UNIVERSITY of WISCONSIN Superior

From Data to Impact:

Understanding Donor Giving Behavior at UWS

Caleb Cuterer, Shannon Hoffman, Krishna Pandari, Jim Pieper, Walgama Jayasekara

SEPTEMBER 12, 2024

INTRODUCTION

Predicting Donor Behavior Using Machine Learning

Project Goal:

Analyze and predict donor behavior using machine learning, with a focus on improving fundraising strategies and donor engagement.

Data Source and Scope:

- Anonymized donor data from the UW-Superior Foundation, stored in the Raiser's Edge database.
- Focused on a subsection of data, including constituent records and gift records.
- The analysis was limited to a subsection of available data due to confidentiality.

INTRODUCTION

Predicting Donor Behavior Using Machine Learning

Key Limitations:

- Incomplete records (e.g., missing alumni information such as student organization involvement and athletic participation).
- Shift in student demographics over time (e.g., increase in online vs. nontraditional students).
- Missing **age/birthdate** data for many constituents, which is a critical feature for donor prediction.

DATA ANALYSIS

The data reveals interesting insights into the composition of donors and their engagement in different campus activities such as 'Student Organizations', 'Arts', 'Fraternal Organizations', and athletics. Additionally, examining the age distribution offers a deeper understanding of donor demographics.

jupyter DonorBehavior Last Checkpoint: 20 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted | Python 3 (ipykernel)

Run Code

In [2]:
```
# Load dataset
data = pd.read_csv('./Resources/constituents.csv')
data
```

Out[2]:

	ConstituentID	City	State	Postcode	Country	Constituent type	Constituent codes	Lifetime giving	First gift amount	First gift type	...	Education	Gender	Marita status
0	60472	Superior	WI	54880-2556	United States	Organization	Business (No start date - No end date)	0	NaN	NaN	...	NaN	NaN	NaN
1	43735	Hermantown	MN	55811-1755	United States	Organization	Other Organizations (8/29/2012 - No end date)	50	50	One-time gift	...	NaN	NaN	NaN
2	60145	Esko	MN	55733-9645	United States	Organization	Other Organizations (No start date - No end date)	50	50	One-time gift	...	NaN	NaN	NaN
3	21332	NaN	NaN	NaN	United States	Organization	Business (No start date - No end date)	25	25	One-time gift	...	NaN	NaN	NaN
4	50410	Superior	WI	54880-1504	United States	Organization	Business (No start date - No end date)	0	NaN	NaN	...	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
50908	46217	Duluth	MN	55808-1945	United States	Individual	Attended Didn't Graduate (No start date - No e...	75	75	One-time gift	...	UW-Superior,Not primary,Business Administratio...	Female	NaN
50909	46390	Duluth	MN	55808-1725	United States	Individual	Former Parent (No start date - No end date)	0	NaN	NaN	...	NaN	Male	Married
50910	50122	Duluth	MN	55808-1737	United States	Individual	Alumni (No start date - No end date)	0	NaN	NaN	...	UW-Superior,Primary,Physical Education,BS,2014	Unknown	NaN
50911	59551	Centuria	WI	54824-7721	United States	Individual	Parent (No start date - No end date)	0	NaN	NaN	...	NaN	Female	NaN
50912	15331	Saginaw	MN	55779-9532	United States	Individual	Alumni (No start date - No end date)	0	NaN	NaN	...	UW-Superior,Primary,Physical Education,BS,2001	Male	Married

50913 rows × 34 columns

Analyzed sample dataset with 50913 rows and 34 columns

- Total Donors: 9889 (19.42%)
- Total Non-Donors: 41024 (80.58%)

Counts and Percentages of Donors by Attribute:

- Age: 2408 donors (24.35%)
- Student Organizations: 594 donors (6.01%)
- Arts: 171 donors (1.73%)
- Fraternal Organizations: 655 donors (6.62%)
- Men's Athletics: 533 donors (5.39%)
- Scholarship Recipient: 1014 donors (10.25%)
- Women's Athletics: 268 donors (2.71%)

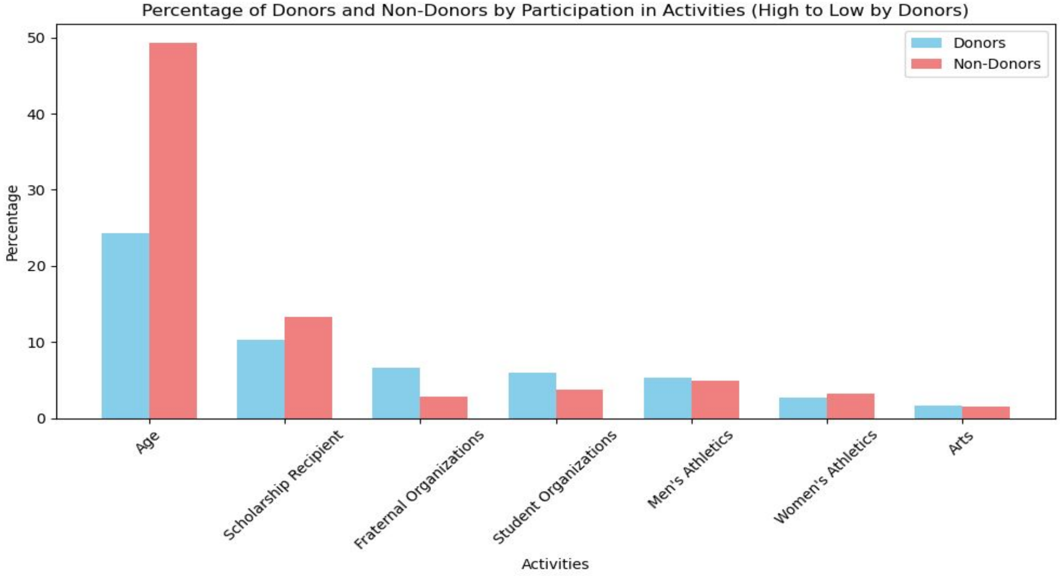
Overlap of Donors across Multiple Attributes:

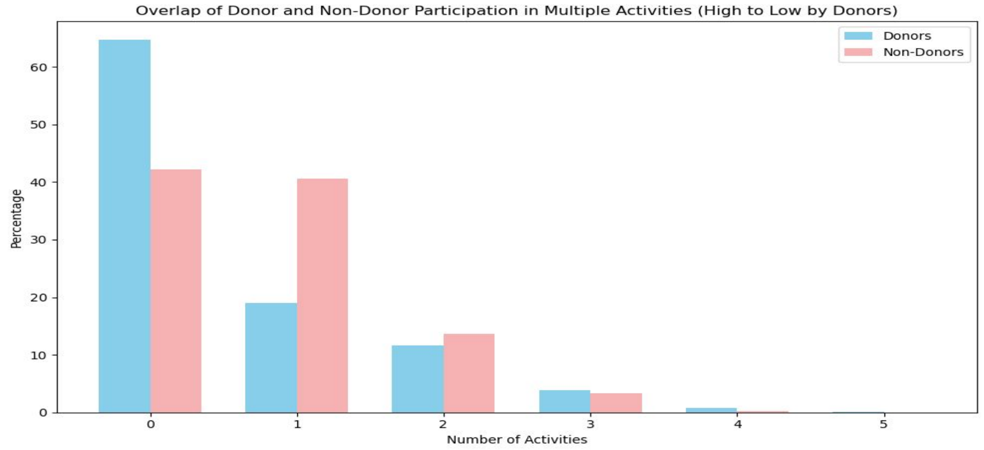
- 0 attributes: 6395 donors (64.67%)
- 1 attributes: 1884 donors (19.05%)
- 2 attributes: 1152 donors (11.65%)
- 3 attributes: 382 donors (3.86%)
- 4 attributes: 71 donors (0.72%)
- 5 attributes: 5 donors (0.05%)

UNIVERSITY of WISCONSIN Superior

VISUAL - Participating Activities

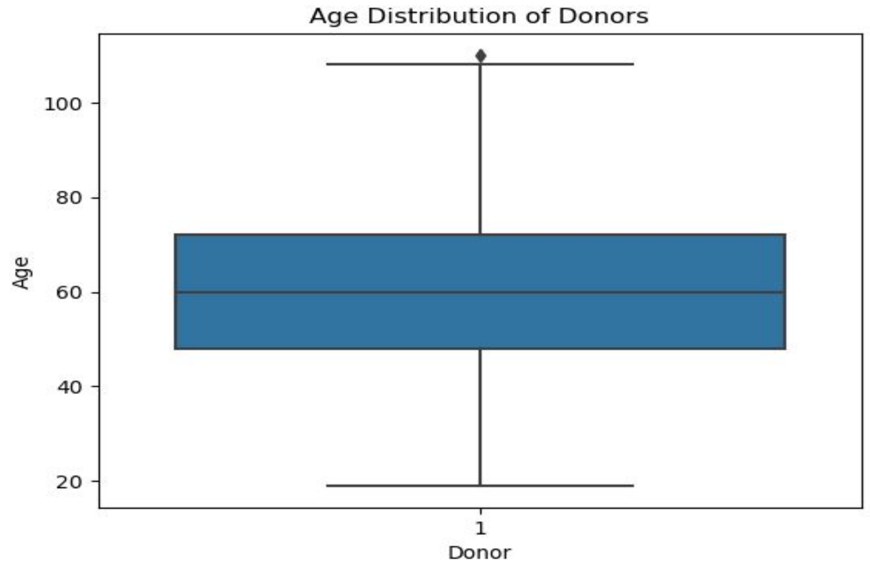
Constituents involved in Student Organizations, Fraternal Organizations, Arts, and similar activities are more likely to become donors compared to those who are not. Additionally, donors tend to engage in multiple activities.

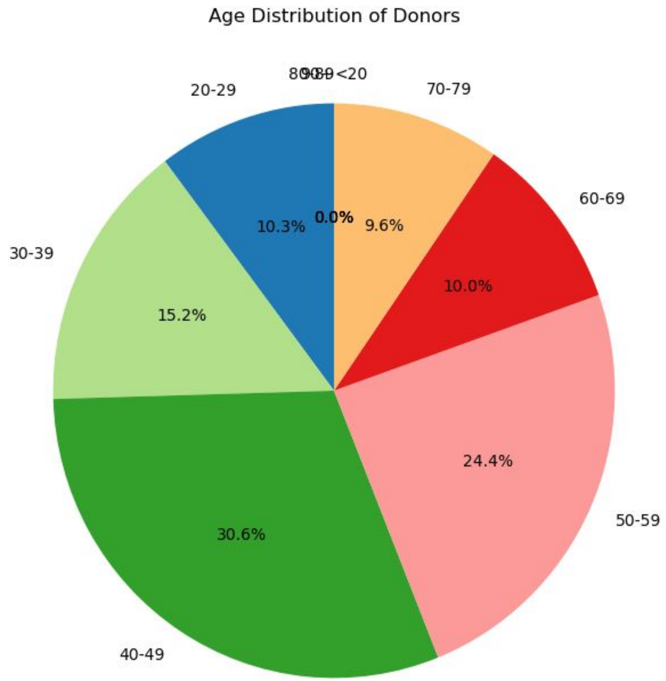


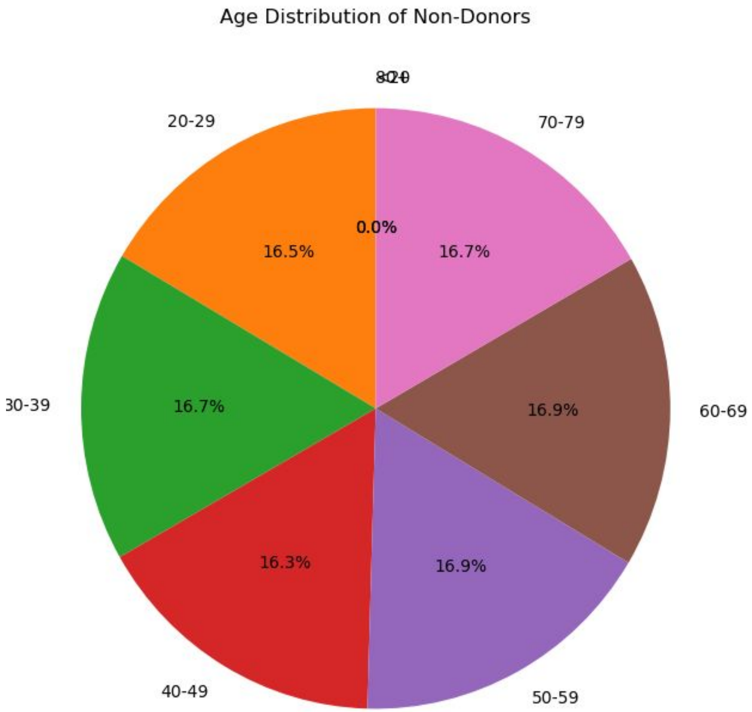


VISUAL - Age Distribution of Donors and Non-Donors

More than half of donors are between 40 and 60.

