



# THE SACRAMENTO BEE

**Predictive Analytics  
Subscriber Data Modeling**

*Life. Captured daily.*  
**The Sacramento Bee**  
[sacbee.com](http://sacbee.com)



# Agenda

---

## 1. Database Structure

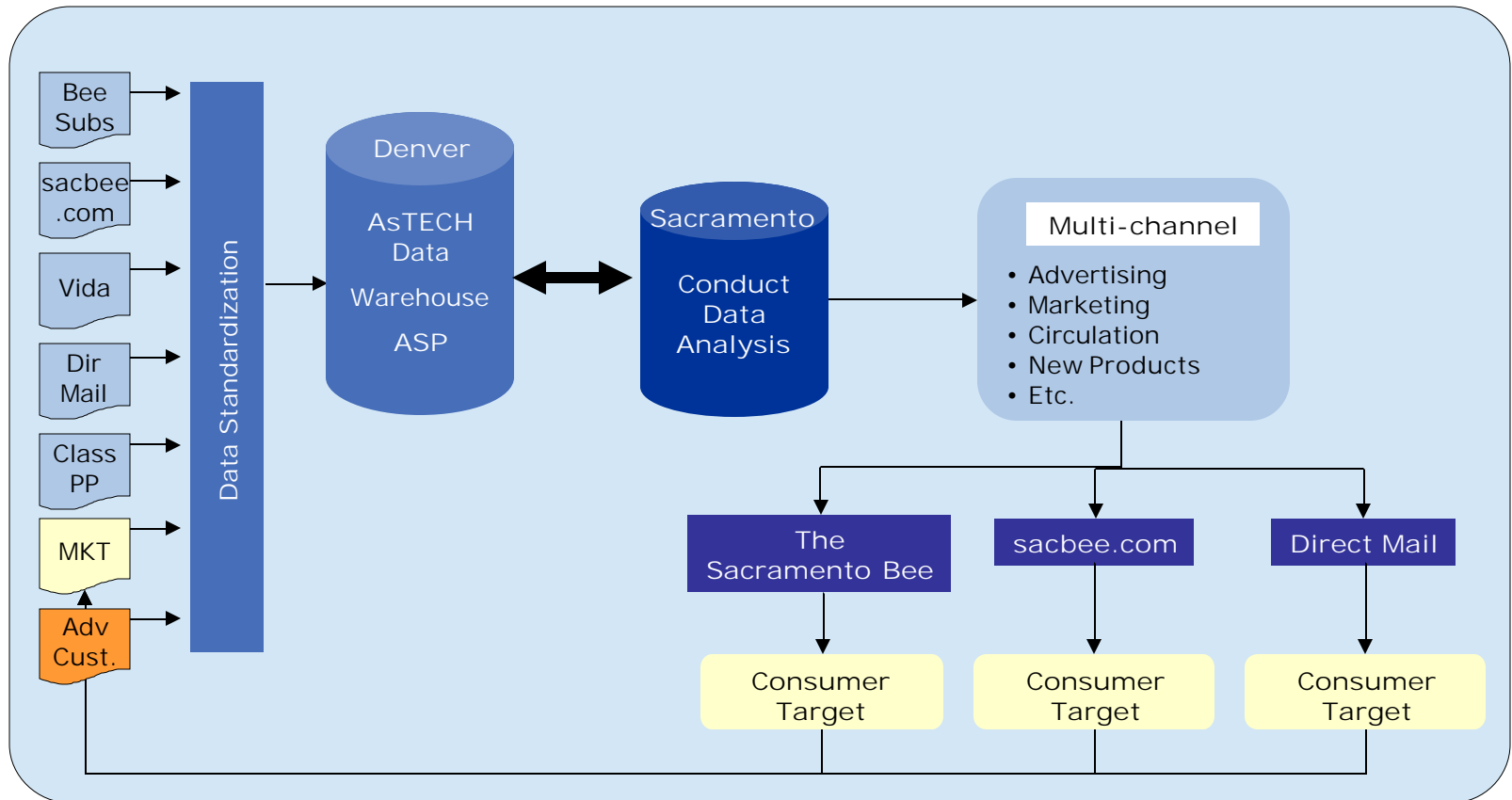
## 2. Example Circulation Application

- Customer transaction data in predictive behavioral model
- Use of multiple communication channels for acquisition/retention

