

Predictability and other Predicament

Claudia Perlich

Chief Scientist

@claudia_perlich

dstillery

“Models tend to go where the signal is”



Certainty Vortex

How ‘predictable’ is a predictive modeling task? (given the data)

Pizza for
Dinner?



Random



Sexual
Orientation



Deterministic

For now let's try to predict who is male ...



Predicting Probability (Male) in Facebook

00000d41ed774823fca142945ec915c0, 1, , , , , , , en_GB,,
00000dee02d70cf8c0d79f96b6d1c59d, 0, , , , , , , en_US,,
00000f232abfe25a80156fe069395460, 0, 1992, 20, 2, 2, , , , , 19,-5
00000f4ba0cff946b1c0e3b051287ede, 0, 1993, 19, 2, , , , , en_US, 310, 8
0000130571654e3afaa62f4e9d2e4f63, 0, , 2, 2, , , , , en_US, 193, 7
00001544469ae9b408869a463a1dd77a, 1, 1984, 28, 2, , , , , en_US,, -4

40% Men

Data:

Facebook public dataset with 200K anonymized users, their demographics and their likes

Methodology:

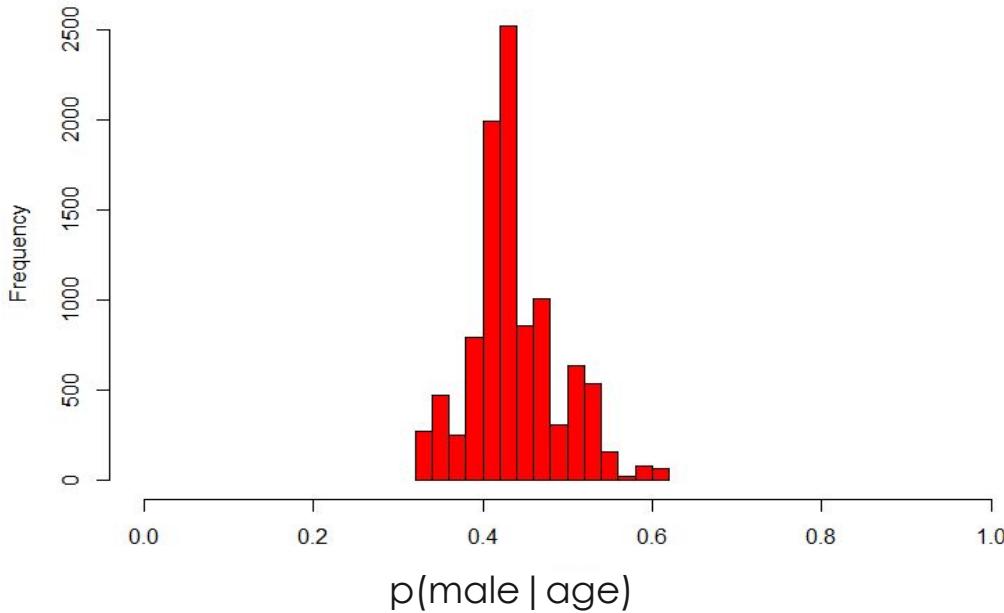
Logistic regression on sparse representation

00001544469ae9b408869a463a1dd77a	100198443380917
00001544469ae9b408869a463a1dd77a	100248613695
00001544469ae9b408869a463a1dd77a	10050726267
00001544469ae9b408869a463a1dd77a	101021248409
00001544469ae9b408869a463a1dd77a	101054236602446
00001544469ae9b408869a463a1dd77a	101425333232551
00001544469ae9b408869a463a1dd77a	10148199466
00001544469ae9b408869a463a1dd77a	10150154095435553
00001544469ae9b408869a463a1dd77a	10161539667
00001544469ae9b408869a463a1dd77a	101844593108
00001544469ae9b408869a463a1dd77a	101936079845392
00001544469ae9b408869a463a1dd77a	101987301816
00001544469ae9b408869a463a1dd77a	102038567018
00001544469ae9b408869a463a1dd77a	102040023230884
00001544469ae9b408869a463a1dd77a	10212595263
00001544469ae9b408869a463a1dd77a	102168219824412
00001544469ae9b408869a463a1dd77a	1022548733

Take 1: Predict Gender Based on age ...



