



# ***Targeting the Promotion Takers***

Brad Barber, Andrew Choi, Michael Kilby, Jim Turner

# Problem

- Company recently offered promotion to members of an airline frequent flyer club via direct mail
- Response rate was less than desired
- In the future, company wants to target consumers most likely to accept offer



***Objective: Create a model that classifies frequent flier club members based on their likelihood of accepting the promotion***

# Data Set

- Team analyzed data set compiled by company that conducted promotion
  - Data contained *stratified sample* of 4,986 observations of 49 variables each, including the binary response variable “SALE”
- Team cleaned data for invalid observations and non applicable variables
  - Cleaned data set contains 4,798 observations and 26 relevant variables

## Sample of Data Set

Member ID	Mileage Balance	Premier YTD Miles	Rewards										NFB Enroll 6months	NFB Enroll 12months	
			MP Credit Miles	Rewards Credit Miles	SB Credit Miles	MP Credit Miles Added	Rewards Credit Miles Added	SB Credit Miles Added	Credit Miles Added	NFB Miles Earned	NFB Transact.	NFB Miles Earned			
			12months	12months	12months	Dummy	Dummy	Dummy	Dummy	12months	12 months	Dummy			Dummy
1	28143	0	1	1	1	0	1	0	1	0	1	174	1	0	0
2	19244	0	1	1	1	0	0	0	0	0	0	215	2	0	0
3	41354	0	1	1	1	1	0	0	0	1	1	4123	4	0	0

Num Online Purch. 12months	Travel Miles 12months	Travel Transact. 12months	Days Since Award Flight	Award Flight Dummy	Days Since Int'l Flight	Days Since Last Dom. Flight	Days Since Last Flight	Sales Region	Days Since Enroll. In FF Program	Club Member Dummy	College+ Member Dummy	E-Mail Dummy	Sale
0	0	0		0	480	29	Partners	7000	0	0	1	0	
0	0	0		0			Partners	6968	0	0	0	0	
0	0	0		0			Partners	7034	0	0	1	0	

# Approach

- Clean data – remove invalid observations and irrelevant or repetitive variables
- Examine data set
  - Look at basic summary measures of the data
  - Qualitatively examine variables
- Pivot data to look for relevant structures
- Model the data with various techniques
  - Perform logistics regression and discriminant analysis and compare results
- Select an appropriate model to classify future consumers



# Initial Examination

- Some summary measures were not useful
  - Single variable measures (mean, standard deviation) did not suggest course of action
  - Correlation matrix did not reveal better than a 0.12 correlation with SALE
- Dialogue with source of data revealed that response rate in data set was greater than in reality
  - In data set, promotion was accepted in 13.2% of cases
  - In reality, promotion was accepted in only 0.018% of cases
- Qualitative examination of data revealed three categories of data collected
  - Flight activity: Freq. flyer club activity related to flying
  - Non-flight activity: Freq. flyer club activity related to non-flying promotions
  - Miscellaneous shopping activity: Other measures of consumer behavior (e.g., online shopping)

# Pivoting Data

- Response variable was compared against predictor averages and percent difference between classes were calculated (>33% highlighted in orange)
- Team developed some insights and identified promising variables
  - Buyers (success class) more often had non-flight bonus (NFB) activity and possessed an airline branded credit card

Sale	Average of		Average of		Average of		Average of		Average of		Average of	
	Mileage Balance	Premier YTD Miles	MP Credit Miles Earned 12months	Rewards Credit Miles Earned 12months	SB Credit Miles Earned 12months	MP Credit Miles Added 12months Dummy	Rewards Credit Miles Added 12months Dummy	SB Credit Miles Added 12months Dummy	Credit Miles Added 12months Dummy	NFB Miles Earned 12months	NFB Transact. 12 months	
0	64275	136	1.9202	1.0168	1.0077	0.5023	0.0401	0.0038	0.5177	14824	10.2573	
1	63625	124	2.2378	1.0220	1.0283	0.6803	0.0520	0.0126	0.7055	22329	14.2441	
% Diff	-1.0%	-9.1%	16.5%	0.5%	2.1%	35.4%	29.5%	227.8%	36.3%	50.6%	38.9%	