## 3. ROI/Value of Database

# Sacramento Database

Database System Diagram

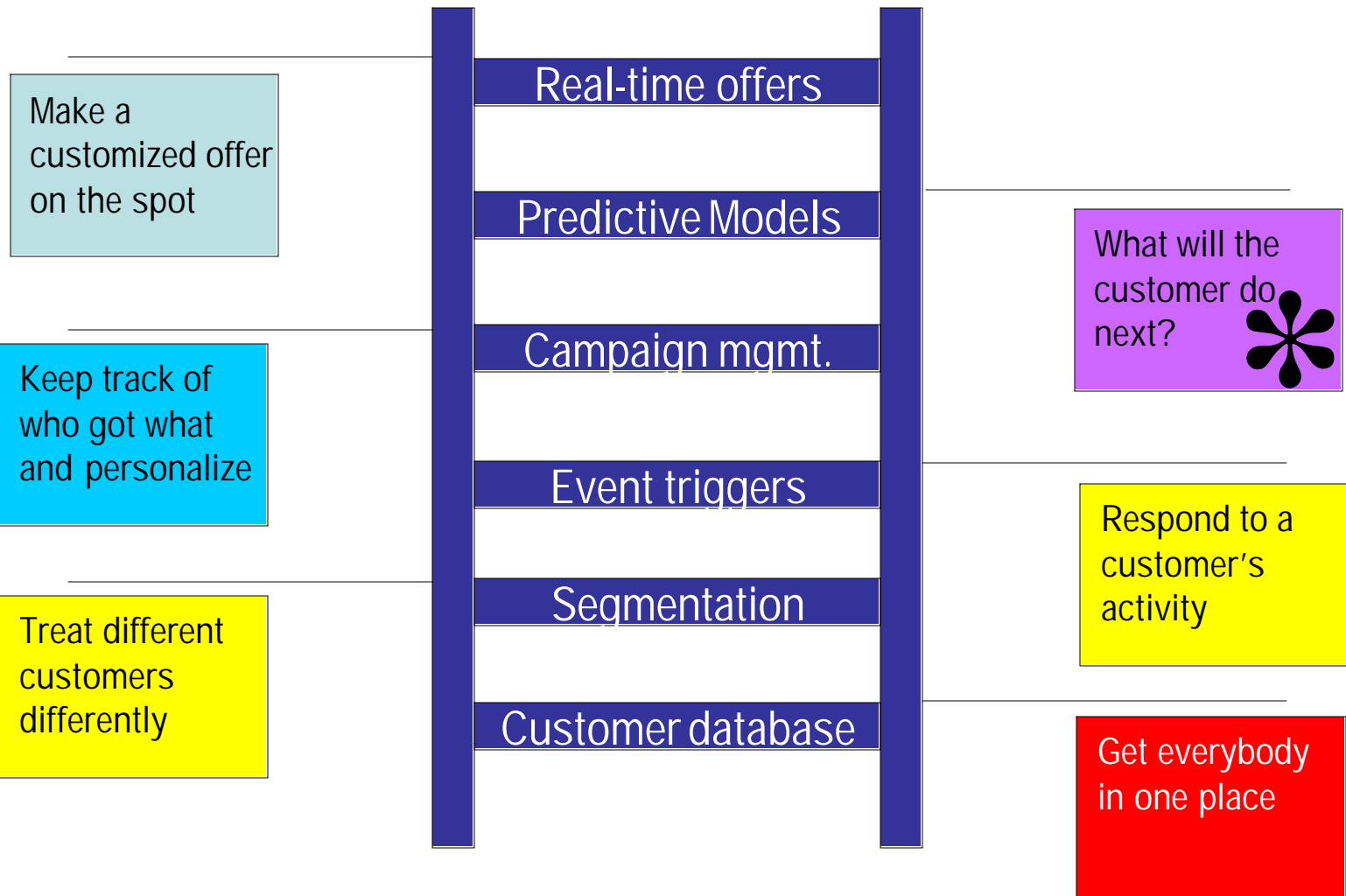


## Other Necessary tools:

- MS Access
- Excel



# Customer Communications Ladder





## Example: Circulation Application

---

### Problem

Have a growing volume of “non-pay” subscription stops



## Goal

---

**Reduce the volume of “non-pay” customer STOPS**



## Project Background

---

### Data

- New subscription starts from 01/01/06 to 06/30/06
- 8 aggregate zip codes selected
- Each customer record received data appends:
  - **Household demographic and behavioral data**
  - **Circulation transaction and payment history**



## Project Deliverables

---

### Resulting Predictive Model to Identify Potential Non-Pays

Model development using Logistic Regression

**70% accurate at predicting non-pay customers**

### Acquisition/Retention

- Retention strategy to focus on highest risk to be non-pay
- Acquisition strategy to wean us from writing non-pay orders



## Non-Pay Stops by Age

---

No Pay Flag	Age						
	18-24	25-34	35-44	45-54	55-64	65-74	75+
No	39%	43%	44%	59%	62%	65%	69%
Yes	61%	57%	56%	41%	38%	35%	31%

---

TOTAL = All 2006 starts that had household append (non-pays and other stop and active)

- **Age breakdowns are fine for use in model (1-7 value)**
- **45+ more likely to pay**
- **18-44 are more likely to be non-pay**