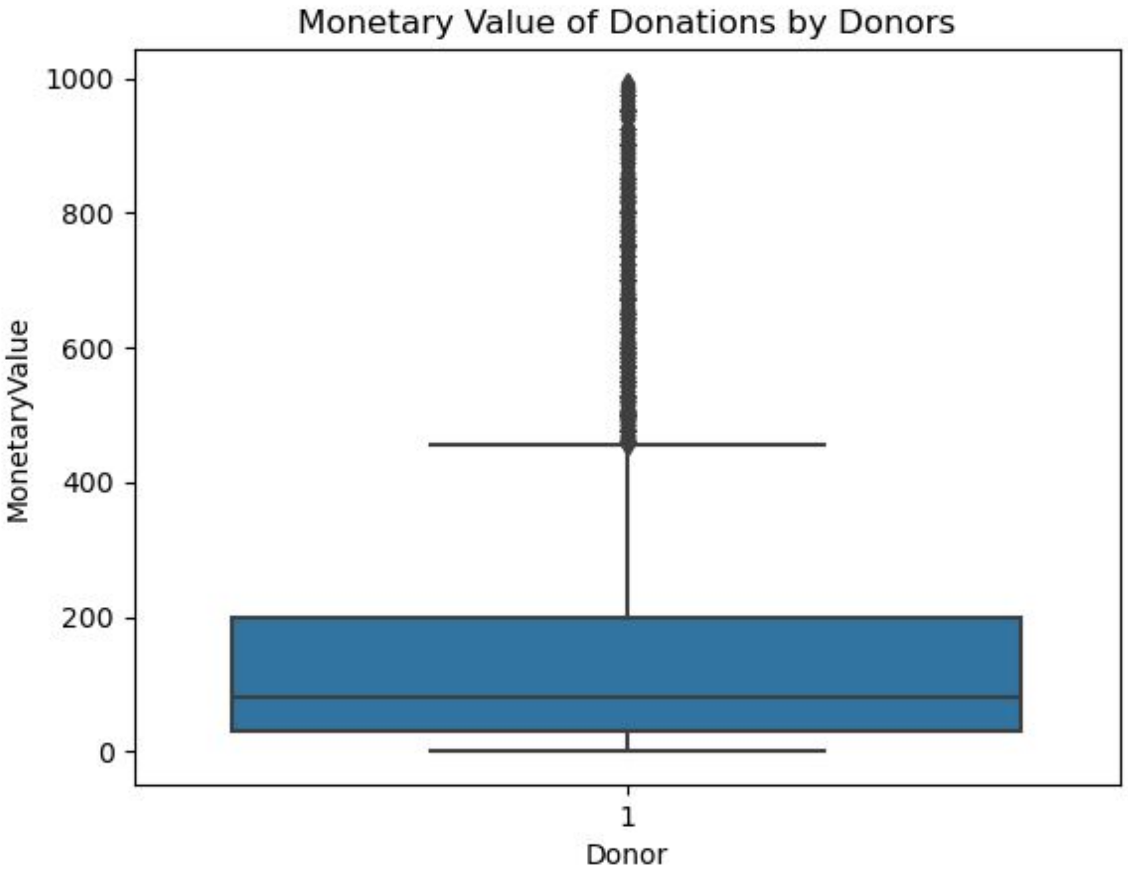




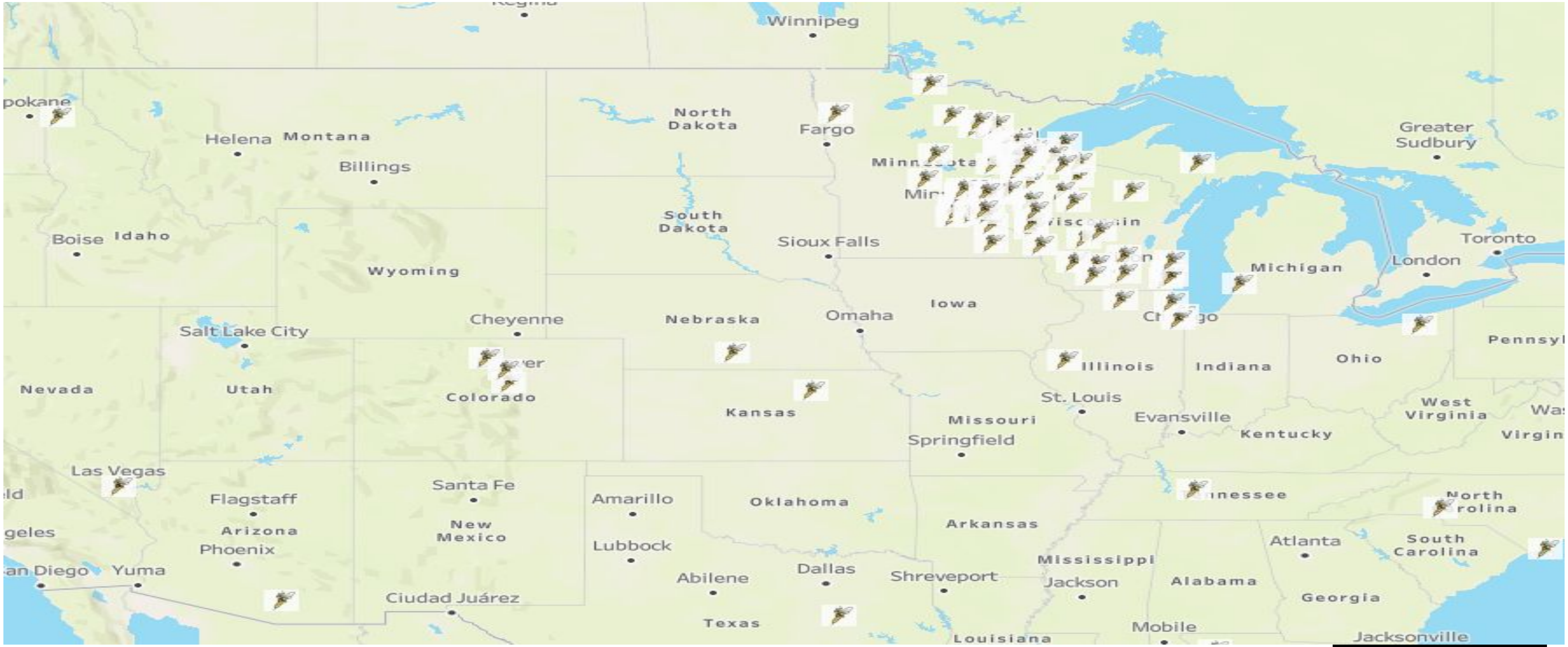


VISUAL - Monetary Value of Donations

The usual donation is $100



VISUAL - Lifetime Gifts by Zipcode- Tableau



UNIVERSITY *of* WISCONSIN Superior

Tableau Visualization Payment type

-

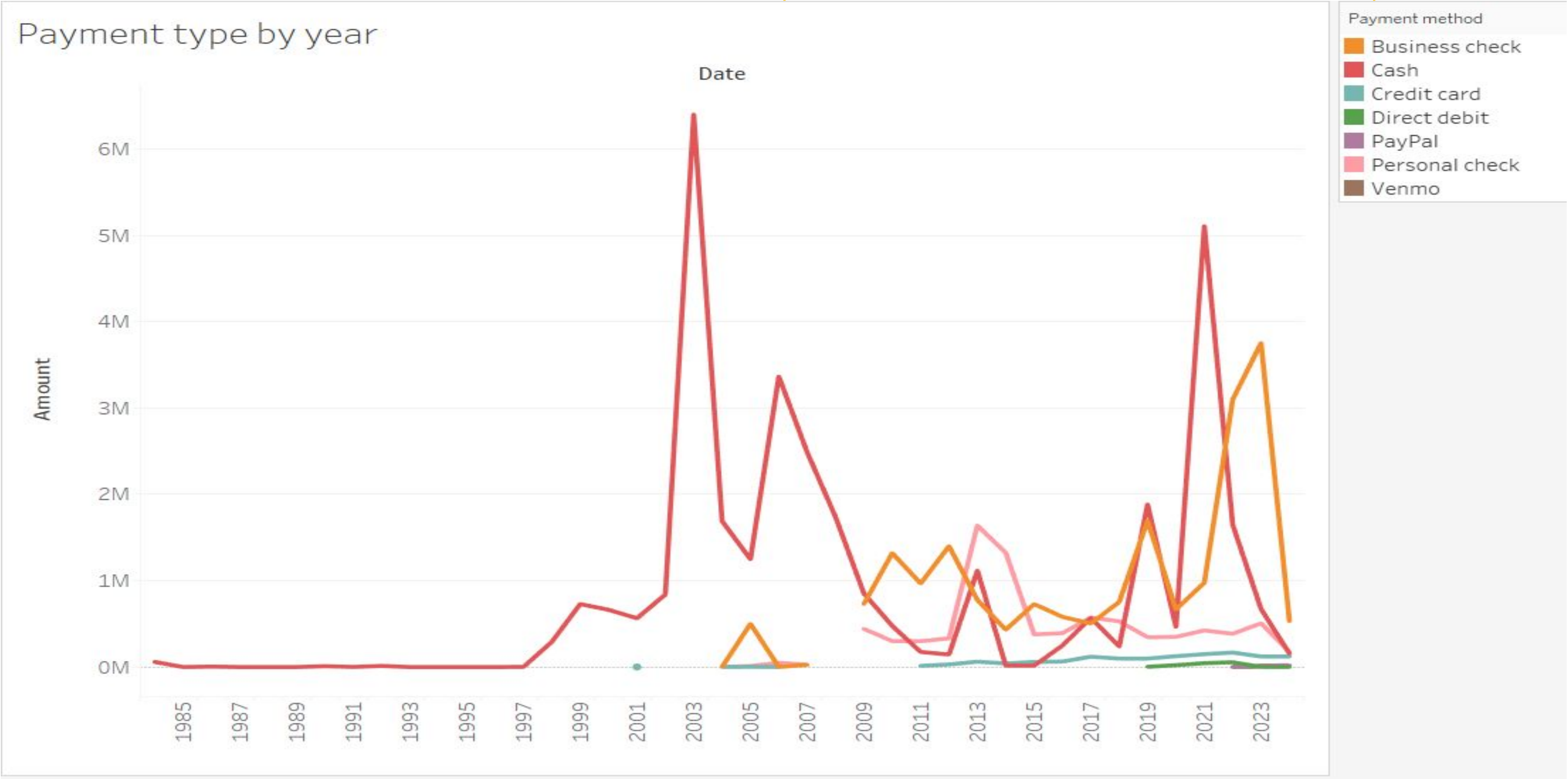
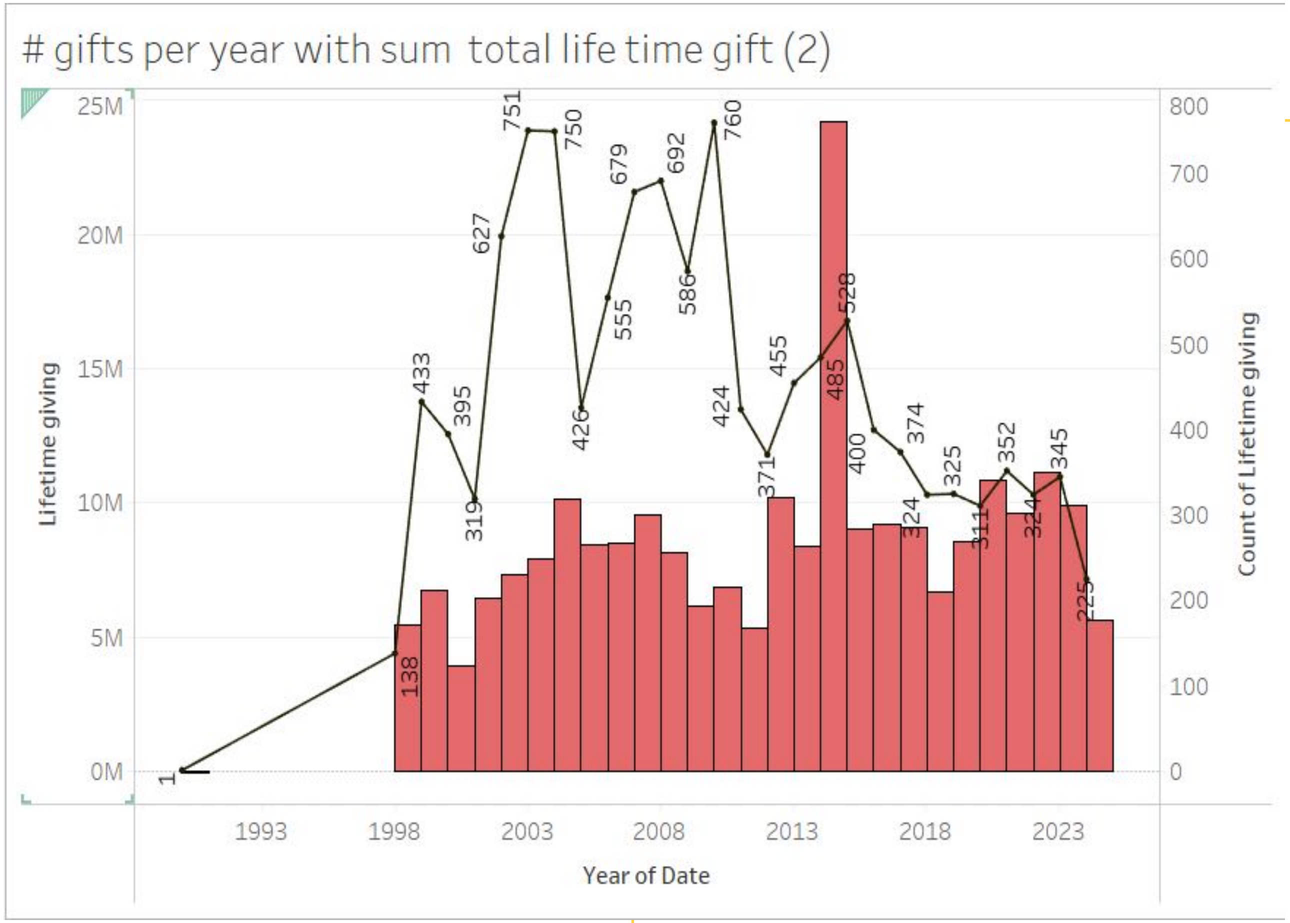
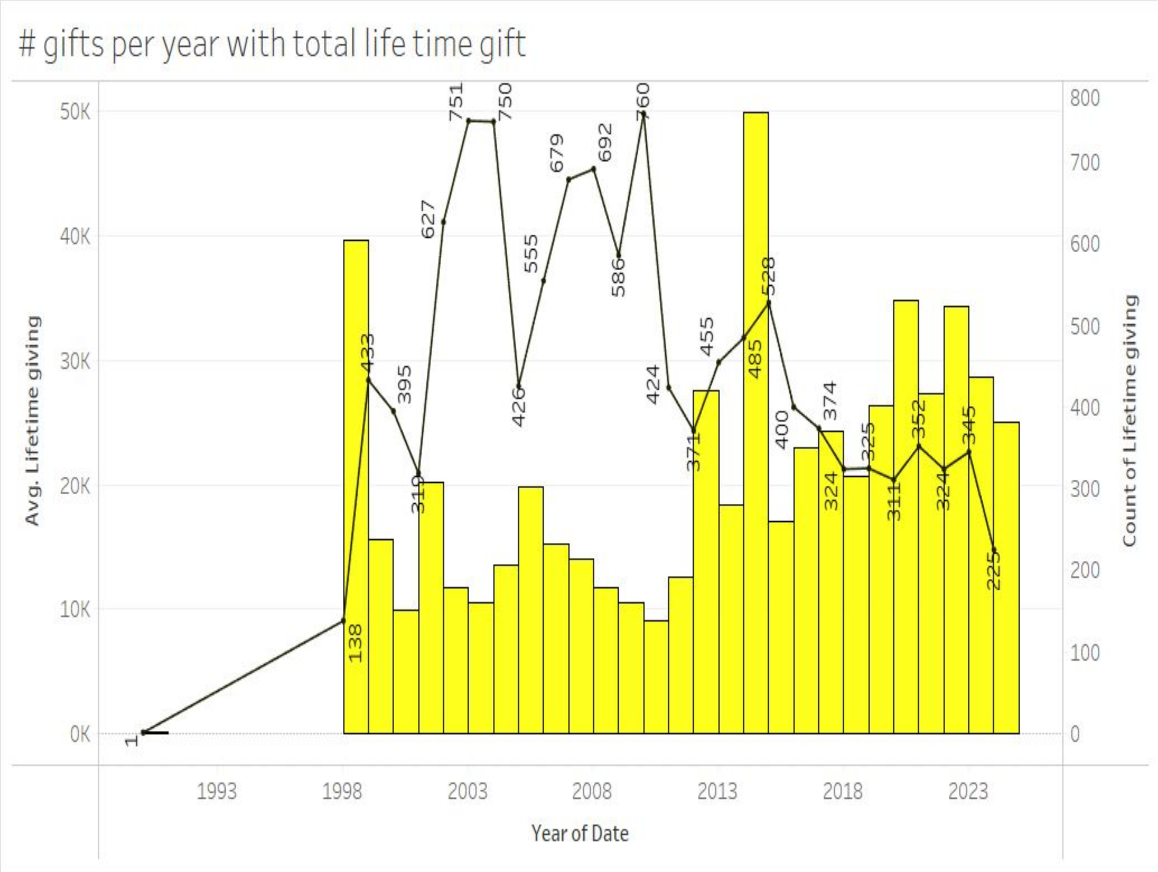


