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?

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 webpage & communications)

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.

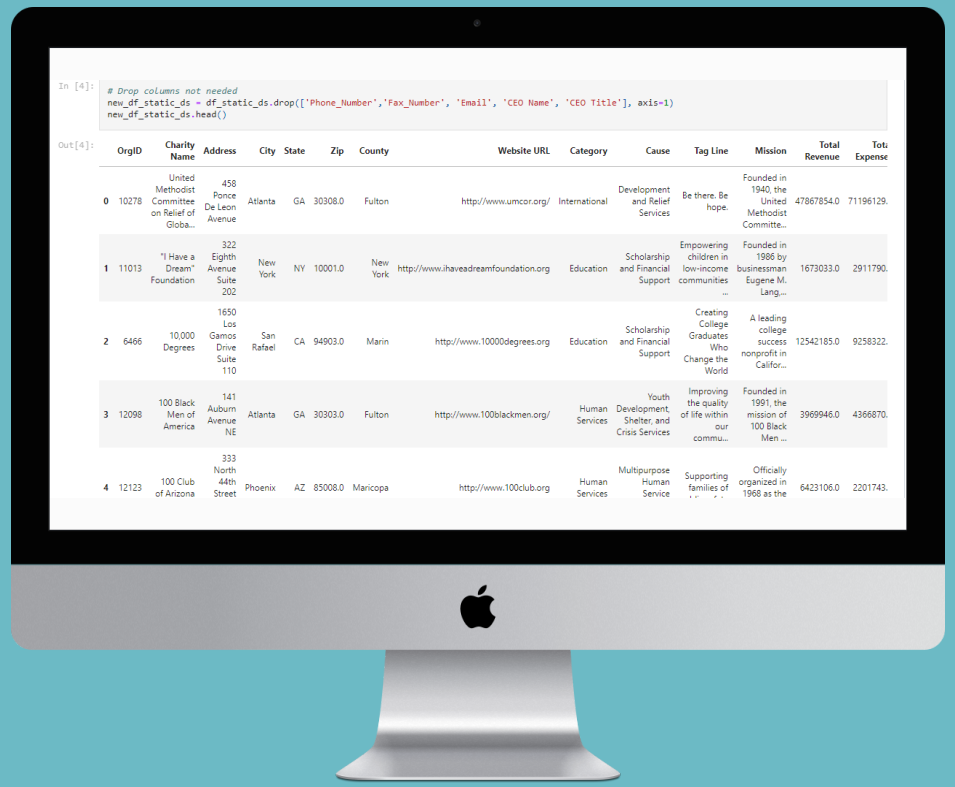
Our Objectives

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 Extraction



```
In [4]: # Drop columns not needed
        new_df_static_ds = df_static_ds.drop(['Phone_Number','Fax_Number', 'Email', 'CEO Name', 'CEO Title'], axis=1)
        new_df_static_ds.head()
```

Out[4]:

	OrgID	Charity Name	Address	City	State	Zip	County	Website URL	Category	Cause	Tag Line	Mission	Total Revenue	Tota Expense
0	10278	United Methodist Committee on Relief of Globa...	458 Ponce De Leon Avenue	Atlanta	GA	30308.0	Fulton	http://www.umcor.org/	International	Development and Relief Services	Be there. Be hope.	Founded in 1940, the United Methodist Committe...	47867854.0	71196129.
1	11013	"I Have a Dream" Foundation	322 Eighth Avenue Suite 202	New York	NY	10001.0	New York	http://www.ihaveadreamfoundation.org	Education	Scholarship and Financial Support	Empowering children in low-income communities ...	Founded in 1986 by businessman Eugene M. Lang,...	1673033.0	2911790.
2	6466	10,000 Degrees	1650 Los Gamos Drive Suite 110	San Rafael	CA	94903.0	Marin	http://www.10000degrees.org	Education	Scholarship and Financial Support	Creating College Graduates Who Change the World	A leading college success nonprofit in Califor...	12542185.0	9258322.
3	12098	100 Black Men of America	141 Auburn Avenue NE	Atlanta	GA	30303.0	Fulton	http://www.100blackmen.org/	Human Services	Youth Development, Shelter, and Crisis Services	Improving the quality of life within our commu...	Founded in 1991, the mission of 100 Black Men ...	3969946.0	4366870.
4	12123	100 Club of Arizona	333 North 44th Street	Phoenix	AZ	85008.0	Maricopa	http://www.100club.org	Human Services	Multipurpose Human Service	Supporting families of	Officially organized in 1968 as the	6423106.0	2201743.