## Non-Pay Stops by Age and Income

---

- **57% of starts age 18-34 stop due to non-pay**
  - Bulk found in lower income (<\$45k) segments
- **42% of all starts are 18-44 with income of less than \$45k**
  - Only segment where more than 50% of stops are non-pay
  - Segment makes up nearly 1/3 of all households in test base
  - Index at 124 for all subscription starts and 147 for non-pay stops (young, less affluent)
- **Subscribers age 55+ are more likely to pay for their subscription (not end up non-pay), regardless of income level**



# Score Geography (where) and People (who)

## Zip Code No Pay Index

ZIP5	Urban Sub Starts %	No Pay %	No Pay Index
95815	14.9%	16.7%	112
95816	8.8%	7.1%	81
95817	10.9%	12.6%	116
95818	9.6%	8.4%	87
95819	6.0%	3.3%	54
95820	16.7%	17.6%	105
95822	20.7%	20.4%	98
95824	12.3%	13.9%	113

Index = (No Pay % / Urban Sub Starts %)\*100

## PrismNE No Pay Index

PRISMNE	Urban Sub Starts %	No Pay %	No Pay Index
04	4.1%	3.3%	80
*07	6.9%	4.0%	58
16	6.1%	5.4%	89
*26	7.0%	4.3%	61
29	13.2%	11.0%	83
31	8.3%	9.3%	112
40	5.5%	4.5%	83
54	14.8%	15.4%	105
59	5.1%	4.9%	95
61	6.1%	4.6%	76
65	7.0%	10.4%	148
66	15.0%	22.1%	148
67	0.3%	0.2%	92

Index = (No Pay % / Urban Sub Starts %)\*100

- The 95819 zip may provide more quality subscribers (ideal zip)
  - It indexes at 64 for starts.
  - Shows a low incidence of non-pay rate.
- Poorest performing zip codes are 95815, 95817 and 95824
  - Score high for subscriber and high for non-pay index.
  - High for subscriber Index
  - To bring down non-pays, focus on better orders.
- PrismNE codes **65 and 66** are fueling non-pay
  - Are a very high portion of the subscriber base as well.



## Predictive Model Learning

---

- Low-level predictive variables:
  - Age and income are important, but lower-level predictors
- High-level predictive variables:
  - Start source
  - Past payment or non-pay history
  - 7-day subscriber or not
  - Prism segment
- Length of residency (LOR):
  - Longer, the subscriber is less likely to end up a non-pay



## Predictive Model Insights

---

### Starts MORE Likely to end up a Non-Pay Stop?

- **New subscribers who had at least one non-pay stop, past year**
  - 4.4 times more likely to repeat as a non-pay stop
- **7-day starts are 2.2 times more likely to end up a non-pay stop**
  - When compared against non 7-day subscribers
- **Telesales, door crews and FSI starts are 1.5 to 2.0 times more likely to end up a non-pay**
- **PrismNE 65 & 66 index high for non-pay starts**
  - 65 & 66 are 1.5 times more likely to be a non-pay stop

All statements are based on all other variables held constant



## Predictive Model Insights

---

### Starts LESS Likely to be a Non-Pay Stop?

- **Starts who make at least one payment within their first 30-days**
  - Are 4.3 times less likely to end up a non-pay stop
- **Starts who made at least 1 payment (on a prior subscription start)**
  - Are 2.7 times less likely to be a non-pay stop
- **Press Club members are 2.0 times less likely to end up non -pay**
- **Starts with a 50% discount rate**
  - Are 1.7 times less likely to be a non-pay stop
- **With each increase in age and income bracket**
  - The subscriber is less likely to end up as a non-pay stop

