SPRING PRIVACY SERIES

Alternative Scoring Products

MARCH 19, 2014

Welcome

Overview of Predictive Analytics

Claudia Perlich

Chief Scientist, Dstillery

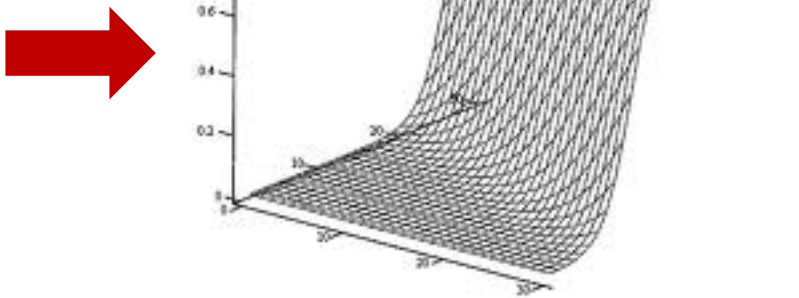
SPRING PRIVACY SERIES

Alternative Scoring Products

MARCH 19, 2014

Predictive Modeling: Algorithms that Learn from Data

```
0.749380154021744 0.00059018398654502 8 779 20
0.798285720814701 0.00034831487664849 1962 20
0.540374430975371 0.00141733796704089 891 20
0.724142820274877 0.00111803594088993 1734 20
0.420231281165794 0.00068347360815481 7 1755 10
1.63577343467107 0.00229268808860358 1546 10
0.2501293492083 0.00843852618368423 1065 10
2.13146775632401 0.00078509352282078 7 779 10
0.819076913250702 0.00818205576664053 1473 20
0.779888840620191 0.0120380305675224 891 20
1.04225917097729 0.00314198703840882 1969 20
1.74645754395112 0.00165522594785467 703 10
0.677900040859134 0.00150284527448034 1236 10
1.05657287551612 0.0382022935590516 1912 20
0.74167007811735 0.0169101519222402 1830 20
0.83606256536671 0.00023750063707252 4 779 20
3.22319651310187 0.00956306677658697 916 10
0.47156025759991 0.00070449374617436 8 1065 10
0.851249341944925 0.055344262295082 1473 10
1.08403941502146 0.00160579161489708 1830 20
0.786523498700205 0.0432004773760777 1775 20
0.193594514913282 0.00049573401130375 5 596 10
```



Sign in | Open an Account | Help

LendingClub

Investing | Personal Loans | How It Works | About Us

PRIME is investing made easy!

Upgrade your account to PRIME today *for free*[1]

NEW Minimum account value required only $5,000

- Set your investment criteria and PRIME places orders as matching inventory becomes available.
- Track your progress toward your target allocation.
- You can place manual orders, pause, or cancel PRIME at any time.

Learn More About PRIME

Looking for a loan?

