Retaining Customers

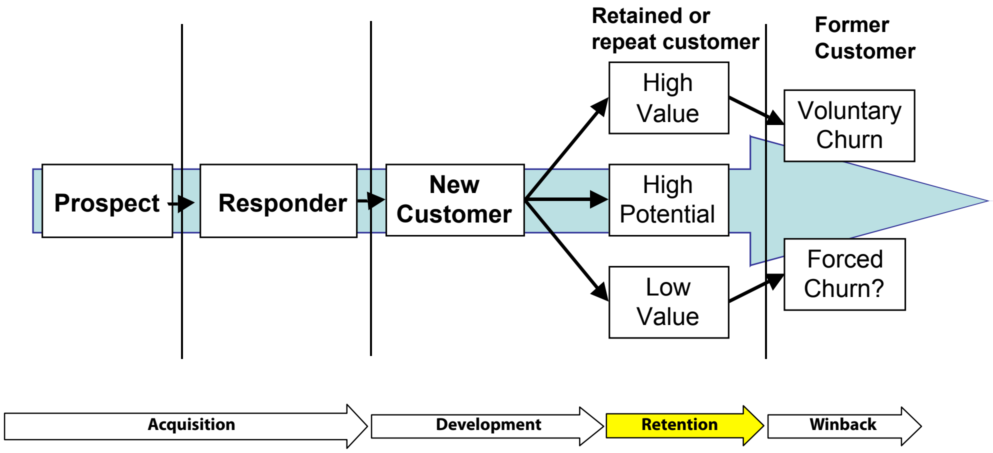
Professor Song Yao
Olin Business School

Customer Analytics

Churn Management Problem

We now consider how we can use information based marketing to postpone the end of the customer lifecycle

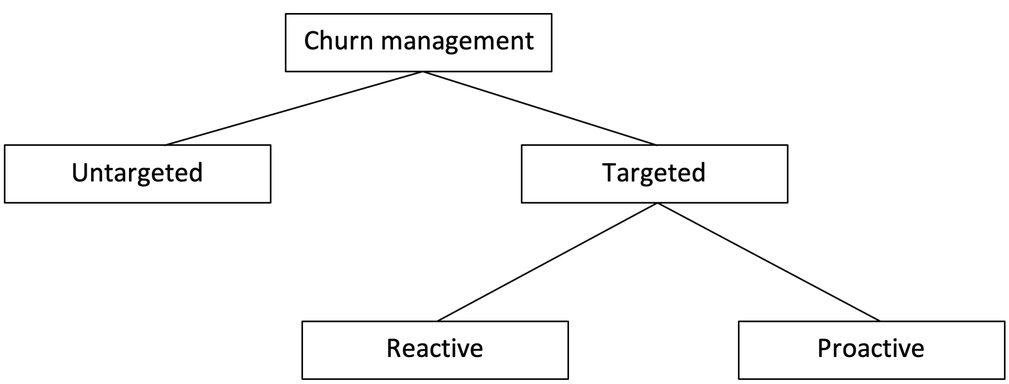
APPLICATIONS OF PREDICTIVE TECHNIQUES



- Modeling and predicting response rates
- Modeling and predicting the success of cross/ up-sell attempts
- Modeling and predicting customer attrition
- Modeling and predicting "reactivation" success

Retention strategies (churn management) have historically been untargeted

APPROACHES TO CHURN MANAGEMENT



- Improvements in product quality
- Mass advertising

- Resolve issues **at the time of or after** the attrition

- Resolve issues **before** the attrition

How can predictive analytics help?