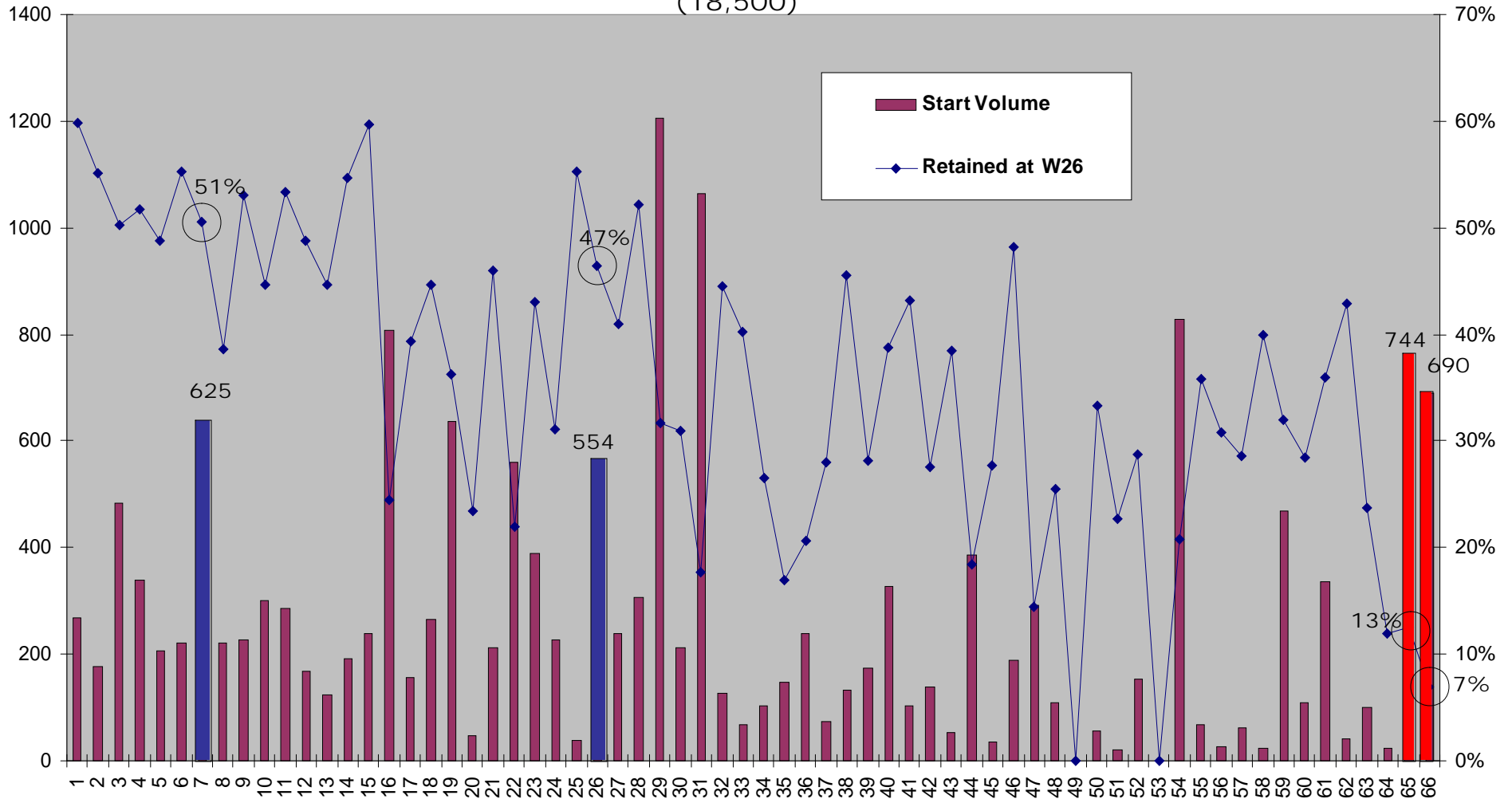
All statements are based on all other variables held constant

# Subscription Starts and Retention by PrismNE Segment

Start Volume

January – April, 2006  
(18,500)

Retention %



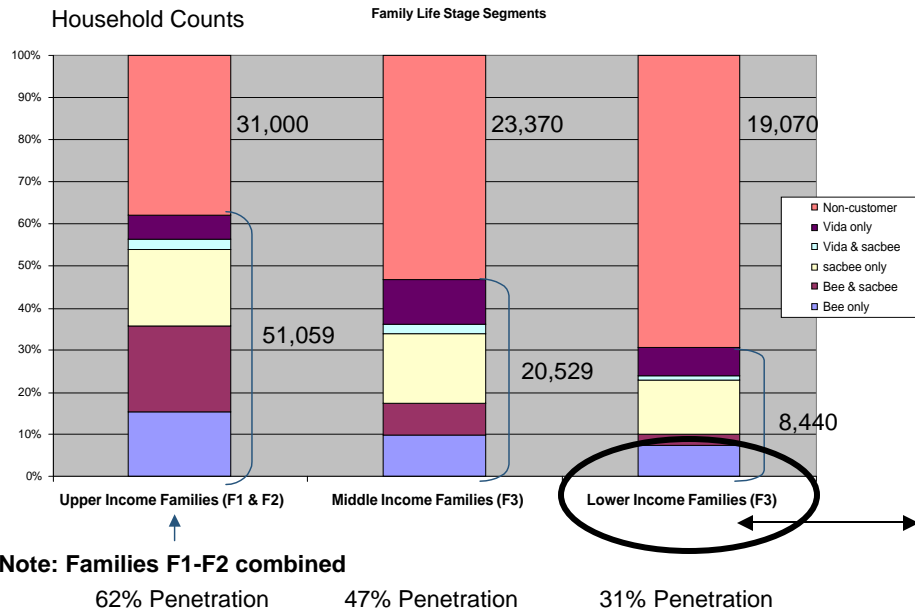
PrismNE Segment

# High Percentage of Non-Pay Segments

## Segments 65 and 66

- Part of Sustaining Families (lower-income families)

