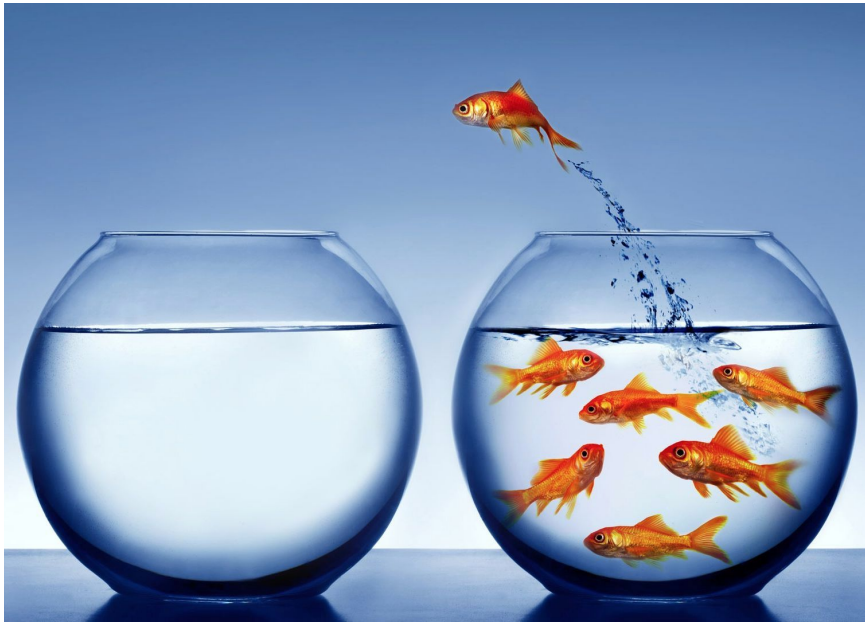
Cutting the Cord: Predicting Customer Churn for a Telecom Company

Brenner Heintz

Why Study Customer Churn?

- Losing customers is the BIGGEST problem that businesses face
- It costs **5x - 25x as much to find a new customer than it does to retain an existing customer**
    - **Source: Harvard Business Review (link)**
- **We can spend $ to save $$$$**