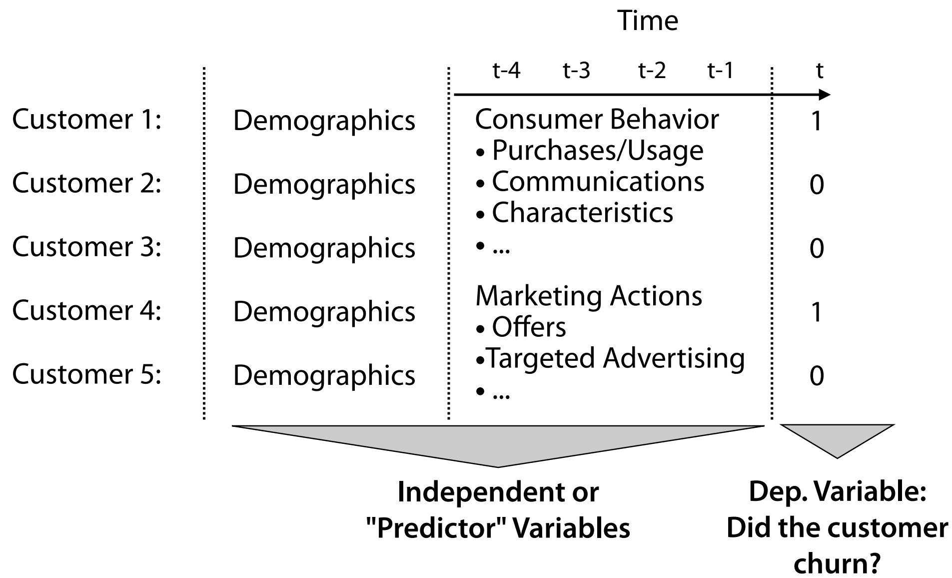
CONTRIBUTIONS OF PREDICTIVE ANALYTICS

- **Predict individual churn probabilities**
- *Example:*
  - Overall (baseline) churn rate is 20% per year
  - With model identify customers who have much higher churn rates than the average (e.g., 30%, 40%, etc.)

- **Determine which factors lead to the higher rate, i.e. are most likely to contribute to the attrition**
- *Example:*
  - Model reveals that customers with a tech support wait time exceeding 5 minutes are more likely to leave
  - Give priority to customers who are likely to churn

We need attrition data to develop a retention model

REQUIRED DATA FOR RETENTION MODEL



Churn Management Model Example: Step 1

CHURN MANAGEMENT STEPS

1. Develop a model to predict customer churn
2. Use model to understand main drivers of churn
3. Use insights to develop actions/offers/incentives
4. Estimate the impact of these actions/offers/incentives on the probability of churn
5. Evaluate the economics

We will design a churn management program for Netflix

NETFLIX EXAMPLE*

NETFLIX

- Largest *subscription* video streaming service provider
- 66 MM customers 2024
- High recurrent revenue
- Churn rate about 2% per month (industry average 5%)

*The numbers used in this example are hypothetical and should not be considered an accurate reflection of churn at Netflix

Three plans at different price levels

NETFLIX PRICE HIKES

Price	Jan 2022	Oct 2020	Jan 2019	Oct 2017	Oct 2015	Apr 2014	Apr 2013	2011
Premium (4K, 4 screens)	**$19.99**	$17.99	$15.99	$13.99	$11.99	$11.99	$11.99	N/A
Standard (HD, 2 screens)	**$15.49**	$13.99	$12.99	$10.99	$9.99	$8.99	$7.99	$7.99
Basic (No HD, 1 screen)	**$9.99**	$8.99	$8.99	$7.99	$7.99	$7.99	N/A	N/A

CHURN PROBLEM

- Churn rate 2% per month
  = 21.5% per year (=1-(1-0.02)^12)

We have a rich set of variables to predict attrition

NETFLIX DATA

- **Dependent variable:**
  Did customer cancel Netflix in July 2024?

- **Independent variables:**
  Collected on June 2024

```
                 Age         Income  SubscriptionLength  SatisfactionScore  \
count  200000.000000  200000.000000       200000.000000      200000.000000
mean       34.718935   65032.108110           23.276055           5.501760
std         9.606132   26863.651008           23.021182           2.599713
min        18.000000   20000.000000            1.000000           1.000000
25%        28.000000   45860.750000            6.000000           3.250000
50%        35.000000   60097.000000           16.000000           5.500000
75%        41.000000   78631.000000           33.000000           7.760000
max        65.000000  200000.000000          120.000000          10.000000

         PremiumPlan         Female   SlowInternet  Low10PercentUsage
count  200000.000000  200000.000000  200000.000000      200000.000000
mean        0.300445       0.499915       0.201090           0.100250
std         0.458453       0.500001       0.400816           0.300334
min         0.000000       0.000000       0.000000           0.000000
25%         0.000000       0.000000       0.000000           0.000000
50%         0.000000       0.000000       0.000000           0.000000
75%         1.000000       1.000000       0.000000           0.000000
max         1.000000       1.000000       1.000000           1.000000
```


We have data on 200,000 customers to predict churn

SAMPLE AND METHOD FOR PREDICTION

- **Training and test** Samples (80/20 split)

