



ANALYZEUP



ANALYZEUP: AN INVESTIGATION INTO PHILANTHROPY

Sailee Joshi - *Master of Ceremonies*, Aimee Decoste - *Model Designer*, Hankai Zhu - *Database Engineer*,
Kylie Kwan - *Dashboard/ UI Manager*

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.

The Nonprofit Problem

- Charities are rated based on:
- Financial health (from IRS 990 forms)
 - Accountability & transparency (from charity's self-reported financial information)

However, a lag of 2-5 years is fairly typical.

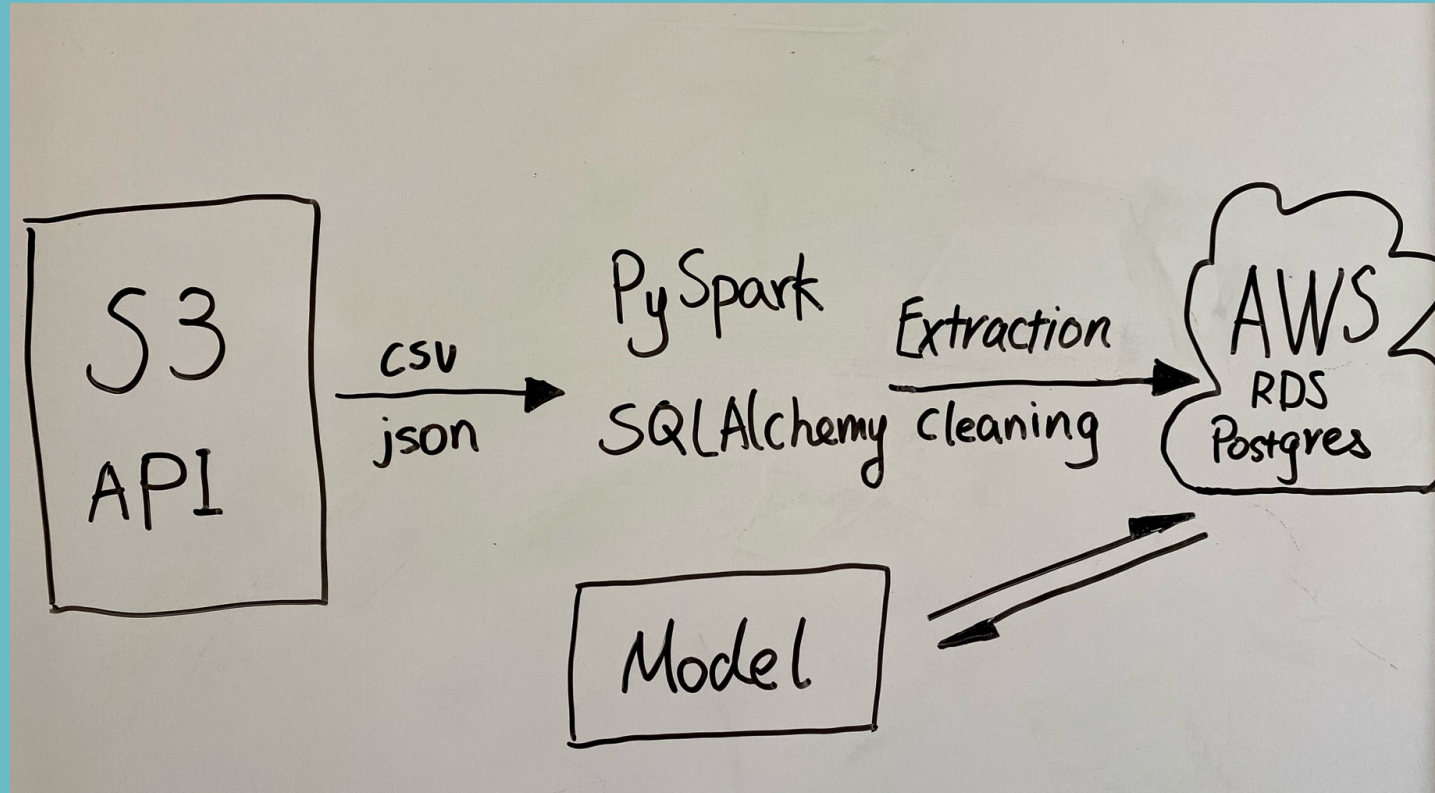
By analyzing data provided by Charity Navigator, we hope to provide updated rankings and help nonprofits maximize their impact.

