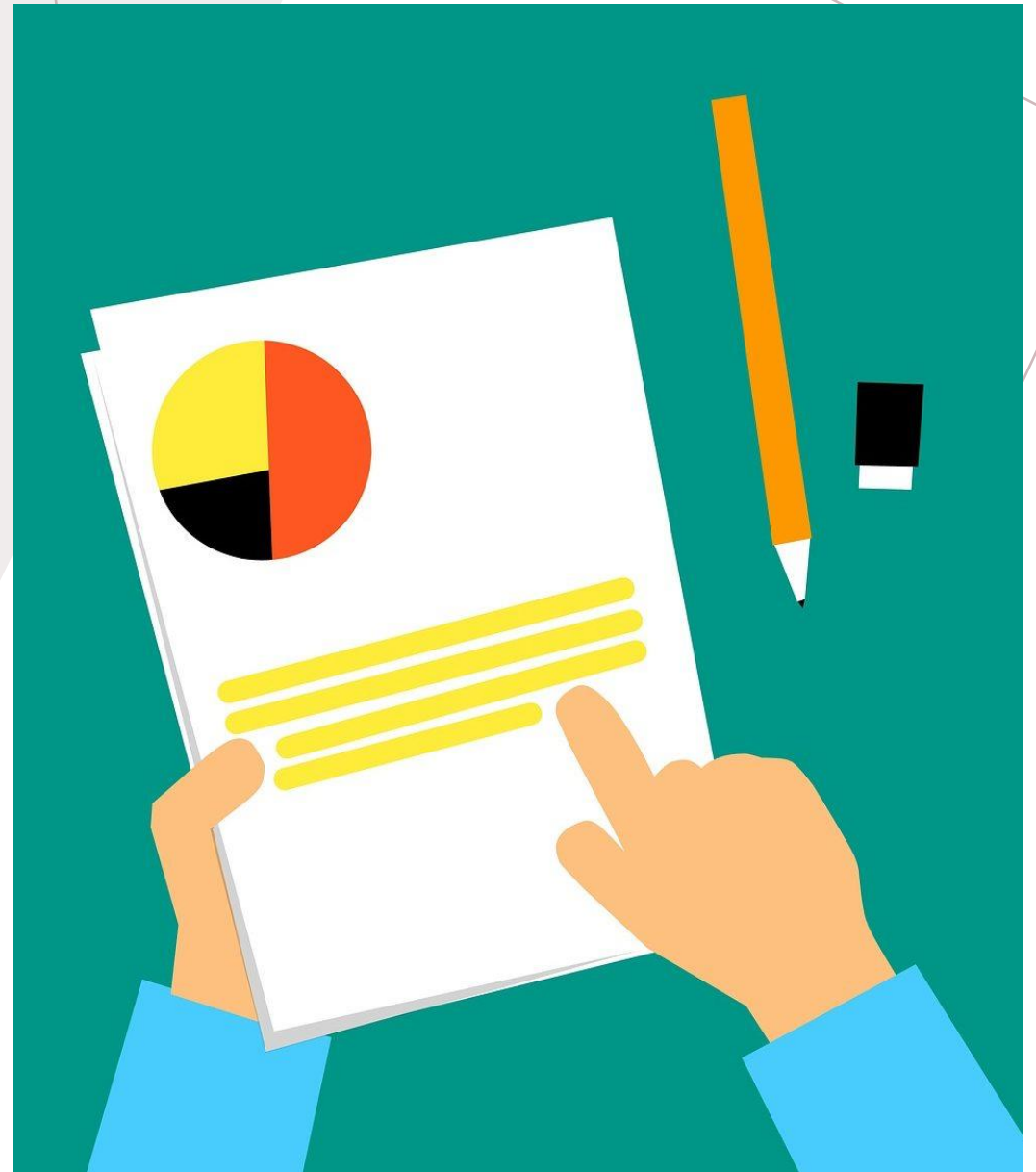
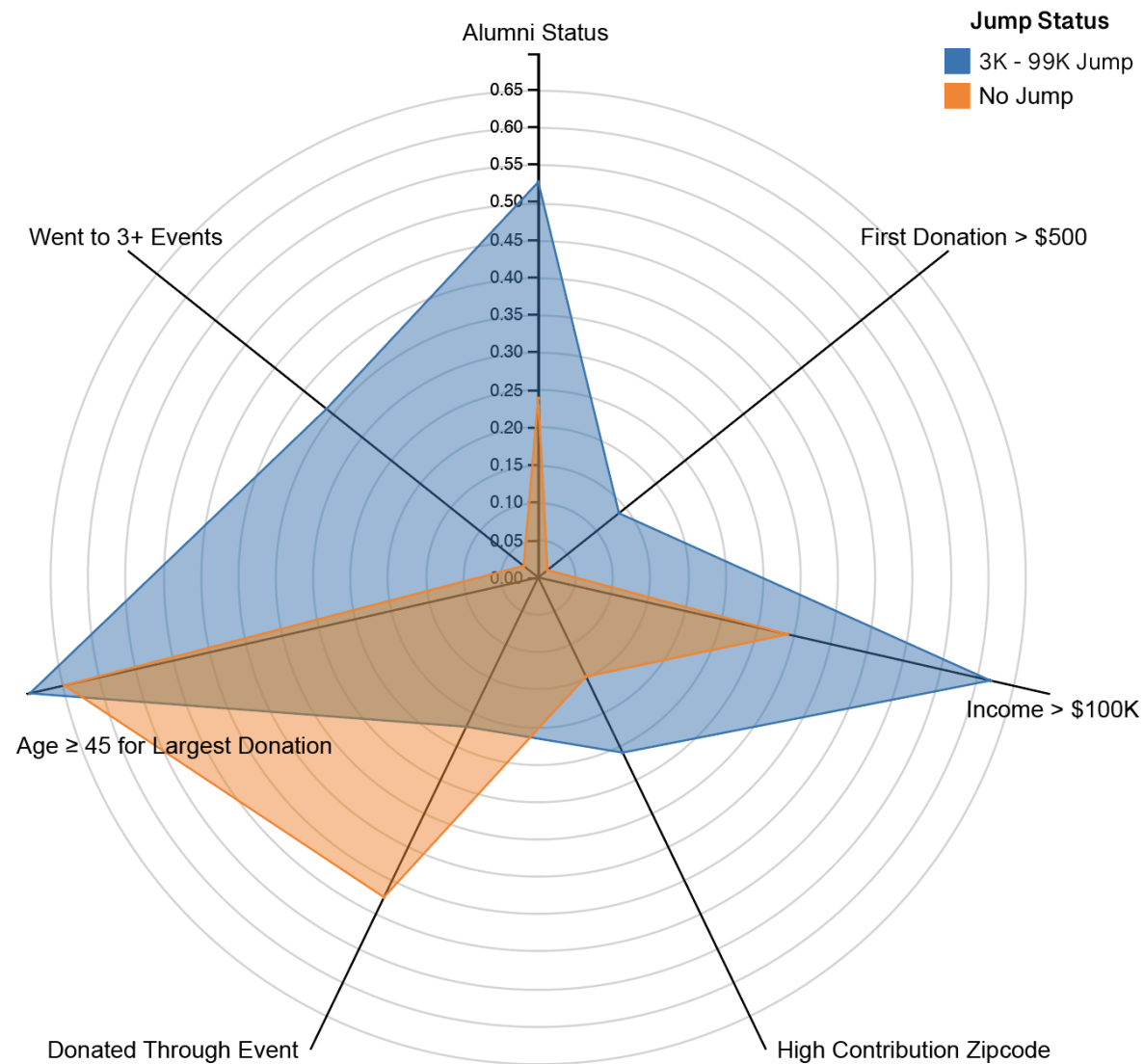


DONOR GIVING INSIGHTS REPORT

PREDICTING A 3K - 99K JUMP

BY GROUP 4:
CAROLINE PIER, XIDAN KOU,
CARLI WERNER TROY STEIN, SONGYUAN
WU, AND KAT HUSAR





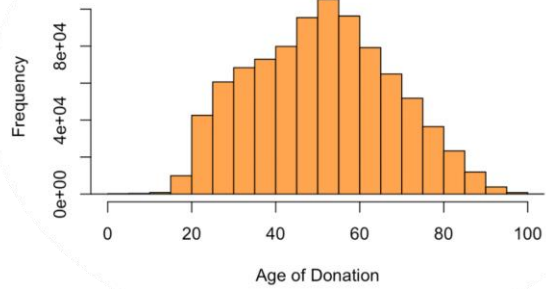
Composition of a Jump

Proportional Compositions by Jump Status

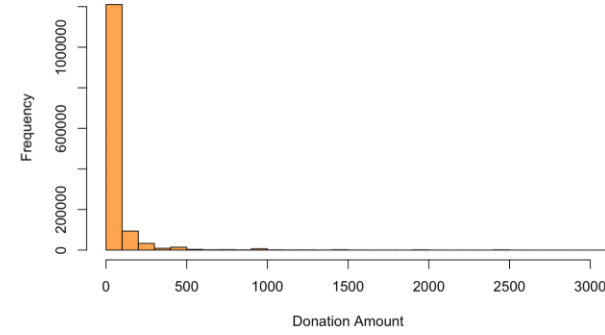
- Notable differences between whether they went to three or more events, income level greater than \$100,000 per year, first donation being greater than \$500, high contribution zip-code (Columbus Area), and OSU alumni status.
 - These factors were later considered in modeling.
- The ages of the largest donation were similar for both groups.