Tableau Visualization





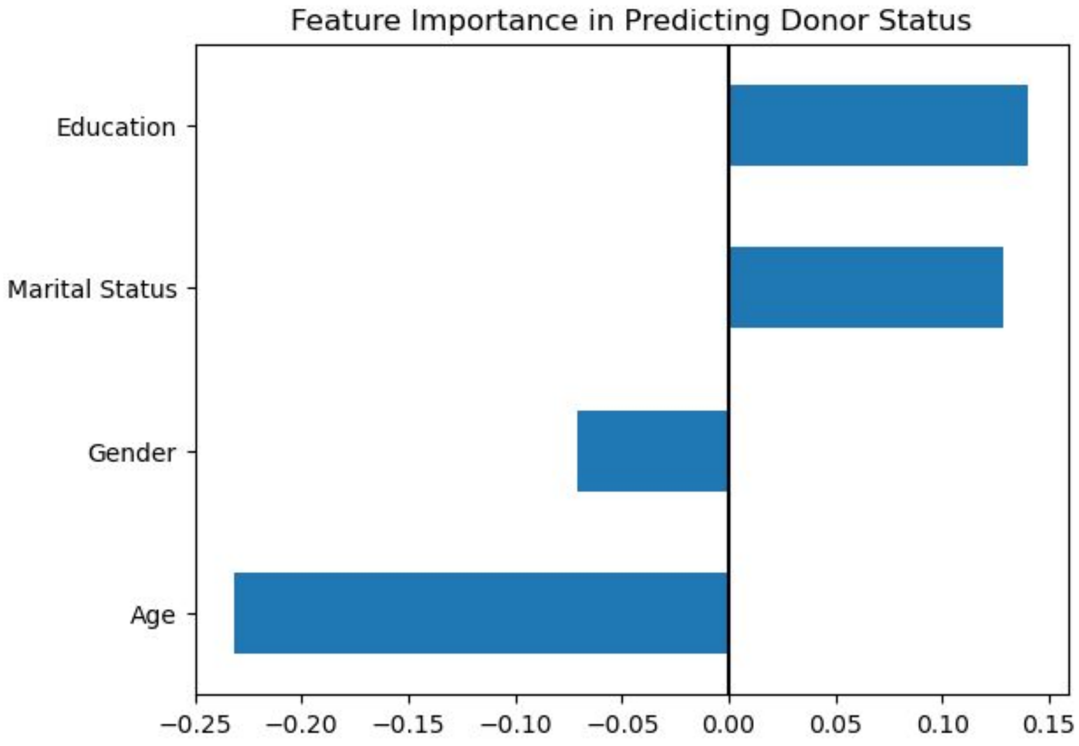
Caption

The trends of Avg. Lifetime giving and count of Lifetime giving for Date Year. Color shows details about Avg. Lifetime giving and count of Lifetime g
data is filtered on count of Date and Age . The count of Date filter ranges from 1 to 4,196. The Age filter has multiple members selected.