Member Sign-in

Account Summary GO

Browse Notes GO

Not yet a member? Join now

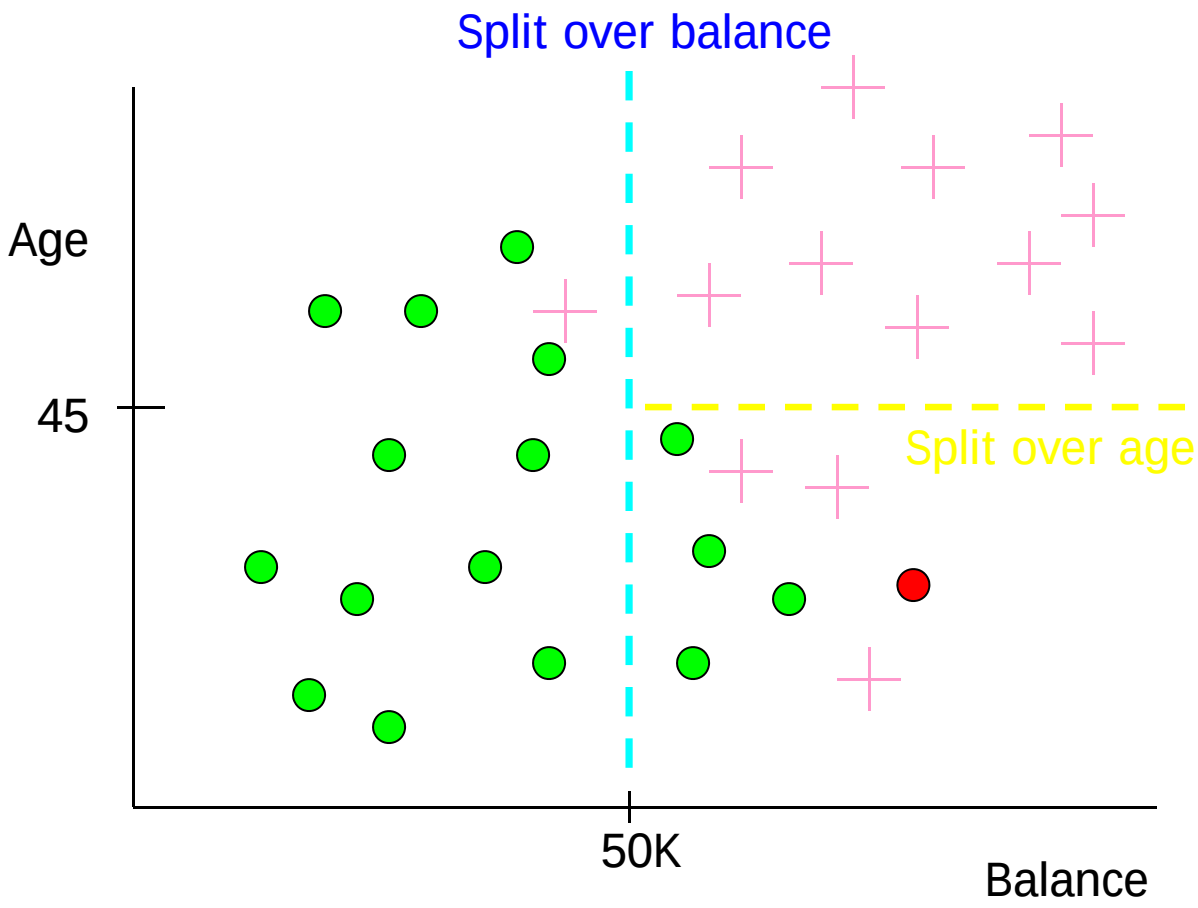
Loans funded to date: $3,873,562,575

Interest paid to Investors: $362,988,792

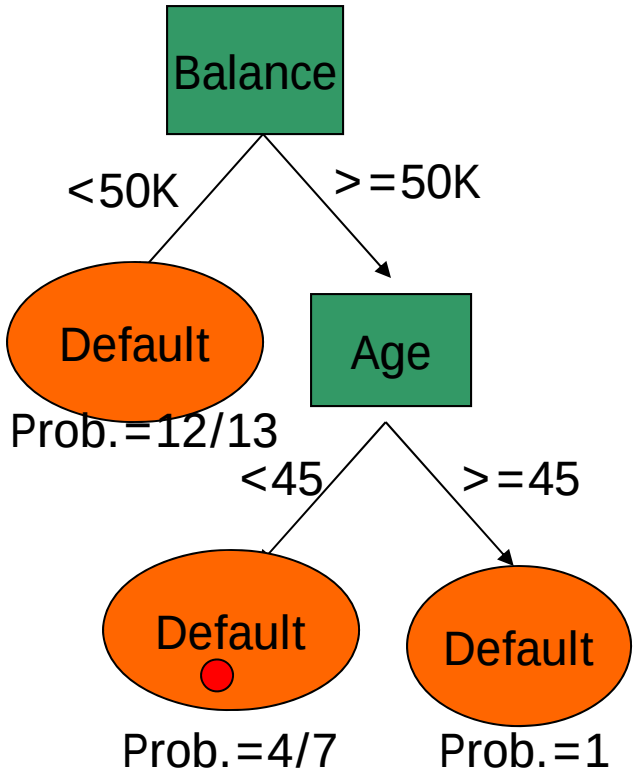
Example: Micro Loans

Age	Income	Default
35	75K	no
68	83K	yes
43	61K	no
71	56K	yes
…	…	…

Learning to Classify



Classification tree



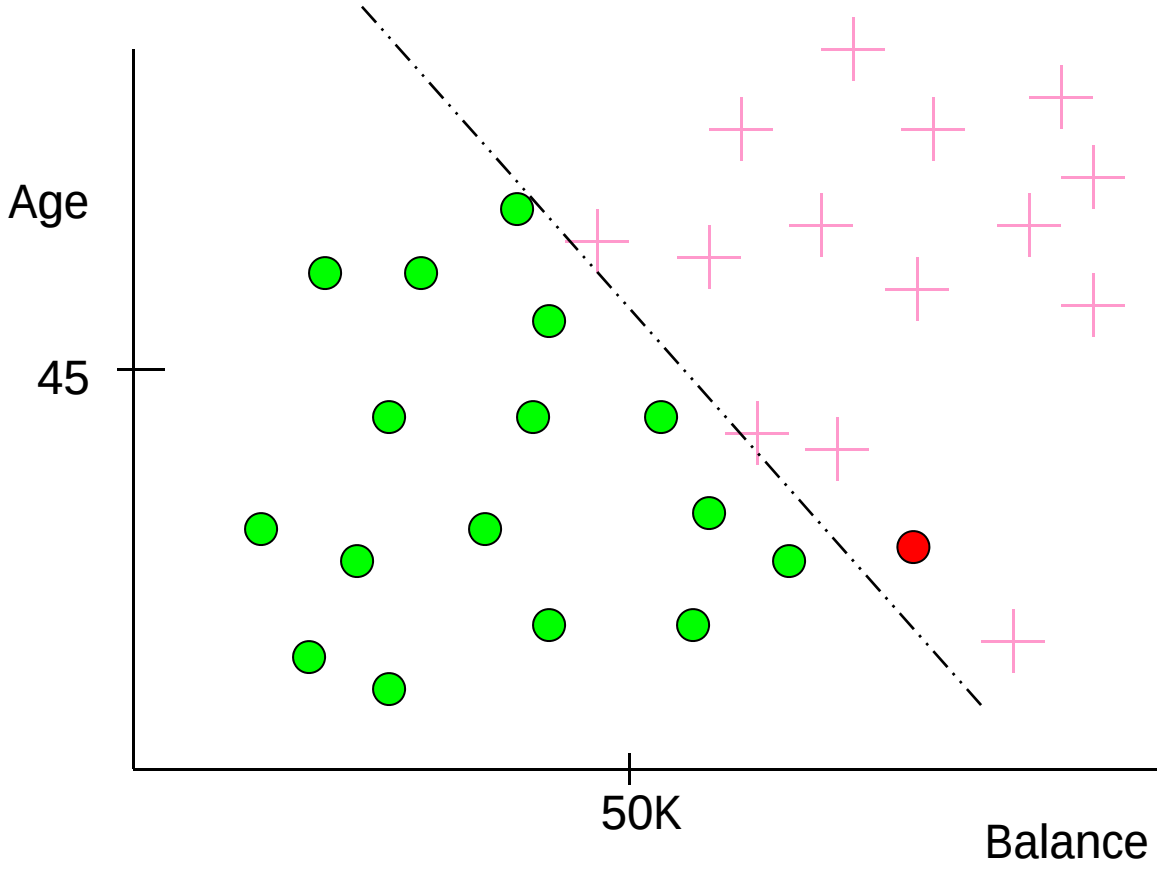
Bad risk (Default) – 16 cases

Good risk (Not default) – 14 cases

Probability of default=4/7

Learning to Classify

Logistic Regression



$$p(+|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$\beta_0 = 123$
$\beta_1 = -1.3$