Composition of a Jump

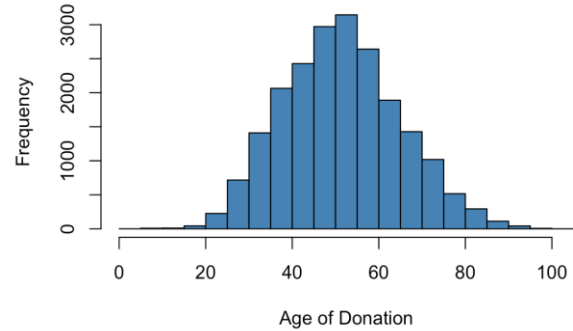
Age of Largest Donation for those who did Not Jump



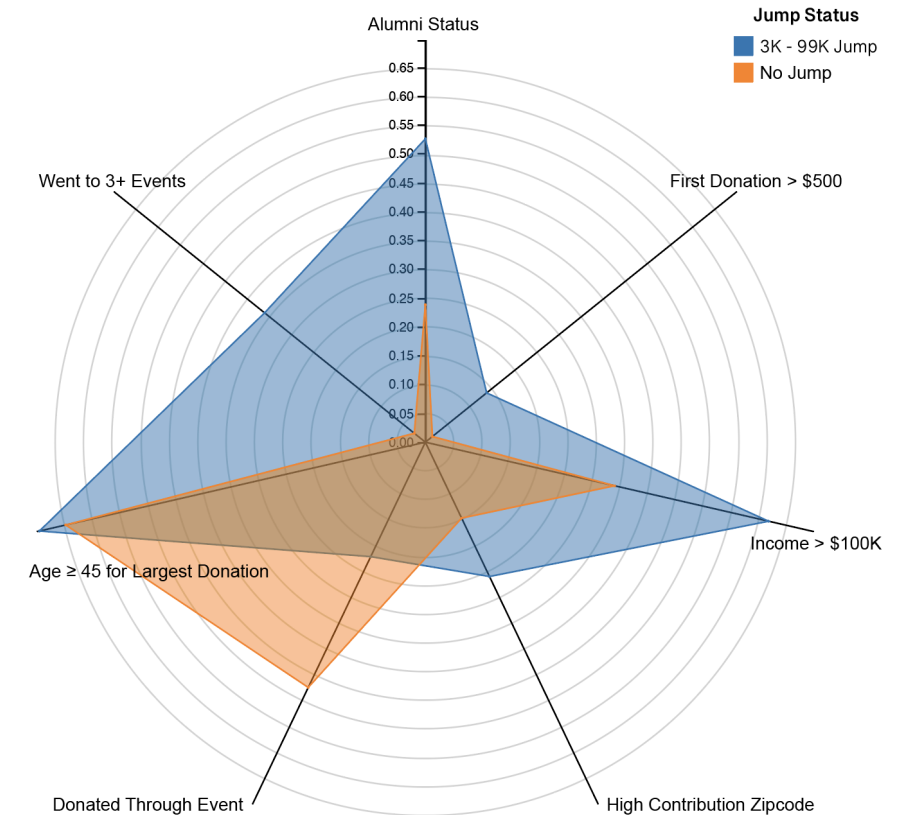
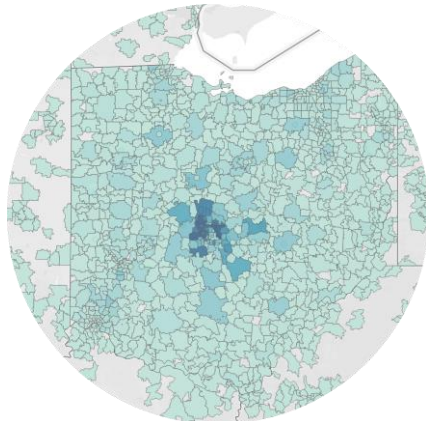
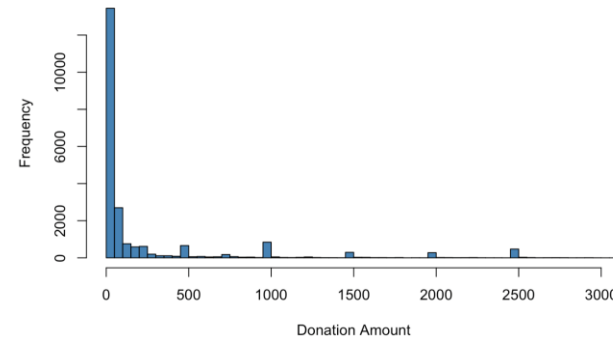
First Gift for No Jump Donors

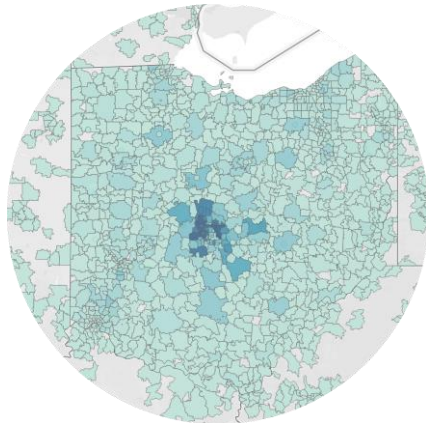
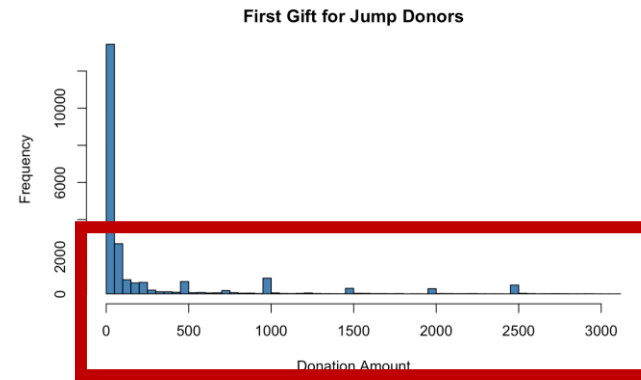
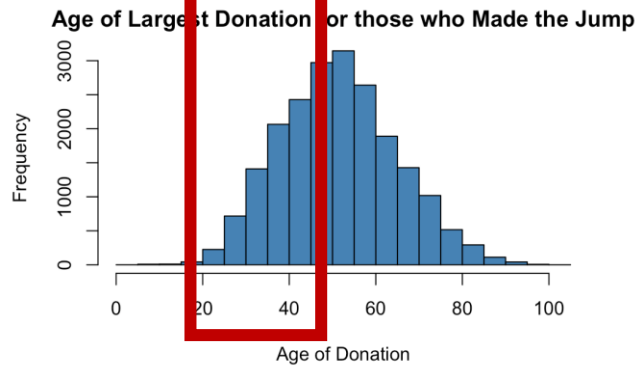
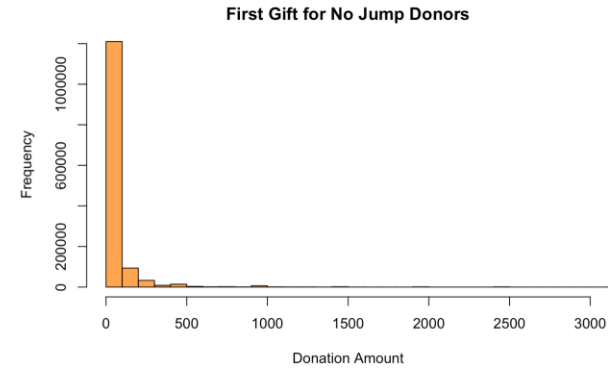
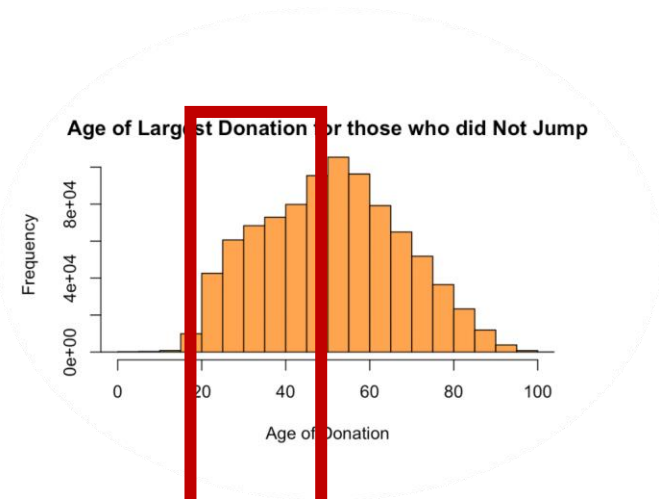


