Northeastern University: Planned Gift Propensity Modeling

Surya, Michael, Trang, Sridhathri

MISM 6214

Northeastern University

Project Overview

Predict planned-gift likelihood using models tailored to an extremely rare donor outcome.

Identify and rank high-value prospects by isolating the top 1–10%

Engineer behavioral and engagement features to strengthen model performance despite demographic missingness.

Create actionable donor segments that Advancement can use to prioritize outreach and optimize cultivation strategy.

Which donors are most likely to make a planned gift of $100,000+?
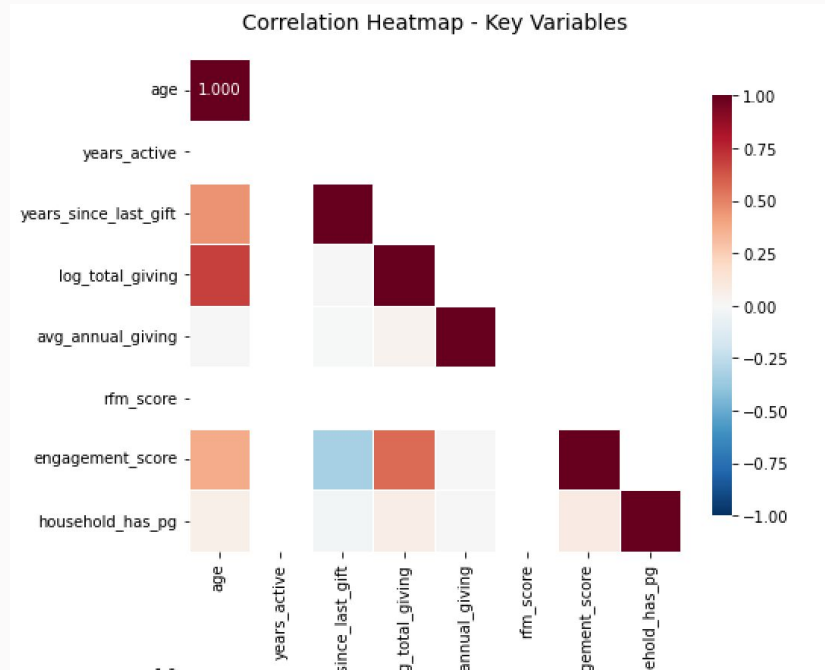
Handling Missing Values

Why Age Cannot Be Imputed

- **53.9% missing** — too incomplete for reliable imputation.
- **Missingness is not random** — PG rate is **5x lower** when age is missing.
- **No strong proxy** — other age-related fields also have high missingness.

Our Solution

- Use **variables highly correlated with age** to capture similar information.
- Key proxies: **years since last gift**, **log total giving**, **engagement score**.
- Together, they reflect **donor maturity and wealth effects** — the behavioral drivers behind age.

Correlation Heatmap - Key Variables


```
Planned Giving Rates:
Category              PG Count      Total       PG Rate    vs Overall
-------------------------------------------------------------------------
Has Age                    245     146,669     0.1670%         1.75x
Missing Age                 59     171,533     0.0344%         0.36x
Overall                    304     318,202     0.0955%         1.00x
```

Feature Engineering

- **Captured donor behavior** through interaction terms that combine capacity, engagement, and relationship signals.
- **Prioritized high-signal features** with low missingness and strong correlation to planned giving.
- **Replaced unusable demographics** with behavioral proxies.

Built 3 feature sets:

- A **transparent, low-missing set** for Logistic Regression
- A **broader behavioral + capacity set** for XGBoost to model complex patterns
- An **expanded set leveraging all engineered variables** to maximize predictive lift for LightGBM

Logistic Regression Model

PR-AUC = 0.0211

- 21 times better than random (0.01% positive cases in data set)