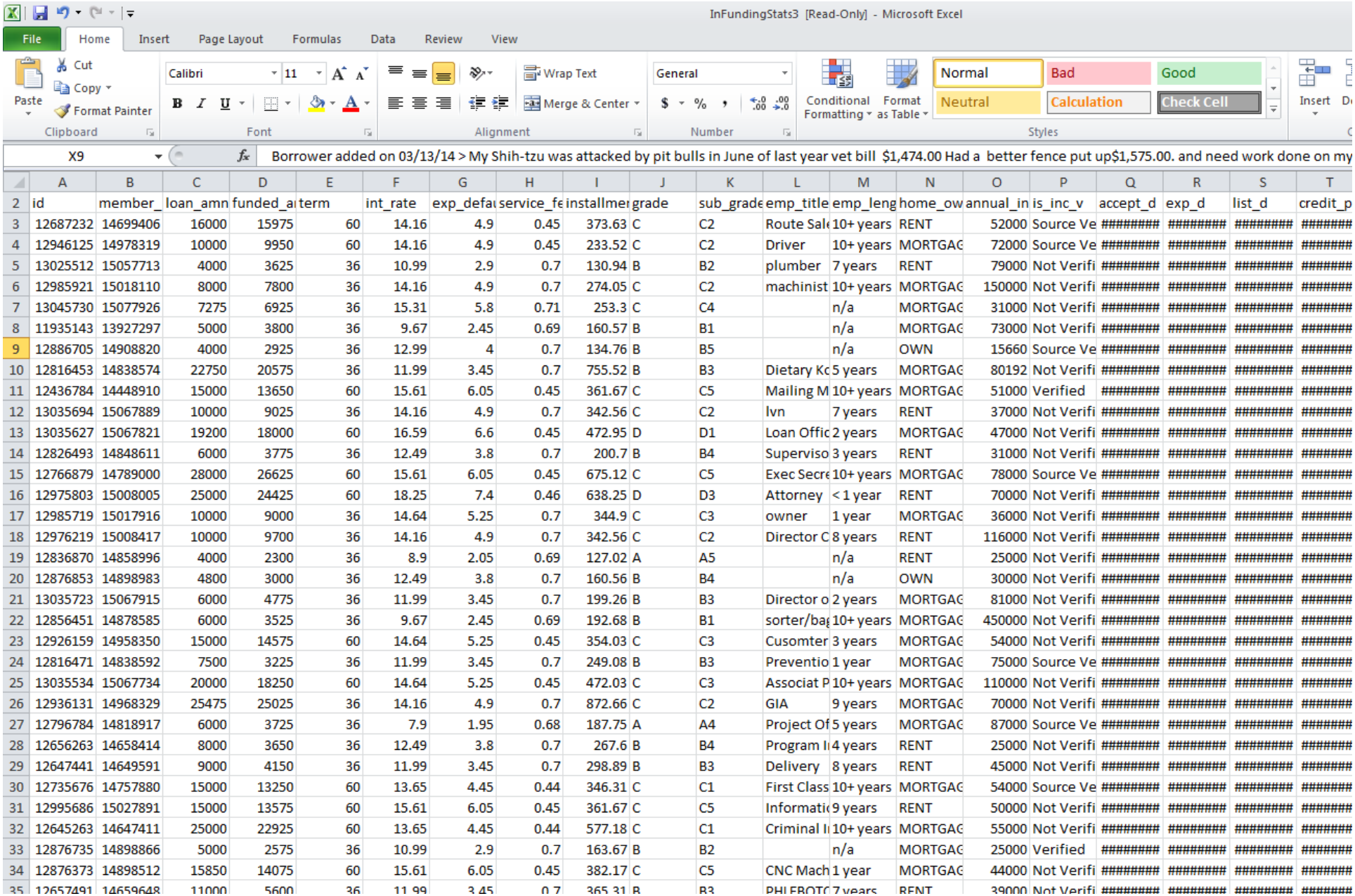
Bad risk (Default) – 16 cases

Good risk (Not default) – 14 cases

$p(+|x) = 0.48$

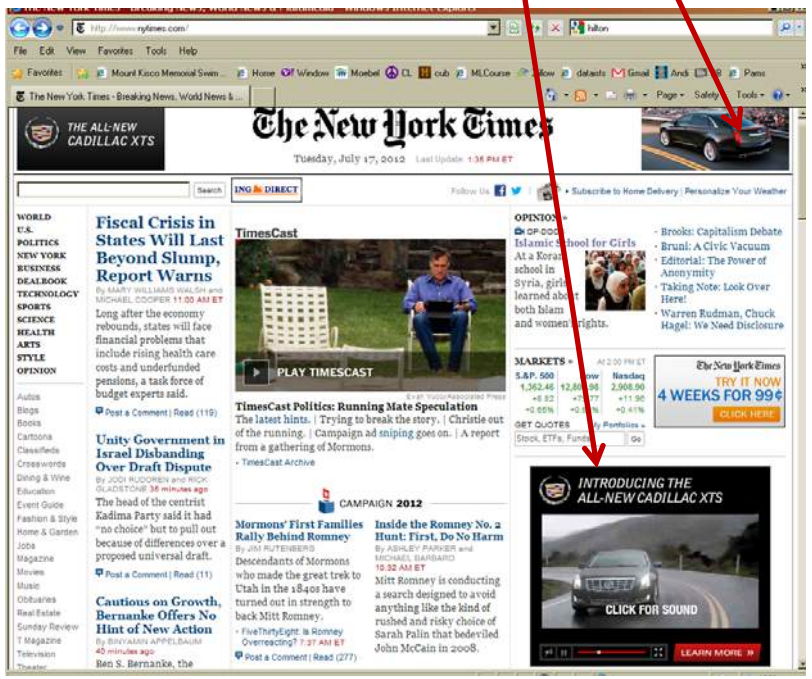
Lending Club Data

- Text
- Loan Category
- Demographic information
- Credit Score



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2	id	member_	loan_amn	funded_a	term	int_rate	exp_defa	service_fe	installmer	grade	sub_grade	emp_title	emp_leng	home_ow	annual_in	is_inc_v	accept_d	exp_d	list_d	credit_p
3	12687232	14699406	16000	15975	60	14.16	4.9	0.45	373.63	C	C2	Route Sal	10+ years	RENT	52000	Source Ve	########	########	########	########
4	12946125	14978319	10000	9950	60	14.16	4.9	0.45	233.52	C	C2	Driver	10+ years	MORTGAG	72000	Source Ve	########	########	########	########
5	13025512	15057713	4000	3625	36	10.99	2.9	0.7	130.94	B	B2	plumber	7 years	RENT	79000	Not Verifi	########	########	########	########
6	12985921	15018110	8000	7800	36	14.16	4.9	0.7	274.05	C	C2	machinist	10+ years	MORTGAG	150000	Not Verifi	########	########	########	########
7	13045730	15077926	7275	6925	36	15.31	5.8	0.71	253.3	C	C4		n/a	MORTGAG	31000	Not Verifi	########	########	########	########
8	11935143	13927297	5000	3800	36	9.67	2.45	0.69	160.57	B	B1		n/a	MORTGAG	73000	Not Verifi	########	########	########	########
9	12886705	14908820	4000	2925	36	12.99	4	0.7	134.76	B	B5		n/a	OWN	15660	Source Ve	########	########	########	########
10	12816453	14838574	22750	20575	36	11.99	3.45	0.7	755.52	B	B3	Dietary Kc	5 years	MORTGAG	80192	Not Verifi	########	########	########	########
11	12436784	14448910	15000	13650	60	15.61	6.05	0.45	361.67	C	C5	Mailing M	10+ years	MORTGAG	51000	Verified	########	########	########	########
12	13035694	15067889	10000	9025	36	14.16	4.9	0.7	342.56	C	C2	lvn	7 years	RENT	37000	Not Verifi	########	########	########	########
13	13035627	15067821	19200	18000	60	16.59	6.6	0.45	472.95	D	D1	Loan Offic	2 years	MORTGAG	47000	Not Verifi	########	########	########	########
14	12826493	14848611	6000	3775	36	12.49	3.8	0.7	200.7	B	B4	Superviso	3 years	RENT	31000	Not Verifi	########	########	########	########
15	12766879	14789000	28000	26625	60	15.61	6.05	0.45	675.12	C	C5	Exec Secre	10+ years	MORTGAG	78000	Source Ve	########	########	########	########
16	12975803	15008005	25000	24425	60	18.25	7.4	0.46	638.25	D	D3	Attorney	< 1 year	RENT	70000	Not Verifi	########	########	########	########
17	12985719	15017916	10000	9000	36	14.64	5.25	0.7	344.9	C	C3	owner	1 year	MORTGAG	36000	Not Verifi	########	########	########	########
18	12976219	15008417	10000	9700	36	14.16	4.9	0.7	342.56	C	C2	Director C	8 years	RENT	116000	Not Verifi	########	########	########	########
19	12836870	14858996	4000	2300	36	8.9	2.05	0.69	127.02	A	A5		n/a	RENT	25000	Not Verifi	########	########	########	########
20	12876853	14898983	4800	3000	36	12.49	3.8	0.7	160.56	B	B4		n/a	OWN	30000	Not Verifi	########	########	########	########
21	13035723	15067915	6000	4775	36	11.99	3.45	0.7	199.26	B	B3	Director o	2 years	MORTGAG	81000	Not Verifi	########	########	########	########
22	12856451	14878585	6000	3525	36	9.67	2.45	0.69	192.68	B	B1	sorter/bag	10+ years	MORTGAG	450000	Not Verifi	########	########	########	########
23	12926159	14958350	15000	14575	60	14.64	5.25	0.45	354.03	C	C3	Cusomter	3 years	MORTGAG	54000	Not Verifi	########	########	########	########
24	12816471	14838592	7500	3225	36	11.99	3.45	0.7	249.08	B	B3	Preventio	1 year	MORTGAG	75000	Source Ve	########	########	########	########
25	13035534	15067734	20000	18250	60	14.64	5.25	0.45	472.03	C	C3	Associat P	10+ years	MORTGAG	110000	Not Verifi	########	########	########	########
26	12936131	14968329	25475	25025	36	14.16	4.9	0.7	872.66	C	C2	GIA	9 years	MORTGAG	70000	Not Verifi	########	########	########	########
27	12796784	14818917	6000	3725	36	7.9	1.95	0.68	187.75	A	A4	Project Of	5 years	MORTGAG	87000	Source Ve	########	########	########	########
28	12656263	14658414	8000	3650	36	12.49	3.8	0.7	267.6	B	B4	Program I	4 years	RENT	25000	Not Verifi	########	########	########	########
29	12647441	14649591	9000	4150	36	11.99	3.45	0.7	298.89	B	B3	Delivery	8 years	RENT	45000	Not Verifi	########	########	########	########
30	12735676	14757880	15000	13250	60	13.65	4.45	0.44	346.31	C	C1	First Class	10+ years	MORTGAG	54000	Source Ve	########	########	########	########
31	12995686	15027891	15000	13575	60	15.61	6.05	0.45	361.67	C	C5	Informati	9 years	RENT	50000	Not Verifi	########	########	########	########
32	12645263	14647411	25000	22925	60	13.65	4.45	0.44	577.18	C	C1	Criminal I	10+ years	MORTGAG	55000	Not Verifi	########	########	########	########
33	12876735	14898866	5000	2575	36	10.99	2.9	0.7	163.67	B	B2		n/a	MORTGAG	25000	Verified	########	########	########	########
34	12876373	14898512	15850	14075	60	15.61	6.05	0.45	382.17	C	C5	CNC Mach	1 year	MORTGAG	44000	Not Verifi	########	########	########	########
35	12657491	14659648	11000	5600	36	11.99	3.45	0.7	365.31	B	B3	PHLEBOTC	7 years	RENT	39000	Not Verifi	########	########	########	########

Targeted Online Display Advertising



Who should we target for a product?

Does the ad have an effect?

What data should we pay for?

Where should we advertise and at what price?

Attribution?

Which request are fraud?

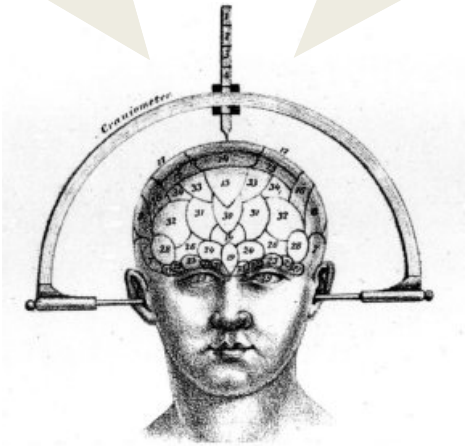
Ad Exchange

Agnostic Data

A consumer's online activity → **gets recorded like this:**

The Non-Branded Web

The Branded Web



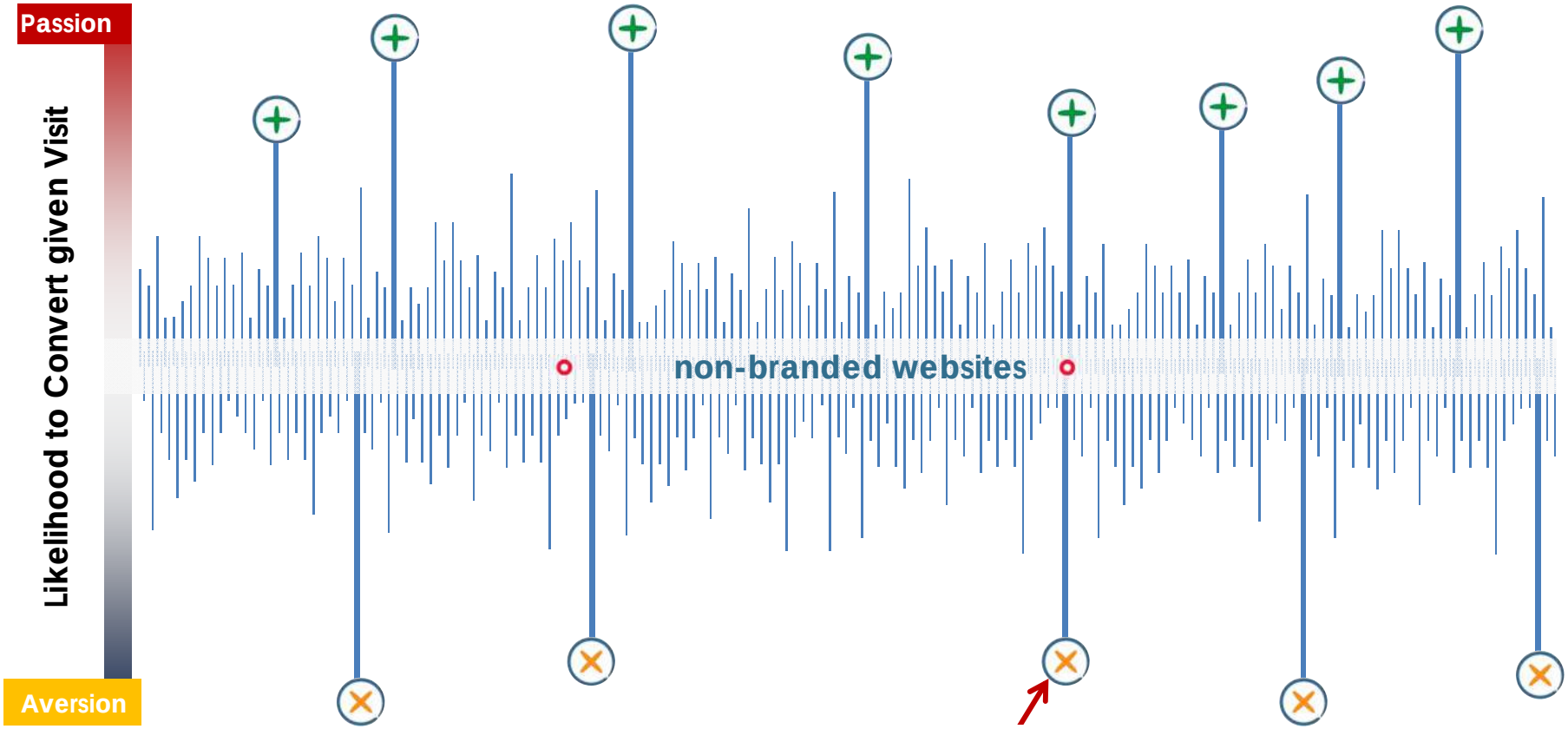
Purchases
Encoded
date1 3012L20
date 2 4199L30
…
date n 3075L50