Age of Largest Donation for those who Made the Jump

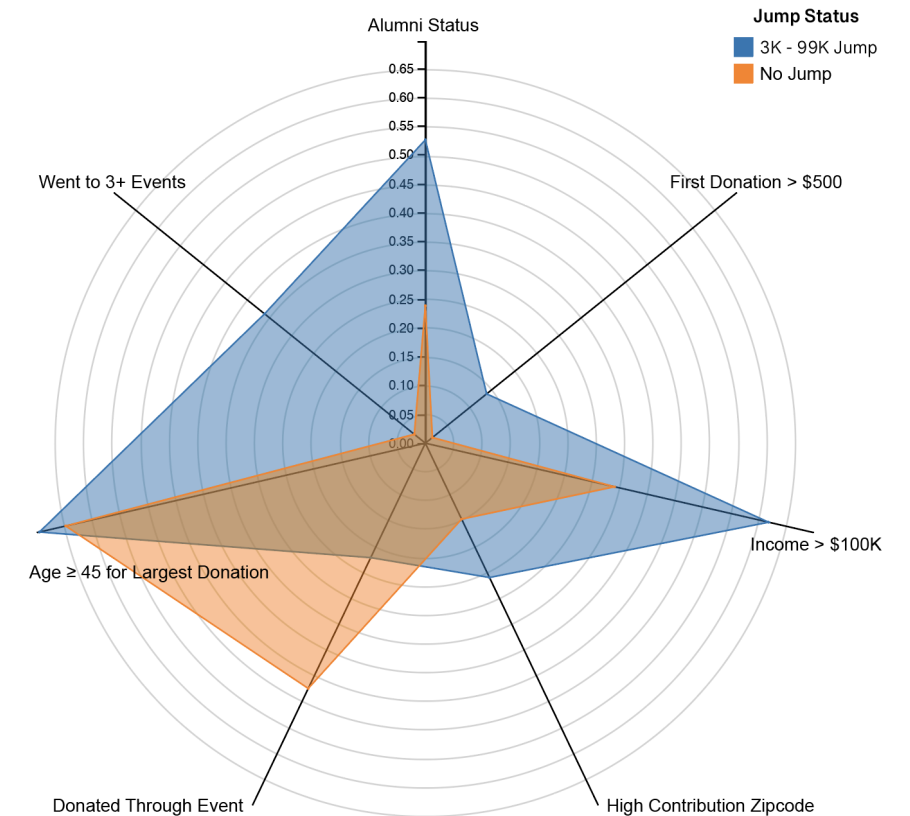


First Gift for Jump Donors

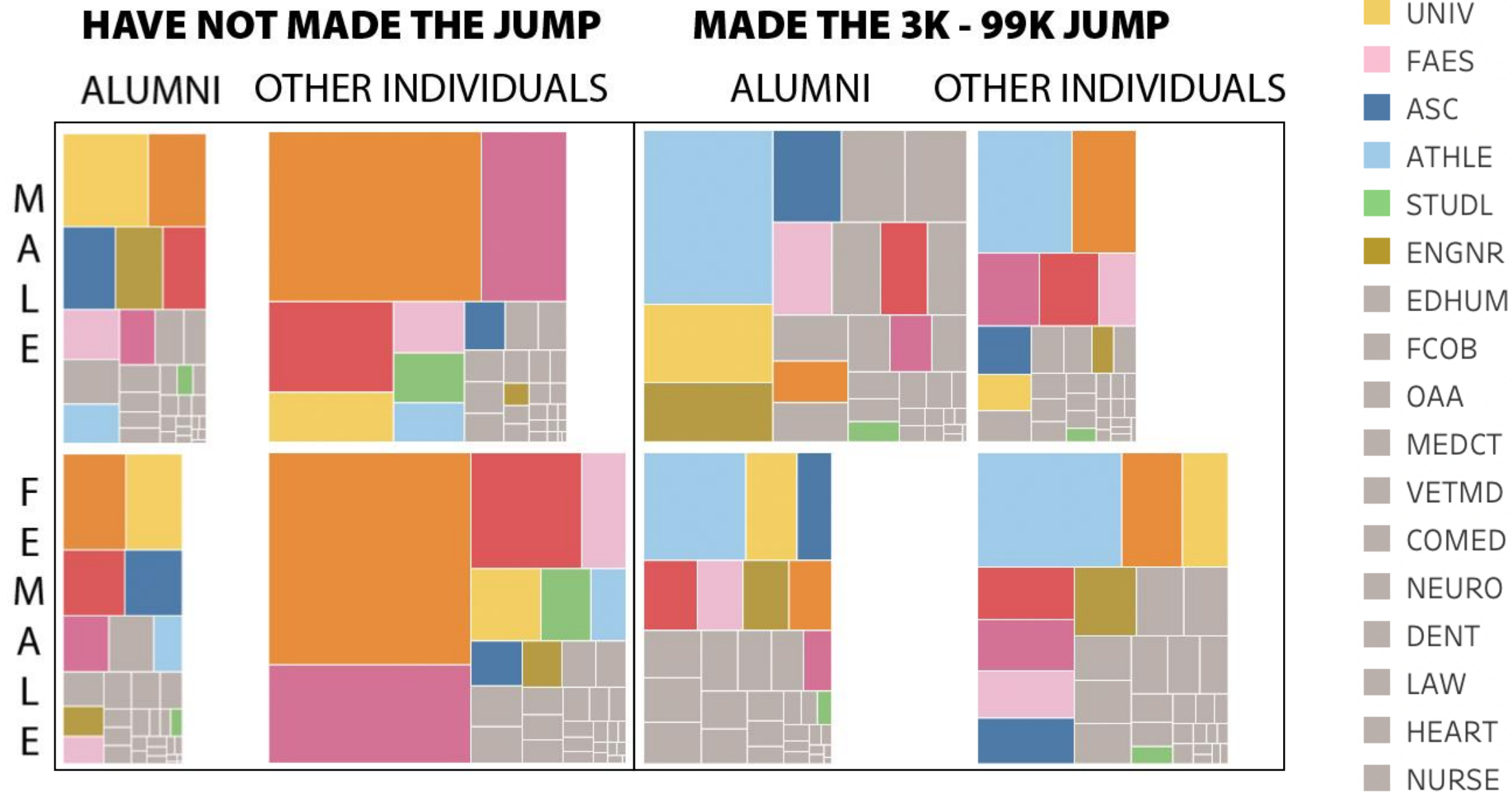




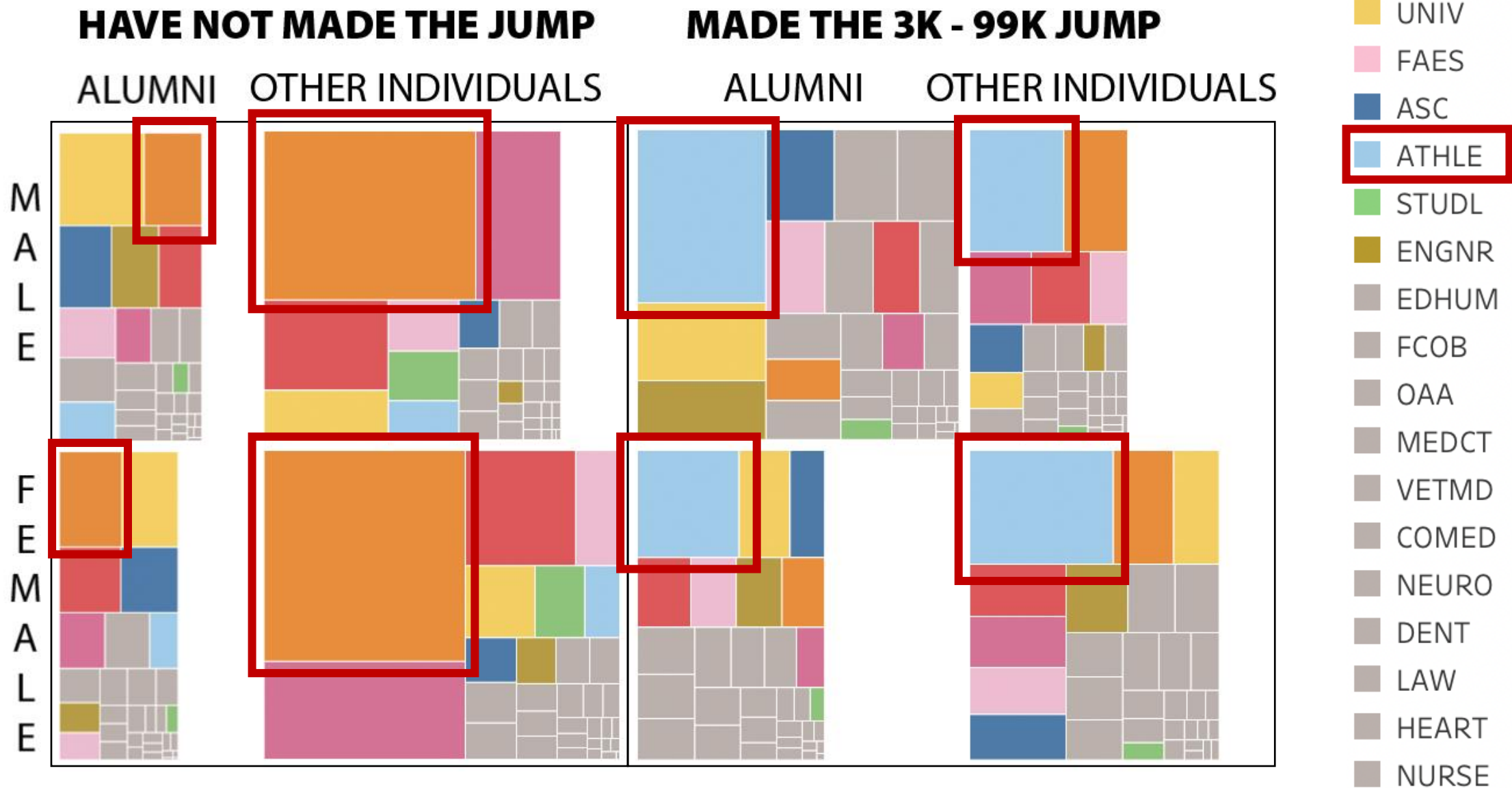
Composition of a Jump



Top Giving Unit by Constituency and Gender



Top Giving Unit by Constituency and Gender



Alumni Feature Selection



THE
MOTIVATION



THE
PROCEDURE



THE RESULTS



THE INSIGHTS

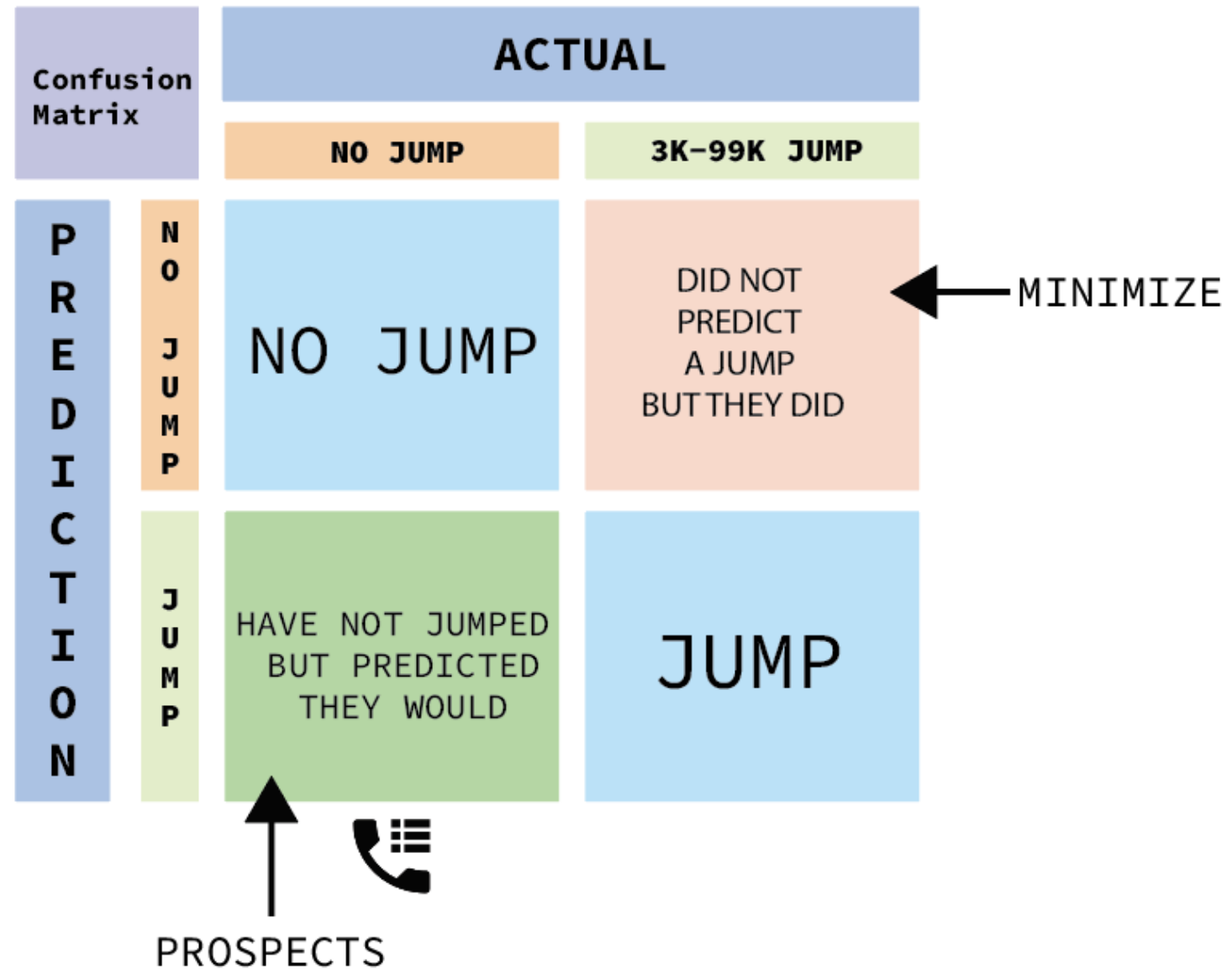
Alumni Model Motivation

- Alumni may be more likely to give because of their personal experiences at OSU
- Considered subset of only Alumni to investigate the impact of these experiences on 3k to 99k donor transitions
- Goal: Identify demographics that play an important role in classifying 3k to 99k donors