Sale	Average of		Average of		Average of		Average of		Average of		Average of	
	NFB Enroll Miles Earned 6months Dummy	NFP Enroll Miles Earned 12months Dummy	Num Online Purch. 12months	Average of Travel Miles 12months	Average of Travel Transact. 12months	Days Since Award Flight 12months	Average of Award Flight Dummy	Days Since Enroll. In FF Program	Average of Club Member Dummy	Average of College+ Member Dummy	Average of E-Mail Dummy	
0	0.0368	0.0725	0.0440	404	1.1960	676968	0.3231	3479	0.0872	0.0034	0.6469	
1	0.1260	0.2488	0.1402	574	1.5685	653596	0.3465	3211	0.1087	0.0157	0.7071	
% Diff	242.8%	243.0%	218.8%	42.0%	31.1%	-3.5%	7.2%	-7.7%	24.6%	368.3%	9.3%	

# Modeling Techniques

- Team had to evaluate best way to model data in the face of certain trade offs
- Logistic regression had the advantage of handling both continuous and categorical data, but can't directly incorporate proportion and costs
- Discriminant analysis can directly incorporate proportion and costs, but may react poorly to categorical data
  - Assumption of multivariate normal distribution violated
- In the end, team decided to perform both techniques and compare results
  - If results were reasonably similar and model contained few dummy variables, then select discriminant analysis

# Logistics Output

Parsimony

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-2.68867	0.13449	0.00000	*
Mileage Balance	0.00000	0.00000	0.00759	1.00000
SB Credit Miles Added 12months Dummy	1.37623	0.57135	0.01601	3.95996
Credit Miles Added 12months Dummy	0.39393	0.14069	0.00511	1.48279
NFB Transact. 12 months	0.03421	0.00643	0.00000	1.03480
NFB Enroll Miles Earned 12months Dummy	0.65797	0.15834	0.00003	1.93086
E-Mail Dummy	0.31798	0.12618	0.01173	1.37434

Cut off Prob.Val. for Success (Updatable)	<b>0.092</b>
---	--------------

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	322	62
0	1559	936

Residual df	2872
Residual Dev.	2153.48755
% Success in training data	13.3379646
# Iterations used	8
Multiple R-squared	0.04776933

Error Report			
Class	# Cases	# Errors	% Error
1	384	62	16.15
0	2495	1559	62.48
<b>Overall</b>	<b>2879</b>	<b>1621</b>	<b>56.30</b>



# Logistics Output



## Profit Maximization

**Revenue:** predicted *actual* successes x revenue/success (\$1143.25)

- **Cost:** (predicted successes + predicted failures)x mailing costs (\$0.30)

**Profit**

## High Cost for Misclassifying Successes (as failures):

- \$1143.25 in lifetime revenue loss for missing a promotion taker
- Only \$0.30 loss for misclassifying a failure (mailing cost)

## Cut-off Value Adjusted to Maximize Profits:

Extrapolated Classification Confusion Matrix			
		Predicted Class	
Actual Class		1	0
1		537	103
0		2598	1560

Cut off Prob.Val. for Success (Updatable)

**0.092**

$537 \times \$1143.25 =$  **\$613,544**

$(537 + 2598/.00185) \times \$0.30 =$  **\$421,167**

**Profit** **\$192,655**

# Discriminant Output

Variables	Classification Function	
	1	0
Constant	-4.3614192	-1.66164255
Mileage Balance	0.00000076	0.00000281
SB Credit Miles Added 12months Dummy	2.05637503	-0.39000067
Credit Miles Added 12months Dummy	1.93270576	1.54490471
NFB Transact. 12 months	0.0748618	0.03778628
NFP Enroll Miles Earned 12months Dummy	0.33540472	0.10360844
E-Mail Dummy	2.88937116	2.58875108

Cut off Prob.Val. for Success (Updatable) **0.068**

Classification Confusion Matrix		
Actual Class	Predicted Class	
	1	0
1	605	29
0	3586	578

Error Report			
Class	# Cases	# Errors	% Error
1	634	29	4.54
0	4164	3586	86.11
<b>Overall</b>	<b>4798</b>	<b>3615</b>	<b>75.34</b>

# Discriminant Output

## Profit Maximization

**Revenue:** predicted successes x revenue/success (\$1143.25)

- **Cost:** (predicted successes + predicted failures)x mailing costs (\$0.30)

### Profit

### High Cost for Misclassifying Successes:

- \$1143.25 in revenue loss for missing a promotion taker
- Only \$0.30 loss for misclassifying a failure (mailing cost)



### Cut-off Value Adjusted to Maximize Profits:

Error Report			
Class	# Cases	# Errors	% Error
1	634	29	4.54
0	4164	3586	86.11
<b>Overall</b>	<b>4798</b>	<b>3615</b>	<b>75.34</b>

Cut off Prob.Val. for Success (Updatable)