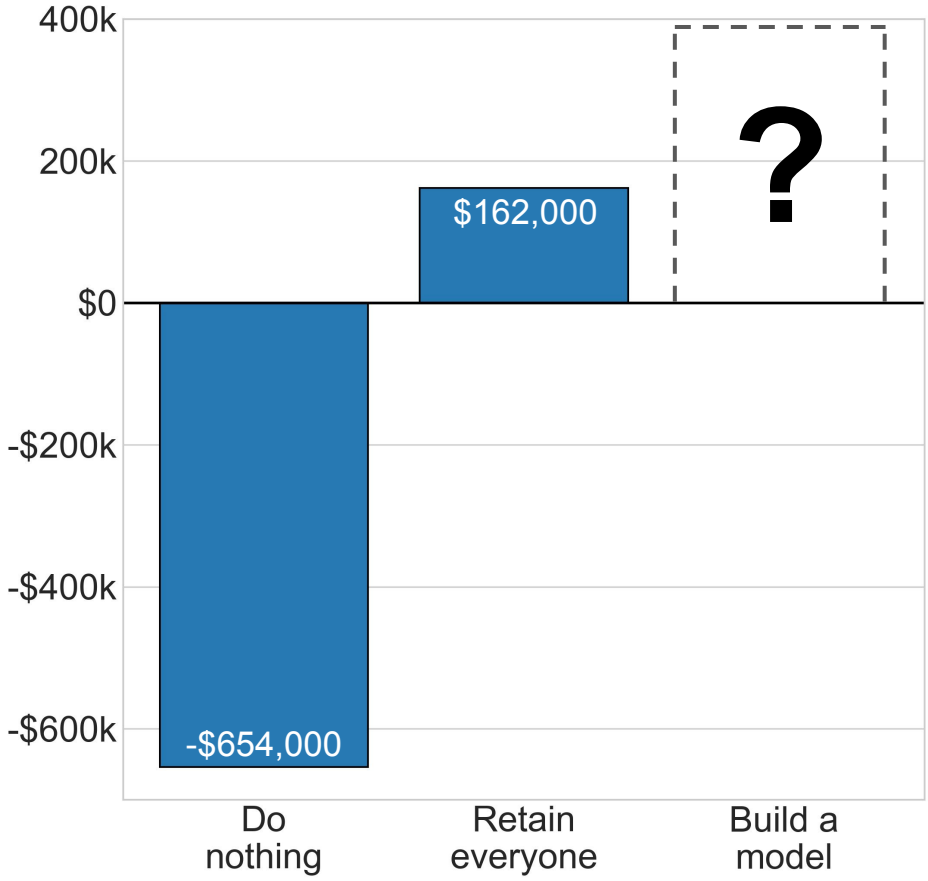
Data & Methodology

- **We want to predict:** Monthly customer “churn”
- Kaggle dataset, n = 7,033
- Independent variables:
  - Which services customers had (internet, phone, cloud…)
  - How much they paid per month
  - How long they had been subscribers
  - Basic demographic info
- Assumptions:
  - It costs the company **$500 to find a new customer**
  - It costs the company **$100 to retain an existing customer**


27%
of customers left our company
within the month

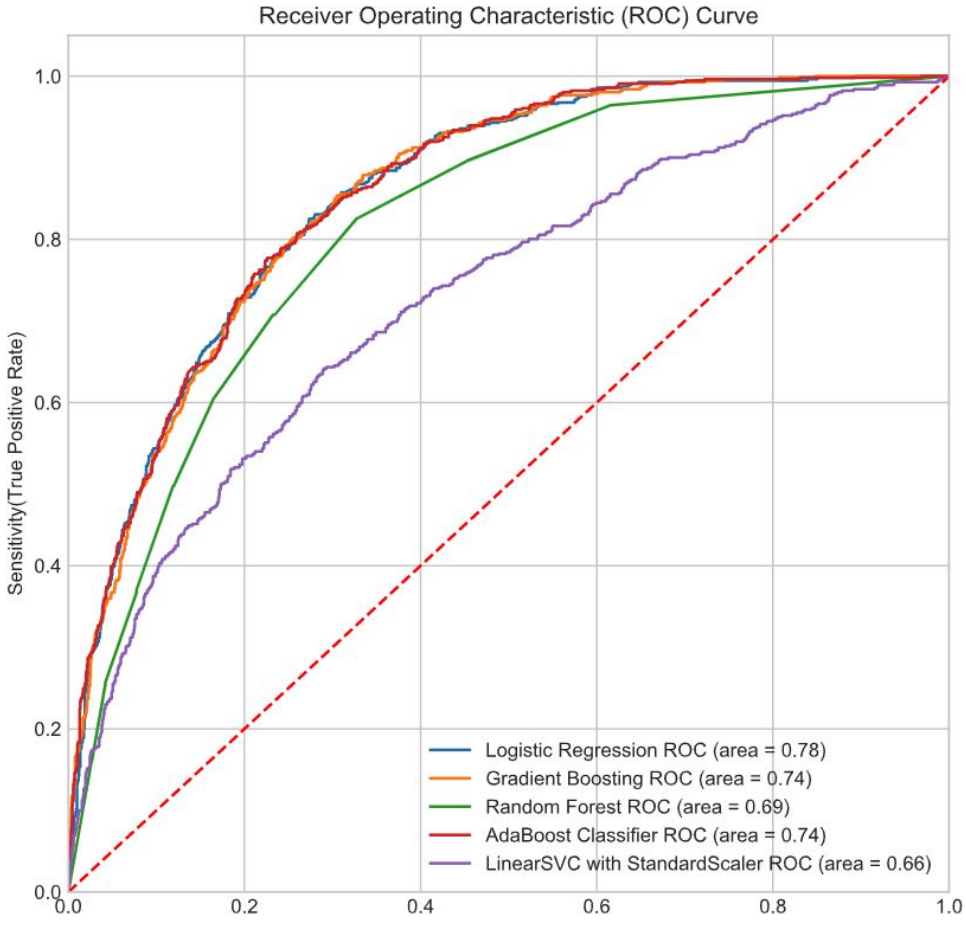
Setting the Bar

- Build a model that's better than doing nothing, or retaining everyone
- Customers get used to paying less and don't want to pay more
- Recurring revenue is highly valuable