Product Touch – Family



**65**  
Big City Blues

F4 Sustaining Families

Inner-City Multi-Ethnic Families  
 Renter / Hi-Rise Multi  
 Age <45  
 WC, Service, BC  
 Elem. School, H.S.  
 Black, Asian, Hispanic

---

Rent videos  
 Eat at Sizzler Steakhouse  
 Essence Magazine  
 All My Children TV  
 Nissan Sentra

**66**  
Low-Rise Living

F4 Sustaining Families

Poor, Young Solo-Parent Families  
 Renter / Hi-Rise Multi  
 Age <35  
 Service, BC  
 Elem. School, H.S.  
 Black, Hispanic

---

Buy gospel music  
 Shop at Footlocker  
 Ebony Magazine  
 Soul Train TV  
 Mitsubishi Gallant

PrizmNE segment descriptions: Claritas Corporation, 2006.



## Applying Predictive Analytics to Retention

---

### **As part of the business practice:**

- Isolate new subscription starts in a file every two-weeks
- Score all new starts by “probability” to end up a non-pay
- Develop communication plan, begin w/highest-risk group:
  - Early and often in customer cycle
  - Seek ‘some’ payment first 30-days (significant)
- Track results

# New Subscription Starts

Score based on probability to be non-pay

New Customer Start	Designation	Probability	Calculated values for each statistically significant variable											EXP()	
		P	0.016	0.571	-0.535	-0.074	-0.114	-0.702	-0.108	1.474	0.006	-0.986	-1.448	0.770	E
		P	CONSTANT	TS	DISC50	AGE	INCOME	PRESS CLUB	LOR	PRIOR NO PAY	PRISMNE NO PAY INDEX	MAKE PAY	PAY 0 TO 30	SUB 7 DAY FLAG	
1	Medium	43%	0.016	0.571	0.000	-0.296	-0.570	0.000	-0.324	0.000	0.531	-0.986	0.000	0.770	0.75
2	High	66%	0.016	0.000	0.000	-0.222	-0.228	-0.702	-0.324	1.474	0.888	-0.986	0.000	0.770	1.9848
3	High	84%	0.016	0.000	0.000	-0.148	-0.114	0.000	-0.216	1.474	0.887	-0.986	0.000	0.770	5.3807
4	High	64%	0.016	0.000	0.000	-0.074	-0.114	-0.702	-0.216	1.474	0.887	0.000	-1.448	0.770	1.8091
5	High	71%	0.016	0.000	0.000	-0.296	-0.228	0.000	-0.432	1.474	0.600	-0.986	0.000	0.770	2.5043
6	High	61%	0.016	0.000	0.000	-0.296	-0.456	0.000	-0.432	1.474	0.348	-0.986	0.000	0.770	1.5496
7	High	45%	0.016	0.000	0.000	-0.296	-0.456	0.000	-0.432	1.474	0.498	-0.986	0.000	0.000	0.8338
8	High	47%	0.016	0.000	0.000	-0.296	-0.342	-0.702	-0.540	1.474	0.498	-0.986	0.000	0.770	0.8978
9	Low	12%	0.016	0.000	0.000	-0.222	-0.570	-0.702	-0.324	0.000	0.498	0.000	-1.448	0.770	0.1378
10	High	60%	0.016	0.000	0.000	-0.222	-0.342	0.000	-0.324	0.000	0.498	0.000	0.000	0.770	1.4862
11	Low	19%	0.016	0.000	0.000	-0.518	-0.114	0.000	-0.432	0.000	0.571	-0.986	0.000	0.000	0.2316
12	Medium	28%	0.016	0.571	0.000	-0.370	-0.114	0.000	-0.540	0.000	0.458	-0.986	0.000	0.000	0.3809
13	Low	16%	0.016	0.000	0.000	-0.296	-0.684	0.000	-0.216	0.000	0.531	-0.986	0.000	0.000	0.195
14	Medium	37%	0.016	0.000	0.000	-0.296	-0.228	0.000	-0.216	0.000	0.887	0.000	-1.448	0.770	0.5974
15	Low	8%	0.016	0.000	0.000	-0.296	-0.684	-0.702	-0.432	0.000	0.348	0.000	-1.448	0.770	0.0882
16	High	52%	0.016	0.000	0.000	-0.296	-0.798	0.000	-0.432	1.474	0.348	-0.986	0.000	0.770	1.1007
17	High	55%	0.016	0.000	0.000	-0.222	-0.570	0.000	-0.324	1.474	0.498	0.000	-1.448	0.770	1.2143
18	Low	5%	0.016	0.000	0.000	-0.518	-0.342	-0.702	-0.540	0.000	0.600	0.000	-1.448	0.000	0.0532

Each new individual customer start

Regression analysis isolates the variables showing significance for predicting a non-pay customer. These variables are now used to score each new subscription start, predicting the probability that it will end up a non-paying subscriber or not.