Browsing History
Hashed URL's:
date1 abkcc
date2 kkllo
date3 88iok
date4 7uiol
…

I do not want/need to 'understand' who you are …

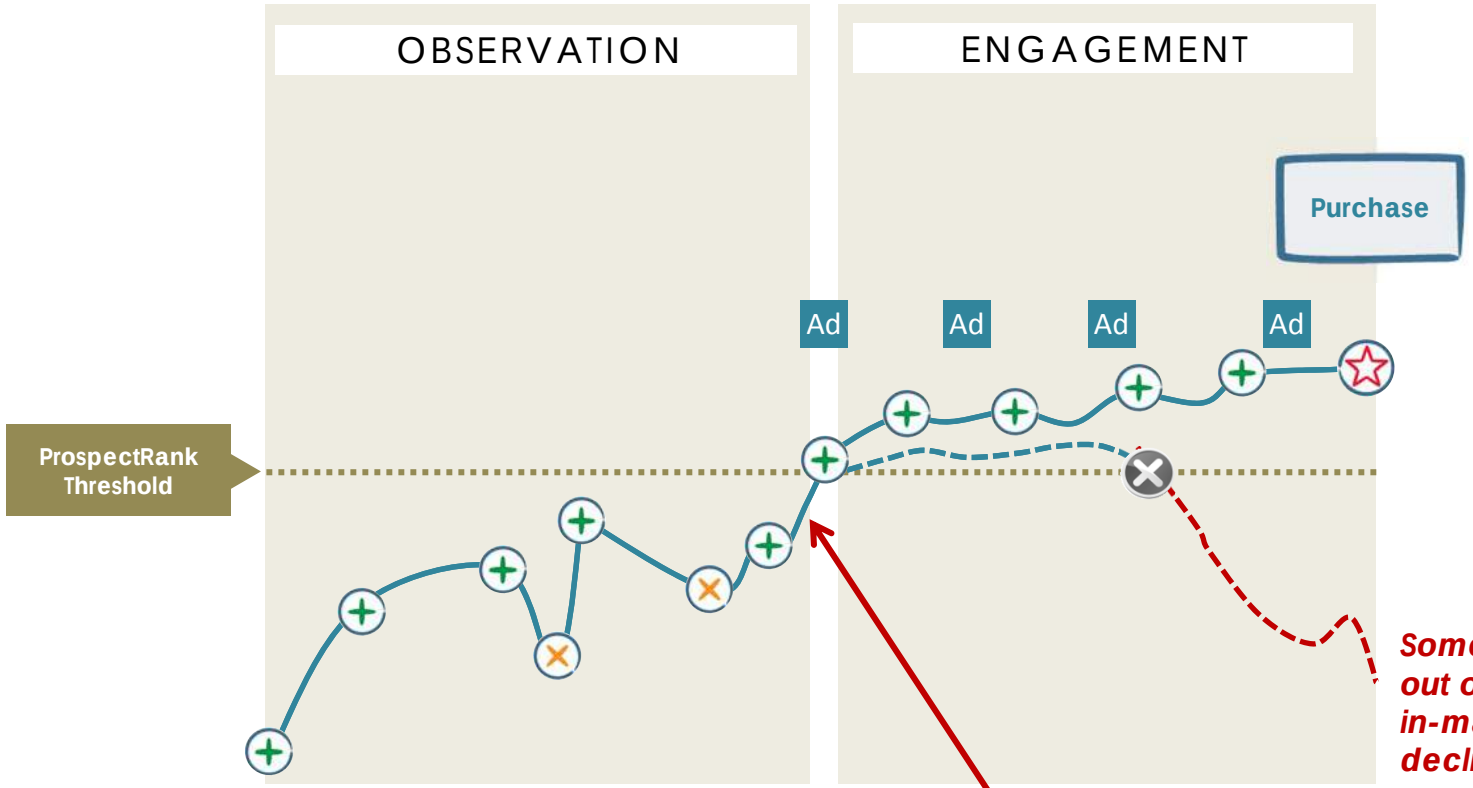
Model in 10 Million Dimensions

Using Naïve Bayes and Stochastic Gradient Decent Logistic Regression, we estimate statistical correlations between 10s of millions of web URLs and 1000s of branded actions.



$$p(\text{buy}|\text{urls}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Real-time Scoring of a Browser



site visit with positive correlation

site visit with negative correlation

$$p(\text{buy}|\text{urls}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Models in Our World

- Spam Detection
- Fraud/Fault Detection
- Financial Trading
- Medial Diagnosis/Quality control
- Sentiment Analysis
- Prioritization in General
- CRM
- Recommender systems
- Advertising/Targeting

Important Takeaways

- The algorithm is secondary
- The data is KEY
- Quality control is HARD
- Model is only as good as the modeler
- Very difficult to really understand the data

Panel Discussion

- **Pamela Dixon**, Founder, World Privacy Forum
- **Edmund Mierzwinski**, Consumer Program Director and Senior Fellow, U.S. Public Interest Research Group
- **Claudia Perlich**, Chief Scientist, Dstillery
- **Stuart Pratt**, President and CEO, Consumer Data Industry Association
- **Ashkan Soltani**, Independent Researcher and Consultant
- **Rachel Nyswander Thomas**, Executive Director of Data-Driven Marketing Institute, and Vice President of Government Affairs, Direct Marketing Association
- **Joseph Turow**, Professor, University of Pennsylvania

SPRING PRIVACY SERIES

Alternative Scoring Products

MARCH 19, 2014

Presentation

Ashkan Soltani

Independent Researcher and Consultant

SPRING PRIVACY SERIES

Alternative Scoring Products

MARCH 19, 2014

whoami

twitter: @ashk4n
ashkan.soltani@gmail.com
independent researcher & consultant

today: alternative scoring

- methodology
- findings
- data sources

THE WALL STREET JOURNAL. | WHAT THEY KNOW

WHAT THEY KNOW

Websites Vary Prices, Deals Based on Users' Information

Email Print Save 144 Comments

By JENNIFER VALENTINO-DEVRIES, JEREMY SINGER-VINE and ASHKAN SOLTANI

December 24, 2012

It was the same Swingline stapler, on the same Staples.com SPLS +0.25% website. But for Kim Wamble, the price was $15.79, while the price on Trude Frizzell's screen, just a few miles away, was $14.29.

A key difference: where Staples seemed to think they were located.

A Wall Street Journal investigation found that the Staples Inc. website displays different prices to people after estimating their locations. More than that, Staples appeared to consider the person's distance from a rival brick-and-mortar store, either OfficeMax Inc. OMX +0.73% or Office Depot Inc. ODP +0.18% If rival stores were within 20 miles or so, Staples.com usually showed a discounted price.

What Price Would You See?

"How can they get away with that?" said Ms. Frizzell, who works in Bergheim, Texas.

In what appears to be an unintended side effect of Staples' pricing methods—likely a function of retail competition with its rivals—the Journal's testing also showed that areas that tended to see the discounted prices had a higher average income than areas that tended to see higher prices.

methodology

```
71    # Permutation logic
72  ▾ for ua in user_agents
73  ▾   for proxy in proxies
74  ▾     for profile in (components[:PROFILES_DIR].entries - [ ".", ".." ])
75          permutation_name = [ profile, proxy[:name], ua[:name], ].join(".")
76          profile_path = File.join(components[:PROFILES_DIR].path, profile)
77          zipfile_path = File.join(ZIP_DIR, permutation_name + ".zip")
78
```

user-agent

older findings: orbitz

TECHNOLOGY

On Orbitz, Mac Users Steered to Pricier Hotels

Email Print Save 258 Comments A A

By DANA MATTIOLI

Updated Aug. 23, 2012 6:07 p.m. ET

Orbitz has found that Apple users spend as much as 30% more a night on hotels, so the online travel site is starting to show them different, and sometimes costlier, options than Windows visitors see. Dana Mattioli has details on The News Hub. Photo: Bloomberg.