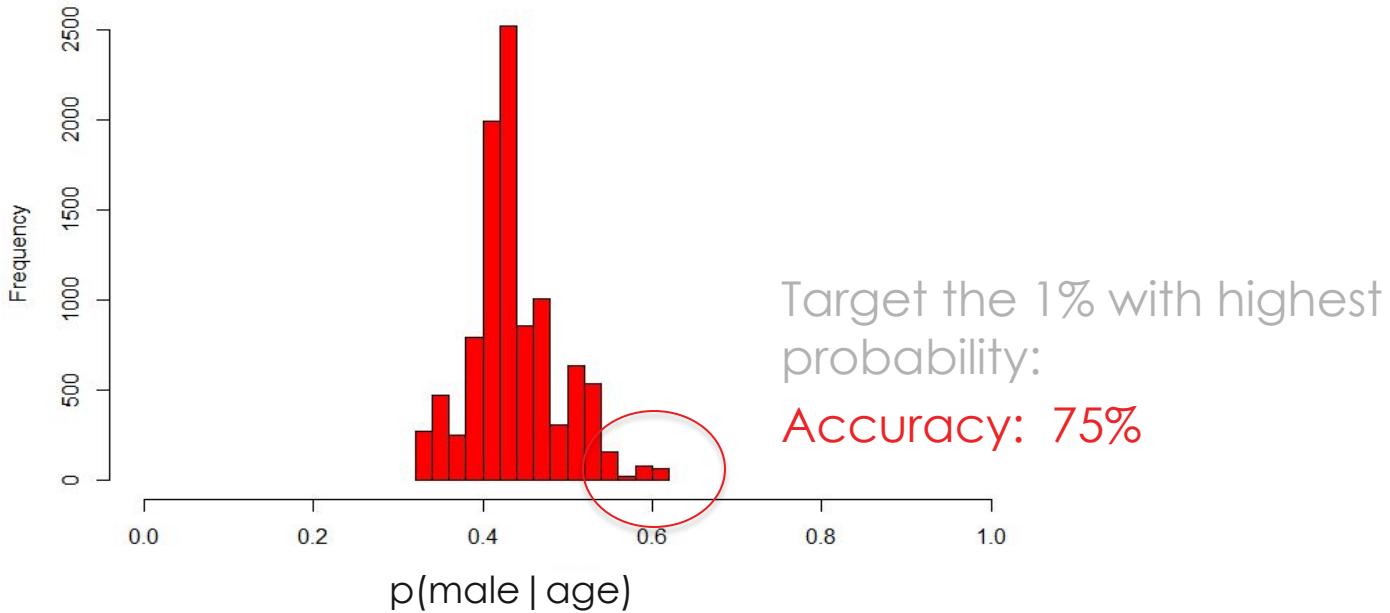
Gender based on age: very little signal ...