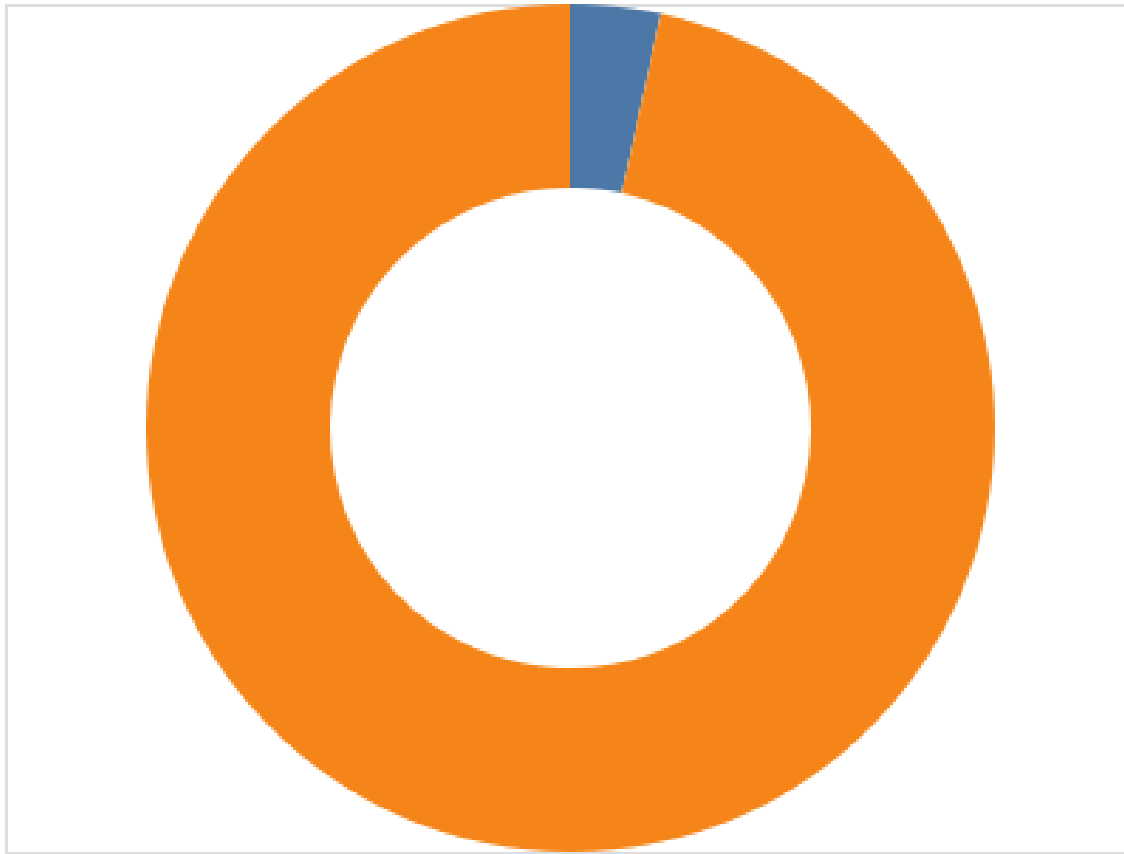
Random Forest Procedure

- Predictor variables: any variable that describes someone's involvement both currently and when he/she was a student
- Start with full predictor list -> eliminate variables with low importance via random forests. Attempted to:
 - Achieve high f1 score in minority group while minimizing false negatives
 - Eliminate variables causing high node impurity (low importance)

Model Building Mentality



3k to 99k Donor Status Class Imbalance



Category

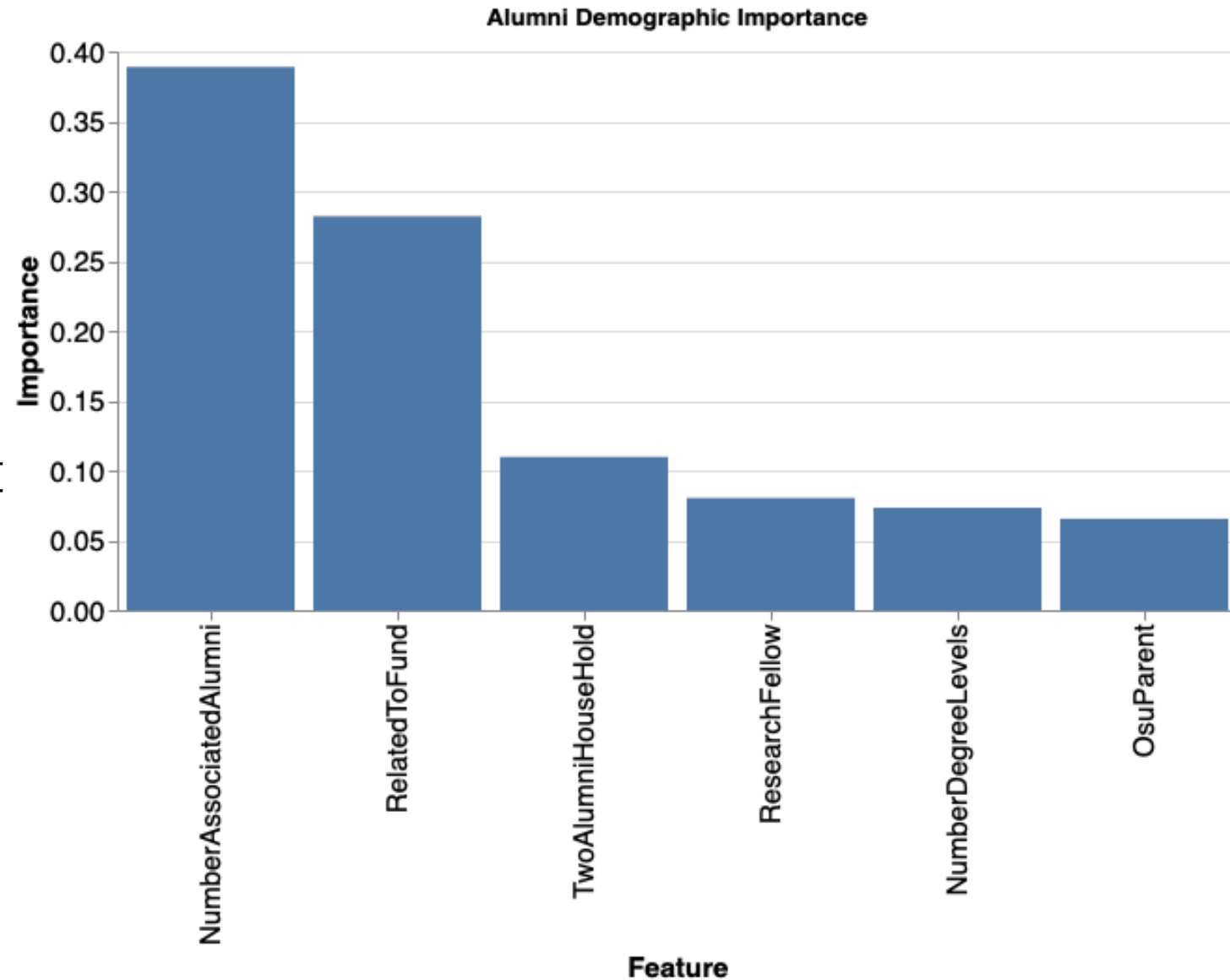
● 3k to 99k ● Not 3k to 99K

Addressing Class Imbalance

- Down-sampled majority class
 - 3:1 ratio
- Trained on down-sampled data
- Tested on original data

Final Alumni Random Forest Model

- Built using the 6 variables with the highest importance



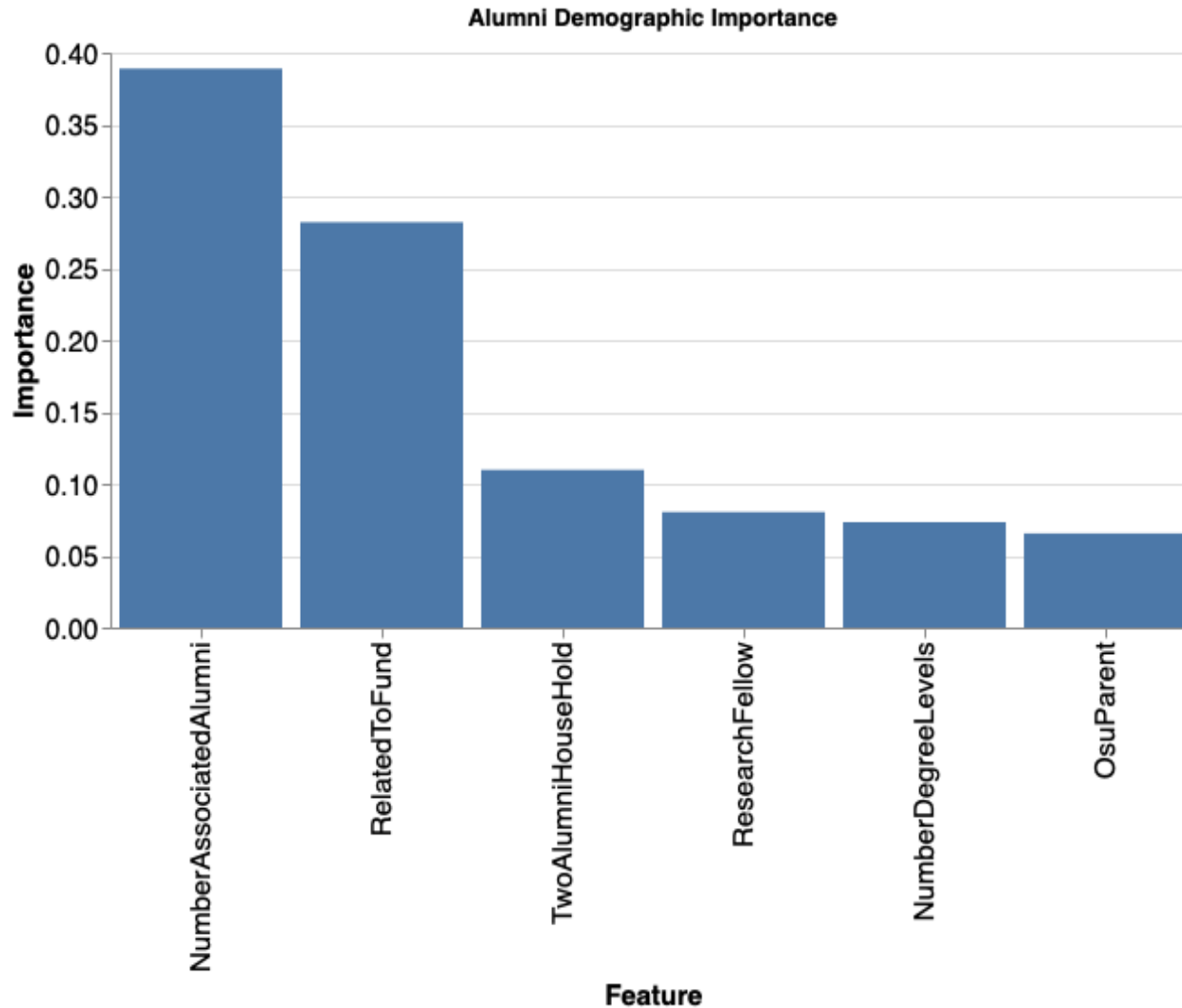
Results

- Potentially Identified over 1000 3k to 99k Donors
- But, also misclassified 873 current 3k to 99k donors
- Final jumper f1 score: 0.25