FEATURE IMPORTANCE
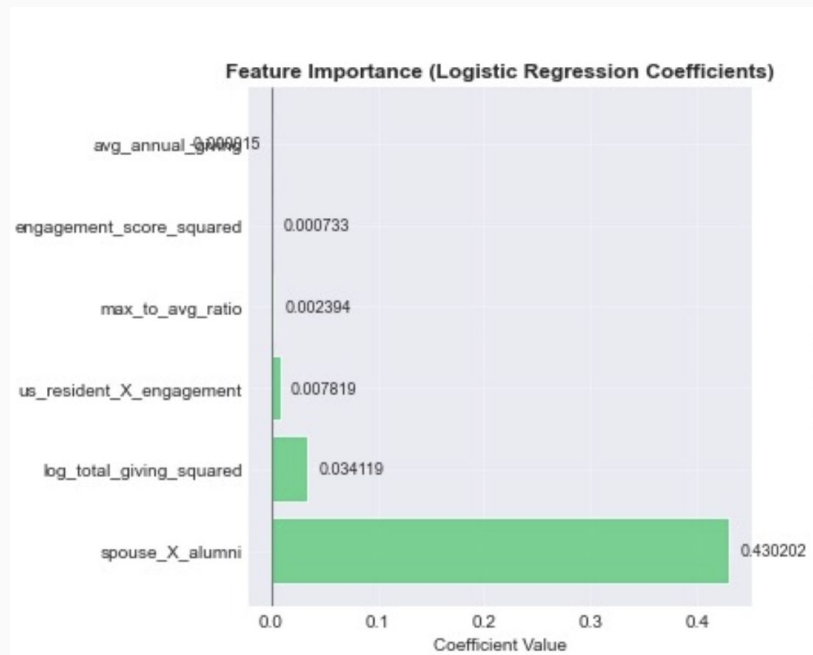
1. spouse_X_alumni (0.43)

- Alumni who have a spouse connection to the university show significantly elevated planned giving propensity
- The dual household connection creates deeper institutional ties

2. log_total_giving_squared (0.034)

- Cumulative wealth effect accelerates at high giving levels

3. us_resident_X_engagement (0.008)

- US residents with high engagement show elevated PG propensity

Feature Importance (Logistic Regression Coefficients)

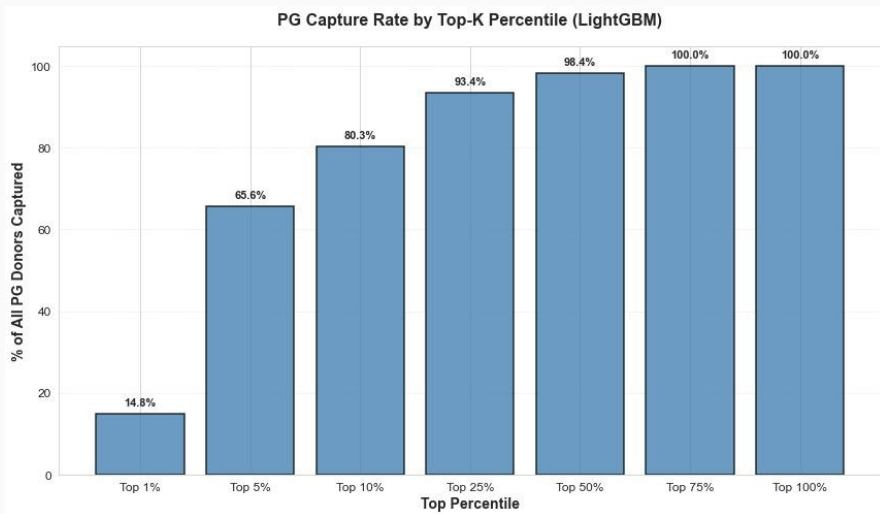| Feature	Coefficient Value
avg_annual_giving	0.000015
engagement_score_squared	0.000733
max_to_avg_ratio	0.002394
us_resident_X_engagement	0.007819
log_total_giving_squared	0.034119
spouse_X_alumni	0.430202

Northeastern University

XGBoost vs LightGBM

XGBoost Model (PR-AUC = 0.051)

- By contacting top 1% (636 people), captures **37.7%** of all PG donors
- By contacting top 5% (6,364 people), captures **67%** of all PG donors
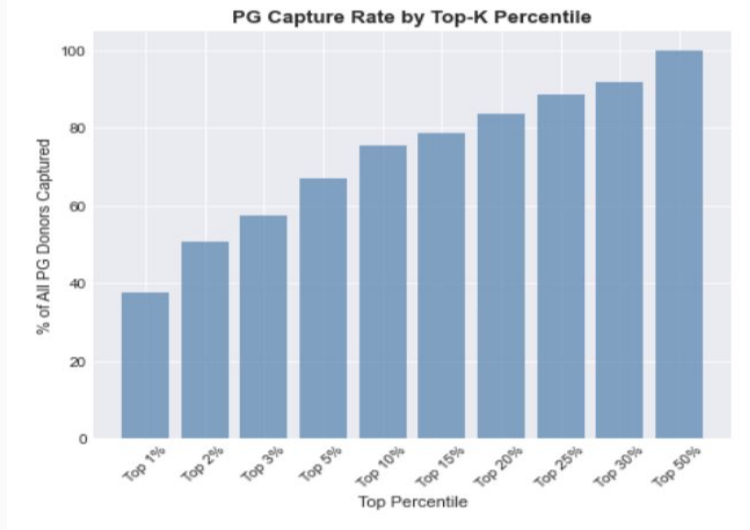
LightGBM Model (PR-AUC = 0.3266)