Demographic Variables and Donor Status

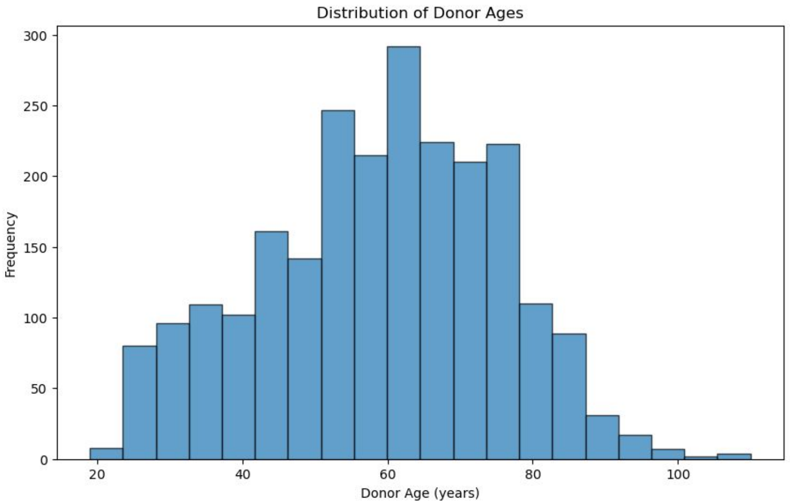
- Logistic Regression Model: investigated the feature importance of several demographic variables on donor status
- Focused subsequent analyses on age and marital status

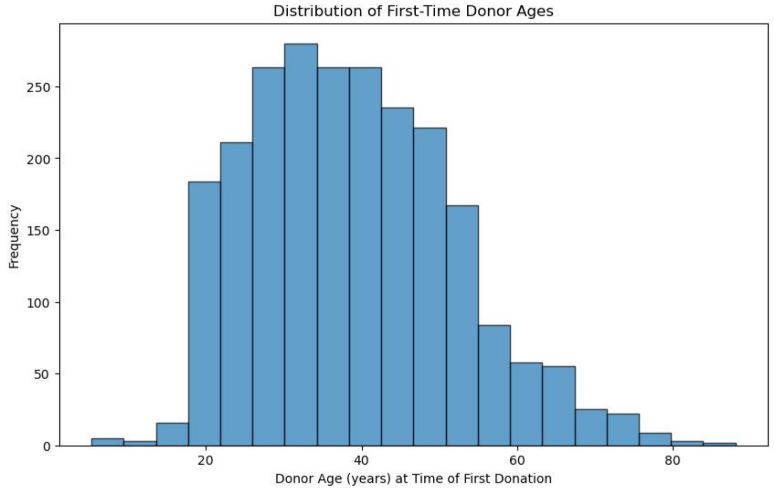
	precision	recall	f1-score	support
0	0.10	0.01	0.01	129
1	0.79	0.98	0.87	480
accuracy			0.78	609
macro avg	0.44	0.49	0.44	609
weighted avg	0.64	0.78	0.69	609



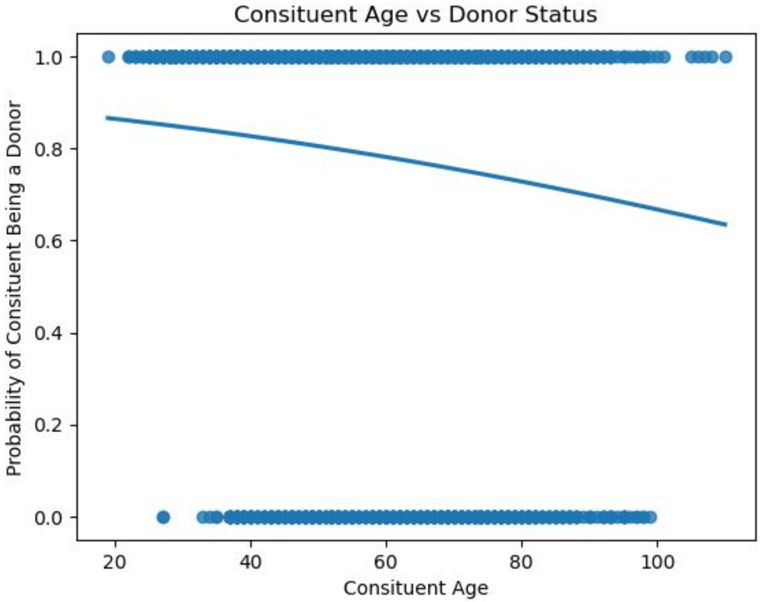
UNIVERSITY of WISCONSIN Superior

Age as a Predictor of Donor Status





- **Age Distributions:** donors overall and first-time donors
- **Logistic Regression Model:** age as a predictor of donor status
- Adjusting for inflation

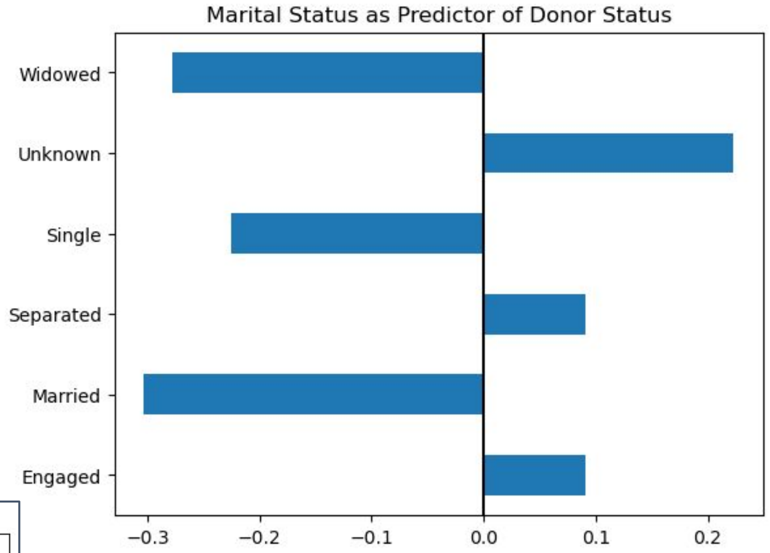


	precision	recall	f1-score	support
0	0.39	0.05	0.10	129
1	0.79	0.98	0.88	480
accuracy			0.78	609
macro avg	0.59	0.52	0.49	609
weighted avg	0.71	0.78	0.71	609

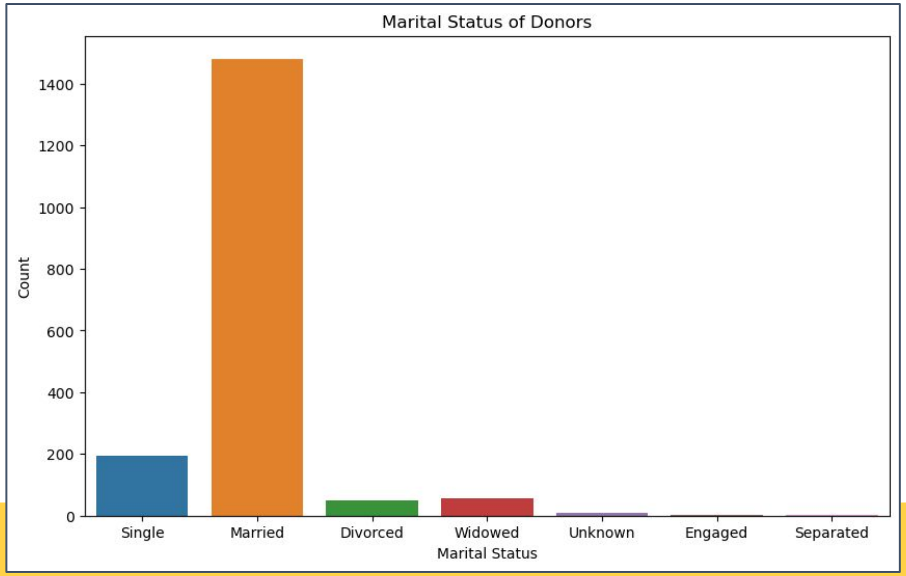
UNIVERSITY *of* WISCONSIN Superior

Marital Status as a Predictor of Donor Status

- Distribution of Marital Statuses amongst Donors
- Logistic Regression Model: looked at the influence on donor status for each of the 6 different marital status categories
- One-Way Anova: marital status and first donation amount
  - Significant difference between groups (test statistic = 7.55, p-value = $4.801e^{-08}$



	precision	recall	f1-score	support
0	0.62	0.08	0.14	129
1	0.80	0.99	0.88	480
accuracy			0.79	609
macro avg	0.71	0.53	0.51	609
weighted avg	0.76	0.79	0.73	609

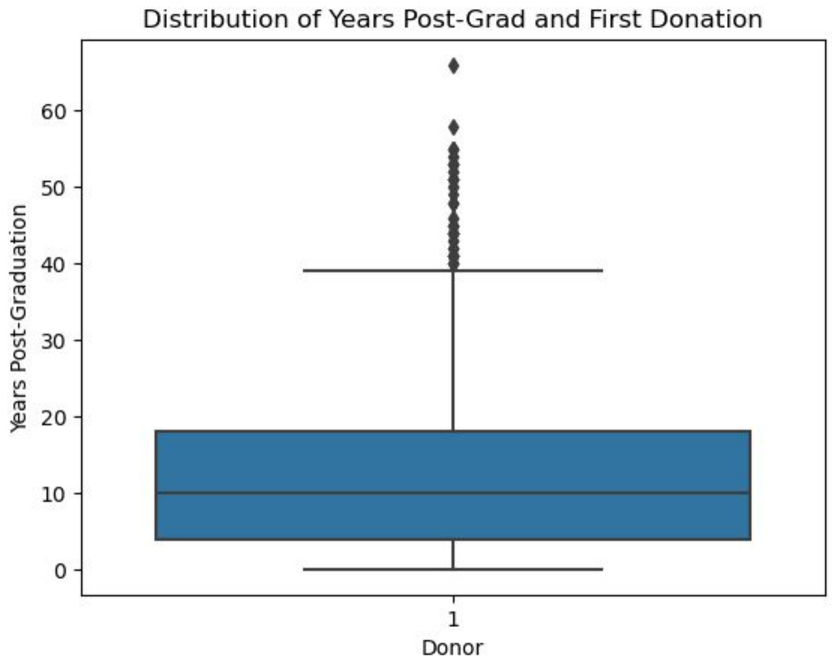


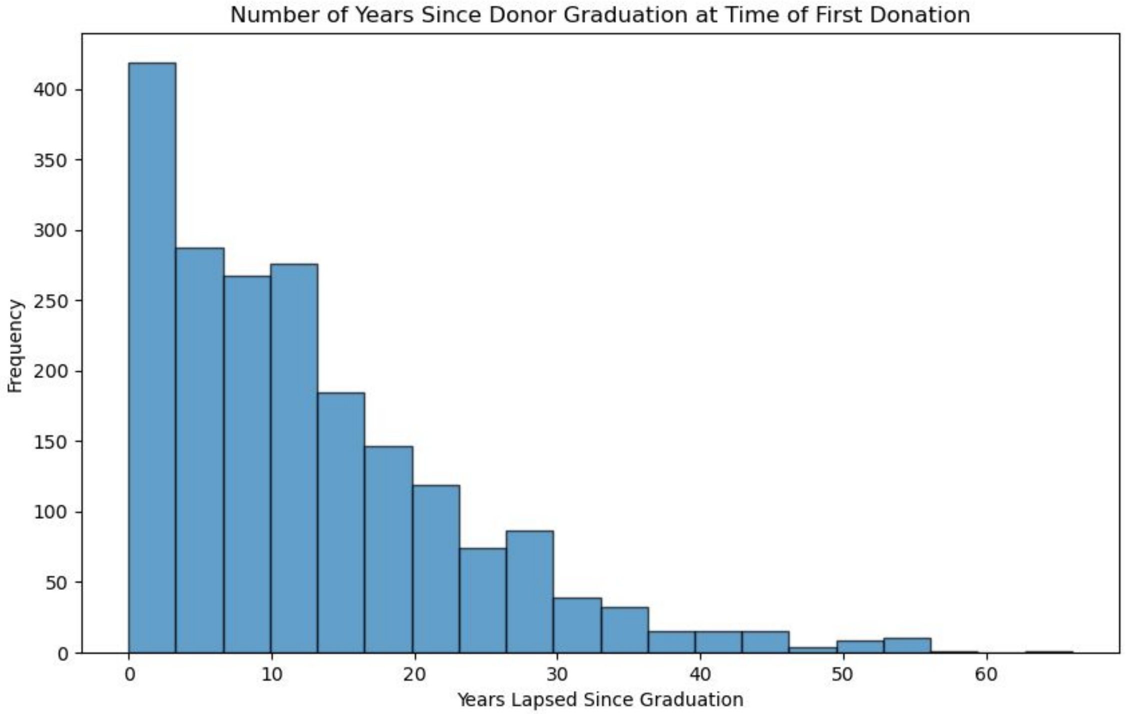

```
#One-Way Anova for Marriage and Donation Amount
grouped = mardf.groupby('Marital status')['First_donation_adjusted']
marriage_donations = {category: donations.values for category, donations in grouped}
Divorced = marriage_donations['Divorced']
Engaged = marriage_donations['Engaged']
Married = marriage_donations['Married']
Separated = marriage_donations['Separated']
Single = marriage_donations['Single']
Unknown = marriage_donations['Unknown']
Widowed = marriage_donations['Widowed']

f_oneway(Divorced, Engaged, Married, Separated, Single, Unknown,Widowed)
```

UNIVERSITY *of* WISCONSIN Superior

Graduation Date to First Donation

- Investigated the amount of time between donors’ graduation and their initial donations
- **Average**: 12.52 years from graduation to first donation





UNIVERSITY *of* WISCONSIN Superior

ALUMNI GIVING

data processing

Objective

Identify patterns among alumni donors by analyzing selected features.