**0.068**

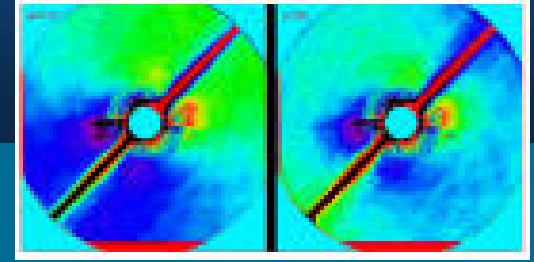
$$634 \times \$1143.25 = \quad \quad \quad \mathbf{\$691,666}$$

$$(634 + 4164/.00185) \times \$0.30 = \quad \quad \quad \mathbf{\$581,191}$$

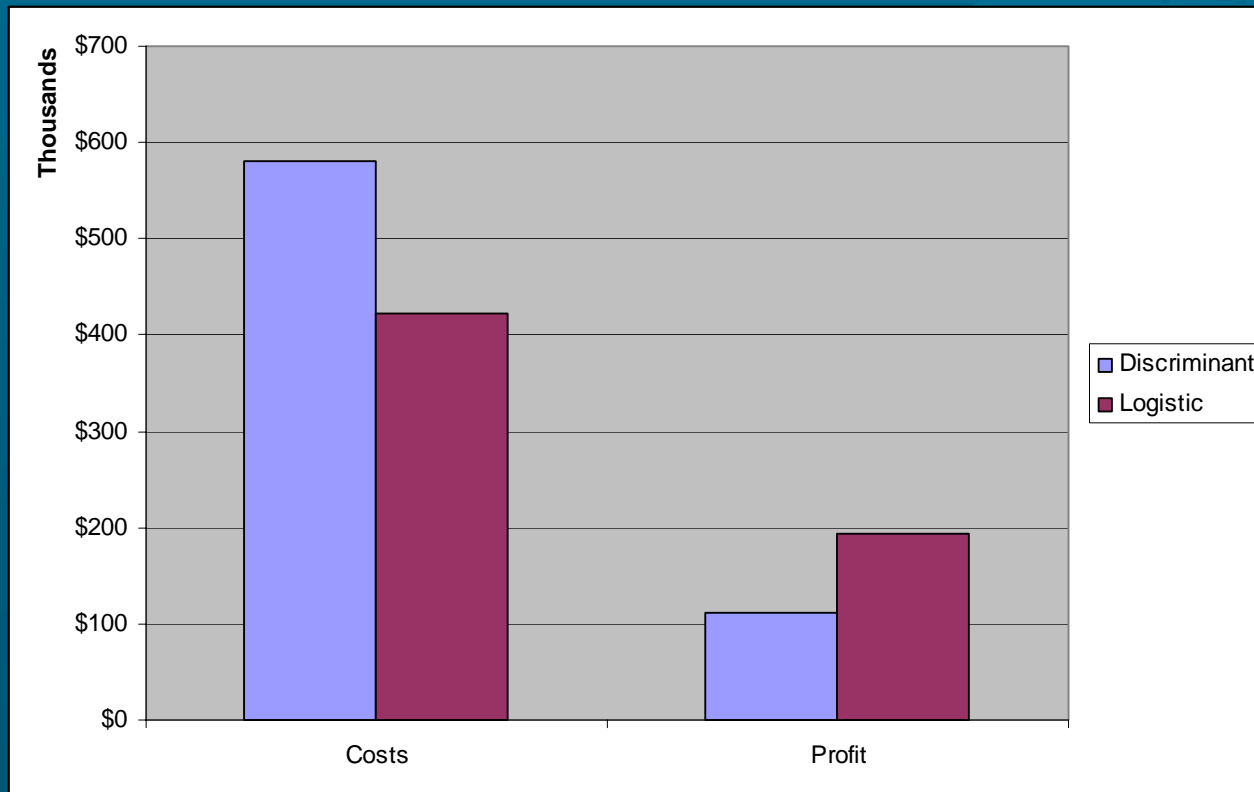
**Profit**

**\$110,475**

# Comparison



## Logistic Regression vs Discriminant Analysis



# Conclusion

## Problem:

Maximizing profits of future frequent flyer direct mailing campaigns by targeting airline customers most likely to accept an offer.

- Very low acceptance rate in the past:  
635 acceptances out of 2.25 million mailings

## Solution:

Distinguish likely buyers from those unlikely to accept a promotion by utilizing a logistic regression model.