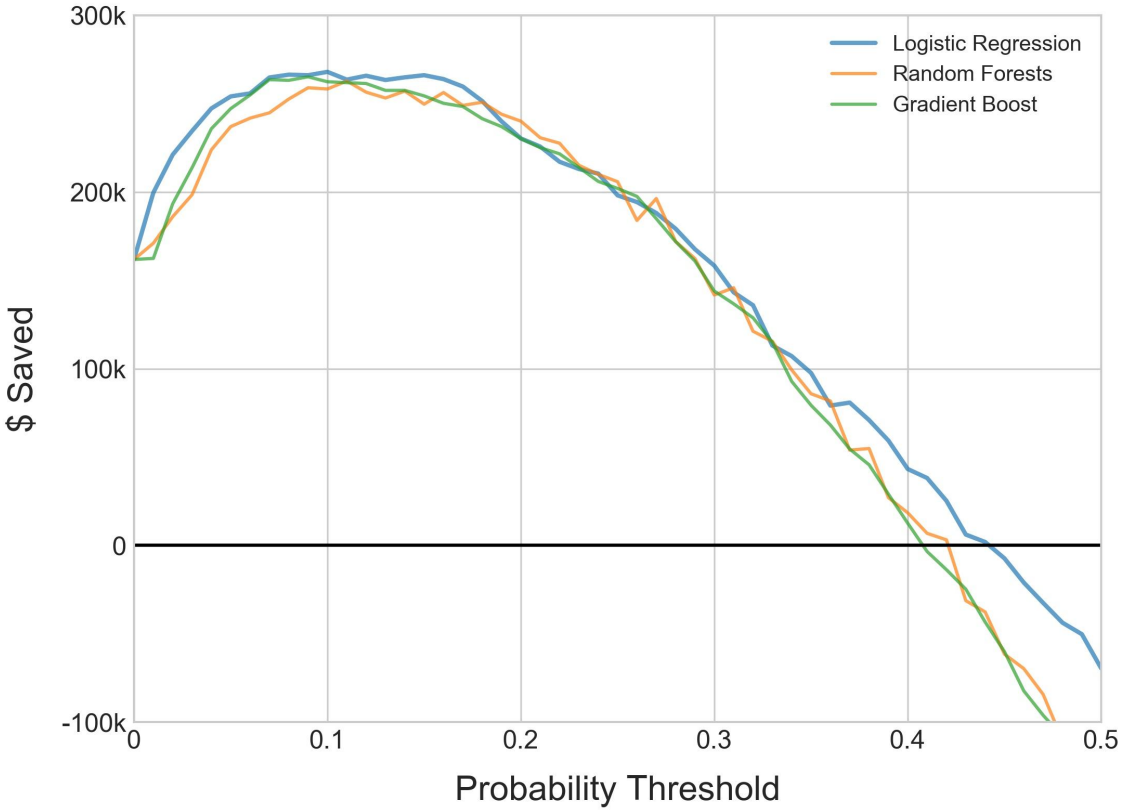
Our Model

- Recall
- SMOTE
- Used **GridSearchCV** to cross-validate our models and tune our hyperparameters