Data Cleaning Overview

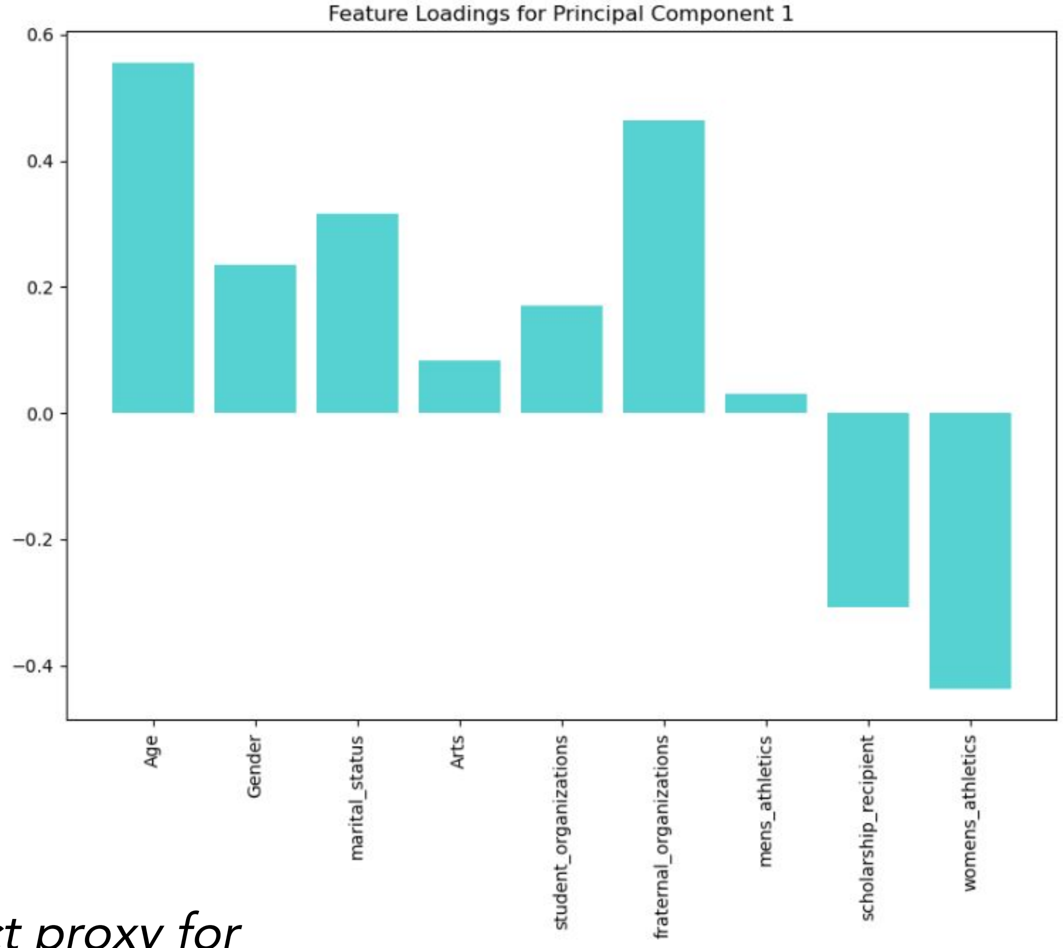
- Merged constituent and gift data to focus on alumni donors.
- Removed missing data (e.g., missing first_gift_date and Age).
- Standardized columns for consistency (e.g., renaming and formatting).
- Coded categorical data (e.g., Gender,) into numeric values.
- Created boolean flags for participation (e.g., student organizations)

ALUMNI GIVING

selected features across models

Rank order of feature importance across all models*:

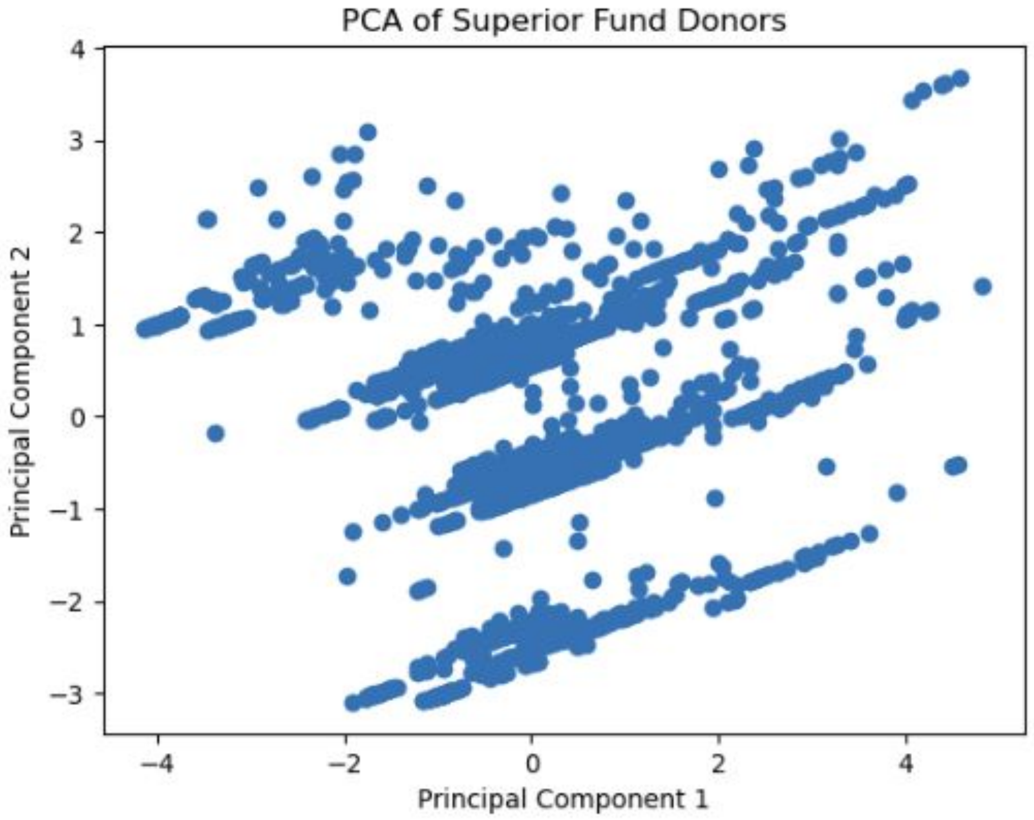
- Age
- Marital Status
- Gender
- Student Organizations
- Scholarship Recipient
- Fraternal Organizations
- Men's Athletics
- Arts
- Women's Athletics



*giving features not included because they act as direct proxy for indicator of giving, by definition, overshadowing any other interesting features. (See "Future Analysis")

ALUMNI GIVING

pca analysis



Explained Variance:

The first 5 components explained approximately 71% of the variance.

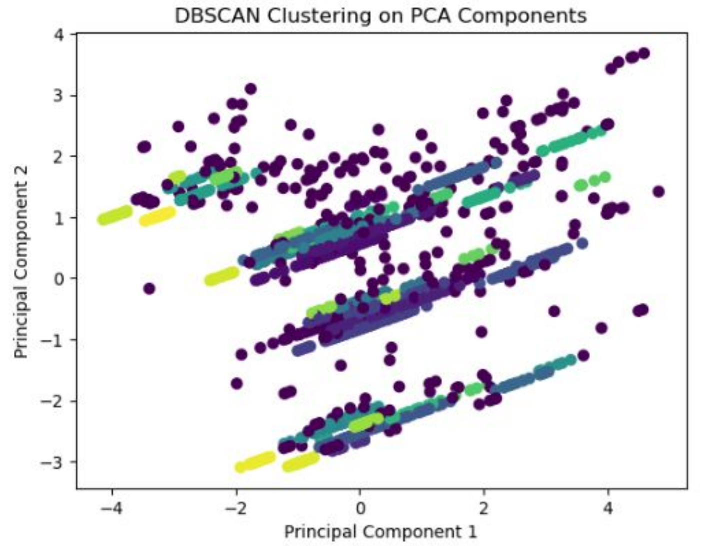
PC1: 20.17%
PC2: 17.04%
PC3: 14.13%
PC4: 10.04%
PC5: 9.76%

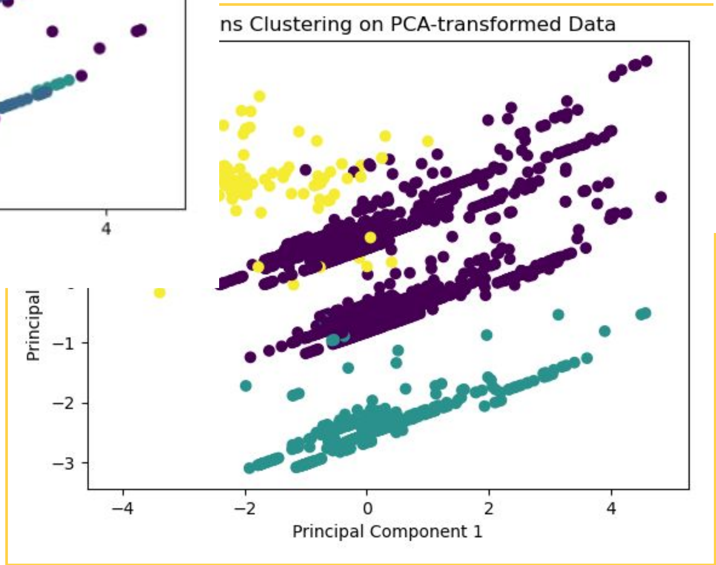
Key Insights:

