Capstone Project

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

Alfredo Perez
Coursera – IBM Data Science

1. Introduction

Business Problem, Scope and Background

The majority of nonprofit organizations rely upon donors as an important source of their funding structure, according to some surveys an average of 34 percent of these organization's revenues come from individual's contributions. In recent years several scandals involving nonprofit mishandling of funds and other misconducts, have impacted the philanthropy domain and individual's faith on how their contributions are being used.

The goal and interest of this study is to explore and provide individuals and potential donors with the necessary information to make educated decisions for their contributions to major nonprofits organizations located in the city of Philadelphia. The information is based on different types of scores processed by an evaluator corporation in this field.

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.

Considering this problem, I will group the organizations, subject to this study, and provide relevant information and graphs of the those that meet best practices both in the financial and accountability areas, thus donors can have more information at the time of planning their donations.

2. Data Source and Description

For this analysis the data was obtained from an organization called Charity Data Navigator, an organization that collects financial and other information for nonprofits organizations. This organization provides APIs, similar as Four Square, to get reference data of nonprofit and charities organization including rating list and other related content.

After executing the API "Get Organizations", 8,713 nonprofits were found in the city of Philadelphia and 84 candidates (revenues of at least 1$ Million for two consecutive year) have registered scores and reviews.

The study is based on these 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:

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

- The individual Financial and Accountability scores are used to slice some relevant information in these two dimensions.

- Also, location coordinates of each organizations were obtained using GeoPy to show the geographical distribution throughout the city of Philadelphia.

- To create some frequency charts, I used another file that I upload it into GitHub with the general description of all the different nonprofit IRS codes. This file is located in the following link:

https://raw.githubusercontent.com/aperezy17/Coursera_Capstone/master/irs_codes.csv

3. Methodology

I used the Charity Navigator APIs to get the research and reference data, covering charitable organizations, ratings, lists, and other related content.

Charity Navigator's rating system examines two broad areas of a charity's performance; their Financial Health and their Accountability & Transparency

The data models of these APIs are based in the industry standard OpenAPI format for interactive documentation and overall interoperability. I created a free account and defined an application (**app_id**) and a key (**app_key**).

I utilized the GET Organizations API to retrieve a list of the organizations available in the Charity Navigator Data Store and it is the main source of data for this study. As I mentioned in the previous paragraph, the candidate for the study are 84 nonprofit organizations in the city of Philadelphia with revenues higher to 1$ Million in two consecutive years.

I also used GeoPy and Folium libraries to obtain geographical locations (latitude and longitude) of these 85 candidates and plot them on a city map. This will help individuals and donors in their analysis process.

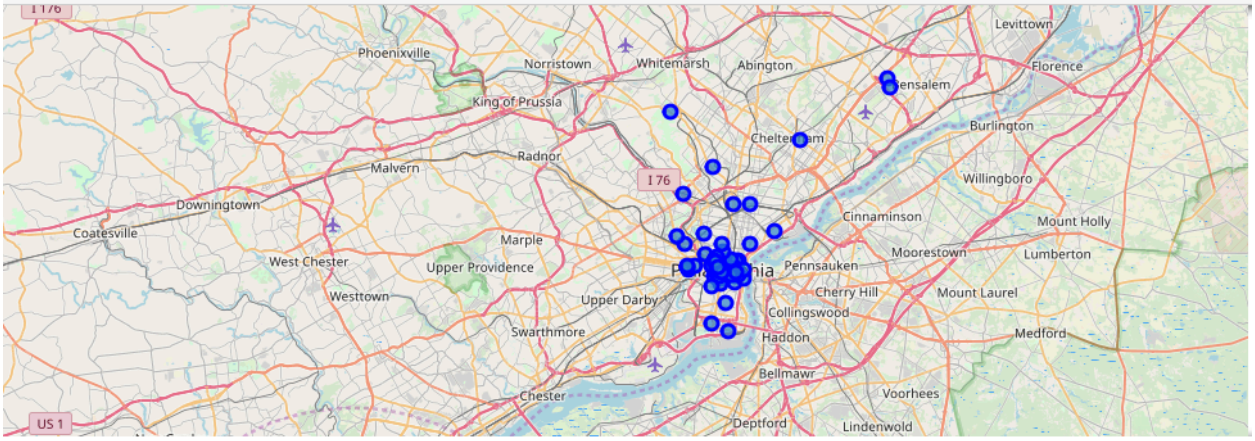
I break my analysis in two different segments. The first one is an exploratory data using descriptive analysis and the second part I will use the unsupervised learning method K-means to find the different clusters of nonprofit organizations based on their Financial and Accountability scores.