Our Objectives

- 1 To build a model that classifies the efficacy of charities based on financial information.
- 2 To provide the most updated ratings on US charities
- 3 To provide an end-user experience that can be used by donors

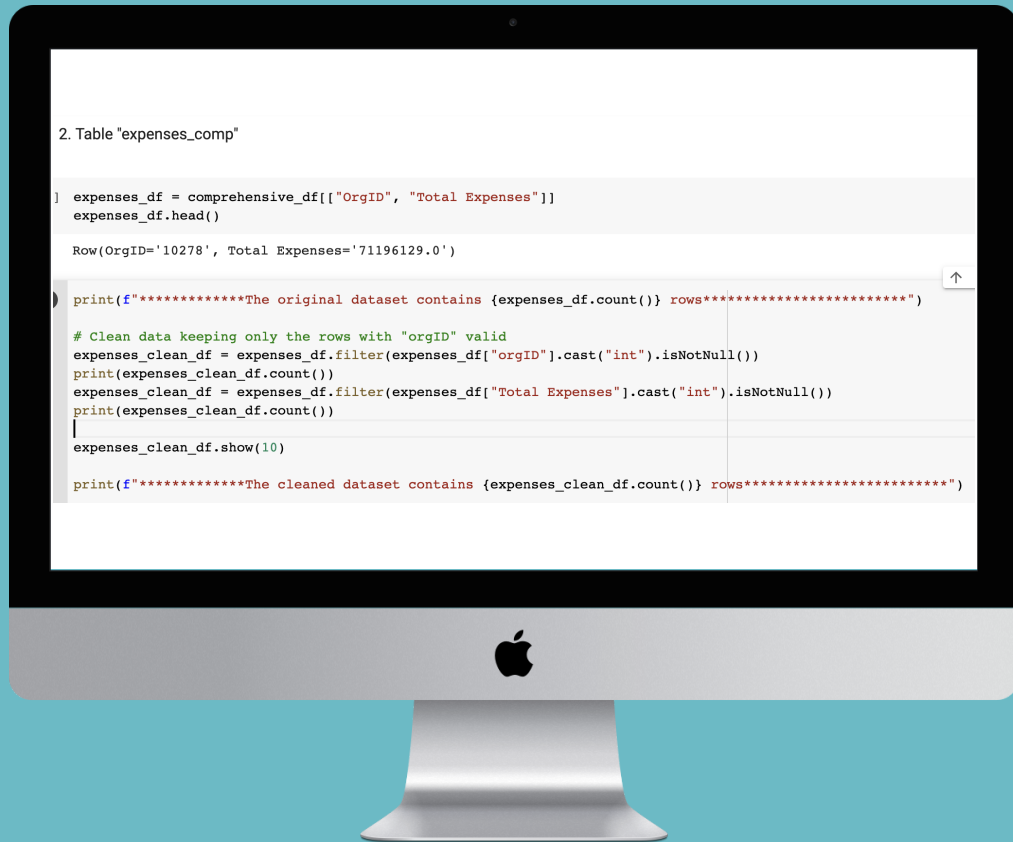


Data Flow



- 1 **AWS**
- 2 **Postgres**
- 3 **Database Tools**
- 4 **Troubleshooting**

Data Extraction, Cleaning and Transformation



- API vs Comprehensive Dataset
- Bad Line Data
 - * Unrecognized Data
 - * NaN Data
 - * Incorrect Data Type
- Selected Features
- Created Tables

Prepare Tables for Modeling

Stored Info in the Database

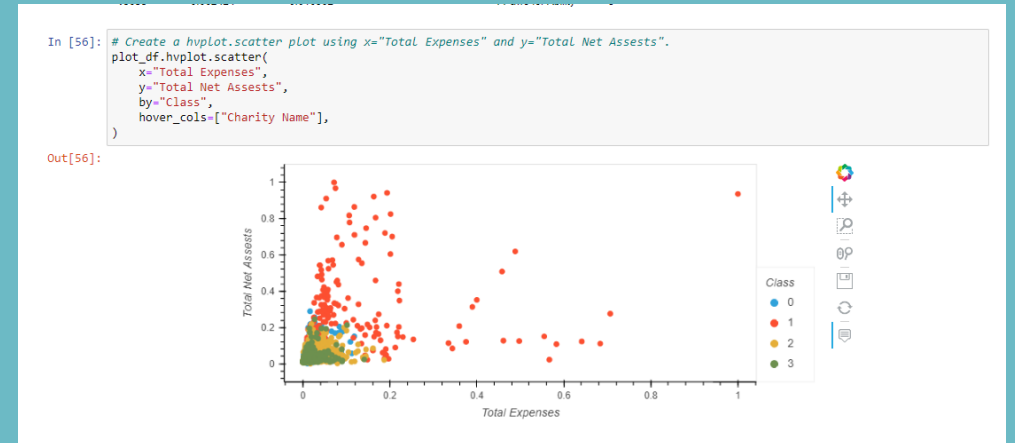
- 1) API extracting data table
- 2) Comprehensive table
- 3) Working table including total charity expenses and total charity assets
- 4) Location of the charity
- 5) Other pertinent background information