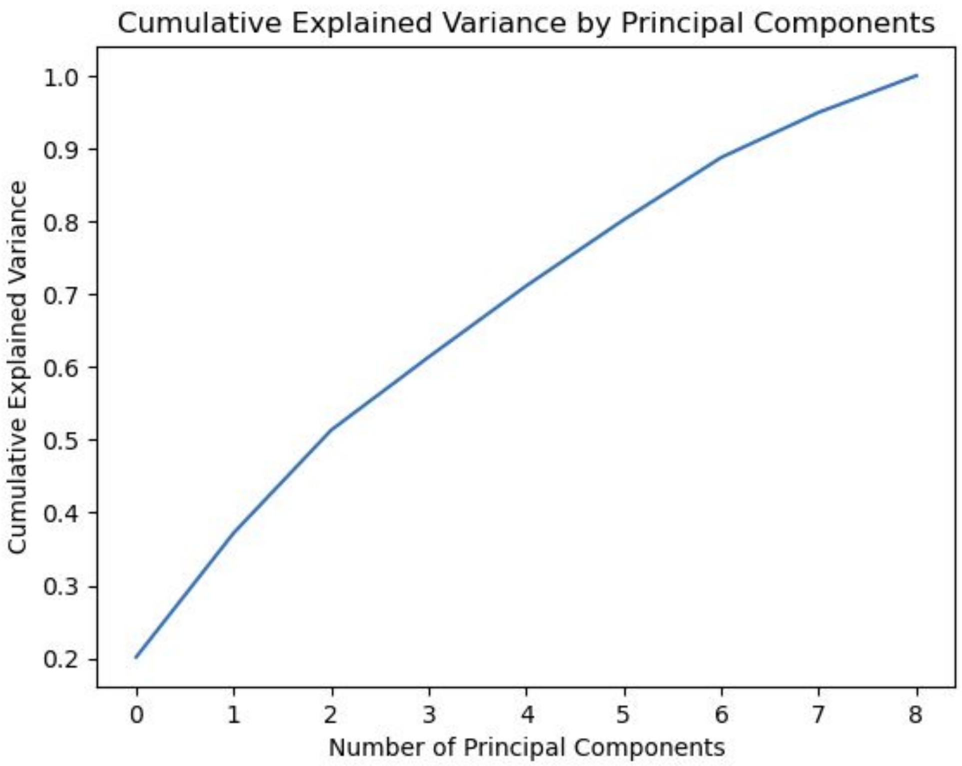
- PC1: Age, Fraternal Organizations, and Marital Status.
- PC2: Gender and Men's Athletics.
- PC3: Participation in Arts and Student Organizations.

ALUMNI GIVING

pca analysis - testing







Outcome:

8 out of the 9 features explained the variance in the data.

ALUMNI GIVING

random tree model

Objective:

Predict high lifetime giving using Random Forest models across different thresholds and funds.

Thresholds:

- $1k,
- $5k, and
- $10k.

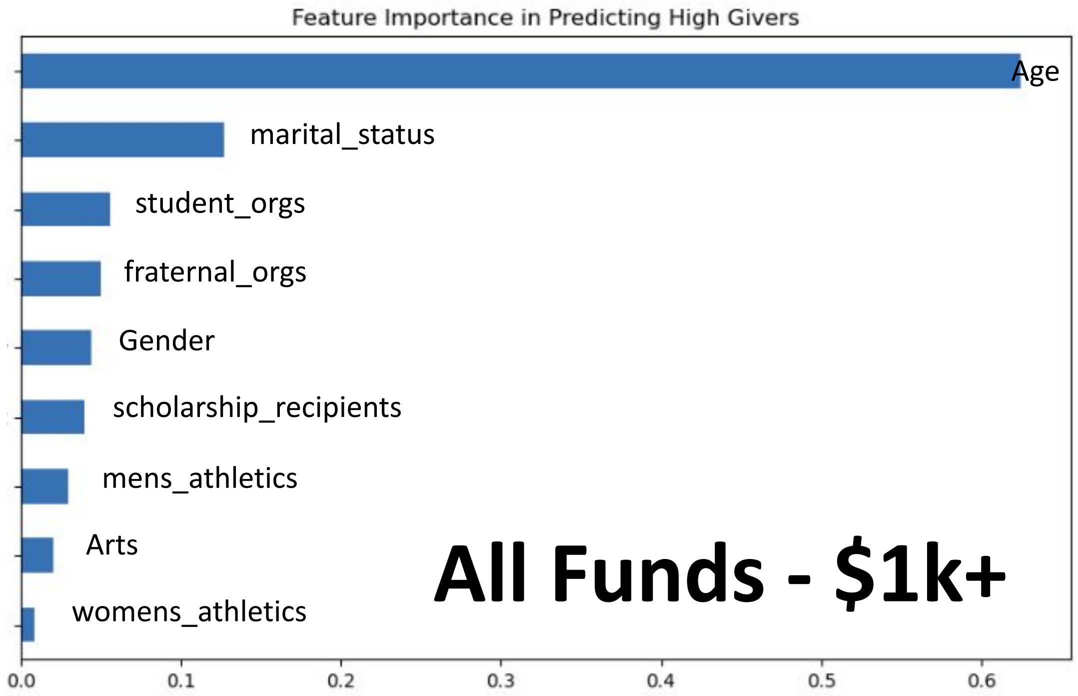
Funds:

Analyzed across both **All Funds** and the **Superior Fund**.

ALUMNI GIVING

model performance - all funds

Model analysis: Accuracy is high but precision, recall, and F1 score are low, indicating an imbalance in the dataset (one class - e.g., low givers - is much more frequent than the other). The model is predicting the majority class (low givers) very well, but it struggles to identify the minority class (high givers).

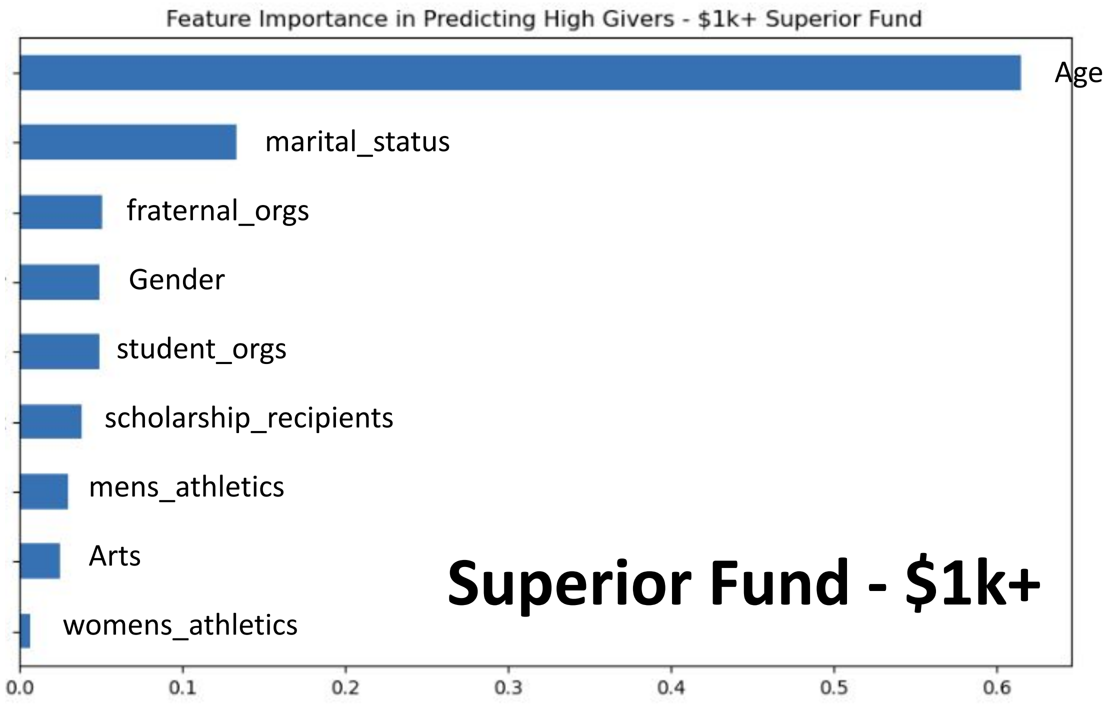


threshold	accuracy	precision	recall	F1
$1k	82%	33%	21%	26%
$5k	94%	31%	10%	15%
$10k	96%	40%	7%	12%

ALUMNI GIVING

model performance - superior fund

Model analysis: Does not meet objective. **For example,** the $5k model correctly predicted 90% of all donors as either high or low givers. However, this high accuracy is mostly predicting the majority class (low givers) correctly. After adjusting weights and thresholds, the model remains biased towards predicting **low givers.** *The model is great at accurately predicting low givers correctly.*



threshold	accuracy	precision	recall	F1 score
$1k	74%	39%	22%	28%
$5k	90%	6%	3%	4%
$10k	94%	0%	0%	0%

ALUMNI GIVING

random tree model

Outcome:

While the Random Forest model was not successful in predicting high-level donors (based on precision, recall, and F1 score), the model did highlight key features that are influential in donor behavior

Top Features (across all models):

- Age (all models)
- Marital Status
- Participation in Student Organizations
- Participation in men's athletics

Key Insights:

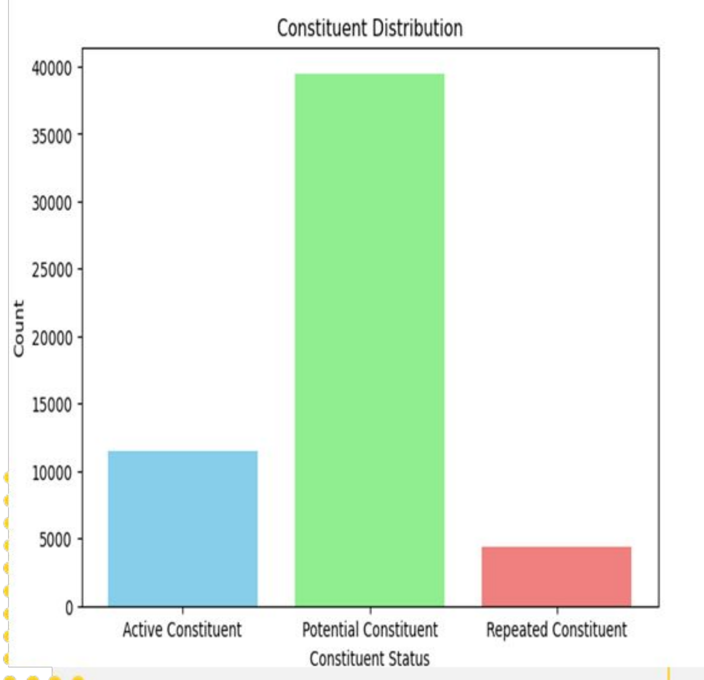
Age emerged as the dominant feature in predicting donation behavior, suggesting that further exploration of age-based donor segmentation may be beneficial.

OBJECTIVE

Move non-active donors to active-donors bucket cost effectively.

NONDONOR BEHAVIOR

Approximately 80% of the dataset consists of non-active donors, representing a substantial portion of the total data. Therefore, it is crucial to implement a targeted campaign strategy to effectively address this segment. This approach will help minimize conversion costs by focusing efforts on turning non-active donors into active ones.



METHODOLOGY

Identify the most influential features of active donors and analyze the distribution of these features within the non-active donor segment.