## Non-pay Risk of New Subscribers

		Probability
<u>New</u>	<u>Customer</u>	<u>Designation</u>
<u>Start</u>		P
1	Medium	43%
2	High	66%
3	High	84%
4	High	64%
5	High	71%
6	High	61%
7	High	45%
8	High	47%
9	Low	12%
10	High	60%
11	Low	19%
12	Medium	28%
13	Low	16%
14	Medium	37%
15	Low	8%
16	High	52%
17	High	55%
18	Low	5%

Customers are grouped into 3 groups of:

- High
- Medium
- Low

probability to end up as a non-pay stop

Files of high and medium-risk customers are forwarded to retention for focus



## Applying Predictive Analytics to Acquisition

### Goal:

- Reduce poor starts through better targeting
- Align offer preference with segment

### How:

- Rank limited resource priorities on:
  - Prism segments
  - Geography
  - Offer

**Folsom's evolving and The Bee is evolving with you.**  
**Pick your package**  
**\$20**

No matter what your lifestyle or available time we have a home delivery package for you.

Pick the package that fits your needs for only \$20...

**7-day for 8-weeks**  
Daily Delivery (Promo Code: D8N 250)

**5-day for 10-weeks**  
Thursday-Monday (Promo Code: TM 200)

**2-day for 12-weeks**  
Saturday & Sunday (Promo Code: SS 127)

**Call 916.321.1111**

CELEBRATING OUR 150<sup>th</sup> ANNIVERSARY  
**The Sacramento Bee**

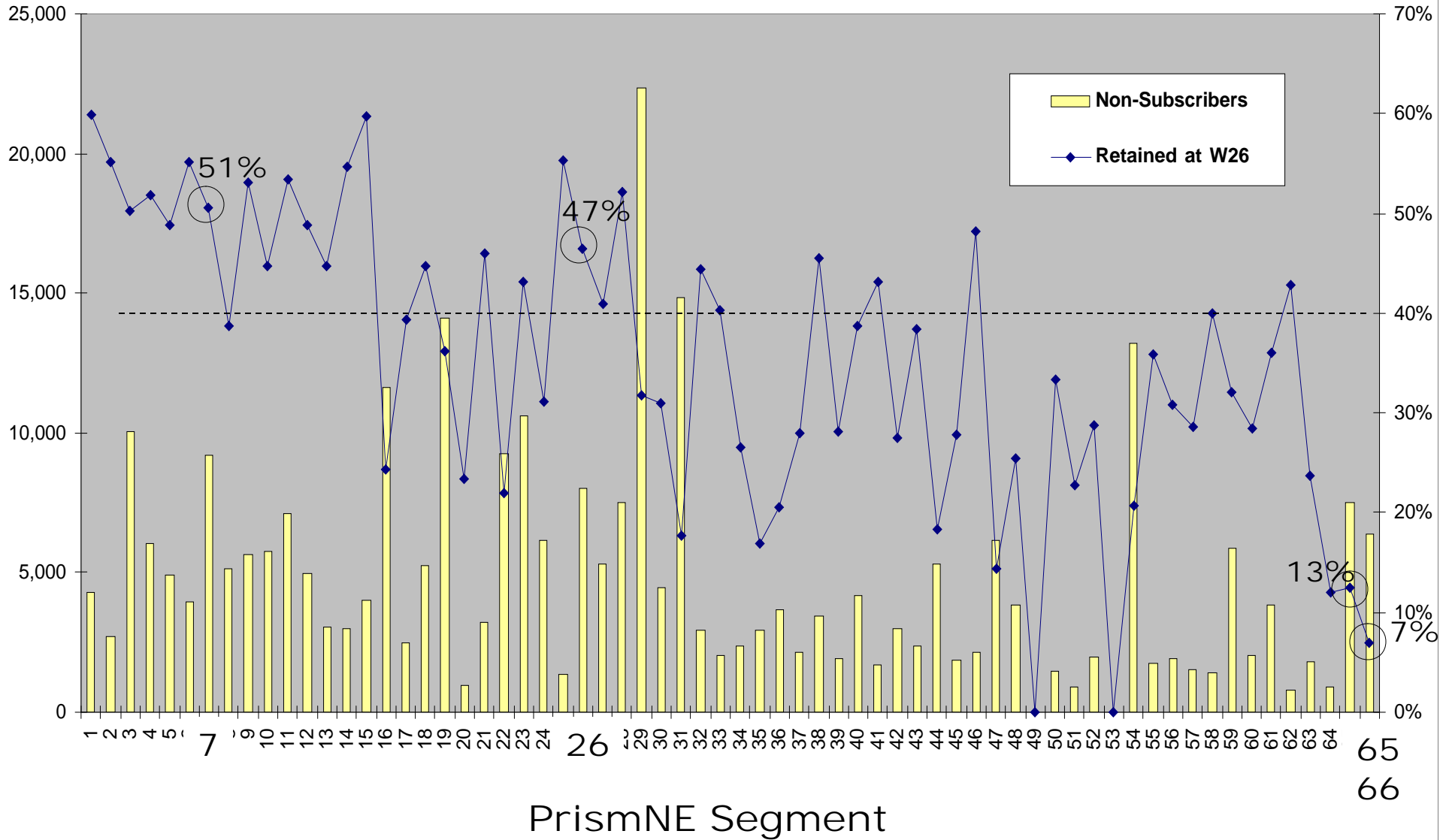
Savings based on newsstand price. Advance payment required. Applicable sales tax included. This offer not good for customers who have subscribed in the 30 days. After introductory offer expires, delivery will continue at the regular rate until the time is notified otherwise. Regular rates may be applied. Offer in participating areas. Mail subscriptions included. Offer expires April 4, 2013.