id	charity_name		cause	total_expenses	total_net_assets
10278	United Methodist Committee on Relief of Globa...	Development and Relief Services		71196129.0	111327442.0
6466	10,000 Degrees	Scholarship and Financial Support		9258322.0	13592921.0
12098	100 Black Men of America	Youth Development, Shelter, and Crisis Services		4366870.0	3407241.0

Selecting a Model

K-Means Model

The AnalyzeUp group decided to pursue a unsupervised model. Ultimately we decided the ideal model for our project would be the K-means clustering model, as it produced a modeling accuracy of 90% and has use-cases of customer segmentation and recommendation systems which align with the scope and goal of our project.



K-means Features

Feature Selection

The features that were selected for modeling are charitable organization's Total Expenses and Total Net Assets as these data points reflect the sum of program expenses, administrative expenses and fundraising expenses, as reported on the income statement (IRS 990) and the difference between assets and liabilities, as reported on the organization's financial statement (IRS 990), respectively.

Connect to database and read the working_table

```
In [2]: engine = create_engine('postgresql://postgres:analyzeup@database-analyzeup.c9mmdeljhxq9.us-west-1.rds.amazonaws.com:5432/analyzeup')

In [3]: query = engine.execute("SELECT * FROM working_table").fetchall()
        query[0:2]

Out[3]: [('10278', ' United Methodist Committee on Relief of Global Ministries', 'Development and Relief Services', '71196129.0', '111327442.0'),
         ('6466', '10,000 Degrees', 'Scholarship and Financial Support', '9258322.0', '13592921.0')]

In [4]: column_names = engine.execute("SELECT * FROM working_table").keys()
        column_names

Out[4]: RMKeyView(['id', 'charity_name', 'cause', 'total_expenses', 'total_net_assets'])

In [5]: working_df = pd.DataFrame(query, columns=column_names)
        working_df.head(2)
```

	id	charity_name	cause	total_expenses	total_net_assets
0	10278	United Methodist Committee on Relief of Globa...	Development and Relief Services	71196129.0	111327442.0
1	6466	10,000 Degrees	Scholarship and Financial Support	9258322.0	13592921.0

The Machine Learning Model

```
Reducing Data Dimensions Using PCA

In [17]: # Using PCA to reduce dimension to three principal components.
pca = PCA(n_components=3)
pca_reduce = pca.fit_transform(X_scaled)
pca_reduce

Out[17]: array([[ 1.49935742,  1.39482189,  1.45913619],
 [ 0.24935478, -1.1069393,   0.19444422 ],
 [-0.54434279,  0.32289496, -0.30278173],
 ...,
 [ 0.910632 ,   0.7703589 ,   0.55868734],
 [ 0.82918002,  0.77182049,  0.55419858],
 [ 0.99961346,  0.77795942,  0.54468432]])

In [18]: # Create a DataFrame with the three principal components.
pcs_df = pd.DataFrame(
    data = pca_reduce, columns=["PC 1", "PC 2", "PC 3"],
    index = X.index
)
print(pcs_df.shape)
pcs_df.head(10)

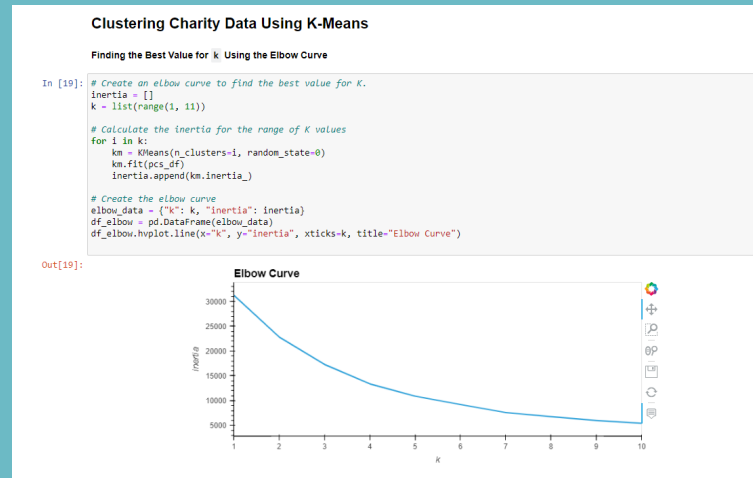
(8143, 3)

Out[18]:
```