Random Forest Confusion Matrix		ACTUAL	
		NO JUMP	3K-99K JUMP
P R E D I C T I O N	N O J U M P	93.45% 32,116	2.54% 873
	J U M P	3.09% 1061	0.92% 316

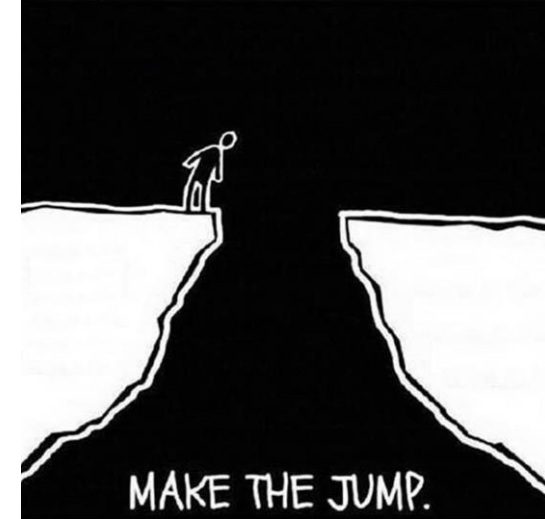


Insights



- Model performance isn't great – but this is still useful
- The model including all alumni demographics was still poor, so let's shift our focus to the few that carried their weight
- Target "OSU Families"
 - Alumni parents + OSU student children + Alumni Friends
- Target research fellows
- Target people related to funds

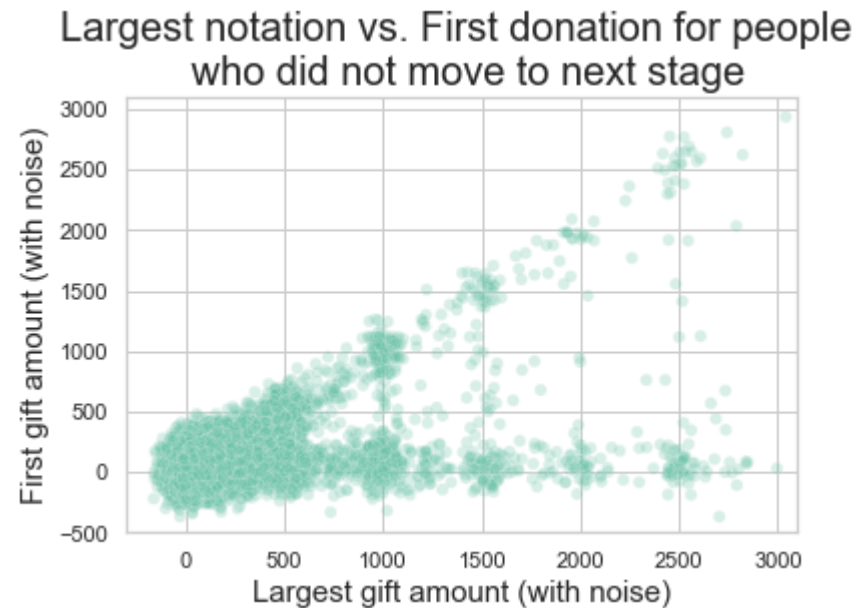
First Donation vs. Largest Donation Motivation



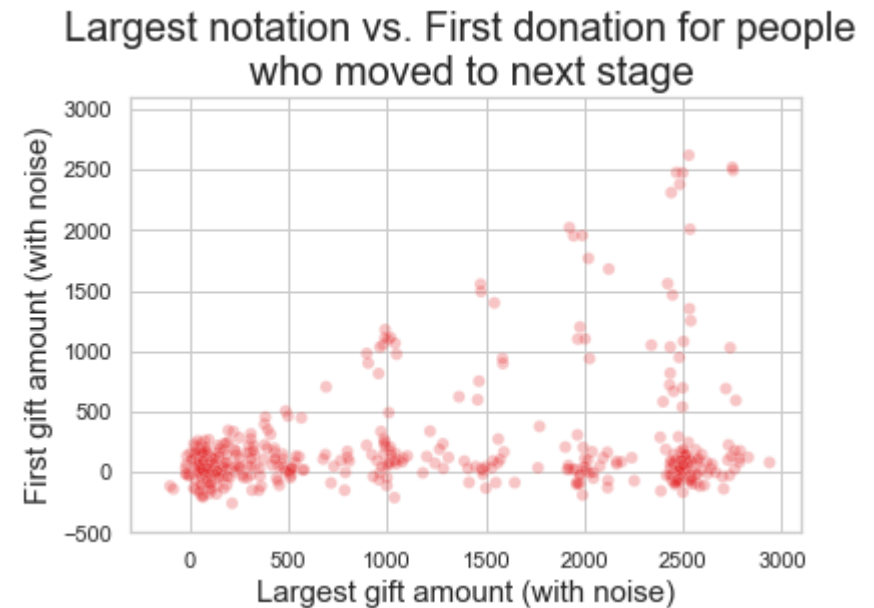
-
- Can we predict that a person is ready to make the jump if they donate more now than what they started with?
 - Is the largest amount donated for people who made the jump close to 3k?