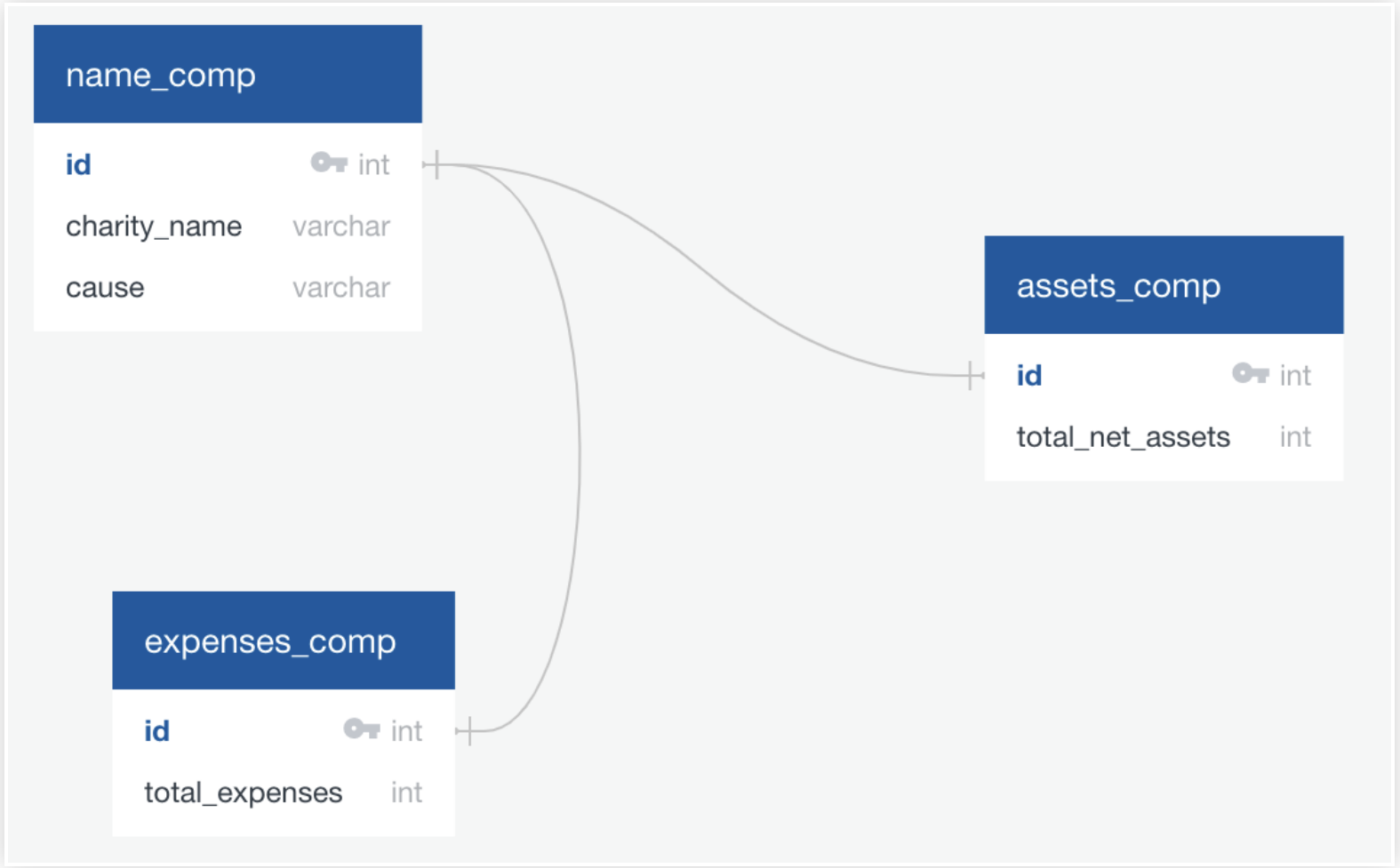
ETL of the Data

For this portion of the analysis, Jupyter Notebook, Pandas, and AWS were used.

The info was stored in a database based on the following attributes:

1) API extracting data table
2) Charity name & cause
3) Total charity expenses, total charity assets
4) Location of the charity
5) Other pertinent background information