	PC 1	PC 2	PC 3
id			
10278	1.499357	1.394821	1.459136
8466	0.249355	-1.106939	0.194442
12098	-0.544343	0.322095	-0.302782
12123	-0.094648	0.065221	0.669248
17473	-0.489340	-0.566606	-2.338448
8770	-0.580834	-0.959028	-0.348520
17318	-0.647536	0.970587	-0.577008
15235	0.266791	0.613196	-1.553920
16289	-0.356657	-0.911833	-0.370960
13055	-0.478276	-0.549649	-2.343371

Feature Engineering

- First the data dimensions were reduced using Principal Component Analysis (PCA).



Best k-value

- Find k value using elbow curve: k=4

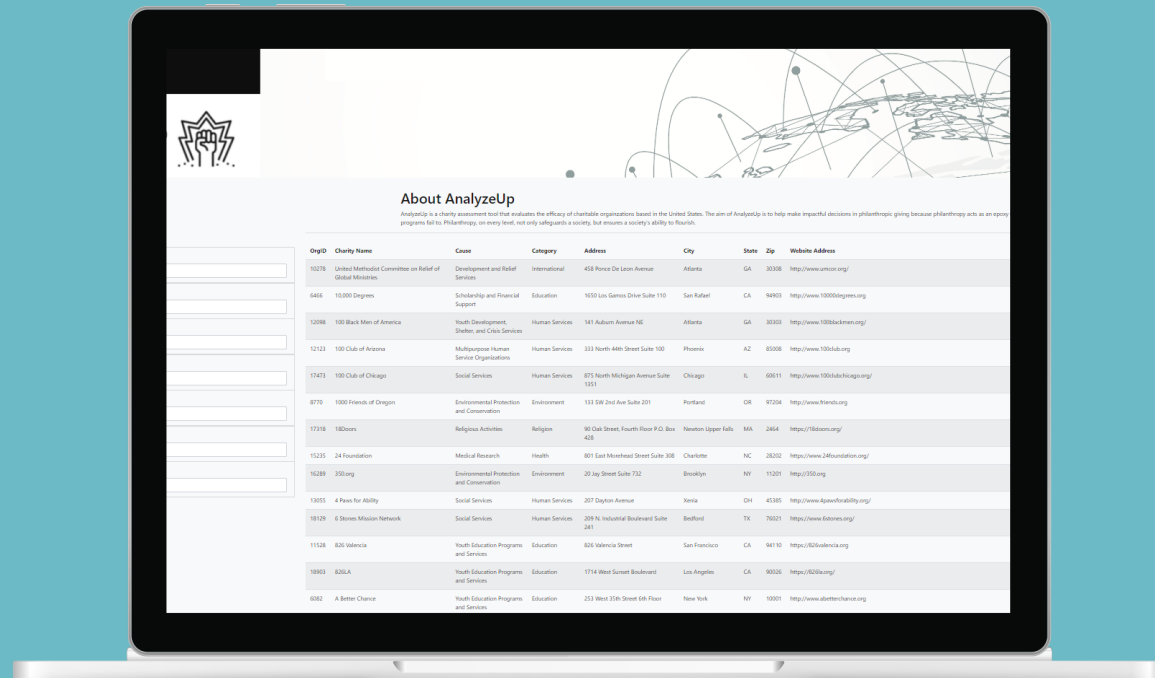
```
# Test the model's clustering performance with Silhouette Coefficient evaluation
kmeans_model = KMeans(n_clusters=4, random_state=1).fit(X)
labels = kmeans_model.labels_
metrics.silhouette_score(X, labels, metric='euclidean')

0.9066127602592019
```

Evaluation

Using Silhouette Coefficient to evaluate the accuracy of the K-means model the results were 90%

Dashboard / User Interface



Dashboard / User Interface

- User Interface to show analysis and search desired data
- Additionally acts as a user interface for potential donors
- By saving our model with Pickle updates can be streamlined by just adding new data.

What the Future Holds

Using the AnalyzeUp model and tools as a jumping off point to come together one a month or so to attempt to use data analytics to solve other social concerns discovered throughout the build of our classification and recommendation system.

Works Consulted

<https://www.artemisg.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>

Thank You



Giving
Hope
Today



Questions