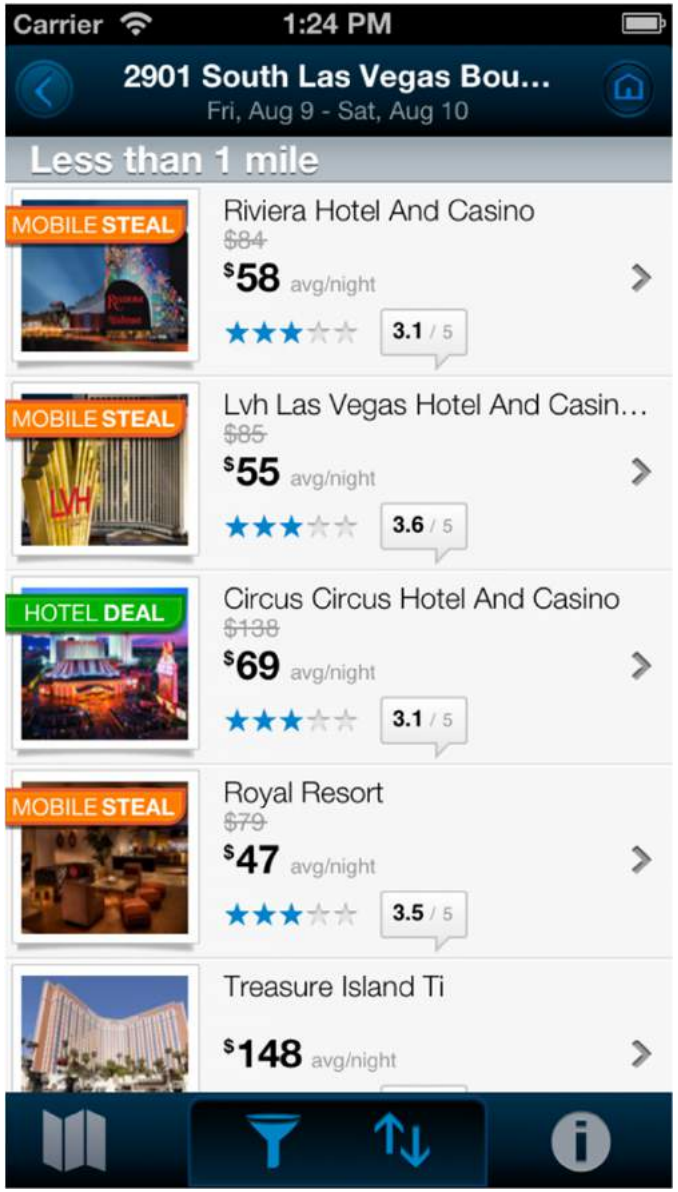
Orbitz Worldwide Inc. OWW -1.66% has found that people who use Apple Inc. AAPL +0.39% 's Mac computers spend as much as 30% more a night on hotels, so the online travel agency is starting to show them different, and sometimes costlier, travel options than Windows visitors see.

The Orbitz effort, which is in its early stages, demonstrates how tracking people's online activities can use even seemingly innocuous information—in this case, the fact that customers are visiting Orbitz.com from a Mac—to start predicting their tastes and spending habits.

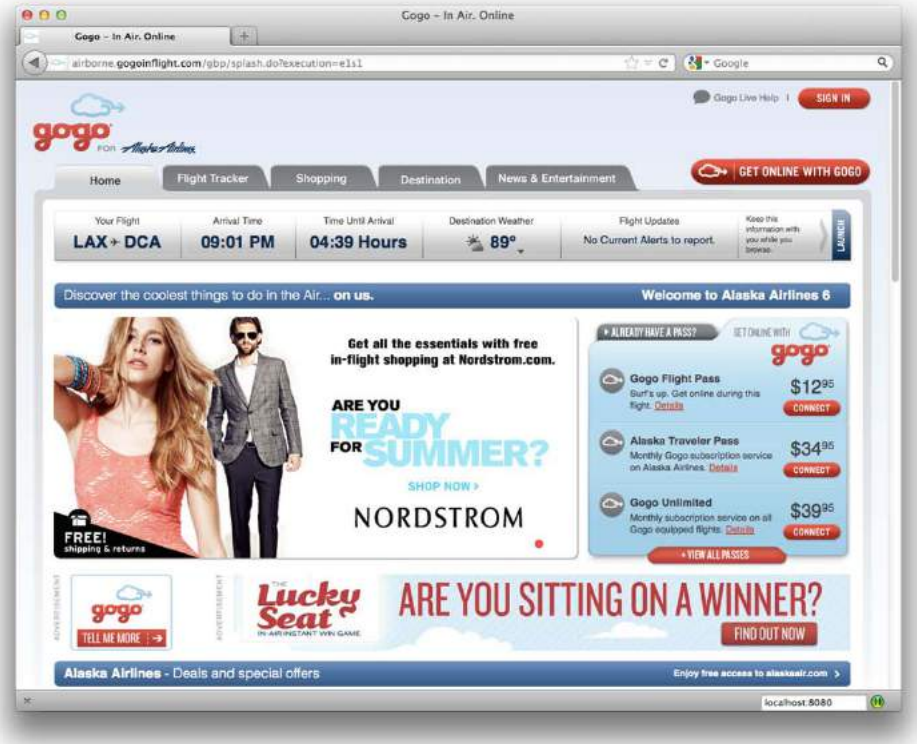
findings: orbitz

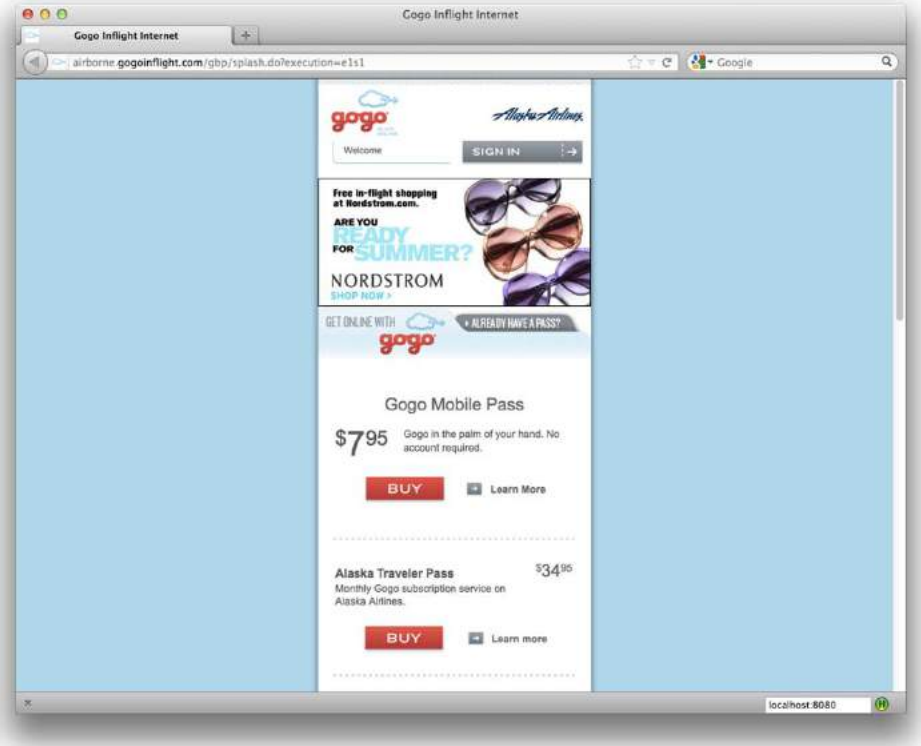
Some sites, for example, gave discounts based on whether or not a person was using a mobile device. A person searching for hotels from the Web browser of an iPhone or Android phone on travel sites Orbitz and CheapTickets would see discounts of as much as **50% off the list price**, Orbitz said.

Both sites are run by Orbitz Worldwide Inc., which in fact markets the differences as "mobile steals." Orbitz says the deals are also available on the iPad if a person installs the Orbitz app.



findings: gogo inflight





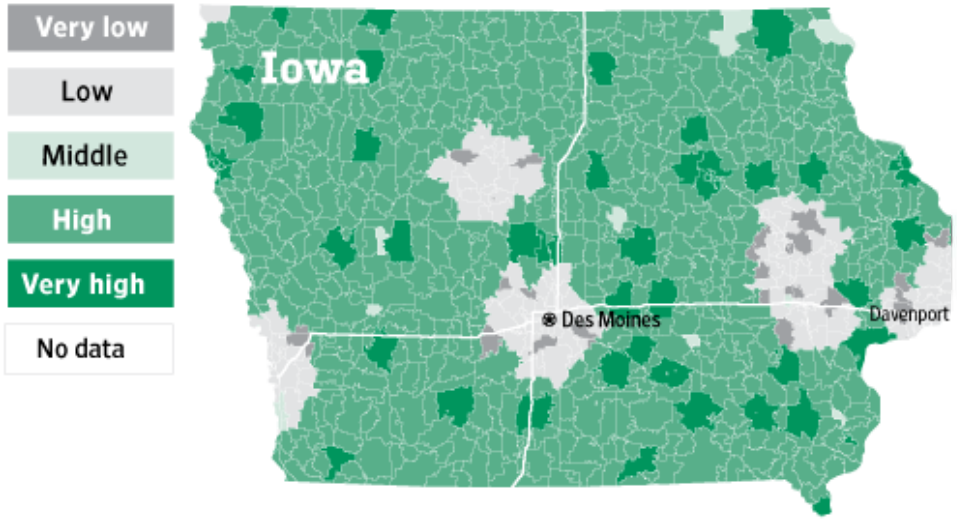
User-Agent: Desktop
$12.95

User-Agent: iPhone
$7.95

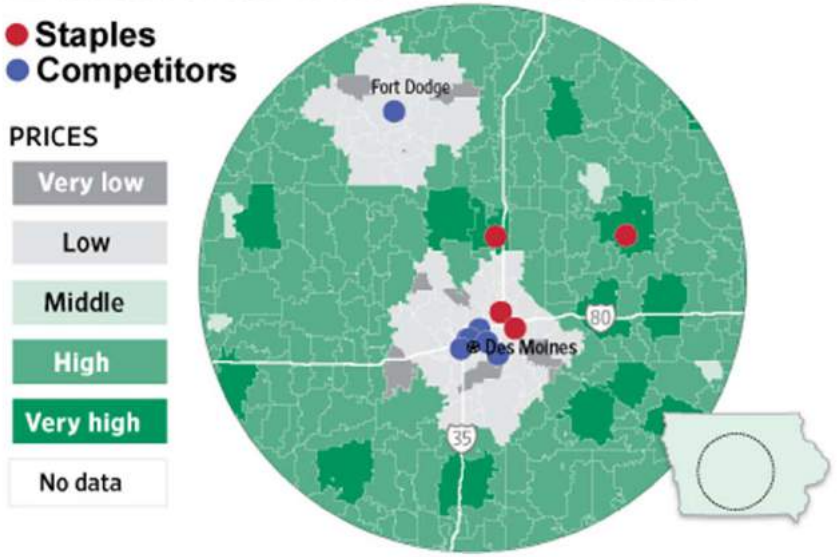
location

findings: staples

Likelihood of receiving higher prices, by ZIP code



Locations of stores relative to price zones



What price are you likely to get?