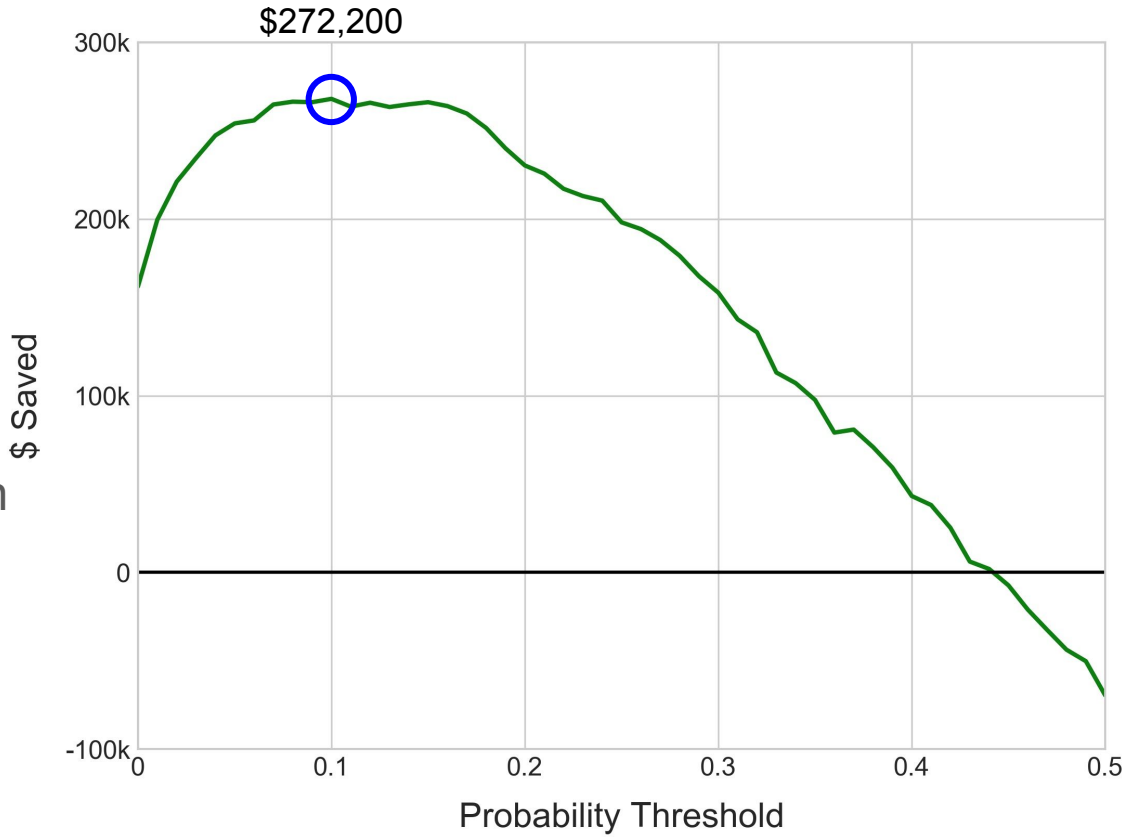
Results

- **“Money Saved”** per model
- Low probability thresholds were ideal

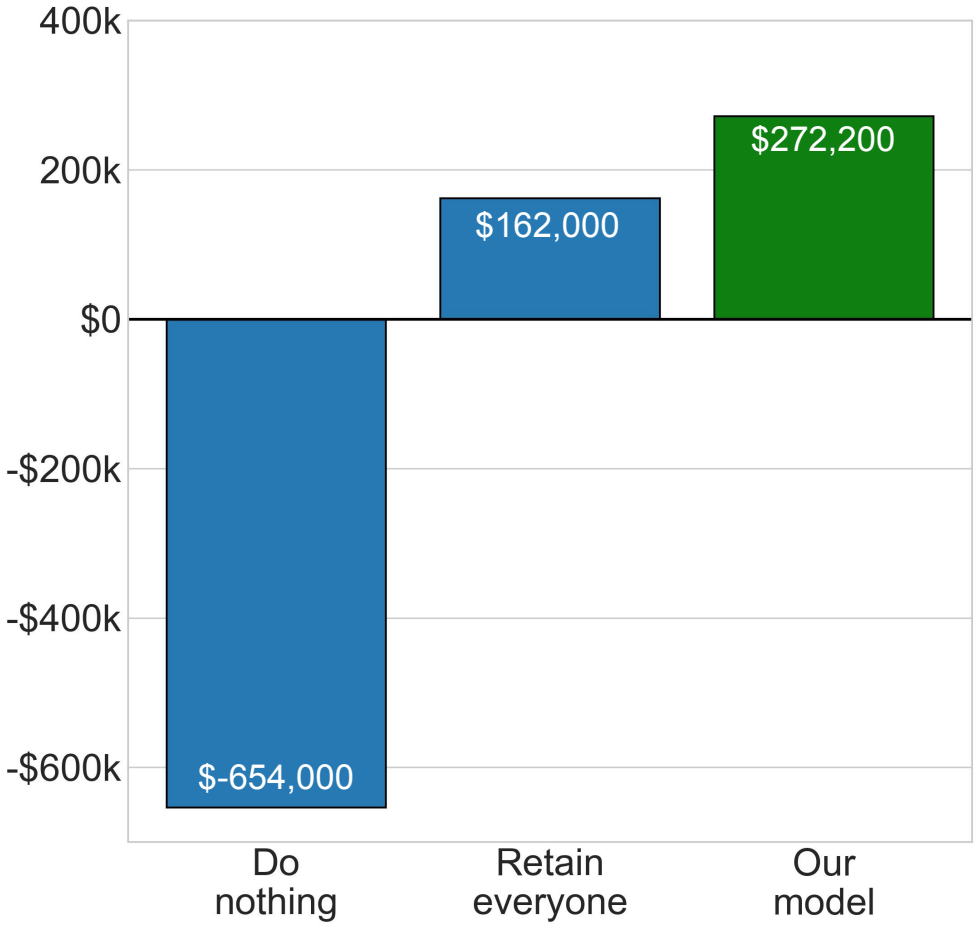


Results

- **“Money Saved”** per model
- Low probability thresholds were ideal
- Logistic Regression won out