# Conclusion



## Recommendations:

**Future direct mailing campaigns should use the above logistic model to:**

- Measure candidates against 6 criteria, including:
  - Frequent flyer mileage balance
  - Level of frequent flyer miles earned via credit cards
  - Level of non-flight bonus miles activity
- Maximize profits by using the optimized cut-off value which takes into account:
  - Prior proportion of promotion takers vs. non-takers
  - Disparate misclassification costs:
    - \$1143.25 for misclassifying taker as a non-taker
    - \$ 0.30 for misclassifying a non-taker as a taker



## ***Targeting the Promotion Takers***

Brad Barber, Andrew Choi, Michael Kilby, Jim Turner

# Backup





# Profit Analysis Comparison

## Validation Data

### Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	<b>0.092</b>
---	--------------

Rev	Cost
1143.25	0.3

Extrapolated Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	538	90
0	2603	1568

# Mailings	2,249,060
Cost	421,842

# Successes	635
# Failures	2,248,425
Revenue	614,497
<b>Profit</b>	<b>192,655</b>

# Sample	4,798
# Sample Failures	4,163
% of Total Failures	0.19%

Log

Error Report			
Class	# Cases	# Errors	% Error
1	627.5	90	14.34
0	4170	2603	62.41
<b>Overall</b>	<b>4798</b>	<b>2693</b>	<b>56.12</b>

D/A

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	608	20
0	3770	400

# Mailings	2,249,060
Cost	611,032

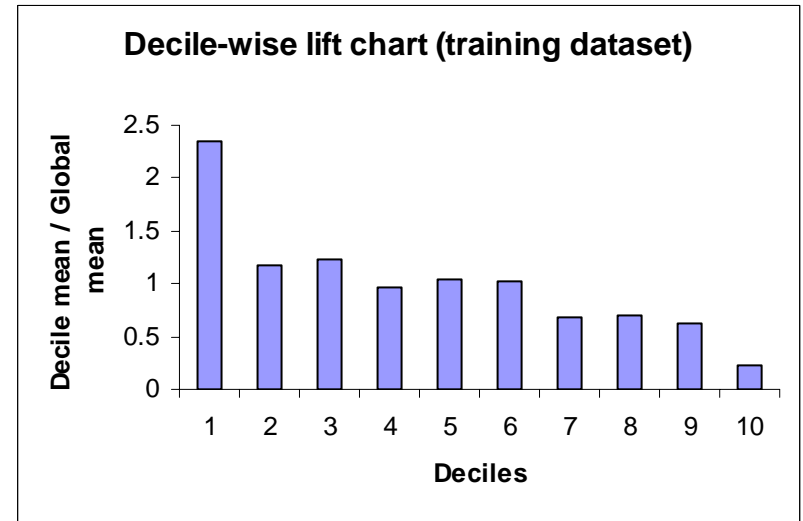
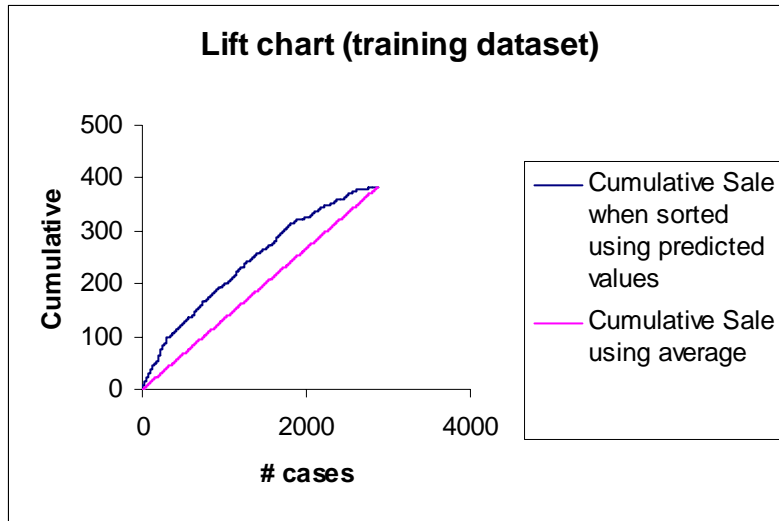
# Successes	635
# Failures	2,248,425
Revenue	694,524
<b>Profit</b>	<b>83,492</b>

# Sample	4,798
# Sample Failures	4,163
% of Total Failures	0.19%

Error Report			
Class	# Cases	# Errors	% Error
1	627.5	20	3.19
0	4170	3770	90.41
<b>Overall</b>	<b>4797.5</b>	<b>3790</b>	<b>79.00</b>

# Lift Charts – Logistic Regression

Tng



Val

