Capstone Project

Classification of nonprofit organizations based on scores in the city of Philadelphia

By Alfredo Perez Coursera – IBM Data Science

Content

- Introduction
- Data Source and Description
- Methodology
- Analysis and Results
- Discussion
- Conclusions

Introduction - Business Problem and Background

- Individual and corporate contributions are an important source of revenues for nonprofit organizations.
- Individual trust in key when deciding where to allocate a donation.
- Misconducts and scandals of some tax-exempt organization has risen the individuals awareness on best practices.
- Relevant information is required not only by individuals but for corporations for their contributions

Introduction - Scope

The scope of this analysis will focus on nonprofit organizations with IRS status 501(c)(3) that generate at least 1\$ millions of dollars in revenue for two consecutive year in the city of Philadelphia.

Data Source and Description

- Charity Navigator API (Evaluator Company)
 - Model are based on OpenAPI format
 - Required a app_id and app_key
 - GET ORGANIZATION API was used to obtain the data for the city of Philadelphia
- IRS Tax-Exempt Organization Code Type
- GeoPy to collect geographical coordinates of each nonprofit location

Data Source and Description

- The study is based on 84 rated organizations using the following features:
 - Score: is the overall score (out of a possible 100) based on Financial Health and Accountability and Transparency, in such a way that nonprofit must excel in both areas in order to score well overall. The company calculation of the total score is:

$$100 - \sqrt{\frac{(100 - \text{Financial})^2 + (100 - \text{A&T})^2}{2}}$$

Data Source and Description

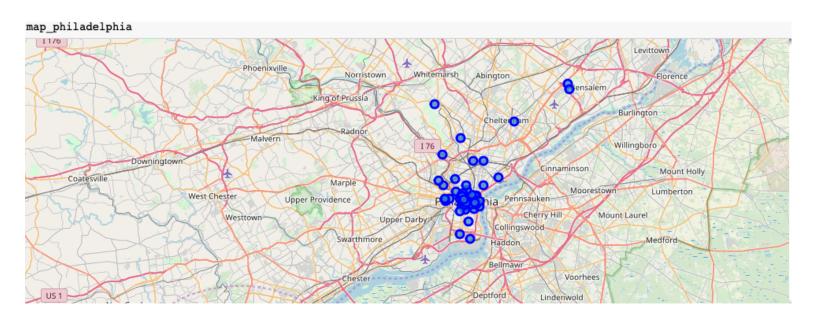
 Another file, upload it into GitHub, with the general description of all the different nonprofit IRS codes is used.

https://raw.githubusercontent.com/aperezy17/Coursera Capstone/master/irs codes.csv

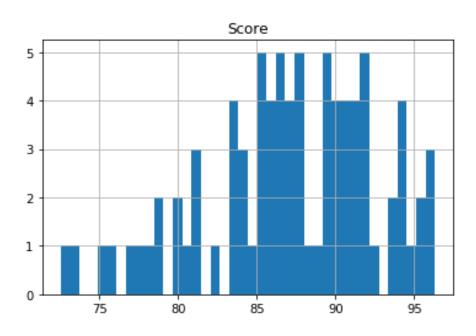
Methodology

- Charity Navigator APIs Get Organization provide reference data, covering charitable organizations, ratings, lists, and other related content.
- Nonprofit locations were mapped using GeoPy and Folium libraries.
- Two Dimension Analysis
 - Data exploration using descriptive analysis
 - K-means unsupervised method to find the different clusters of nonprofit organizations based on their Financial and Accountability scores.

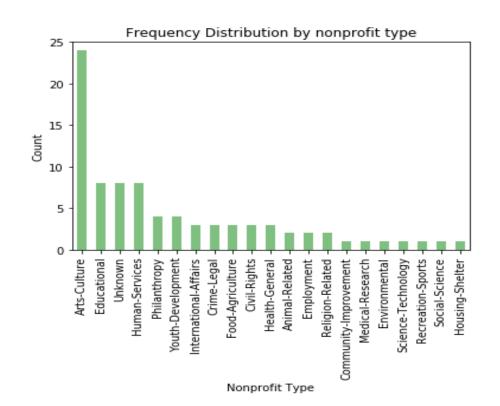
Distribution of selected nonprofit organizations in the city of Philadelphia



Histogram by Overall Score



Distribution by nonprofit type

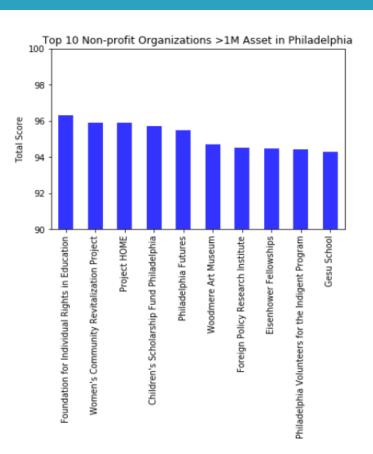


irsCode		codeCount	Description
0	Α	24	Arts-Culture
1	В	8	Educational
2	Z	8	Unknown
3	Ρ	8	Human-Services
4	Τ	4	Philanthropy
5	0	4	Youth-Development
6	Q	3	International-Affairs
7	Ι	3	Crime-Legal
8	K	3	Food-Agriculture
9	R	3	Civil-Rights
10	Е	3	Health-General
11	D	2	Animal-Related
12	J	2	Employment
13	Х	2	Religion-Related
14	S	1	Community-Improvement
15	Н	1	Medical-Research
16	С	1	Environmental
17	U	1	Science-Technology
18	Ν	1	Recreation-Sports
19	V	1	Social-Science
20	L	1	Housing-Shelter