The Journal tested the price of a basic Swingline stapler 20 times for every U.S. ZIP Code. Sometimes, visitors would receive a price of $14.29, while other times they saw a discounted price of $15.79. Some ZIP codes saw the high price much more often than others did. **Enter your ZIP Code below** to see how it fared.

20009 Search

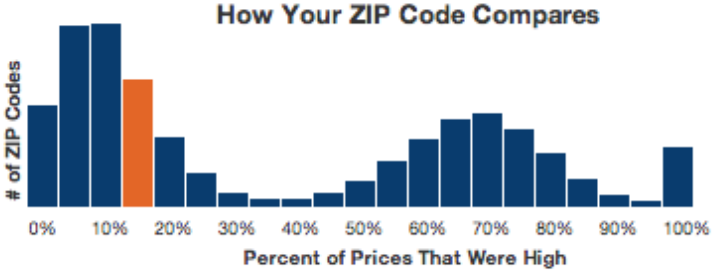
Caveat: Staples might change its pricing strategy at any time, and also might not guess your ZIP Code precisely.

ZIP Code 20009 received the

high price 15% of the time,

out of 20 tests.

How Your ZIP Code Compares



This chart shows the distribution of prices across the U.S. The taller the bar, the more ZIP Codes saw the high price the number of times listed.

findings: staples

What are the prices on other products?

In addition to testing the Swingline stapler across all ZIP Codes, the Journal also tested more than 1,000 randomly selected products from 10 ZIP Codes.

Roughly one-third of the products showed different prices. Among those, the high and low prices differed by about 8%, on average. Some examples:

	Product	Discount Price*	High Price	Dollar Difference	Percent Difference
	Swingline Stapler	$14.29	$15.79	$1.50	10.5%
	BIC Rollerball Pens, 12-Pack	$25.99	$28.49	$2.50	9.6%
	Staples-Brand Mailing Tubes, 15-Pack	$99.99	$109.99	$10.00	10.0%
	Intertape Masking Tape, 24-Pack	$289.99	$319.99	$30.00	10.3%
	SnapSafe-Brand Safe	$1,099.99	$1,199.99	$100.00	9.1%

* Here, the main low price for a product. In a small fraction of cases, roughly 1% overall for the stapler, Staples displayed an extra-low price, $13.57. Percentage differences are expressed as the dollar difference divided by the main low price.

findings: more geography

Home Depot's website offered price variations that appeared to be based on the nearest brick-and-mortar store as well. A 250-foot spool of electrical wiring fell into six pricing groups, including $70.80 in Ashtabula, Ohio; $72.45 in Erie, Pa.; $75.98 in Olean, N.Y and $77.87 in Monticello, N.Y.

Location also seemed to be important for some international companies. The Journal saw Rosetta Stone, which sells software for learning languages, offering discounts of as much **as 20% for people who bought multiple levels of its German lessons from certain locations in the U.S. or Canada**, but not others from the U.K. or Argentina.

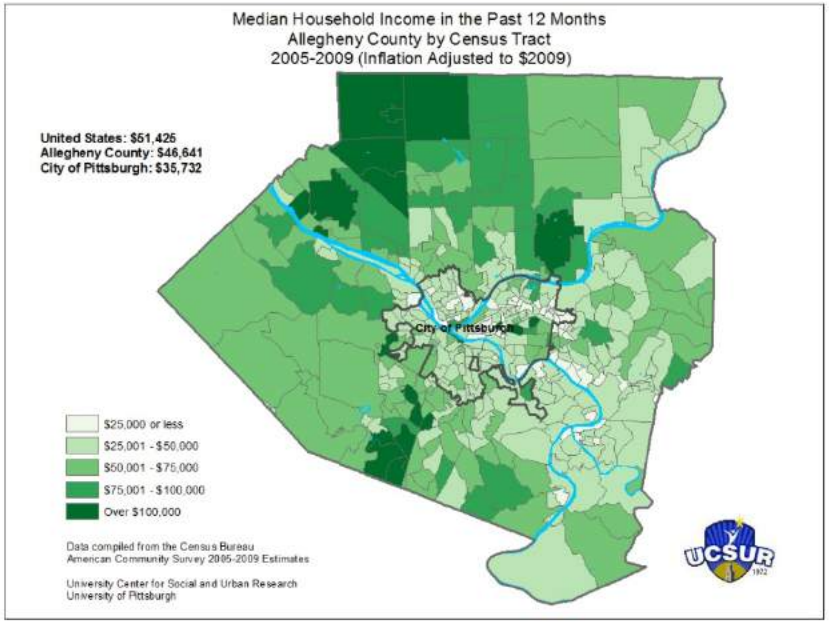
findings: discover

In the tests, Discover, for instance, showed a prominent offer for the company's new "it" card to computers connecting from cities including Denver, Kansas City, Mo., and Dallas, Texas. Computers connecting from Scranton, Penn., Kingsport, Tenn., and Los Angeles didn't see the same offer.

A Discover spokeswoman said that the company was testing the card, but that for competitive reasons, it wouldn't comment further on its "acquisition strategy" for new customers.

