



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."



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.

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

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

Overall Rating:	★★★★	★★★☆☆	★★★☆☆	★★☆☆☆	★☆☆☆☆	0 Stars	Score Advisory
Overall Score:	≥ 90	80 - 90	70 - 80	60 - 70	≤ 50	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However, thousands of impactful charities have outdated ratings or aren't rated at all.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

```
# Drop columns not needed
new_df_201412_20 = df_201412_20.drop(['Phone_Number', 'Fax_Number', 'TollFree', '1800_1800', '1800_1800'], axis=1)
new_df_201412_20.head()
```

	OrgID	Charity Name	Address	City	State	Zip	County	Website URL	Category	Cause	Tag Line	Mission	Total Revenue	Total Expense	
0	10279	United Methodist Committee on Relief of Atlanta	400 Peach Street	Atlanta	GA	30308-0	Fulton	http://www.unmc.org/	International	Development and Relief Services	Be there. Be hope.	Founded in 1942, the United Methodist Committee...	479878546	71190138	
1	11071	"Theresa Dejeu" Foundation	322 Eighth Street Suite 200	Roseburg	OR	97470-0	Van Hook	http://www.theresadejeufoundation.org	Education	Scholarship and Financial Support		Empowering children in low-income communities	Founded in 1988 by Therese Dejeu, M.Ed., M.P.S.	19730310	2917790
2	9488	1000 Lim College	1000 Lim College Drive Suite 110	San Rafael	CA	94903-0	Marin	http://www.1000limcollege.org	Education	Scholarship and Financial Support	Creating College Graduates	A leading college access nonprofit in California	125421610	8258022	
3	12096	100 Back in America	100 Back in America Ave	Atlanta	GA	30303-0	Fulton	http://www.100backinamerica.org/	Human Services	Health Development, Shelter and Crisis Services	Improving the quality of life for our community	Founded in 1991, the mission of 100 Back in America...	38898640	4088870	
4	10122	100 Club of America	310 South 48th Street	Phoenix	AZ	85008-0	Maricopa	http://www.100club.org	Human Services	Multipurpose Human Services	Supporting families of...	Officially organized in 1988 as the	84021000	2201740	

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.

Clustering Charity Data Using K-Means

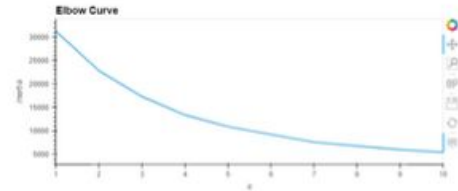
Finding the Best Value for k Using the Elbow Curve

```
In [19]: # Create an elbow curve to find the best value for K.
inertia = []
k = list(range(1, 11))

# Calculate the inertia for the range of K values
for i in k:
    km = KMeans(n_clusters=i, random_state=0)
    km.fit(pcs_df)
    inertia.append(km.inertia_)

# Create the elbow curve
elbow_data = [{"k": k, "inertia": inertia}]
df_elbow = pd.DataFrame(elbow_data)
df_elbow.plot.line(x="k", y="inertia", xticks=k, title="Elbow Curve")
```

Out[19]:



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 $K=1 \rightarrow K=10$ 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.

Evaluating the Model

```
# Evaluate the model's clustering performance with Silhouette Coefficient evaluation
k_model = KMeans(n_clusters=6, random_state=1).fit(X)
is = k_model.labels_
sco = silhouette_score(X, is, metric='euclidean')
```

0.9127682592019

Source: <https://scikit-learn.org/stable/modules/clustering.html#clustering-evaluation> 2.2.10.5. 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>