1. Simple linear relationships (Pearson's correlation).
2. Feature importance (Logistic Regression coefficients).

"Knowing which features have a significant impact allows you to tailor marketing strategies or interventions."

```python
#Find out the correlation between features available and active donors
# Convert categorical data to numeric with `pd.get_dummies`
clean_donors_df_dummy = pd.get_dummies(clean_donors_df)

# Convert any boolean columns (True/False) to integers (1/0)
clean_donors_df_dummy = clean_donors_df_dummy.astype(int)

#Find correlation between activeDoner and other features
correlation_matrix = clean_donors_df_dummy.corr()

# Display the top 20 correlations with 'activeDoner', dropping NaN values and sorting in descending order
top_20_correlations = correlation_matrix['activeDoner'].dropna().sort_values(ascending=False).head(20)

print(top_20_correlations)
```

```
activeDoner                                   1.000000
First gift type_One-time gift                 0.849833
FirstgiftBlock_11-50                          0.692489
First gift fund_Other                         0.434020
First gift fund_Superior Fund                 0.424565
ConstituentCode_Friend                        0.389078
FirstgiftBlock_51-100                         0.384136
FirstgiftBlock_0-10                           0.336650
FirstgiftBlock_101-500                        0.316629
First gift type_Pledge                        0.306444
Marital status_Married                        0.247831
First gift fund_Men's Baseball Fund           0.243317
First gift type_Soft credit one-time gift     0.200729
```

```
Accuracy: 0.8322639780018332
              precision    recall  f1-score   support

           0       0.86      0.94      0.90     11799
           1       0.69      0.48      0.56      3475

    accuracy                           0.83     15274
   macro avg       0.77      0.71      0.73     15274
weighted avg       0.82      0.83      0.82     15274

                                              Feature  Coefficient
0                                            State_AK    -0.551521
1                                            State_AL    -0.056995
2                                            State_AP    -0.412590
3                                            State_AR    -0.002712
4                                            State_AZ     0.125185
..                                                ...          ...
87                               Marital status_Single   -1.757569
88  Marital status_UW-Superior,Primary,Social Work...     0.070917
89                              Marital status_Unknown   -0.596186
90                              Marital status_Widowed    0.060130
91                                 Marital status_None   -1.907107
```

DATA PREPROCESSING

1.Reduce cardinality by consolidating infrequent data values.
2.Convert continuous variables into categorical data to further reduce cardinality.
3. Eliminate feature columns with minimal data points.
4. Add new attributes to the data frame.

```
#Add new column for repeated constituents
constituents_df = constituents_df.withColumn('repeatedConstituents', when(col('Lifetime giving') > col('First gift amount'), 1).otherwise(0))
#Add a new columns to maintain the active donor status
constituents_df = constituents_df.withColumn('activeDoner', when(col('First gift amount') > 0, 1).otherwise(0))

# Extract the ZIP code (before the hyphen) and create a new column 'ZipCode'
constituents_df = constituents_df.withColumn("ZipCode", F.split(constituents_df["Postcode"], "-").getItem(0))
#Create a new column for Constituent codes
# Extract the text before the first "(" and create a new column 'Constituent_Code'
constituents_df = constituents_df.withColumn("ConstituentCode", F.split(constituents_df["Constituent codes"], "\\(").getItem(0))
```

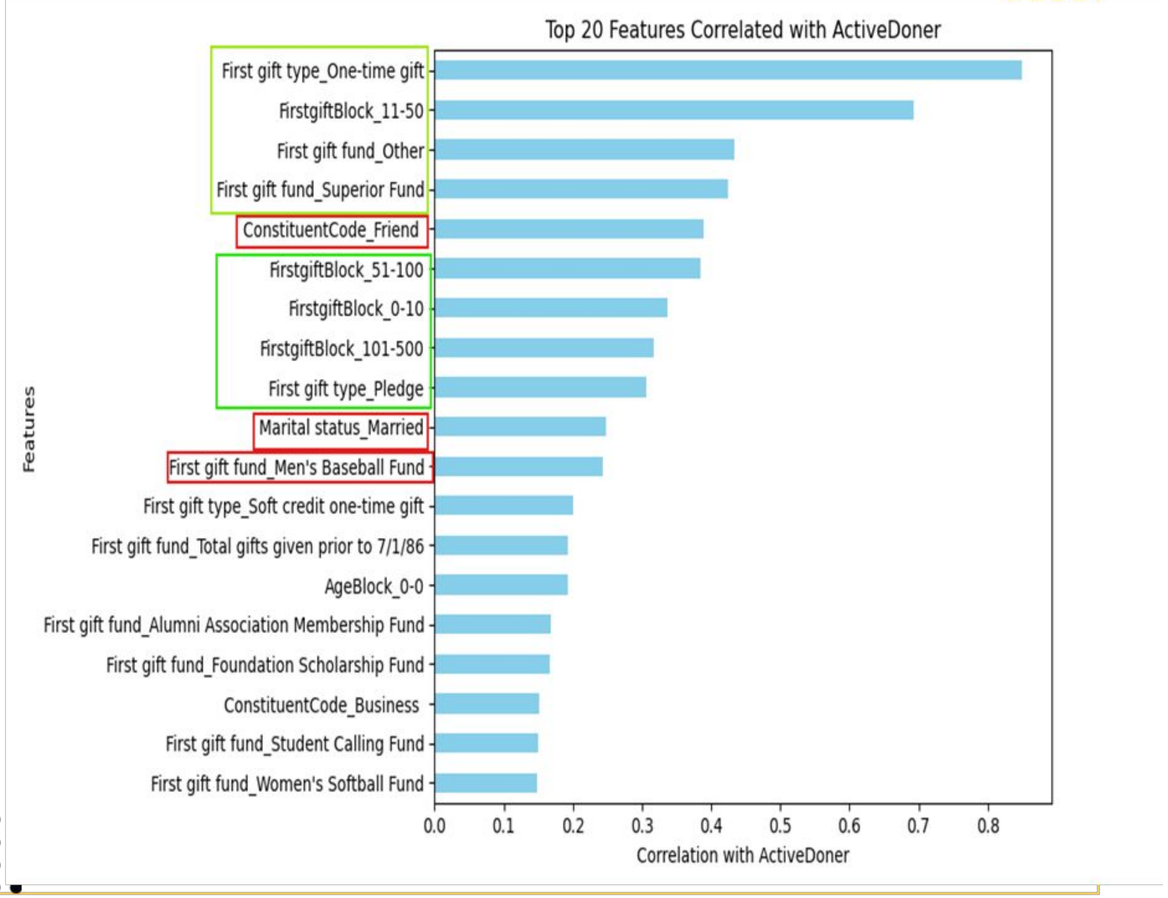
```
# Create a new column 'LifetimegivingBlock' based on the conditions
constituents_df = constituents_df.withColumn(
    "LifetimegivingBlock",
    F.when((F.col("Lifetime giving").isNull()) | (F.col("Lifetime giving") == 0), "0-0")
    .when((F.col("Lifetime giving") > 0) & (F.col("Lifetime giving") <= 10), "0-10")
    .when((F.col("Lifetime giving") > 10) & (F.col("Lifetime giving") <= 50), "11-50")
    .when((F.col("Lifetime giving") > 50) & (F.col("Lifetime giving") <= 100), "51-100")
    .when((F.col("Lifetime giving") > 100) & (F.col("Lifetime giving") <= 500), "101-500")
    .when((F.col("Lifetime giving") > 500) & (F.col("Lifetime giving") <= 1000), "501-1000")
    .when((F.col("Lifetime giving") > 1000) & (F.col("Lifetime giving") <= 5000), "1001-5000")
    .when((F.col("Lifetime giving") > 5000) & (F.col("Lifetime giving") <= 10000), "5001-10000")
    .otherwise("Over 10000")
)


# Create a new column 'AgeBlock' based on the conditions
constituents_df = constituents_df.withColumn(
    "AgeBlock",
    F.when((F.col("Age").isNull()) | (F.col("Age") == 0), "0-0")
    .when((F.col("Age") > 0) & (F.col("Age") <= 20), "0-20")
    .when((F.col("Age") > 20) & (F.col("Age") <= 30), "21-30")
    .when((F.col("Age") > 30) & (F.col("Age") <= 40), "31-40")
    .when((F.col("Age") > 40) & (F.col("Age") <= 50), "41-50")
    .when((F.col("Age") > 50) & (F.col("Age") <= 60), "51-60")
    .when((F.col("Age") > 60) & (F.col("Age") <= 70), "61-70")
    .when((F.col("Age") > 70) & (F.col("Age") <= 80), "71-80")
    .otherwise("Over 80")
)
```

	Column	Null_Count	Null_Percentage
22	AgeInMonths	50909	99.992143
29	Arts	50120	98.442441
33	Women's Athletics	49314	96.859348
30	Fraternal Organizations	49088	96.415454
28	Student Organizations	48785	95.820321
31	Men's Athletics	48342	94.950209
32	Scholarship Recipient	44434	87.274370
27	Solicit codes	44367	87.142773
23	PrimaryEmployment	41070	80.667020
16	Greatest gift amount	39890	78.349341

```
# Look at First gift fund value counts to identify and replace with "Other"
value_counts=clean_donors_df['First gift fund'].value_counts()
value_counts[value_counts > 50]

First gift fund
Superior Fund                              2626
Men's Baseball Fund                         866
Total gifts given prior to 7/1/86           615
Foundation Scholarship Fund                 418
Alumni Association Membership Fund          414
Student Calling Fund                        332
Women's Softball Fund                       326
Women's Basketball Fund                     310
Men's Hockey Fund                           273
Men's Soccer Fund                           270
Women's Volleyball Fund                     269
Men's Basketball Fund                       268
Gigliotti, Tony, Scholarship                247
Women's Soccer Fund                         232
Alumni Association Fund                     185
Women's Athletics Fund                      184
Track and Cross Country Fund                168
Athletics Fund-General                      162
Women's Hockey Fund                         146
0-Unassigned For Use In History              67
Women's Tennis Fund                          65
Nicholson, G.W., Scholarship                 64
Education Department Development Fund        56
Men's Cross Country/Track Fund               55
Alumni Association Registration Fees Fund    53
```

PEARSON CORRELATION

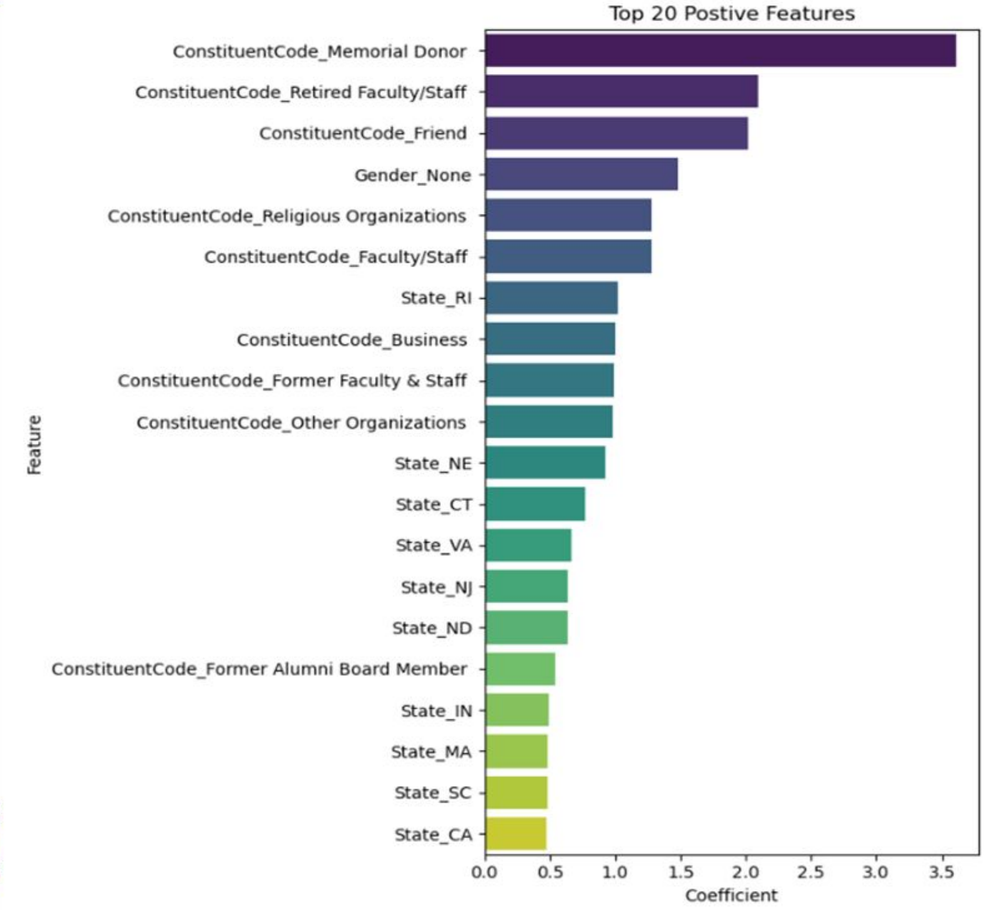
1.Pearson correlation coefficient considering all the attributes.