4. Analysis and Results

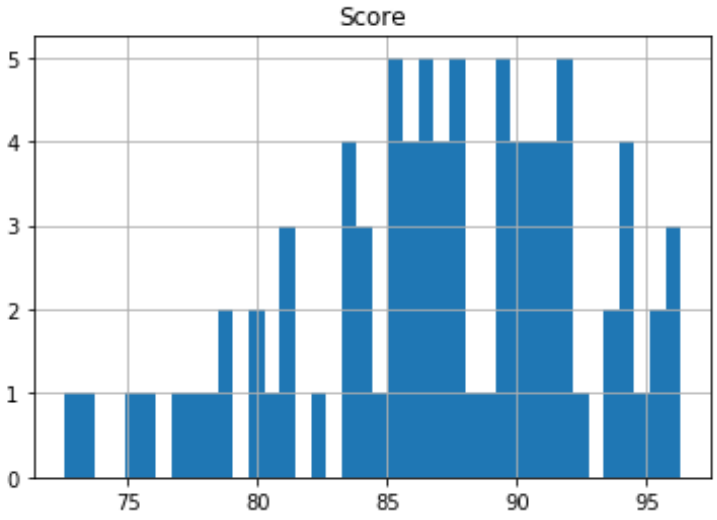
4.1. Data Exploration and Descriptive Analysis