Promo Code: Marketing / Folsom / D8N250 - TM200 - SS127

# Current Non-Subscriber Household Volume by PrismNE

Non-Subscriber  
Volume

Segment  
Retention %

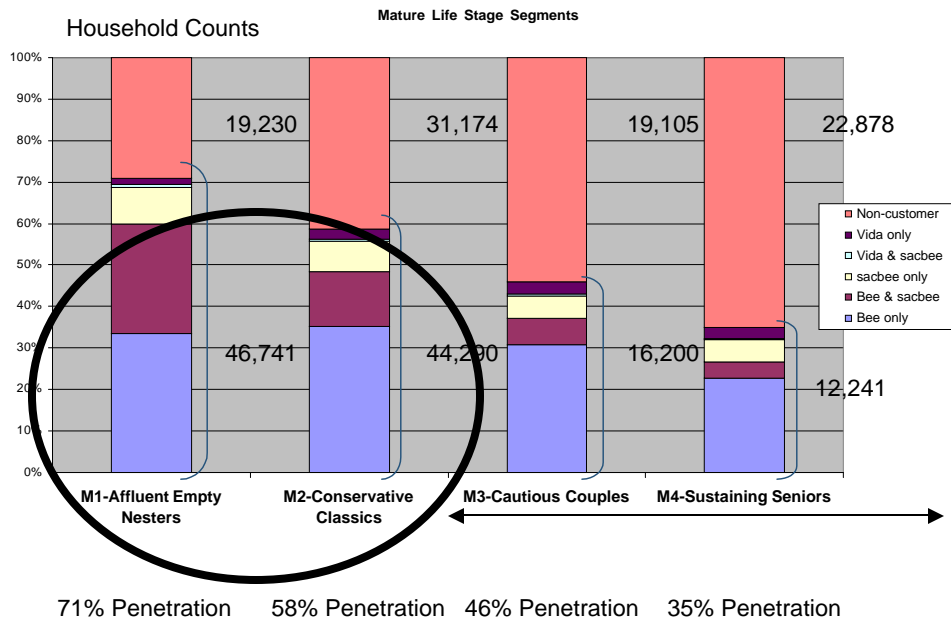


# Good Acquisition Targets With Potential

## Segments 07 and 26

- Part of Mature Years (affluent and upper-middle income empty nesters)

Product Touch – Mature



<p><b>07</b> Money &amp; Brains</p> <p>M1 Affluent Empty Nesters</p>	<p><b>26</b> The Cosmopolitans</p> <p>M2 Conservative Classics</p>
<p>Older, Upscale Urban Sophisticates Owner / SFDU Age 45+ Executive, Professional, WC College Graduate+ White, Asian</p> <p>Shop at Nordstrom Support the arts Business Week Magazine All-news radio BMW 5 Series</p>	<p>Established, Midscale Urban Couples Owner / SFDU, HI-Rise Multi Age 55+ WC Some College White, Black, Asian, Hispanic</p> <p>Go to the theatre Play the lottery Vanity Fair Magazine People's Court TV Lincoln Town Car</p>

PrizmNE segment descriptions: Claritas Corporation, 2006.