findings: staples

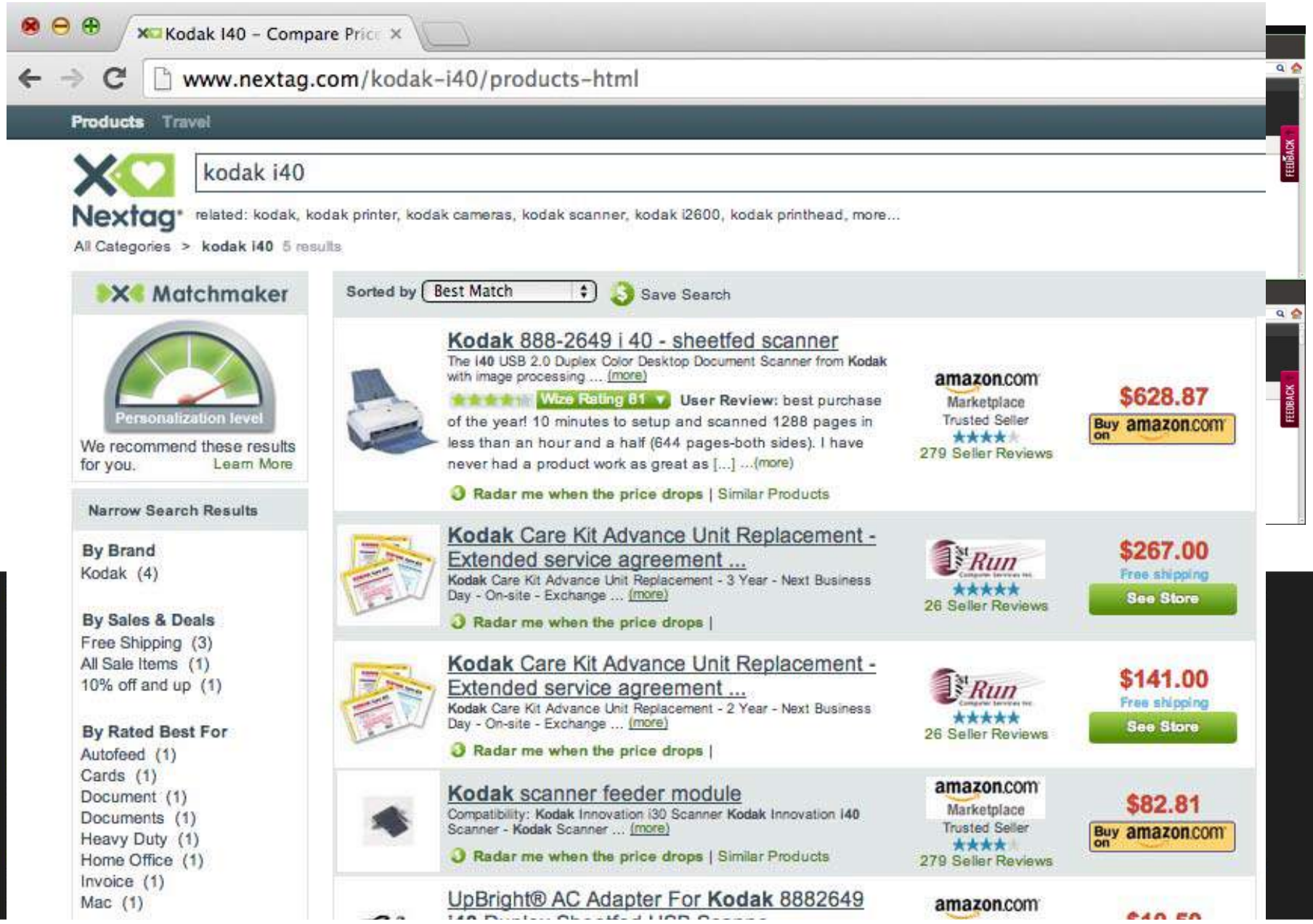
higher income = lower price



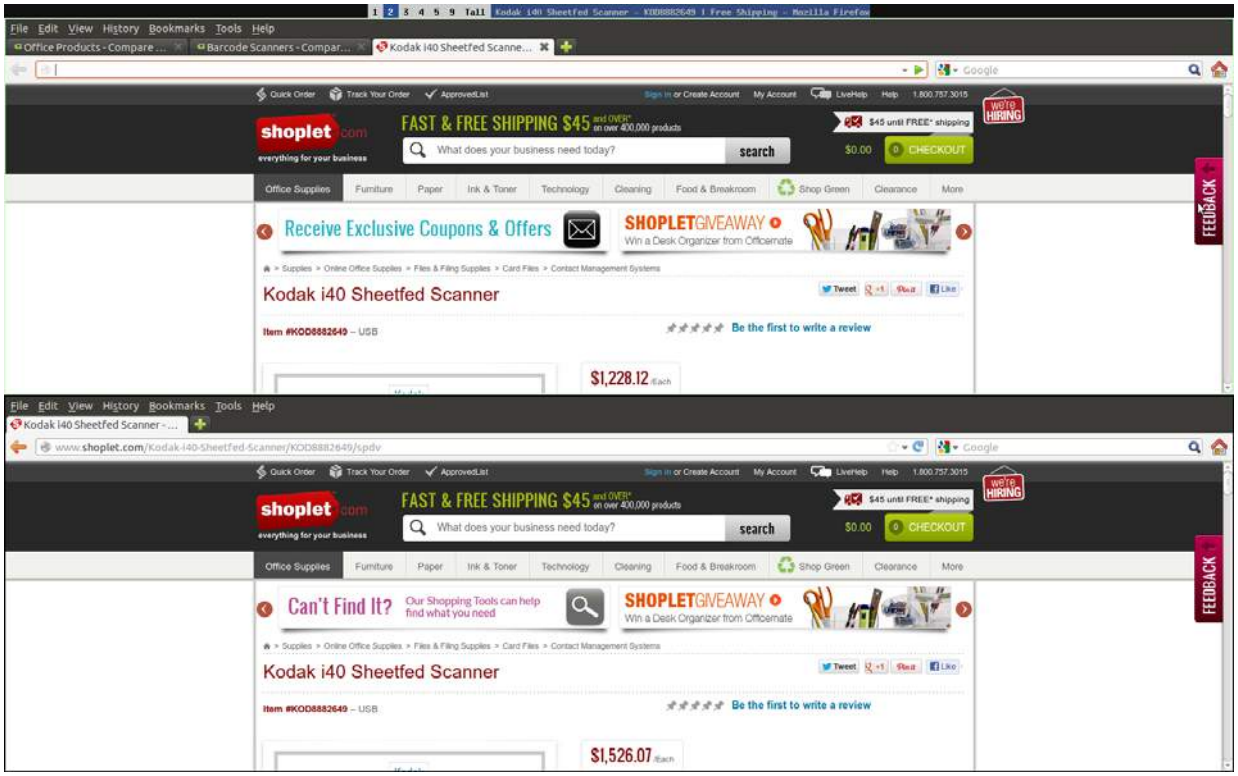
In the Journal's examination of Staples' online pricing***, the weighted average income*** among ZIP Codes that mostly received discount prices was roughly $59,900, based on Internal Revenue Service data. ZIP Codes that saw generally high prices had a lower weighted average income, $48,700.

profiles*

findings: nextag / shoplet



findings: nextag / shoplet




```
27  http://www.shoplet.com/Franklin-Merriam-Webster-s-Collegiate-Electronic-Speaking-Dictionary-amp-Thesaurus/FRKSCD2100/spdv
28  104,73
29  131,28
30
31  http://www.shoplet.com/Royal-ATB-3-AUDIO-BIBLE/ROY39130T/spdv
32  44,19
33  53,46
34
35  http://www.shoplet.com/Franklin-Explorer-14-Lang-Speaking-Global-Translator-w-Dictionary/FRKEST5014/spdv
36  148,06
37  184,00
38
39  http://www.shoplet.com/Franklin-Merriam-Webster-s-Collegiate-Electronic-Speaking-Dictionary-amp-Thesaurus/FRKSCD2100/spdv
40  104,73
41  131,28
```

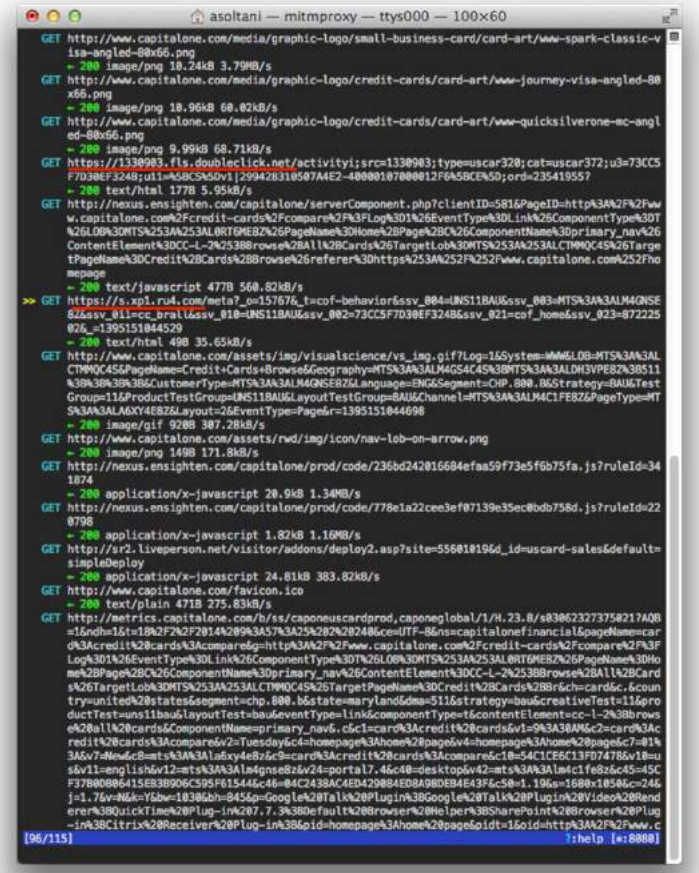
findings: capital one

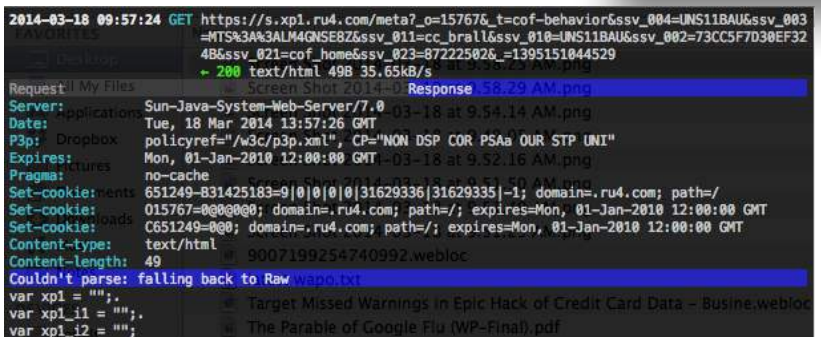
Capital One was showing different users different cards first—either those for "excellent credit" or "average credit."

findings: capital one

Showing 12 Cards for All Credit Levels

Primary Benefit	Purchase APR	Transfer Info	Fee	Credit Level
Venture® Rewards Unlimited 2X miles on every purchase. Plus, earn 20,000 bonus miles. ☐ Compare	13.9%-20.9% variable APR	13.9%-20.9% variable APR; No Transfer Fee	$0 intro annual fee for the first year $59 after that	Excellent Credit Apply Now Card Details ★★★★★ Read all 3,959 reviews
VentureOne® Rewards Unlimited 1.25 miles on every purchase. Plus, earn 20,000 bonus miles. ☐ Compare	0% intro APR until March 2015 11.9%-19.9% variable APR after that	11.9%-19.9% variable APR; No Transfer Fee	$0 Annual fee	Excellent Credit Apply Now Card Details ★★★★½ Read all 2,117 reviews
Platinum Prestige Save money with a 0% intro APR on purchases and transfers ☐ Compare	0% intro APR until June 2015 10.9%-18.9% variable APR after that	0% intro APR until June 2015 10.9%-18.9% variable APR after that; 3% fee on the amount transferred until June 2015	$0 Annual fee	Excellent Credit Apply Now Card Details ★★★★☆ Read all 1,066 reviews
Quicksilver® Rewards Earn unlimited 1.5% cash back on every purchase, plus a one-time $100 bonus	0% intro APR until December 2014 12.9%-22.9%	0% intro APR until December 2014 12.9%-22.9%	$0 Annual fee	Excellent Credit Apply Now Card Details ★★★★★