4.1.1. Distribution of selected nonprofit organizations in the city of Philadelphia

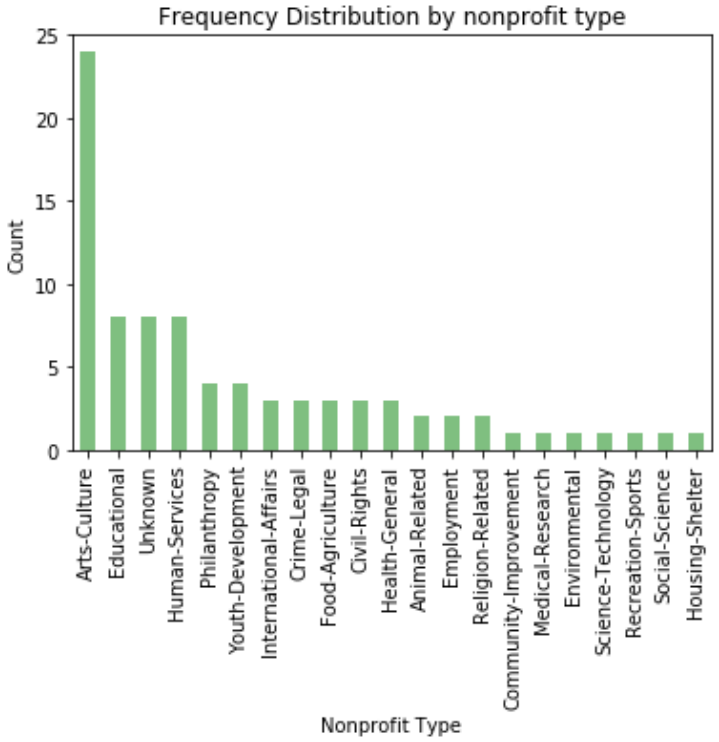
```
map_philadelphia
```



4.1.2. Histogram by Overall Score



4.1.3. Distribution by nonprofit type

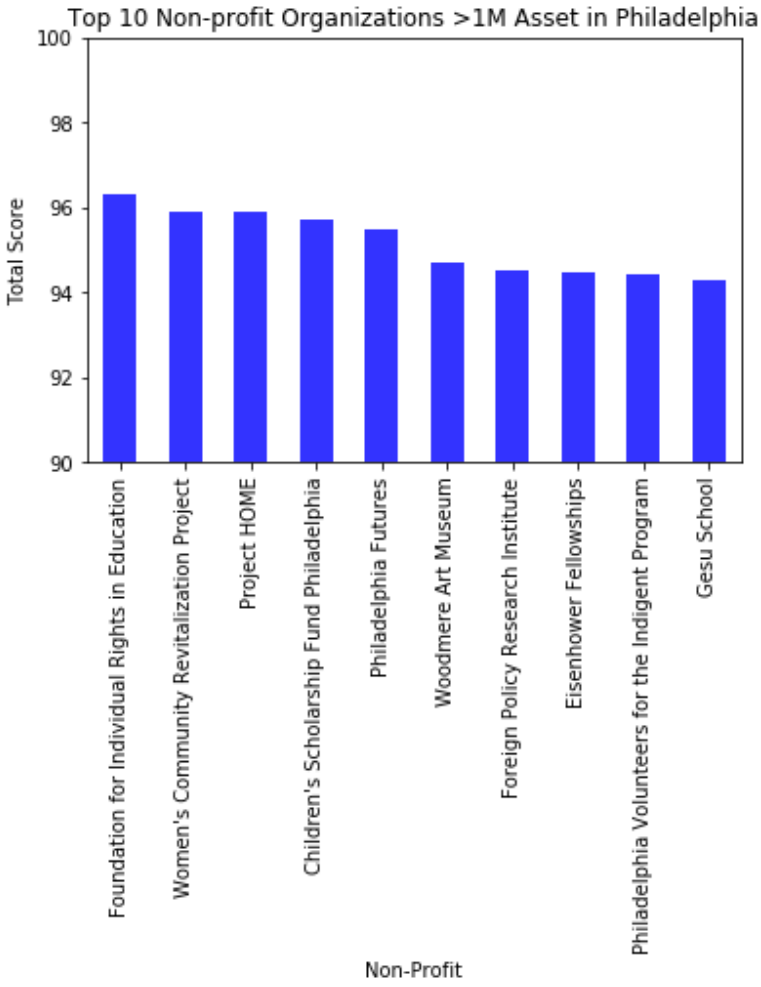


	irsCode	codeCount	Description
0	A	24	Arts-Culture
1	B	8	Educational
2	Z	8	Unknown
3	P	8	Human-Services
4	T	4	Philanthropy
5	O	4	Youth-Development
6	Q	3	International-Affairs
7	I	3	Crime-Legal
8	K	3	Food-Agriculture
9	R	3	Civil-Rights
10	E	3	Health-General
11	D	2	Animal-Related
12	J	2	Employment
13	X	2	Religion-Related
14	S	1	Community-Improvement
15	H	1	Medical-Research
16	C	1	Environmental
17	U	1	Science-Technology
18	N	1	Recreation-Sports
19	V	1	Social-Science
20	L	1	Housing-Shelter