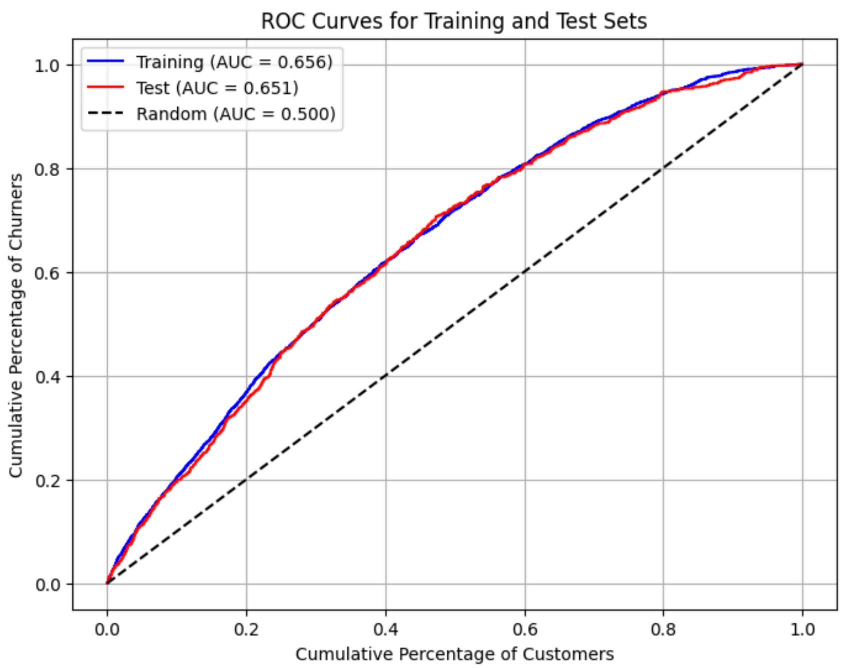
- **Churn Rate** in whole sample (both training and test): 1.85%

- **Method: Logistic Regression**

  • We can also use alternative ML models (e.g., Neural Network). But Step 2, identifying important churn drivers, becomes trickier (see demo code on Canvas).

```python
# create predictors and target and split data
### this line creates a list of all columns except CustomerID and Churn
predictors = [col for col in netflix.columns if col not in ['CustomerID', 'Churn']]
X = netflix[predictors]
y = netflix['Churn']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Fit model on training data
train_formula = 'Churn ~ ' + ' + '.join(predictors) # the .join() joins the predictors with ' + '
train_model = smf.logit(formula=train_formula, data=pd.concat([X_train, y_train], axis=1)).fit()

# Get predictions for both training and test sets
train_pred = train_model.predict(X_train)
test_pred = train_model.predict(X_test)

# Calculate AUC for both sets
train_auc = roc_auc_score(y_train, train_pred)
test_auc = roc_auc_score(y_test, test_pred)
```

Check for overfitting: AUC Training vs. Test Samples



Churn Management Model Example: Step 2

CHURN MANAGEMENT STEPS

1. Develop a model to predict customer churn
2. Use model to understand main drivers of churn
3. Use insights to develop actions/offers/incentives
4. Estimate the impact of these actions/offers/incentives on the probability of churn
5. Evaluate the economics

The AME from the model suggest that a number of factors drive people away

Average Marginal Effects of the Logit Regression

```
Average Marginal Effects:
        Logit Marginal Effects
=====================================
Dep. Variable:                  Churn
Method:                          dydx
At:                           overall
======================================================================================
                       dy/dx    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------
Age               -3.821e-05   3.52e-05     -1.085      0.278      -0.000    3.08e-05
Income            -7.456e-10   1.26e-08     -0.059      0.953   -2.54e-08     2.4e-08
SubscriptionLength -9.929e-06  1.48e-05     -0.671      0.502   -3.89e-05    1.91e-05
SatisfactionScore    -0.0024      0.000    -16.965      0.000      -0.003      -0.002
PremiumPlan           0.0106      0.001     15.051      0.000       0.009       0.012
Female               -0.0002      0.001     -0.245      0.807      -0.001       0.001
SlowInternet          0.0095      0.001     12.537      0.000       0.008       0.011
Low10PercentUsage     0.0119      0.001     12.949      0.000       0.010       0.014
======================================================================================
```

We now need to determine the relative importance of factor for predicting churn

DETERMINING IMPORTANCE

- Select statistically significant variables
- For non-dummy variables:
  - Calculate effect for one standard deviation change in variable (enables comparison with unadjusted dummy variable)
  - Because of logistic: Multiply AME by SD
  - The "Importance" measure is the |AME * SD|
- For **dummy** variables:
  - The "Importance" measure is |AME|
- Sort variables by importance

The AME from the model suggest that a number of factors drive people away

Average Marginal Effects of the Logit Regression

	AME	p-val	STD Dev. (SD)	Importance	Rank of Importance
Age	-3.82E-05	0.278			
Income	-7.46E-10	0.953			
SubscriptionLength	-9.93E-06	0.502			
SatisfactionScore	-0.0024	0	2.598736	0.006237	4
PremiumPlan	0.0106	0		0.0106	2
Female	-0.0002	0.807			
SlowInternet	0.0095	0		0.0095	3
Low10PercentUsage	0.0119	0		0.0119	1

Churn Management Model Example: Step 3 and Step 4

CHURN MANAGEMENT STEPS

1. Develop a model to predict customer churn
2. Use model to understand main drivers of churn
3. Use insights to develop actions/offers/incentives
4. Estimate the impact of these actions/offers/incentives on the probability of churn
5. Evaluate the economics

What should Netflix do to reduce customer churn?