- By contacting top 1% (636 people), captures **14.8%** of all PG donors
- By contacting top 10% (6,364 people), captures **80.3%** of all PG donors

Recommendation

- LightGBM performs better on entire dataset
- XGBoost performs better for top 1% and 5% which is more useful for targeted segmentation

PG Capture Rate by Top-K Percentile (LightGBM)

PG Capture Rate by Top-K Percentile

Model Performance

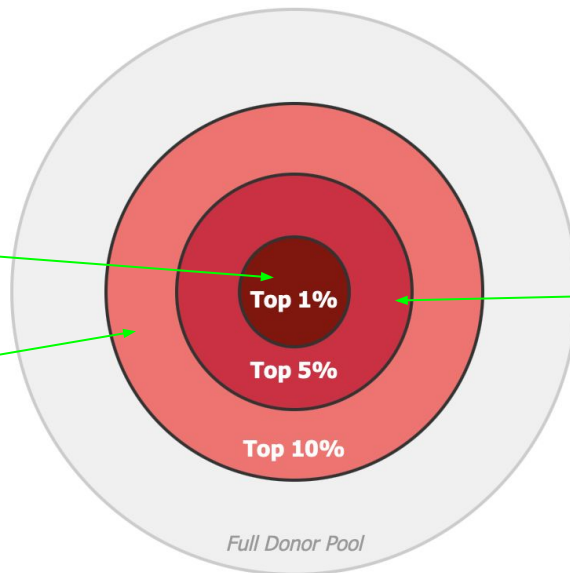| Metric	XGBoost	LightGBM	Winner
PR-AUC Score	0.051	0.327	LightGBM
Improvement over Random	53x	341x	LightGBM
Top 1% Capture Rate	37.7%	14.8%	XGBoost
Top 5% Capture Rate	67%	65.6%	XGBoost
Top 10% Capture Rate	75%	80.3%	LightGBM

Donor Segmentation

Top 1% (3,182 donors)

Contains 37.7% of all likely planned-gift donors and should be prioritized for personal outreach and estate-planning conversations.

Top 5% (15,910 donors)

Capturing 67% of all likely PG donors, this segment is ideal for targeted events, legacy-society invitations, and focused cultivation.

Top 10% (31,820 donors)

Accounting for 75% of likely PG donors, this group represents emerging prospects suited for educational content and early cultivation.

Top 1%
Top 5%
Top 10%
Full Donor Pool

Recommendations

Top 1% - Top Priority

- Senior officer 1:1 outreach
- Start estate-planning conversations
- Provide personalized cultivation + high-touch stewardship

Top 5% - High Potential Donors
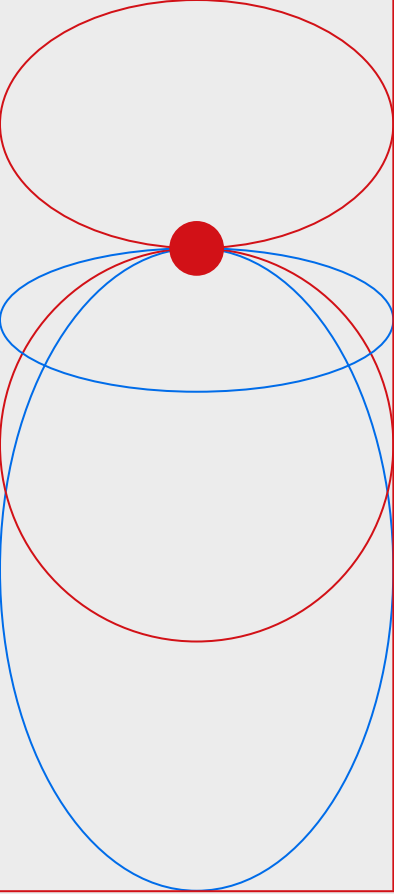
- Invite to targeted PG events/workshops
- Send personalized impact messaging
- Monitor triggers and elevate rising donors to Tier 1

Top 10% - Scalable Cultivation

- Use automated PG education campaigns
- Deliver scalable digital stewardship
- Review engagement quarterly to surface emerging prospects

Northeastern University

Thank You!

Q&A

Northeastern University