LINER REGRESSION MODLE AND COEFFICIENT

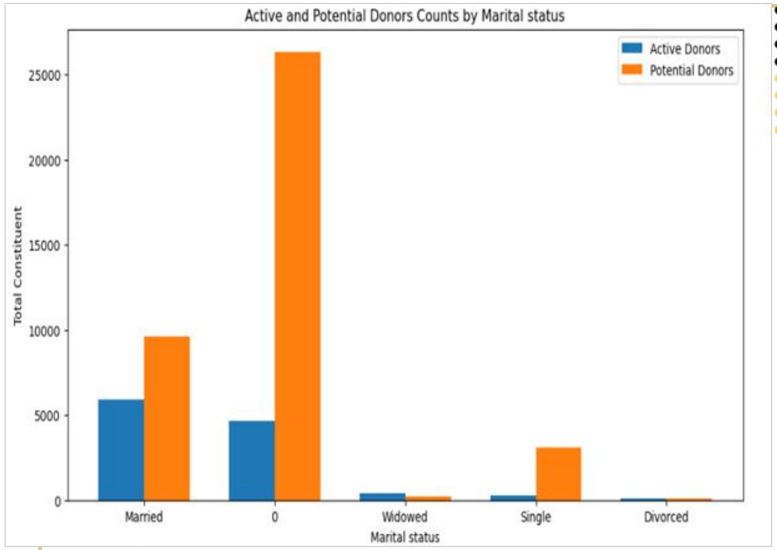
3.Pearson correlation coefficient considering all the attributes.

```
Accuracy: 0.8322639780018332
              precision    recall  f1-score   support

           0       0.86      0.94      0.90     11799
           1       0.69      0.48      0.56      3475

    accuracy                           0.83     15274
   macro avg       0.77      0.71      0.73     15274
weighted avg       0.82      0.83      0.82     15274

                                          Feature  Coefficient
0                                        State_AK    -0.551521
1                                        State_AL    -0.056995
2                                        State_AP    -0.412590
3                                        State_AR    -0.002712
4                                        State_AZ     0.125185
..                                            ...          ...
87                           Marital status_Single   -1.757569
88  Marital status_UW-Superior,Primary,Social Work...  0.070917
89                          Marital status_Unknown   -0.596186
90                          Marital status_Widowed    0.060130
91                             Marital status_None   -1.907107
```

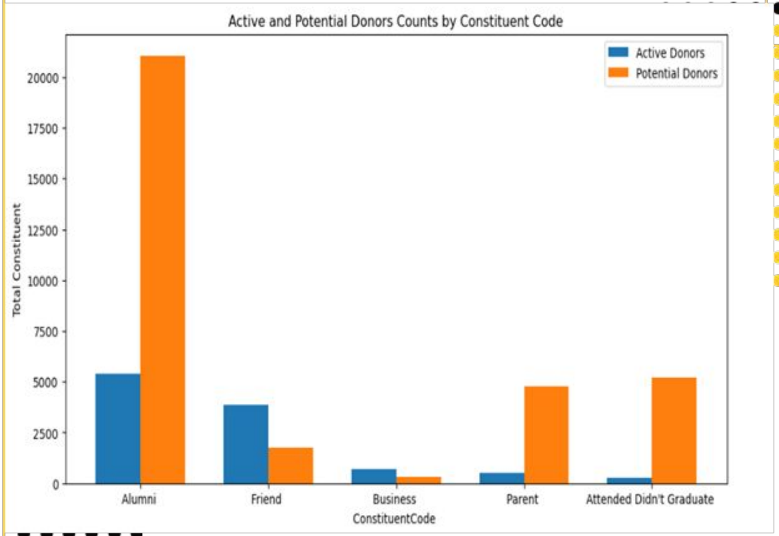



```python
# Selecting features based on correlation analysis
# Adjust the feature list based  correlation analysis
features = ['State','ConstituentCode','AgeBlock','Gender','Marital status']
target = 'activeDoner'
```

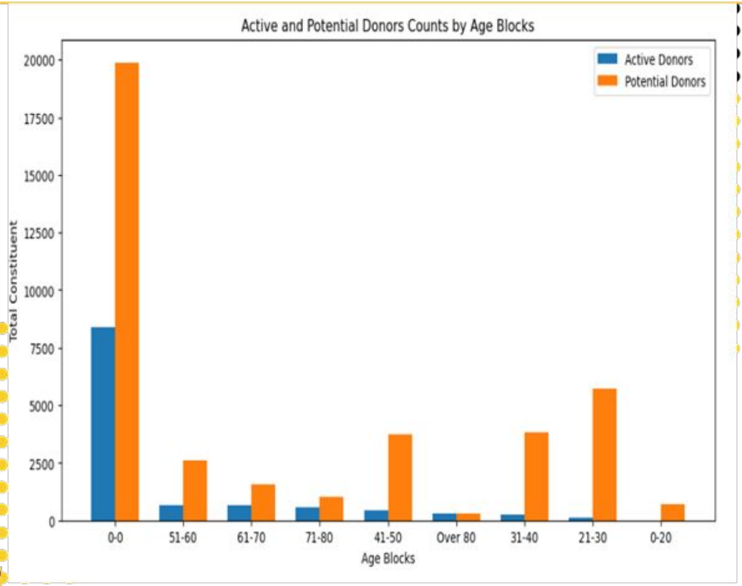
VISUALS



Active and Potential Donors Counts by Marital status
Active Donors
Potential Donors
Total Constituent
25000
20000
15000
10000
5000
0
Married
0
Widowed
Single
Divorced
Marital status



Active and Potential Donors Counts by Constituent Code
Active Donors
Potential Donors
Total Constituent
20000
17500
15000
12500
10000
7500
5000
2500
0
Alumni
Friend
Business
Parent
Attended Didn't Graduate
ConstituentCode



Active and Potential Donors Counts by Age Blocks
Active Donors
Potential Donors
Total Constituent
20000
17500
15000
12500
10000
7500
5000
2500
0
0-0
51-60
61-70
71-80
41-50
Over 80
31-40
21-30
0-20
Age Blocks

CONCLUSION

Key Insights and Next Steps

- Donor Engagement: The data highlights that involvement in campus activities, such as Student Organizations, Arts, Fraternal Organizations, and Athletics, influences the likelihood of alumni giving. Age also emerged as an important factor in donor behavior.
- Targeted Campaigns: By focusing on these key donor attributes, the institution can tailor campaigns to specific donor groups, leading to more sustained and impactful support.
- Misc. Recommendations:
  1. Data Gaps: Address missing or incomplete data (e.g., age and donor attributes) to enhance the accuracy of future models.
  2. Data Cleanup: Refine and update the dataset, then rerun the analysis to improve model performance.
  3. Ongoing Updates: Regularly refresh models as new or updated data becomes available to ensure strategies remain aligned with current trends.

Strategies to consider

- **Target Alumni** aged 40-60 who participated in campus activities for a focused engagement campaign
- **Target Mid-Career Alumni**: The data suggests that donors aged 40-60 are the most engaged group. This cohort is likely at a point in their careers where they have the financial means to contribute, making them an ideal group for targeted fundraising campaigns.
- **Leverage Student and Cultural Groups**: Since **Student/Fraternal Organizations** and **Arts** have high participation rates, marketing efforts for donations should emphasize the legacy of these groups and how donations can help support them for future students.
- **Athletics Appeal**: Alumni who participated in athletics may be more likely to contribute to sports-focused fundraising efforts, such as facility upgrades or team sponsorships.
- **Gratitude from Scholarship Recipients**: Since a significant portion of donors were scholarship recipients, highlighting the impact of scholarships in campaigns can encourage more giving from alumni who benefited from similar programs
- **Monitor and Adjust**: Regularly monitor the effectiveness of the strategies and adjust based on new data, feedback, and results.

Target non-donor alumni aged 40-60

Target alumni aged 40-60 who participated in campus activities for a focused engagement campaign:

There are 6971 (13% of total data) non-donor Alumni between the age 40 - 60 and engaged in different campus activities such as 'Student Organizations', 'Arts', 'Fraternal Organizations', and athletics, and this could be an ideal target group for further analysis.

```
        Age  Student Organizations Arts  Fraternal Organizations  \
24     50.0                    306    0                       43
28     43.0                    306    0                       43
47     41.0                    306    0                       43
56     60.0                    306    0                       43
57     44.0                    306    0                       43
...     ...                    ...  ...                      ...
50900  58.0                    306    0                       43
50901  56.0                    306    0                       43
50904  53.0                    306    0                       43
50905  57.0                    306    0                       43
50907  49.0                    306    0                       43

           Men's Athletics                                          Education
24                       0   UW-Superior,Primary,EdAd: Principalship,MSE,2019
28                       0  UW-Superior,Primary,Computer Science,Economics...
47                       0     UW-Superior,Primary,Communicating Arts,BS,2013
56                       0  UW-Superior,Primary,Guidance and Counseling,No...
57                       0   UW-Superior,Primary,Elementary Education,BS,2003
...                    ...                                                ...
50900                    0  UW-Superior,Primary,Physical Education,No degr...
50901                    0  UW-Superior,Primary,Comm Arts: Journalism,Mass...
50904  Baseball (No date)  UW-Superior,Primary,Physical Education,No degr...
50905  Baseball (No date)  UW-Superior,Primary,Physical Education,Special...
50907                    0   UW-Superior,Primary,Elementary Education,BS,2017

[6971 rows x 6 columns]
```

UNIVERSITY of WISCONSIN Superior

Appendix: Future Analysis - Parking Lot

In order to deepen understanding of donor behavior, next steps include analyzing:

- Detailed Gift Analysis - Frequency of Gifts, Gift Size, Gift Recency, Gift Type, lifetime giving, gift growth over time, or specific combinations therein.
- Education - For alumni, analyzing degree level, major, and gift details.
- Solicitation Type Preferences (analyze the success rate of those solicitation types - e.g. online, mail, postcard, email.
- Participation in Multiple Activities - we analyzed participation in multiple activities, but perhaps analyzing specific combinations of activities might be more strongly correlated with donation behavior.
- Loyalty - total years of giving, total number of gifts, etc.