First Donation vs. Largest Donation

GIVEN THE FIRST DONATION WAS UNDER \$500, 0.3% DONATED OVER \$2,000

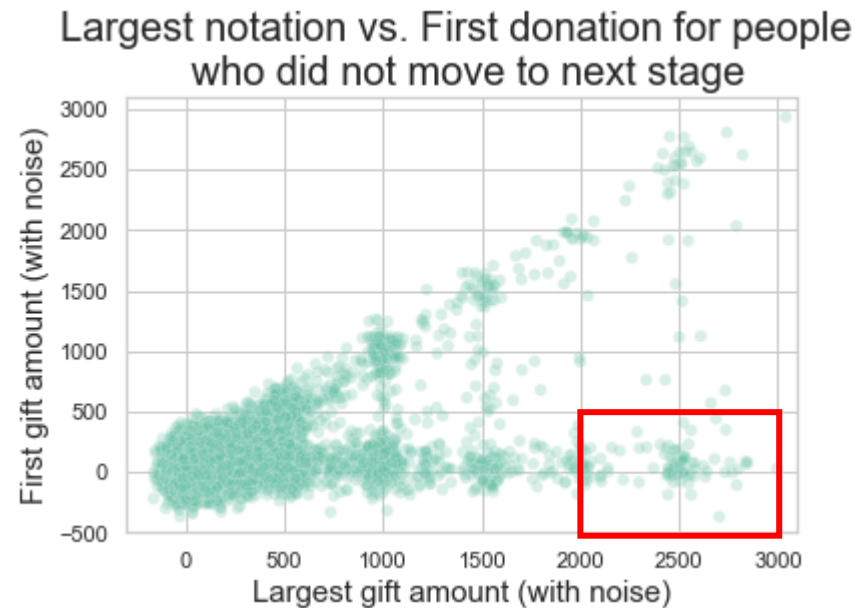


GIVEN THE FIRST DONATION WAS UNDER \$500, 23.5% DONATED OVER \$2,000

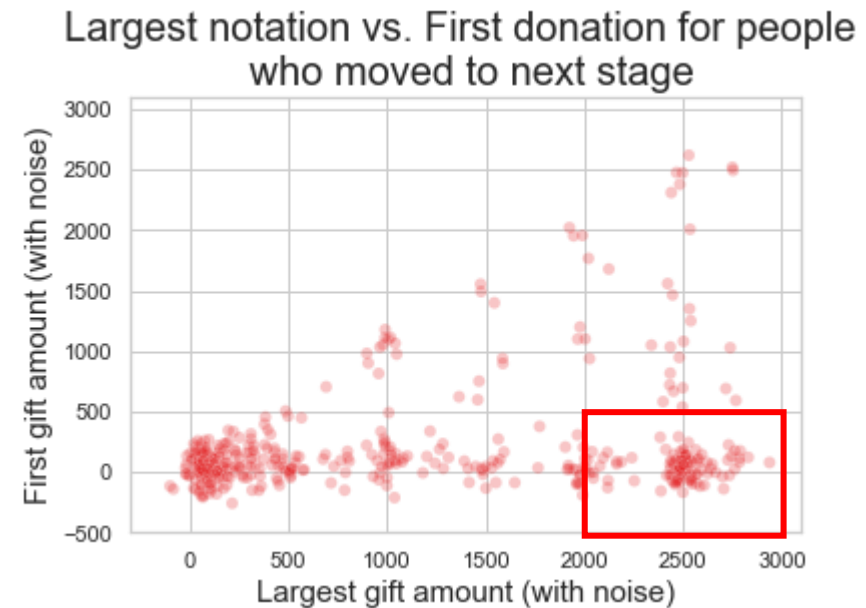


First Donation vs. Largest Donation

GIVEN THE FIRST DONATION WAS UNDER \$500, 0.3% DONATED OVER \$2,000



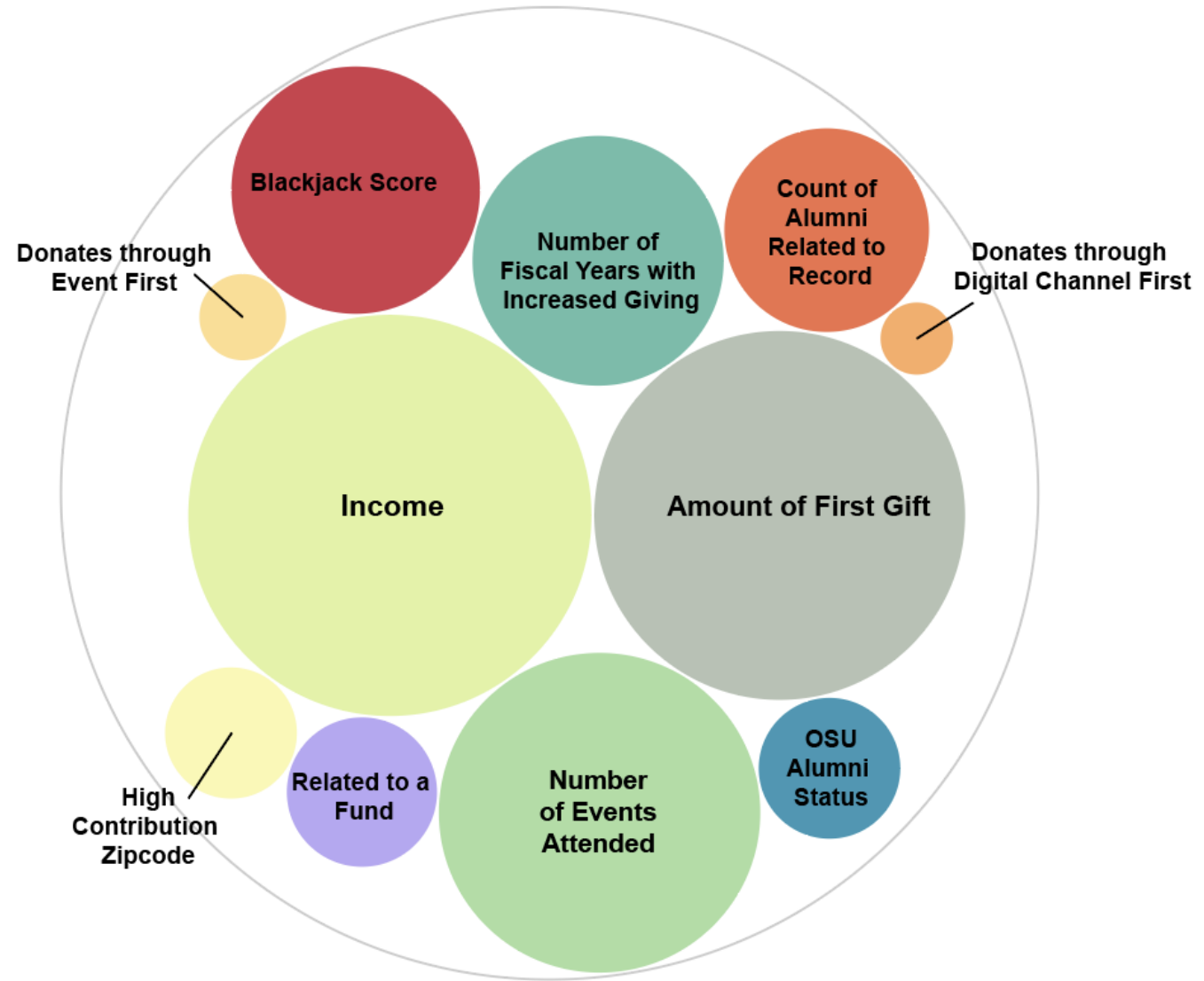
GIVEN THE FIRST DONATION WAS UNDER \$500, 23.5% DONATED OVER \$2,000



Random Forest Predictive Variables

Scaled by Decrease in Model Accuracy
if the Variable was Removed from the
Model

- *Mean Decrease Gini*
 - Model Purity



Logistic Regression Model

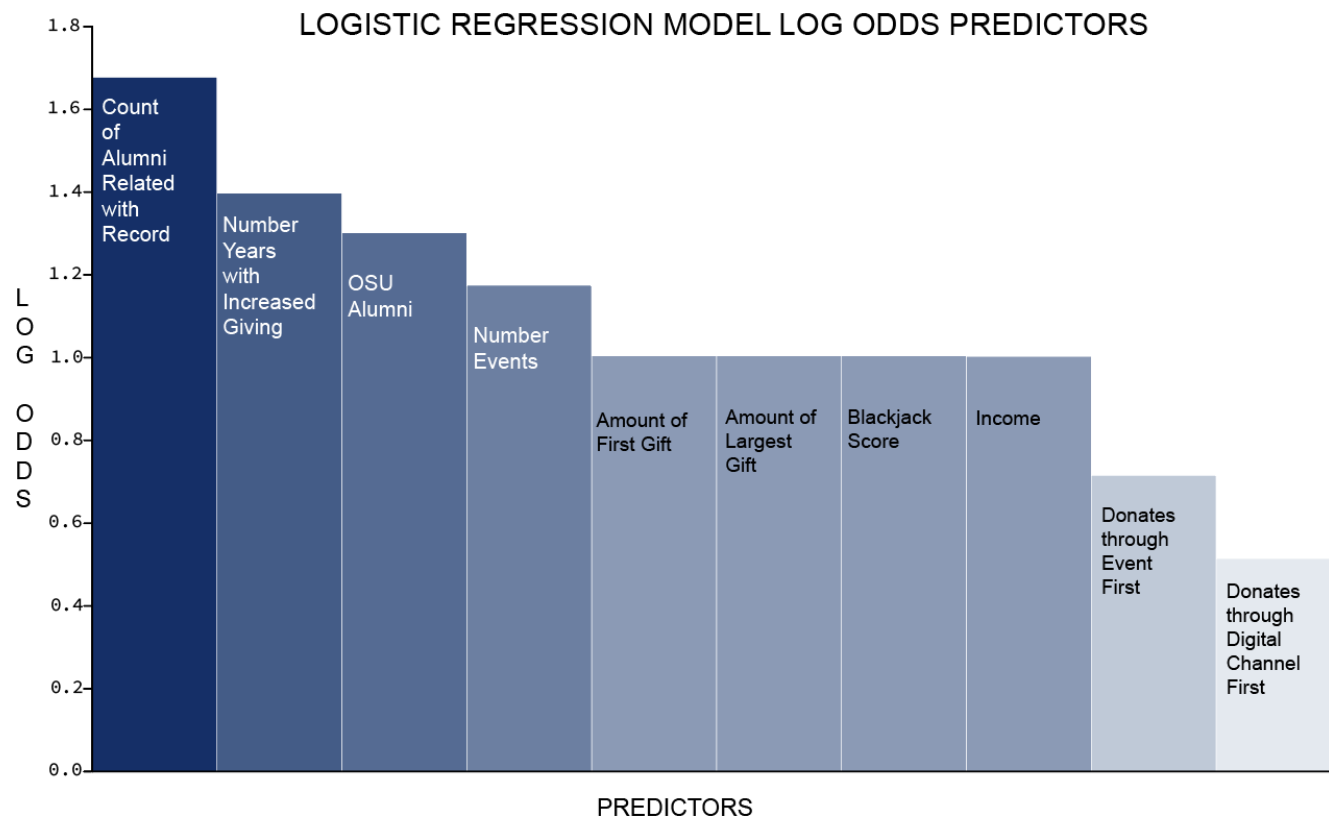
1:3 ratio Jump:No Jump

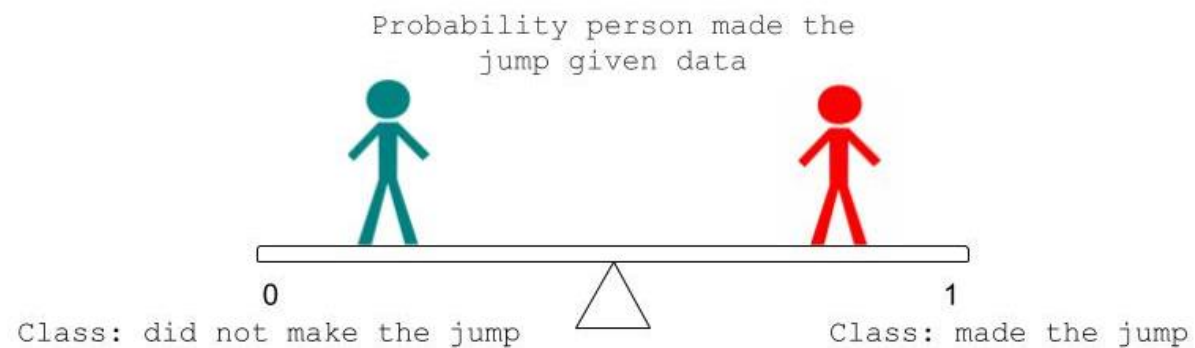
Trained on a 10,000 Jump to 30,000 No Jump Ratio

Tested on a 3,000 Jump to 9,000 No Jump Ratio

Labeled a Correct Prediction if Scaled Probability was >0.5

Logit Model Confusion Matrix		ACTUAL	
		NO JUMP	3K-99K JUMP
P R E D I C T I O N	N O J U M P	72.23% 8679	10.56% 1267
	J U M P	2.68% 321	14.44% 1733





Given the characteristics of a person

- What is the probability they make a jump?
- What is the probability they do not make a jump?

Classify them based on higher probability

Naïve Bayes Model

Naïve Bayes Model

1:3 ratio (oversampling)


Goal: low false negatives (those who made the jump should not be misclassified)

How? Aim for higher recall

Our model results:


- average cross validation recall: 67%
- average cross validation accuracy: 95%

Naive Bayes Confusion Matrix		ACTUAL	
		NO JUMP	3K-99K JUMP
P R E D I C T I O N	N O J U M P	93.26% 4827	0.05% 26
	J U M P	5.18% 268	1.06% 55



Linear SVM model for Donations by Fiscal Years(Good but not good enough)

Linear SVM Confusion Matrix		ACTUAL	
		NO JUMP	3K-99K JUMP
P R E D I C T I O N	N O J U M P	676,132	5,046
	J U M P	700,000	17,289



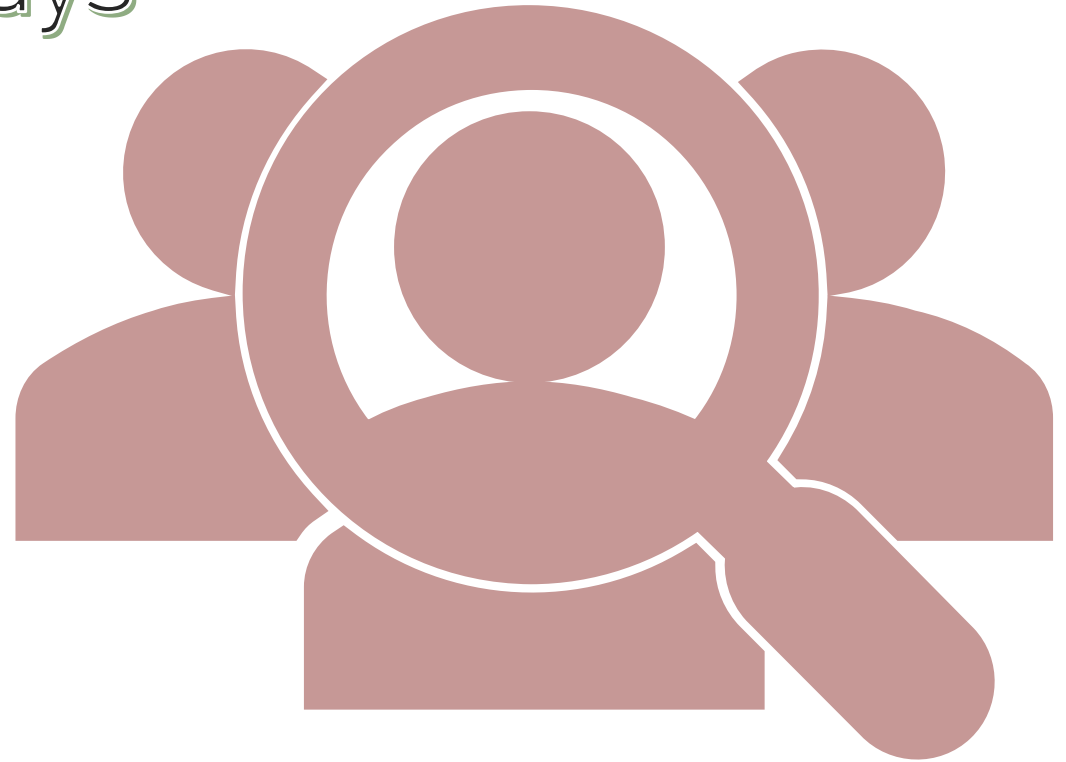
- Very low accuracy (18%)
- f1-scores .03 for people who made jump .29 for people who didn't make jump
- 700k have been identified as having potential to move up