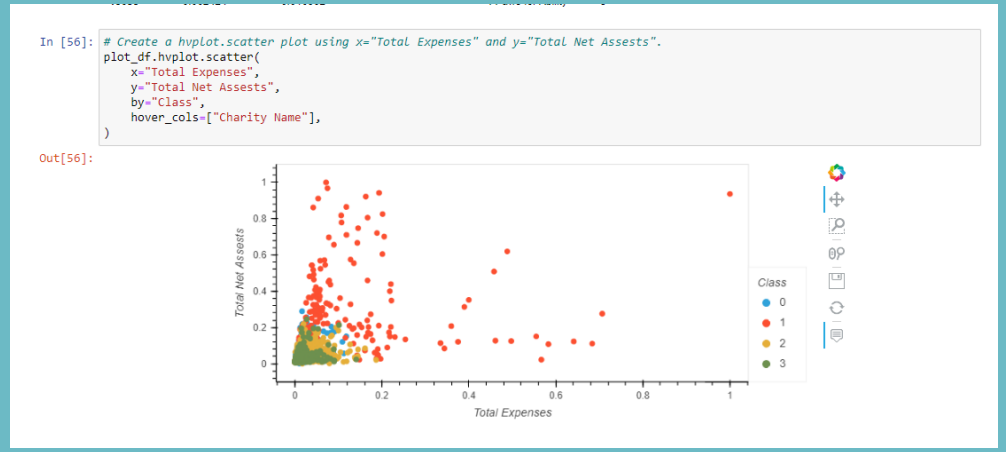
ETL of the Data



Selecting a Model

K-Means Model

After researching the best machine learning models that would adequately produce the best classifications of US charity's general efficacy, 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 a charity'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.c9mmdejuhxq9.us-west-1.rds.amazonaws.com:5432/analyzeu
```

```
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', '1113
        27442.0'),
         ('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)
```

Out[5]:

	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,  1.39482109,  1.45913619],
       [ 0.24935478, -1.1069393 ,  0.1944422 ],
       [-0.54434279,  0.32209496, -0.30278173],
       ...,
       [ 0.910632  ,  0.7703589 ,  0.55068734],
       [ 0.82910002,  0.77102049,  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]:

id	PC 1	PC 2	PC 3
10278	1.499357	1.394821	1.459136
6466	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	0.266781	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).

Clustering Charity Data Using K-Means

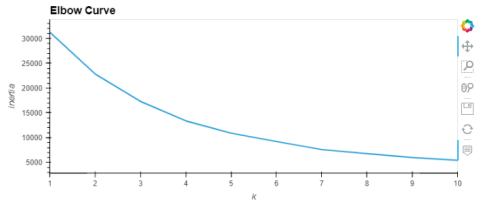
Finding the Best Value for `k` Using the Elbow Curve

```
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.hvplot.line(x="k", y="inertia", xticks=k, title="Elbow Curve")
```

Out[19]:



Best k-value

- Find k value using elbow curve: k=4

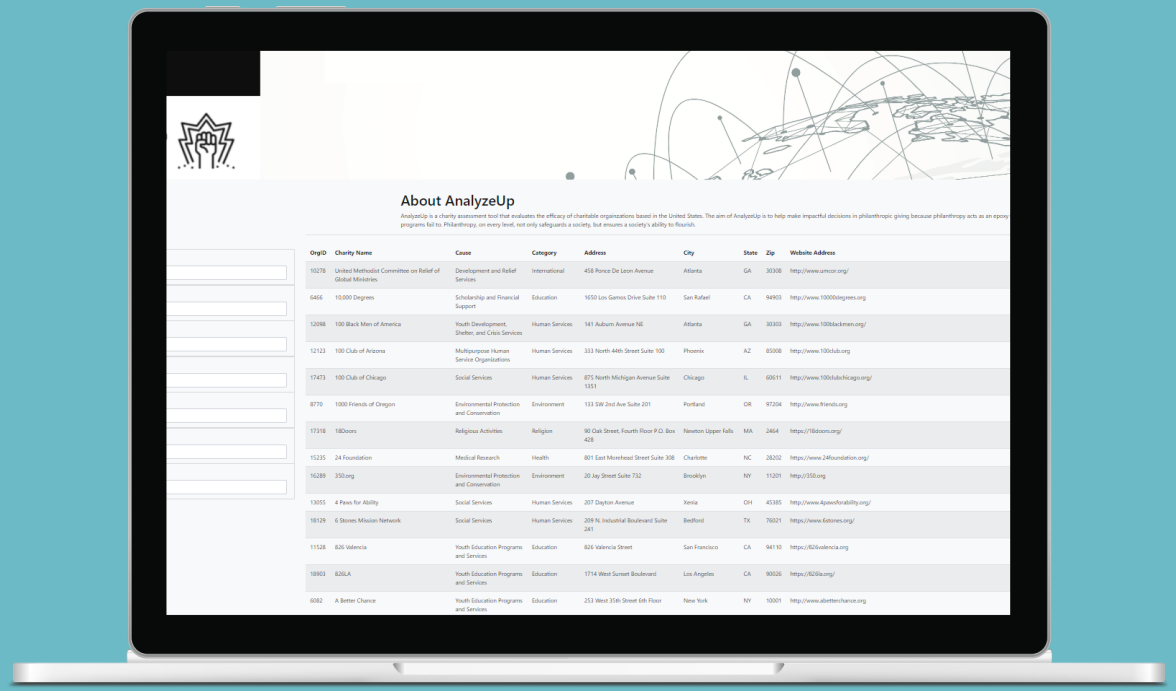
```
# Test the model's clustering performance with Silhouette Coefficient evaluation
kmeans_model = KMeans(n_clusters=4, random_state=1).fit(X)
labels = kmeans_model.labels_
metrics.silhouette_score(X, labels, metric='euclidean')
```

```
0.9066127602592019
```

Evaluation

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

Database



Database / 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.

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