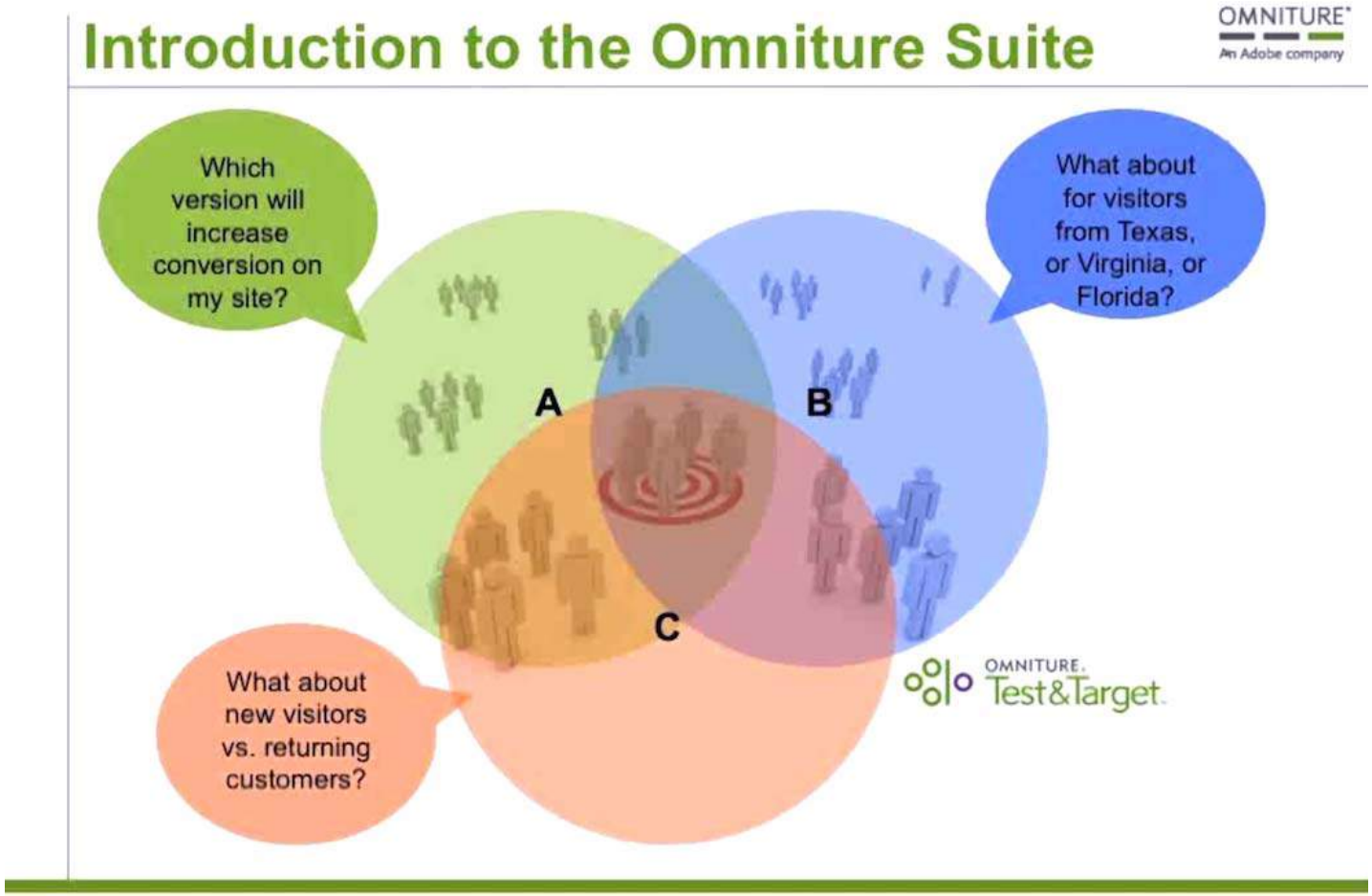


[x+1]

MAKE EVERY INTERACTION COUNT

data sources

data sources



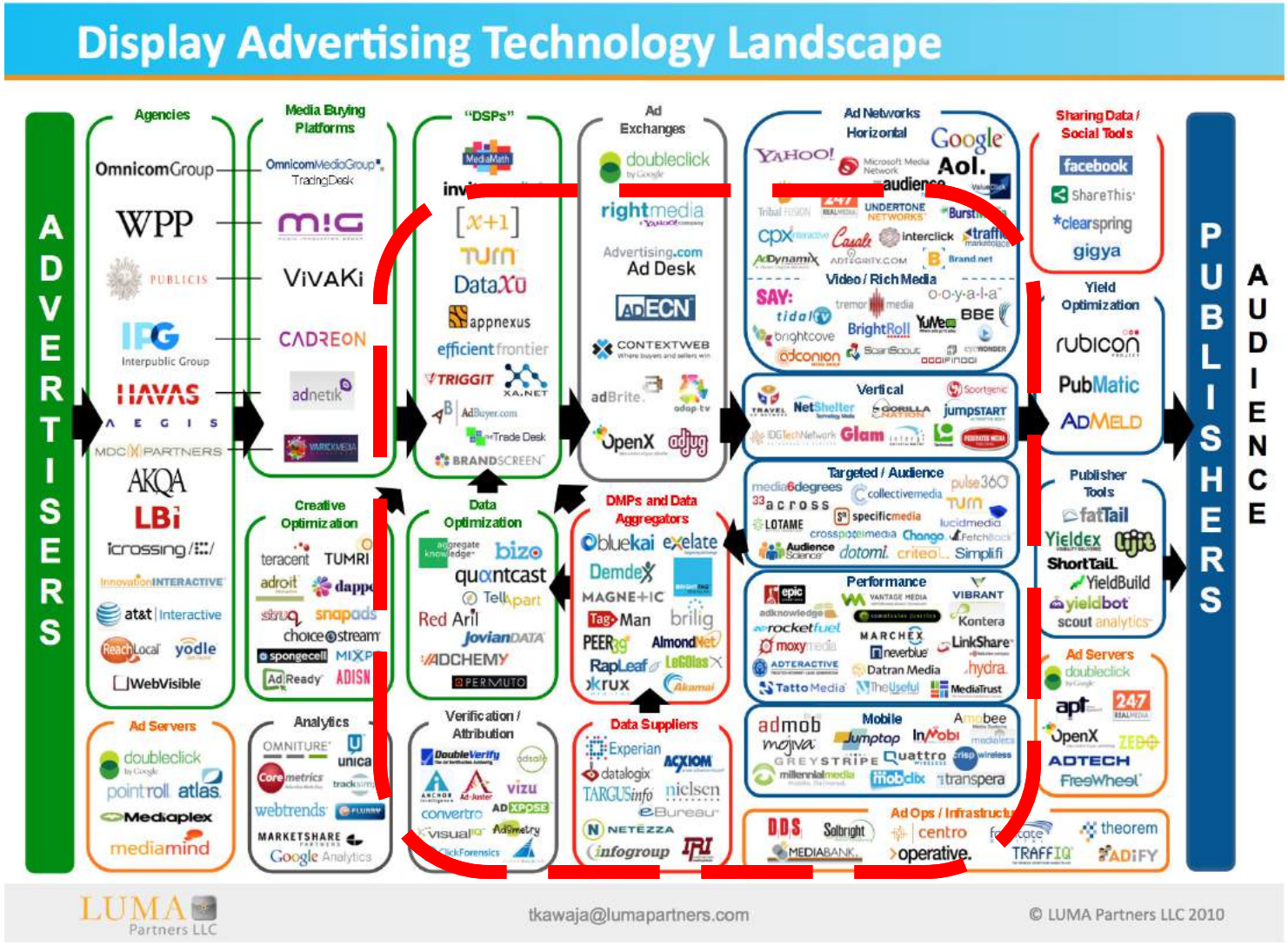
data sources

data sources

© 2010 Omniture, Inc., Confidential & Proprietary

OMNITURE UNIVERSITY

data sources



conclusion

conclusion: staples

As a final test, the Journal ordered two separate Swingline staplers from Staples.com, from two nearby ZIP Codes—one costing $14.29 and the other one $15.79. The staplers arrived the same day. They appear to be indistinguishable from one another and do an equally thorough job of stapling.

Panel Discussion

- **Pamela Dixon**, Founder, World Privacy Forum
- **Edmund Mierzwinski**, Consumer Program Director and Senior Fellow, U.S. Public Interest Research Group
- **Claudia Perlich**, Chief Scientist, Dstillery
- **Stuart Pratt**, President and CEO, Consumer Data Industry Association
- **Ashkan Soltani**, Independent Researcher and Consultant
- **Rachel Nyswander Thomas**, Executive Director of Data-Driven Marketing Institute, and Vice President of Government Affairs, Direct Marketing Association
- **Joseph Turow**, Professor, University of Pennsylvania

SPRING PRIVACY SERIES

Alternative Scoring Products

MARCH 19, 2014