False-Positive: 50.1% (Target)
False-Negative: 0.4%
True-Positive: 1.2%
True-Negative: 48.3%

Donor's First Donation ways

Questions to answer :

- How are donor's first donation ways affect the jump in the future ?
- What suggestions can we make to let more people to jump

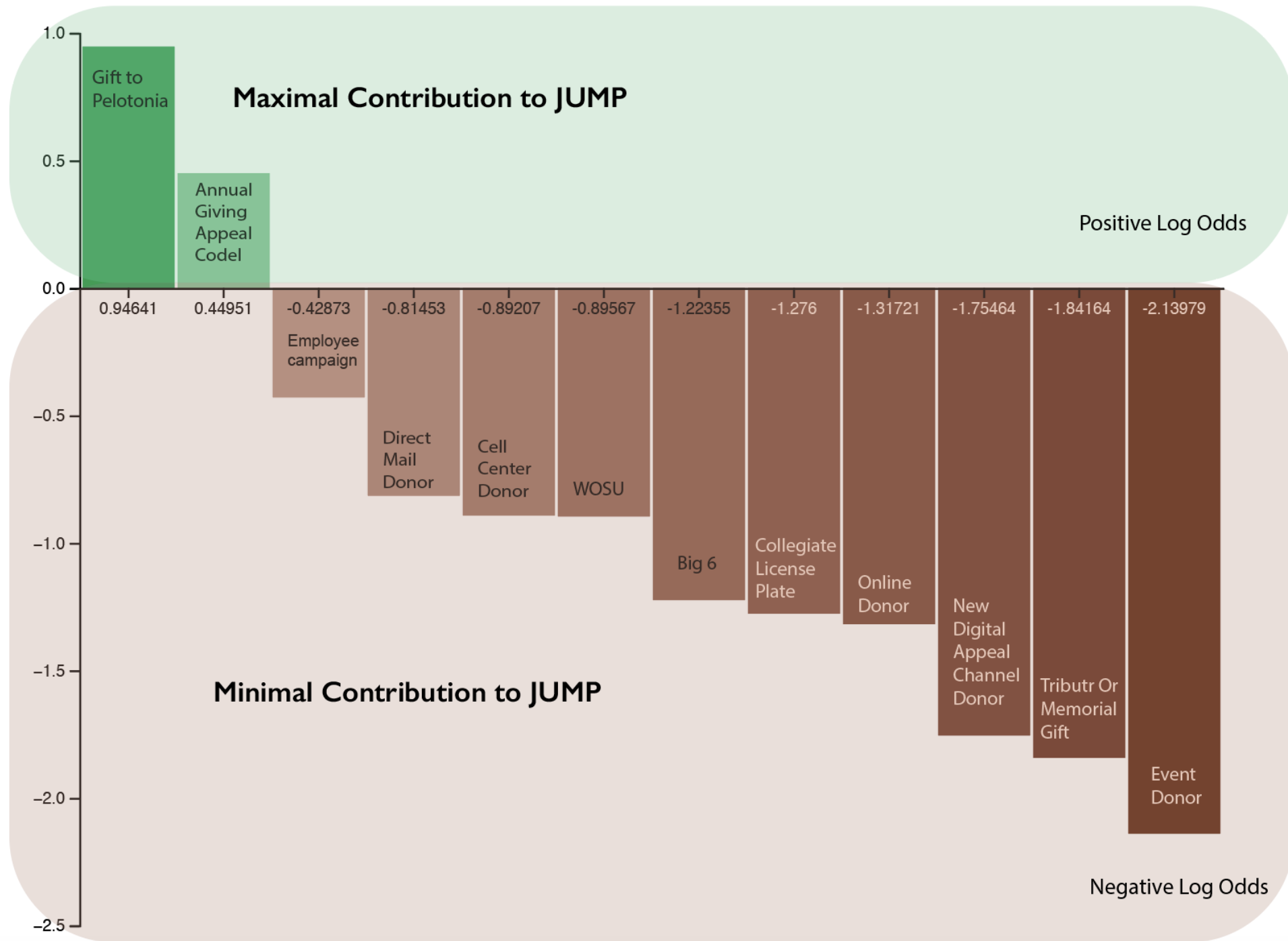


Maximal
Contribution
Variables:

- Gift to Pelotonia
- Annual Giving Appeal Code

Minimal
Contribution
Variables:

- Event Donor
- Tribute Or Memorial Gift
- ...



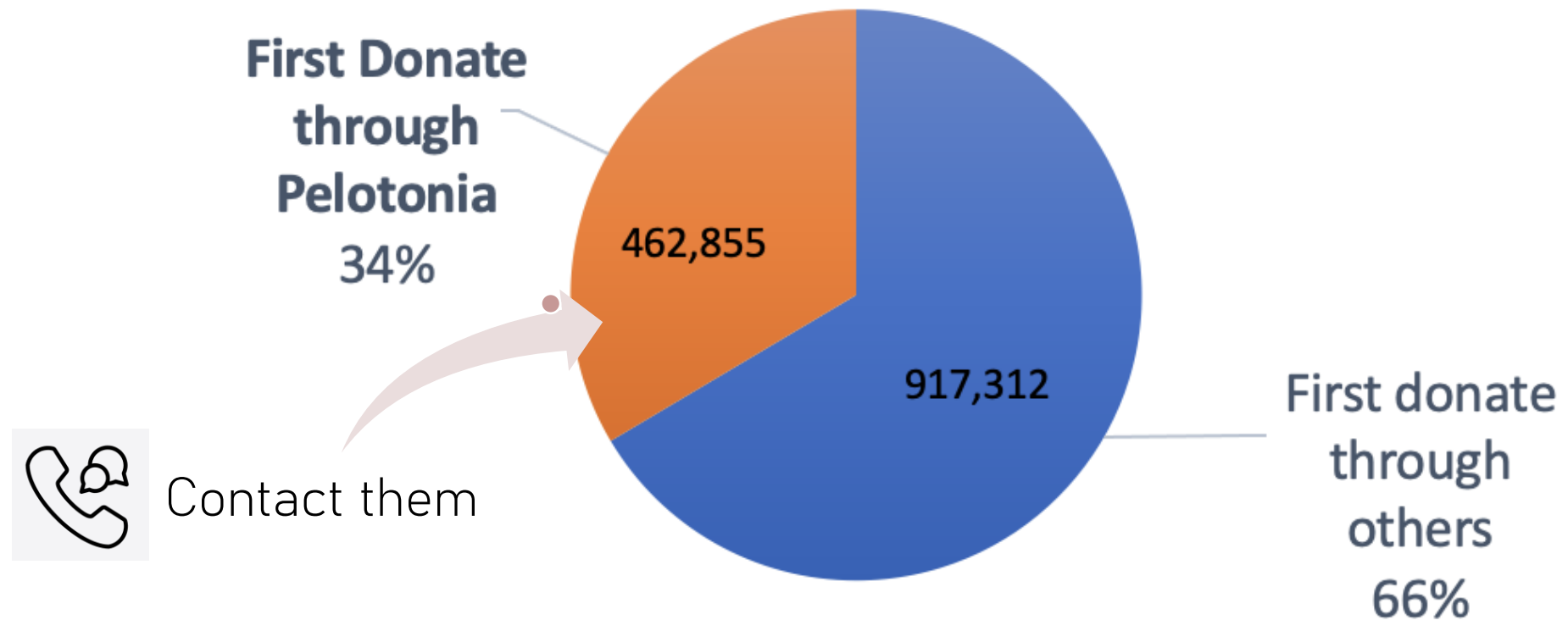


Pelotonia began in 2008 as a two-day bike ride in Columbus, Ohio to raise funds for cancer research at the Ohio State University Comprehensive Cancer Center

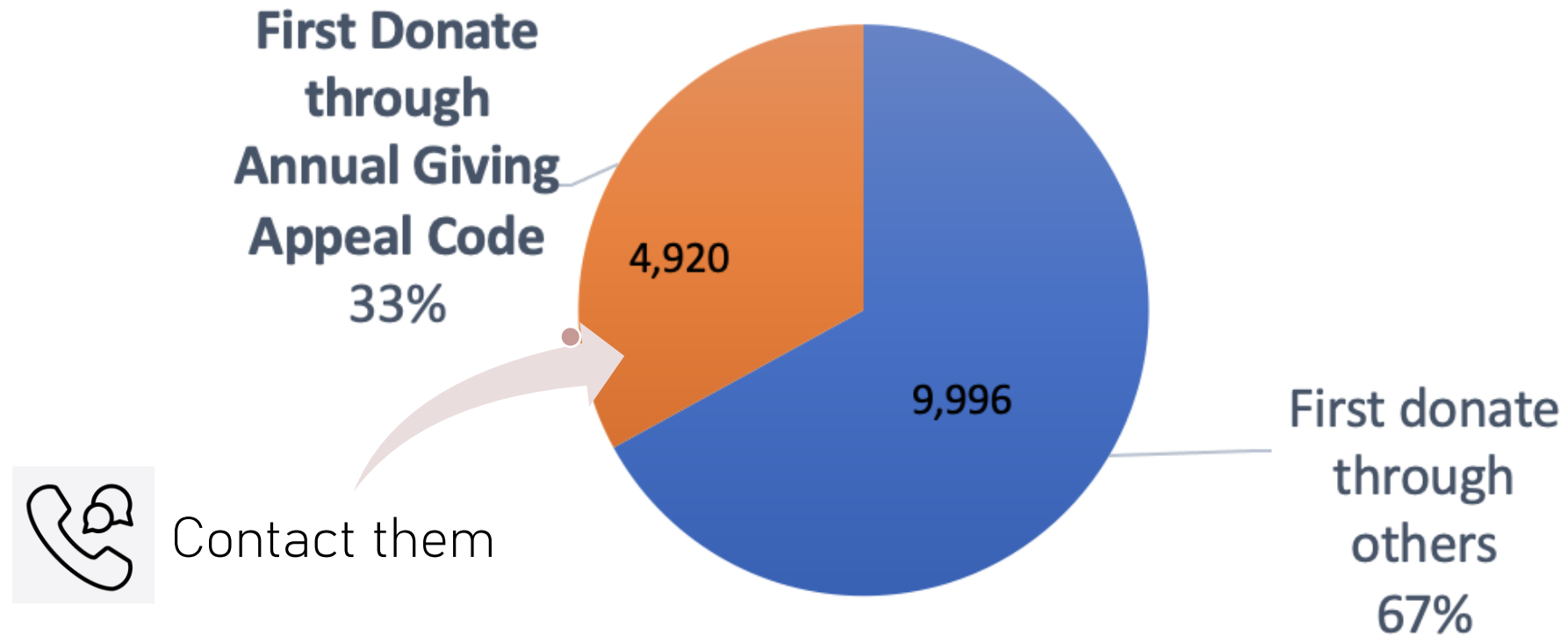
Analysis



Donors who didn't made the jump



Donors who didn't made the jump



Clustering(Hierarchical&K-means)

