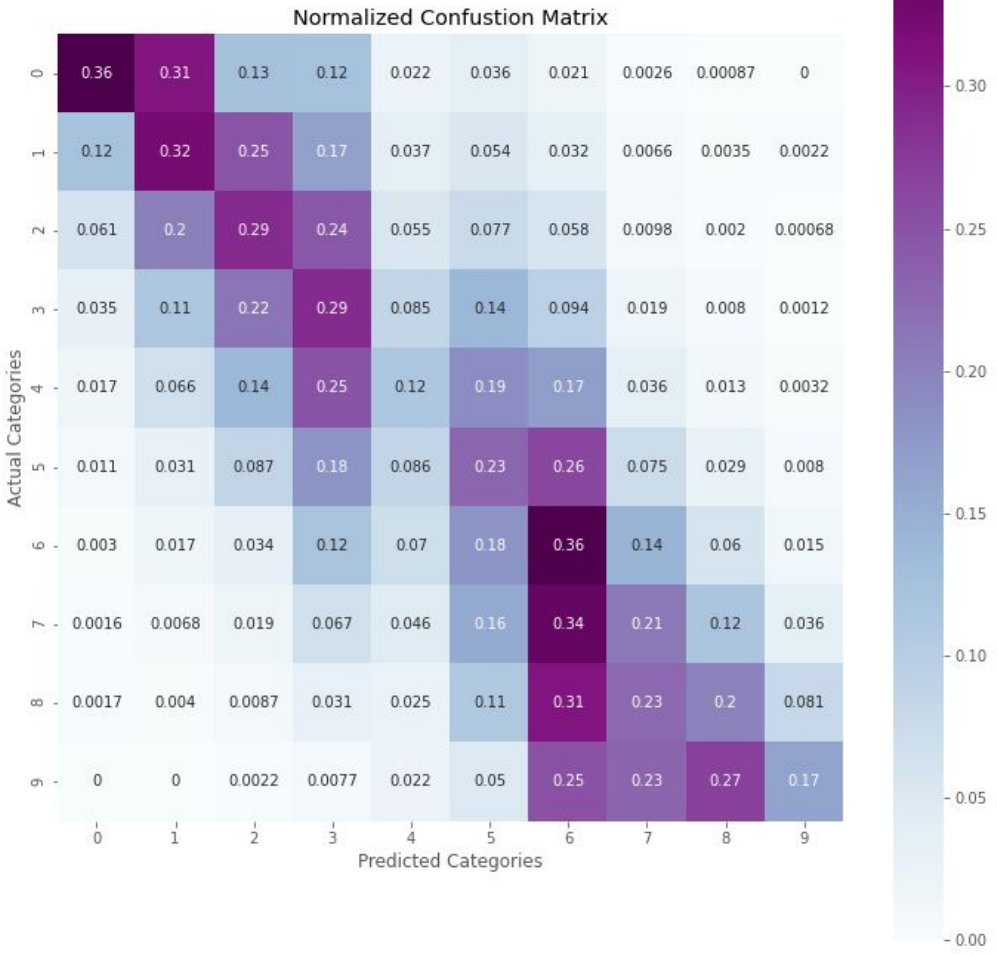


# Model 2:
# Gradient Boost Classifier

Train, Test Score: **.283, .25**

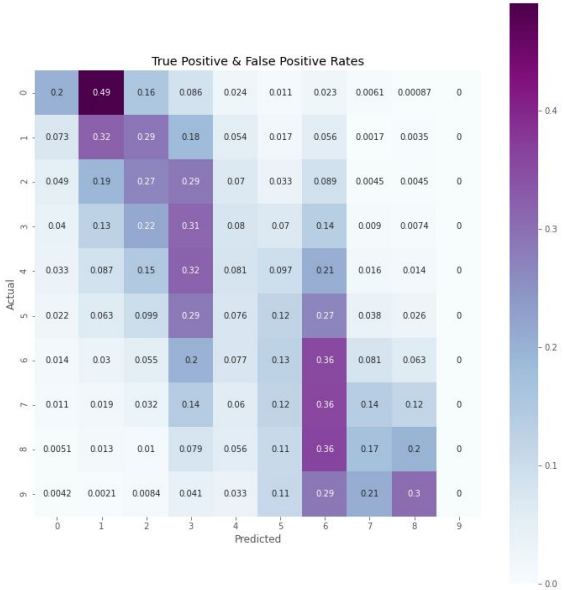


# Less Successful Models

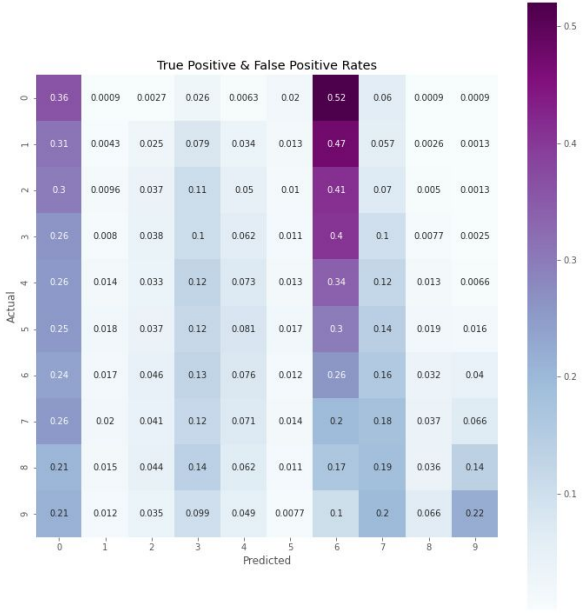
## Logistic Classifier:

Train, Test Score: **.215, .215**




## Support Vector Classifier:

Train, Test Score: **.112, .109**



# Model Application : Finding an Outlier

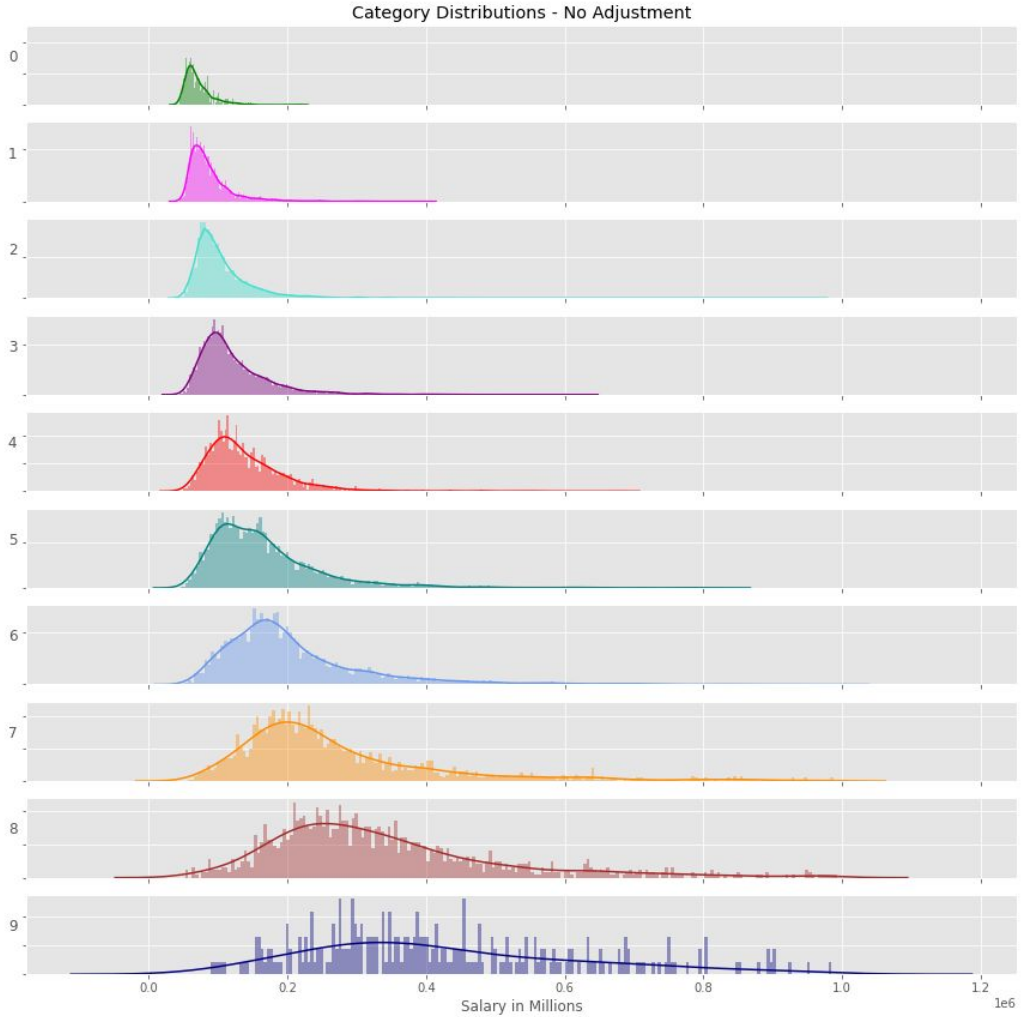
Examining the probability of prediction gives a sense of the confidence of the classifier.

As an application for **Charity Navigator**, this could be incorporated as part of its grading process. If the actual salary falls farther than two categories from that predicted by the model, this could become a trigger for further investigation of qualitative information that might justify an outlier salary. Otherwise it might result in a hit to the not-for-profit for an unreasonably compensated or under compensated executive.

