ANALYZEUP: AN INVESTIGATION INTO PHILANTHROPY

Aimee Decoste, Hankai Zhu, Sailee Joshi, Kylie Kwan

The Philanthropic Ecosystem

According to the most recent data available, there are more than 1.54 million charitable organizations in the United States.

In 2020, 80% of charitable dollars went to religion, education, human services, or public society benefit.

Americans gave $484.85 billion in 2021 - a 4% increase from 2020.

Why Investigate Nonprofits?

"I consider myself a philanthropist."

Total	Baby Boomers	Gen X	Millennials
45%	35%	48%	74%

Verifying Charities

- As the internet has enabled nonprofits to magnify their reach, donors can connect with impactful charities globally and make a larger difference.

- A much higher portion of millennial and gen Z adults consider themselves philanthropists than gen X and boomer adults.

- Trustworthy charities should be able to easily reach donors and continue impacting lives.

Enter Charity Navigator

Charity Navigator ranks organizations in the following aspects:
- Financial health (from IRS 990 forms)
Accountability & transparency (from charity's webpage & communications)

The ranking is converted into an overall score, then a star rating for potential donors.

How Is the Overall Score and Star Rating Calculated?

In our two-dimensional rating system, the overall score is not a sum but rather a measurement of the distance of two component scores from a perfect score of 100 and 100. The smaller the distance to the perfect score, the better the overall score.

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

Although the calculation for Overall Score differs from the calculation for Financial Health and Accountability & Transparency, the Overall Rating has the same rating table.

Overall Rating:	★★★★	★★★	★★	★	0 Stars	Donor Advisory
Overall Score:	≥ 90	80 - 90	70 - 80	55 - 70	< 55	N/A

However, thousands of impactful charities have outdated ratings or aren't rated at all. With Charity Navigator's database, we can change that.

Our Objectives

1. To build a model that classifies the efficacy charities based financial information.

2. To provide the most up to date rating information on US charities.

3. To provide an end user experience that can be used by giving entities (ie. donors)

Data Exploration

ETL of Charity Navigator Data

As a best practice, the data was cleaned to remove null values and duplicate entries.

The charity_name dataframe was then created to be reindexed on the modeled dataframe.

Finally the data dimensions were reduced using PCA.

This ensured we were working with clean data that is optimized for any data analysis or model fitting.

Selecting a Model

ETL of Charity Navigator Data

After researching various ML models, we chose to use an Unsupervised model.

After performing analysis using Kmeans clustering and Hierarchical clustering models, we decided that Kmeans was the ideal model as it resulted in a 90% accuracy rate and also has use cases of customer segmentation and recommendation systems. Given the above, Kmeans aligns with the scope and goal of our project.
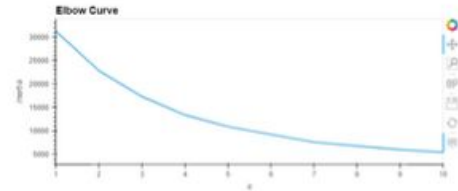
Our Analysis

PCA Reduction + K-Means Clustering

We used Principle Component Analysis (PCA) to reduce the dimensions of our dataset before ultimately running K-Means clustering on the data.

We reduced to 3 principle components, as this improves the clustering results due to noise reduction, as well as decreasing computation cost.

For our K-Means clustering, we iterated from K=1 -> K=10 to find the optimal number of clusters. We found it was 6, as there was an elbow at that value when we graphed the inertias for each K-value.

Finally, we evaluated the model using a Silhouette Score. We obtained a ~91.3% accuracy score, indicating a successful analysis.

aluating the Model

```
t the model's clustering performance with Silhouette Coefficient evaluation
s_model = KMeans(n_clusters=4, random_state=1).fit(X)
ls = kmeans_model.labels_
ics.silhouette_score(X, labels, metric='euclidean')
```
6127602592019

urce: https://scikit-learn.org/stable/modules/clustering.html#clustering-evaluation 2.3.10.5. Silhouette Coefficient

Sources

https://www.artemissg.com/2021/05/future-of-philanthropy-fidelity-charitable/

https://www.nptrust.org/philanthropic-resources/charitable-giving-statistics/

https://www.irs.gov/newsroom/irs-joins-international-organizations-in-fighting-charity-fraud-during-special-awareness-week

https://www.charitynavigator.org/index.cfm?bay=content.view&cpid=5593#rating