4.1.4. Top 10 nonprofit with higher overall scores

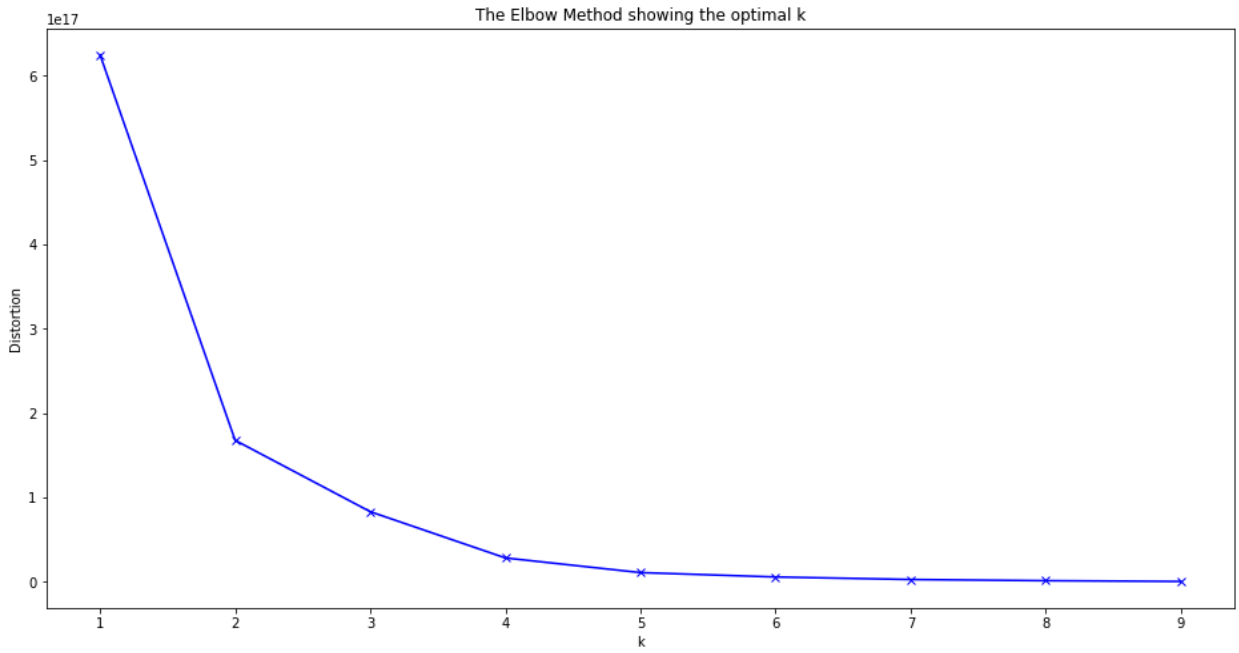
Top 10 nonprofit with higher overall scores			
	npOrganization	Score	irsCodeA
0	Foundation for Individual Rights in Education	96.32	R
10	Women's Community Revitalization Project	95.90	L
59	Project HOME	95.89	P
72	Children's Scholarship Fund Philadelphia	95.72	B
27	Philadelphia Futures	95.47	B
31	Woodmere 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	B

4.1.5. Top 10 nonprofit organizations assets >1$ M



4.2. K-Means Method for Clustering

I used this unsupervised method to model the clusters. The first step was to use the Elbow-Method to get the optimal number of cluster.



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)**:

0 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.

Centroid values in each Cluster

Clus_Score	Score	Rating	finScore	finRating	acctScore	acctRating
0	91.753684	3.763158	90.270000	3.552632	95.526316	3.842105
1	80.372632	2.526316	75.041053	1.947368	89.526316	3.473684
2	85.252963	2.925926	88.167778	3.296296	83.962963	2.814815

Note: in the companion notebook you can view the reports and totals for each cluster.

5. Discussion

This basic classification system together with the map location and reports, provides some insights to potential donor and individuals about the selected nonprofit organization in the city of Philadelphia. Following some discussion bullets:

- 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.
- 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.
- Following the percentual breakdown of organizations on each cluster:

EXCEPTIONAL	22%
GOOD	44%
NEED IMPROVEMENT	34%

6. Conclusion

The philanthropy world is very complex and dynamic. There many reasons why a nonprofit organization exist but one common purpose is to solve a specific community needs with product and services. These 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.

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.

7. 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