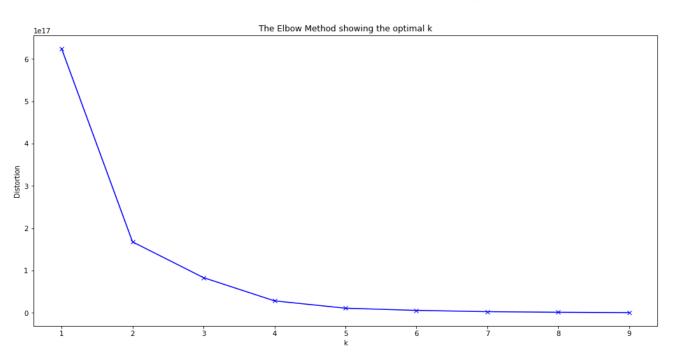
Top 10 nonprofit with higher overall scores

Top 10 nonprofit with higher overall scores							
	npOrganization Score irsCodeA						
0	Foundation for Individual Rights in Education	96.32	R				
10	Women's Community Revitalization Project	95.90	L				
59	Project HOME	95.89	P				
72	Children's Scholarship Fund Philadelphia	95.72	В				
27	Philadelphia Futures	95.47	В				
31	Woodmere Art Museum	94.69	A				
41	Foreign Policy Research Institute	94.52	Q				
36	Eisenhower Fellowships	94.49	Q				
55	Philadelphia Volunteers for the Indigent Program	94.43	I				
63	Gesu School	94.27	В				

Top 10 nonprofit organizations assets >1M



K-Means Method for Clustering



K-Means Method for Clustering

The optimal number of clusters is the value in the point which the distortion - inertia start decreasing in a linear fashion. For this data and iteration, I selected that the optimal number of clusters for the data is 3.

After partitioned the score data into mutually exclusive groups (3 clusters). I created the following profiles for each group or cluster, considering best practice by scores (Clus_Score):

O **EXCEPTIONAL** Exceeds industry standards and outperforms most nonprofit.

1 GOOD Exceeds or meets industry standards and performs as well as or better than other nonprofits.

2 **NEED IMPROVEMENT** Meets or nearly meets industry standards but underperforms most nonprofits.

K-Means Method for Clustering

The optimal number of clusters is the value in the point which the distortion - inertia start decreasing in a linear fashion. For this data and iteration, I selected that the optimal number of clusters for the data is 3.

After partitioned the score data into mutually exclusive groups (3 clusters). I created the following profiles for each group or cluster, considering best practice by scores (Clus_Score):

O **EXCEPTIONAL** Exceeds industry standards and outperforms most nonprofit.

1 GOOD Exceeds or meets industry standards and performs as well as or better than other nonprofits.

2 **NEED IMPROVEMENT** Meets or nearly meets industry standards but underperforms most nonprofits.

Discussion

- The nonprofit locations are concentrated primarily in the downtown areas. This is an important aspect due to real-state overhead needed to maintain these organizations.
- The data showed that the largest group of nonprofits organization in the city of Philadelphia are related to Arts and Culture (approx. 29%).
- The two components that forms the overall Score, Financial Score and Accountability Score, are not necessarily correlated in the cluster classification. Another analysis can be done to correlate these two attributes.

Discussion

There is lack of data of small and mid-size nonprofit organization, and this is out of the scope set for this assignment. Small size nonprofit is important on a rating system and extending the score to other qualitative data, such as individual experience could be beneficial to tune and implement a broader classification system

Discussion

Organization Breakdown on each cluster

EXCEPTIONAL	22%
GOOD	44%
NEED IMPROVEMENT	34%

Conclusions

- Nonprofit organizations are tax-exempt and requires to follow a strict code of ethics and best practices for managing and maintaining their exempt status.
- Nonprofit revenues largely depend on individuals and corporate contributions; thus, this money have to be allocated to the specific programs they served. Due to recent scandals and misconducts, donors need to have a reliable source of information at the time of deciding their contributions.

Conclusions

• In this study we achieved, using a unsupervised K-Means clustering method), a basic classification system for nonprofits with rating (revenues greater than 1M for two consecutive years) from ranking data provided recognized evaluator organization in the city of Philadelphia. Also, a series of reports were generated to add more information to potential donors.

References

- BRAY, Illona: Effective Fundraising for Nonprofits. Nolo. 2010
- DeWITT, Brydon: The Nonprofit Development Companion. John Wiley
 & Sons, Inc., Hoboken. 2010
- Charity Navigator Blogs and API Documentation