	AME	p-val	STD Dev. (SD)	Importance	Rank of Importance
Age	-3.82E-05	0.278			
Income	-7.46E-10	0.953			
SubscriptionLength	-9.93E-06	0.502			
SatisfactionScore	-0.0024	0	2.598736	0.006237	4
PremiumPlan	0.0106	0		0.0106	2
Female	-0.0002	0.807			
SlowInternet	0.0095	0		0.0095	3
Low10PercentUsage	0.0119	0		0.0119	1

E.g., encouraging Low10PercentUsage users to downgrade

CHURN MANAGEMENT STEPS

1. Develop a model to predict customer churn
2. Use model to understand main drivers of churn
3. Use insights to develop actions/offers/incentives
4. Estimate the impact of these actions/offers/incentives on the probability of churn
5. Evaluate the economics

There are two ways to predict the effect of an action/ incentive/offer on the predicted probability of churn

PREDICTING IMPACT

1. **Simulate the effect of a churn driver** by changing the value of the variable and using the estimated logistic regression model to obtain new churn predictions (see Pentathlon Cross-sell and Upsell Assignment)

2. **Test the action/incentive/offer** in the field

In the assignment, we will be using method 1. But in practice, methods 1 and 2 are often applied together!

There are two ways to predict the effect of an action/ incentive/offer on the predicted probability of churn

PREDICTING IMPACT USING SIMULATION

```
## Predict the churn probability if we downgrade the subscription to non-premium
## for those customers who have Low10PercentUsage and have higher churn rate
## than the average (1.85%).

# First check the current churn rate of these customers
low_usage_customers = netflix[(netflix['Low10PercentUsage'] == 1) &
                              (netflix['PremiumPlan'] == 1) &
                              (test_pred > 0.0185)]
print("Current churn rate of premium customers with Low10PercentUsage:")
print(low_usage_customers['Churn'].mean())

# Predict the churn probability for these customers
# if we downgrade the subscription to non-premium

# Create a new dataframe with the selected features and the new subscription plan
new_data = low_usage_customers.copy()
new_data['PremiumPlan'] = 0

# Predict the churn probability for the new data
new_pred = train_model.predict(new_data)

# Print the predicted churn probability
print("Predicted churn probability for the new subscription plan:")
print(new_pred.mean())
```

Using the test sample

```
Current churn rate of premium customers with Low10PercentUsage:
0.030833333333333334
Predicted churn probability for the new subscription plan:
0.026887379053490767
```

Churn Management Model Example: Step 5

CHURN MANAGEMENT STEPS

1. Develop a model to predict customer churn
2. Use model to understand main drivers of churn
3. Use insights to develop actions/offers/incentives
4. Estimate the impact of these actions/offers/incentives on the probability of churn
5. Evaluate the economics

A plan specifies an action/incentive/offer and a rule for selecting customers

EXAMPLE OF CHURN PLAN FOR NETFLIX EXAMPLE

Action: *Downgrade premium users who have the lowest 10% usage*
Targeting rule: Premium, lowest 10% usage
Expected churn benefit: Baseline churn: 3.1%, projected churn: 2.7%

What is it worth to downgrade premium users who have lowest 10% usage ($20 vs. $15.5 monthly rate)?

LTV Comparison

	Now	Year 1	Year 2	Year 3	Year 4	Year 5	
Revenue	$0	$240	$240	$240	$240	$240	
Product/Service Cost	$0	$0	$0	$0	$0	$0	
Action/Offer/Incentive Cost	$0	$0	$0	$0	$0	$0	
Profit	$0	$240	$240	$240	$240	$240	
Prob. active at end of period	100%	69%	47%	32%	22%	15%	
Expected Profit	$0	$164	$113	$77	$53	$36	
Present Value of Exp. Profit	$0	$150	$93	$58	$36	$23	$359

Baseline churn: 3.1%
Yearly: 1-(1-0.031)^12 = 31.5%

Projected churn: 2.7%
Yearly: 1-(1-0.027)^12 = 28%

	Now	Year 1	Year 2	Year 3	Year 4	Year 5	
Revenue	$0	$186	$186	$186	$186	$186	
Product/Service Cost	$0	$0	$0	$0	$0	$0	
Action/Offer/Incentive Cost	$0	$0	$0	$0	$0	$0	
Profit	$0	$186	$186	$186	$186	$186	
Prob. active at end of period	100%	72%	52%	37%	27%	19%	
Expected Profit	$0	$134	$96	$69	$50	$36	
Present Value of Exp. Profit	$0	$122	$80	$52	$34	$22	$310

Value of Churn Program:
$49 per subscriber for 5 years