## Focus on Best Acquisition Targets – Field Test

Zip	Area	Churn	Current Bee Circ.	Current Non-Subs	Best Retaining Target Segs/ Current Non-sub HHlds	Worst Retaining Segments/ Current Non-sub HHlds	Revenue per Circ Unit
95630	Folsom	47%	9,526	15,776	4,452	965	\$142.03

All non-subscribing households

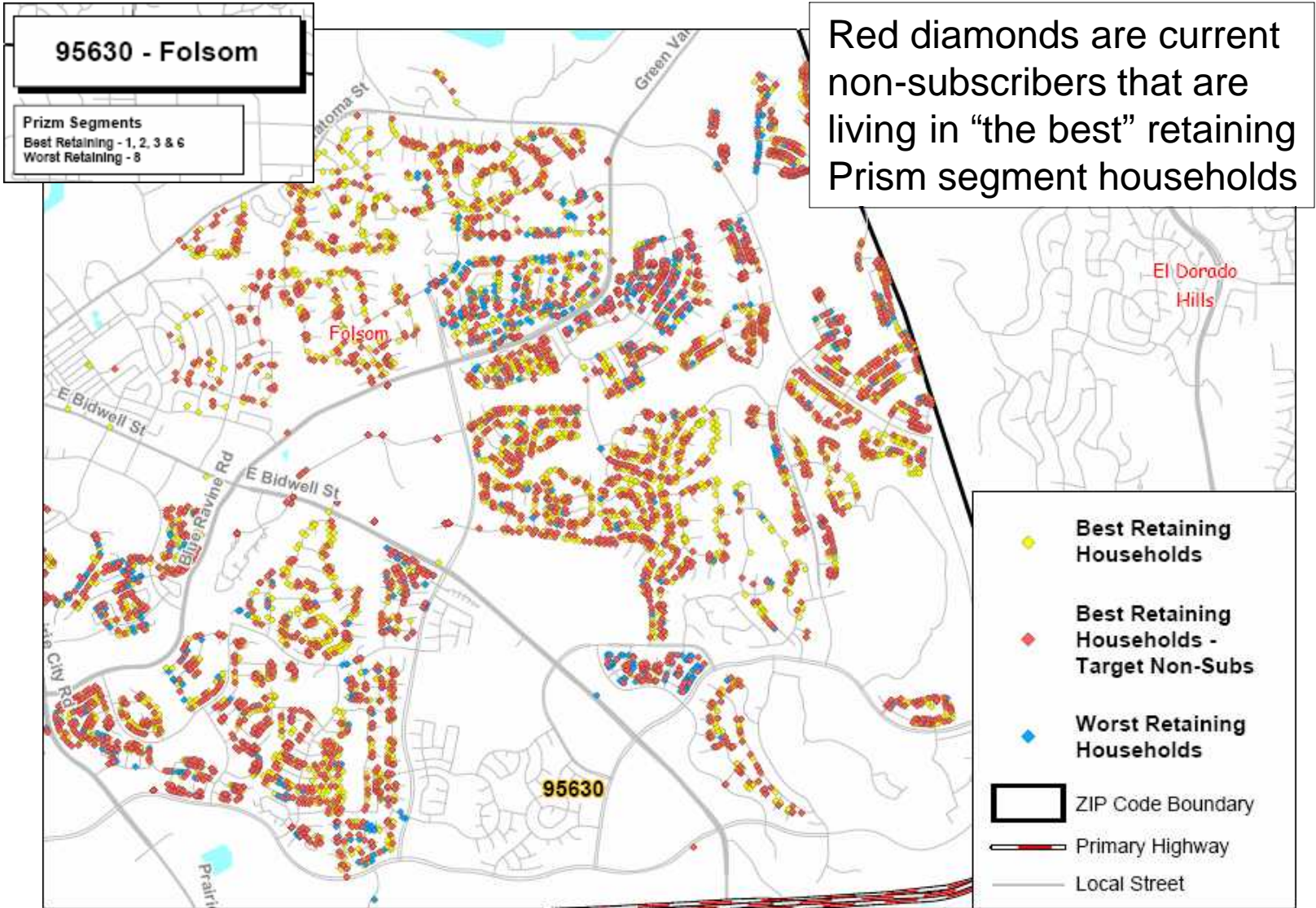
Seg.	95630	
	non-subs	subs
1	714	1,285
2	470	742
3	2,439	1,670
4	-	-
5	85	56
6	829	730
7	-	-
8	965	352
9	38	48
10	130	69
11	192	93

Best target segments  
(Seg. 1, 2, 3, 6)

Segment (8) desirable target  
but harder to acquire ...

By looking at who they are,  
Make a different appeal?

# Best Non-subscriber Targets





## Database - ROI

---

### **Money (\$\$\$):**

- email program alone = \$500,000+
  - Manage targets and integration through MAAX database
- Improved DM response rate and reduced overall spend

### **Strategy/Intelligence:**

- View of product(s) penetration in the market
- View of product portfolio and cross-product relationships
- Coordinate marketing programs across multiple channels

### **New Product Development:**

- Know where we touch the market and where opportunity exists
- View market by lifestyle, demographic, product segments

### **Culture:**

- Cross-division work teams