Variable Name:

- VAL_DONOR_UPORDOWN: count of change from one FY to the next
- N_DONOR_INCREASEDGIVINGYEARS_TIMEPERIOD: Blackjack Score

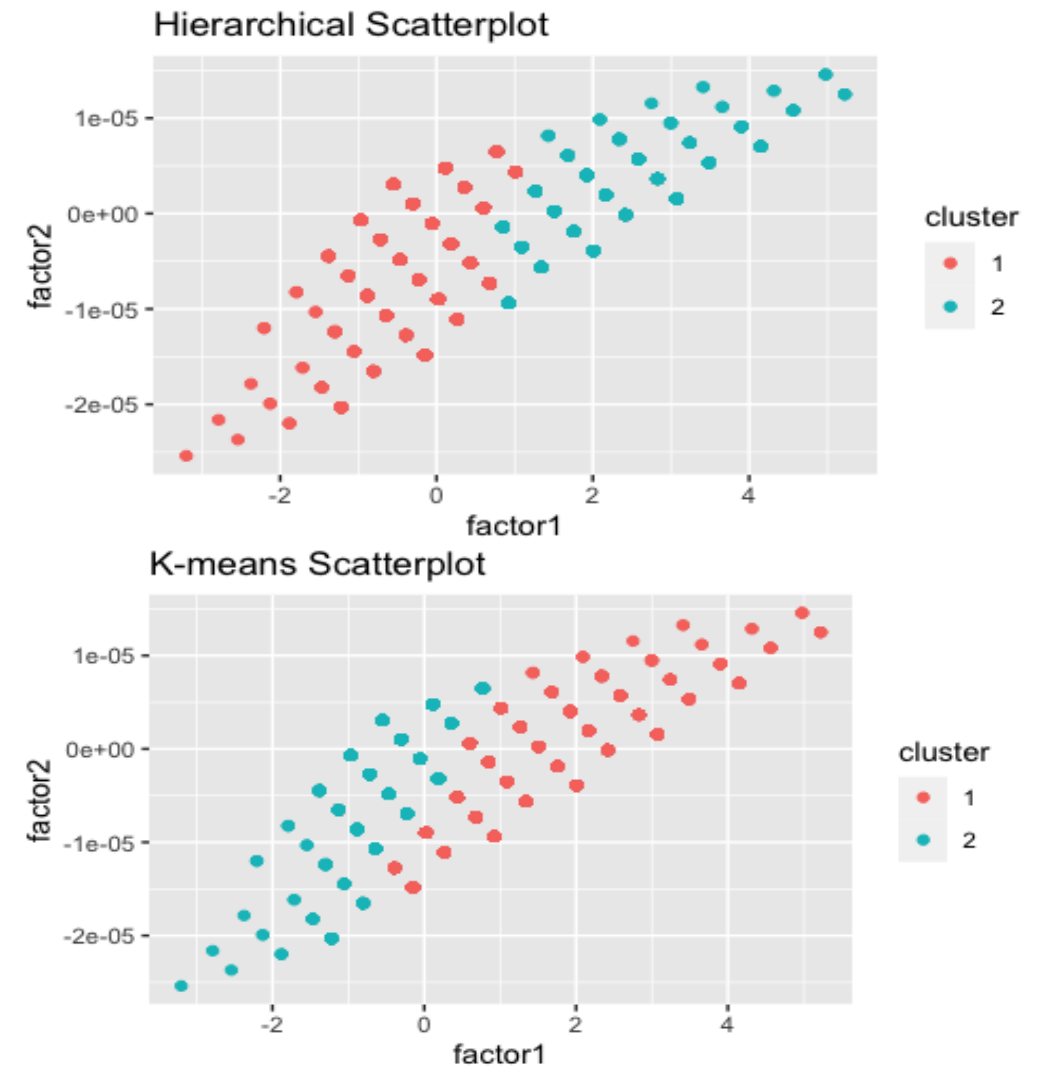
-We could trace those factors back to the original variables and then get the centers for each cluster in the plot.

- The center for cluster 1 in the K-means plot would be 0.04053386 and 1.230845 for VAL_DONOR_UPORDOWN and N_DONOR_INCREASEDGIVINGYEARS_TIMEPERIOD:

- The center for cluster 2 in the K-means plot would be 1.50822281 and 4.358621 for VAL_DONOR_UPORDOWN and N_DONOR_INCREASEDGIVINGYEARS_TIMEPERIOD:

-Do we just use the center to speculate that constituents jump or don't jump?

-Could we connect our unsupervised learning model's results with supervised learning models?



Cluster-Then-Predict

- Create the dataset with cluster labels, without clustering information and two with only one cluster feature
- Loops through these 4 datasets and runs cross validation on each of them

Base-dataset

	VAL_DONOR_UPORDOWN	N_DONOR_INCREASEDGIVINGYEARS_TIMEPERIOD
3805	-0.202942	-0.450563
11367	-0.202942	-0.450563
1540	1.542376	1.116713
684	0.669717	1.639138
9736	0.669717	0.071863

Cluster-feature

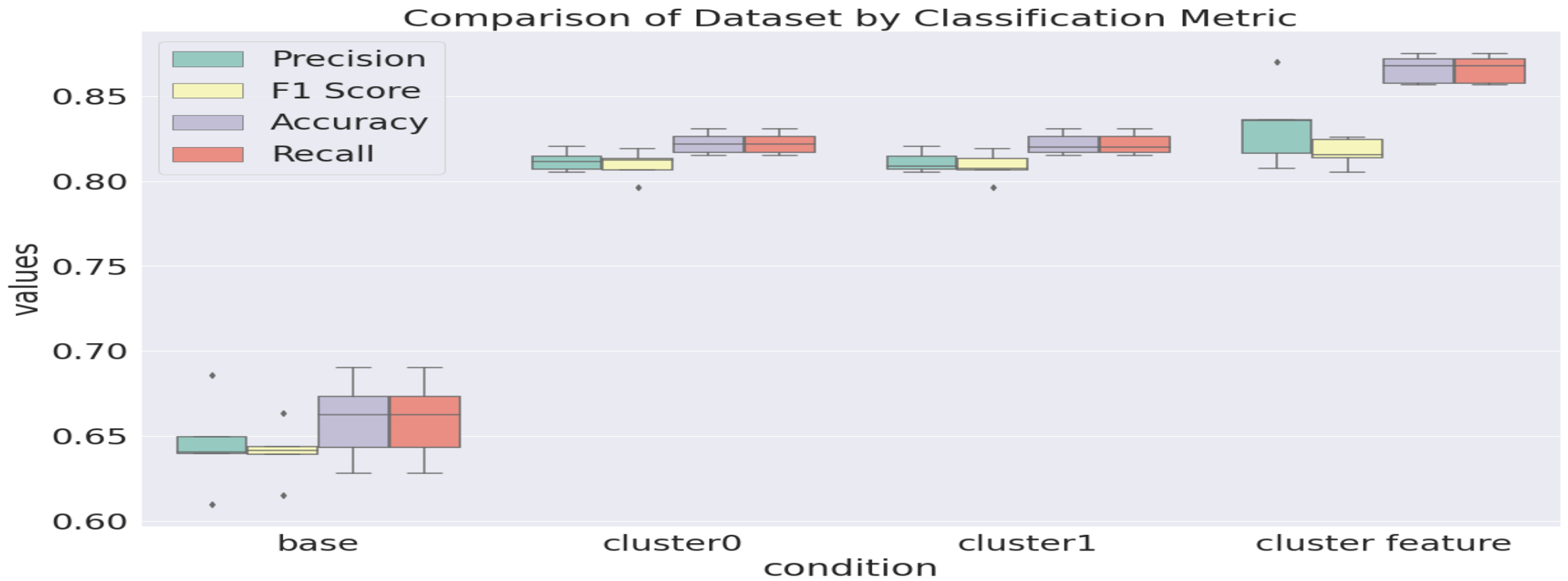
	VAL_DONOR_UPORDOWN	N_DONOR_INCREASEDGIVINGYEARS_TIMEPERIOD	clusters
3805	-0.202942	-0.450563	-0.532460
11367	-0.202942	-0.450563	-0.532460
1540	1.542376	1.116713	1.878074
684	0.669717	1.639138	1.878074
9736	0.669717	0.071863	-0.532460

Cluster-0

	VAL_DONOR_UPORDOWN	N_DONOR_INCREASEDGIVINGYEARS_TIMEPERIOD	clusters
3805	-0.202942	-0.450563	-0.53246
11367	-0.202942	-0.450563	-0.53246
9736	0.669717	0.071863	-0.53246
3659	-0.202942	0.071863	-0.53246
1192	0.669717	-0.450563	-0.53246

Cluster-1

L_DONOR_UPORDOWN	N_DONOR_INCREASEDGIVINGYEARS_TIMEPERIOD	clusters
1.542376	1.116713	1.878074
0.669717	1.639138	1.878074
-0.202942	1.116713	1.878074
5.033011	2.683989	1.878074
3.287693	1.639138	1.878074
...
1.542376	2.683989	1.878074
0.669717	0.594288	1.878074
2.415035	1.639138	1.878074
-0.202942	2.161564	1.878074
-0.202942	2.161564	1.878074



Random Forest Model

There is a surge in the overall metric from the base model to models incorporating clustering information



Overall Suggestions

Focus on Targeting...

- Alumni
- Event Frequentists
 - Jump Donors Tend to Make the Jump when they have Attended an Event Recently
- Those with Large First Donations (>\$500)
- Largest Donation is Greater than \$2,000
- High Contribution Zip-Codes
- Six Figure Incomes
- First Donation Experience
- Cluster-then-Predict for Blackjack Scores and Number of Years of Increased Giving

Thank You!

Any Questions?