# Model Application : Salary Distributions 1

The information about not-for-profits is generalized to a range of possible values for each category.  Although strong skews exist, each category has a somewhat normal distribution.

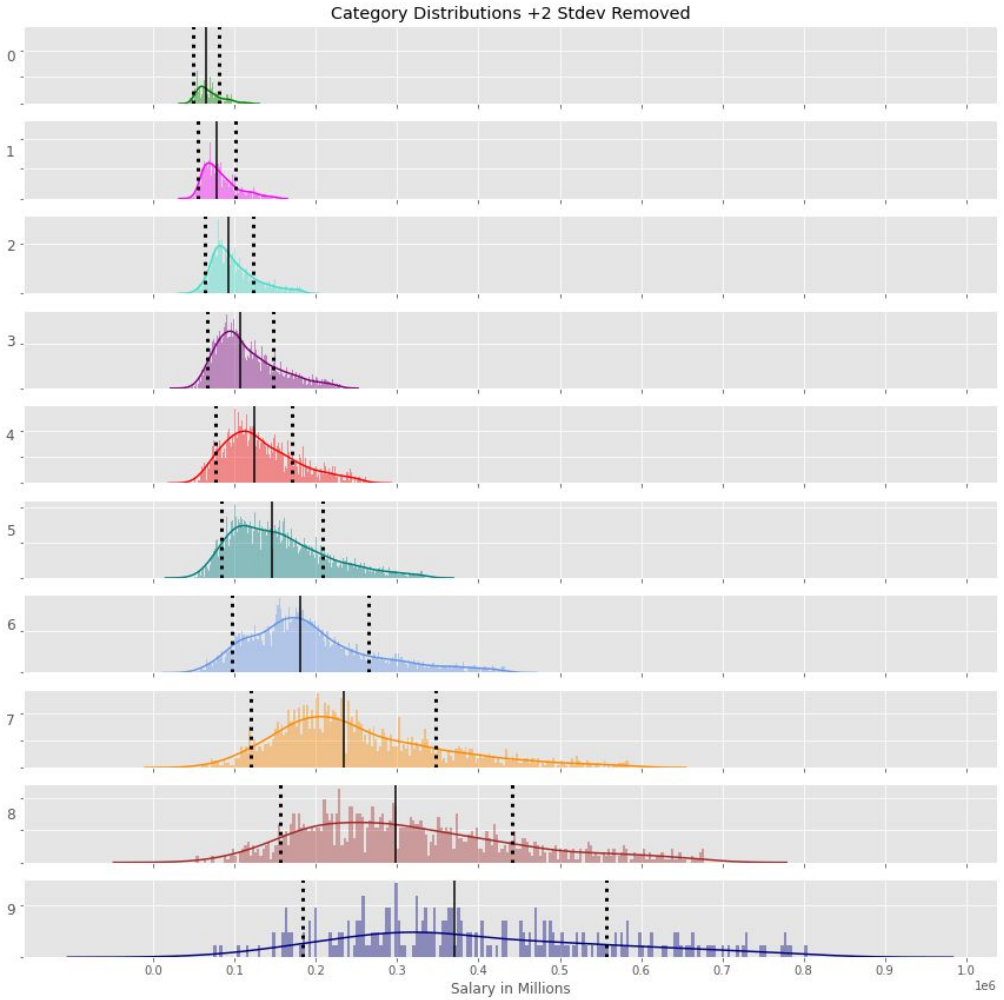
As a tool for a **Board of Directors** trying to determine what to pay an executive, the distributions can seen as ranges of possible values based on the models classifications.



# Model Application : Salary Distributions 2

To make the distributions more effective as a range of possible values:

- removed outliers above 2.5 standard deviations
- assigned median salary point for each category
- assigned standard deviation points encompassing approximately 65% of salaries for each category.



# Why Couldn’t I Get Better Results?

1. Motivations for involvement in a not-for-profit are often intangible.
    a. It represents a cause they care deeply about
    b. Ego boost is hard to quantify
2. Not-for-profits often raise a large share of their money through the connections of its top managers.
    a. The value of connections can sometimes be hard to quantify.
3. Executive title on the tax return is currently a fill in the blank field. A more specific description could reduce ambiguity.
4. Features that I thought would be highly beneficial for classification purposes proved to be minimally effective.
    a. NTEE category
    b. Governance cluster labels

# Next Steps/Other Considerations

- Quantify not-for-profits that provide the most prestige to executives. Provide a survey to not-for-profit executives asking them to rank the reasons for their involvement in that particular organization.
- Obtain data to calculate how much of the not-for-profits income is derived directly from the connections of its executives.

**Questions?**