Results


400k
200k
$0
-$200k
-$400k
-$600k
$-654,000
$162,000
$272,200
Do nothing
Retain everyone
Our model

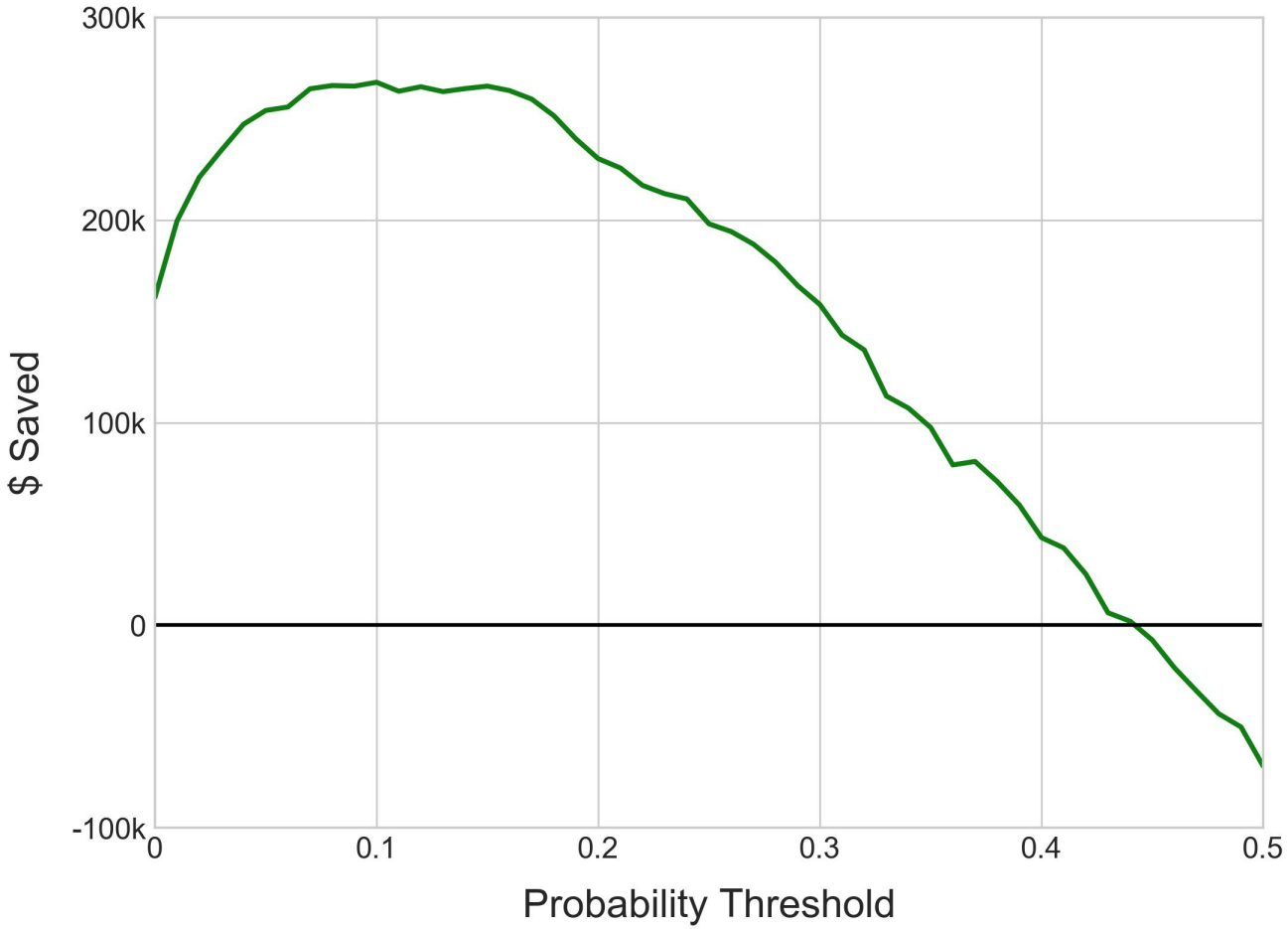
Future Work

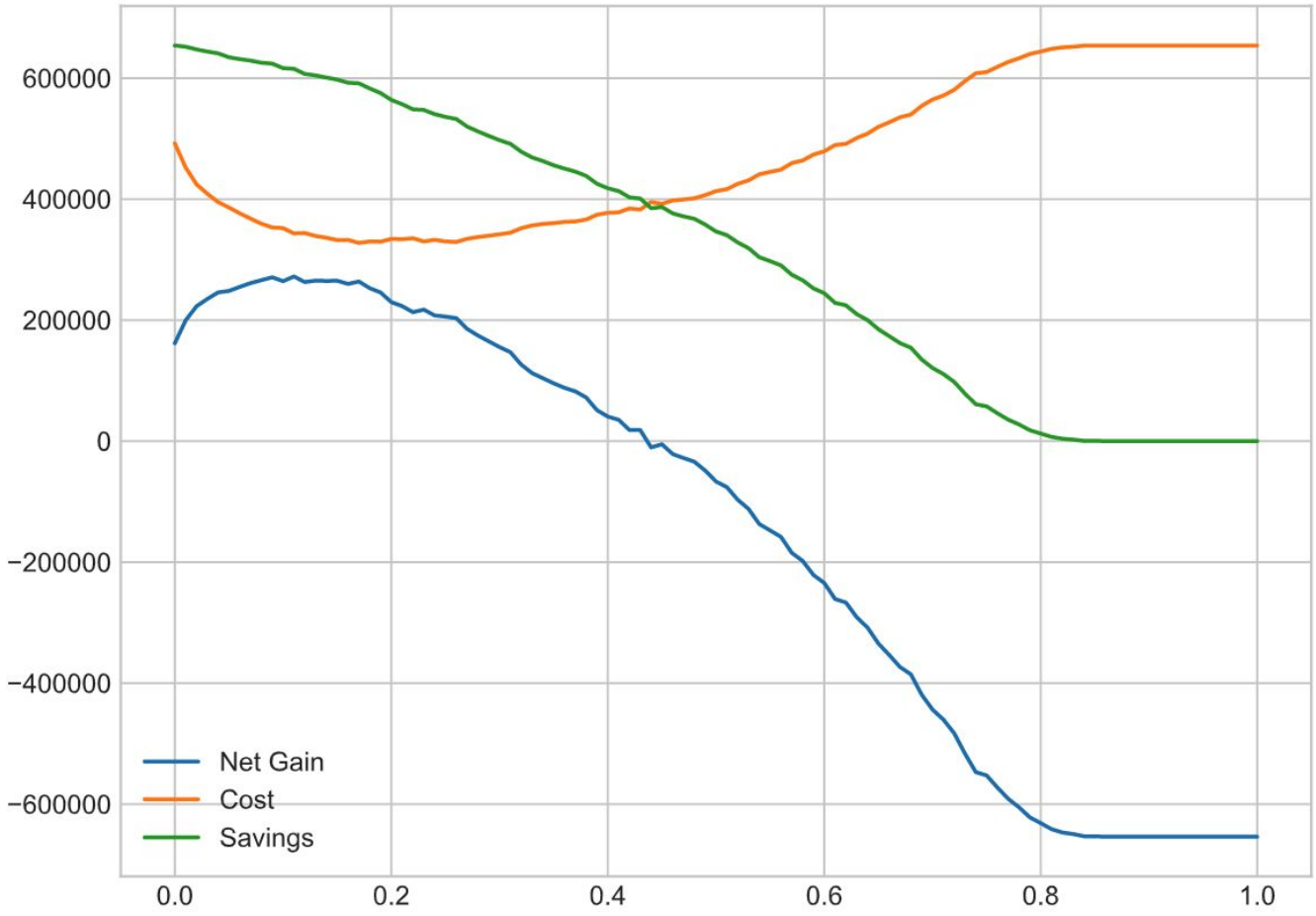
- Focus efforts on retaining high-revenue customers
- Create bundling options for customers that cost less, yet allow the company to keep more customers than one-time interventions
- Explore the ability of new products to make customers more reluctant to switch


Thank You!
github.com/athena15
brenner.heintz@gmail.com

Results

Model	Average Recall	Net Gain
Logistic Regression	81%	$272,200
Random Forest	72%	$262,800
Gradient Boosting	68%	$263,600
AdaBoost	74%	$257,200


300k
200k
100k
0
-100k
$ Saved
0
0.1
0.2
0.3
0.4
0.5
Probability Threshold


600000
400000
200000
0
−200000
−400000
−600000
0.0
0.2
0.4
0.6
0.8
1.0
Net Gain
Cost
Savings