Overall Accuracy: 60%

AUC: 58%

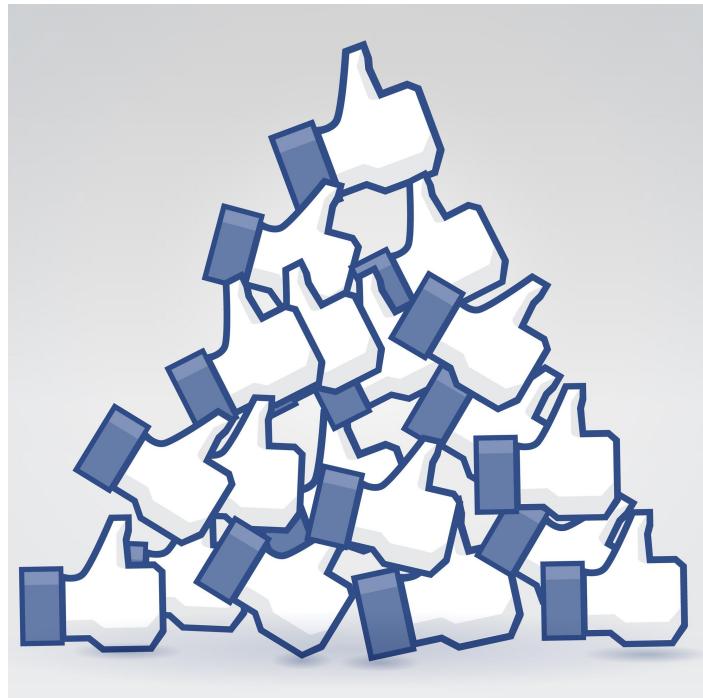
Gender based on age: very little signal ...



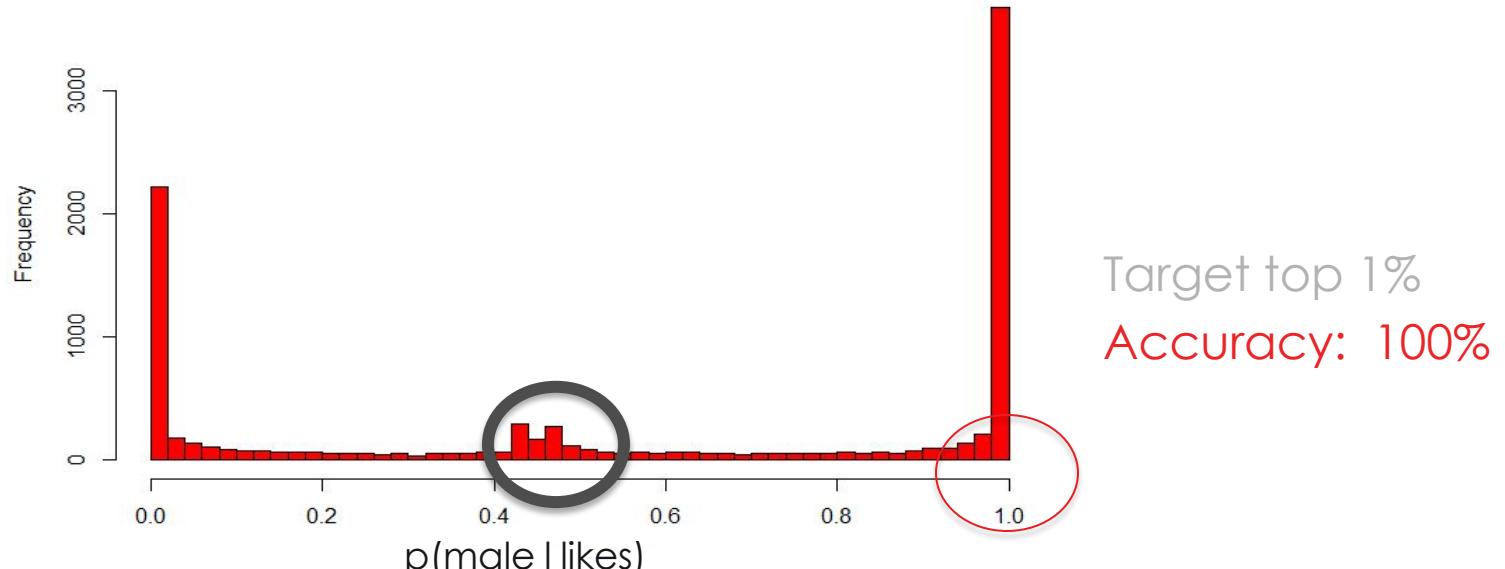
From small to bigger data ...



Take 2: Gender based all your likes

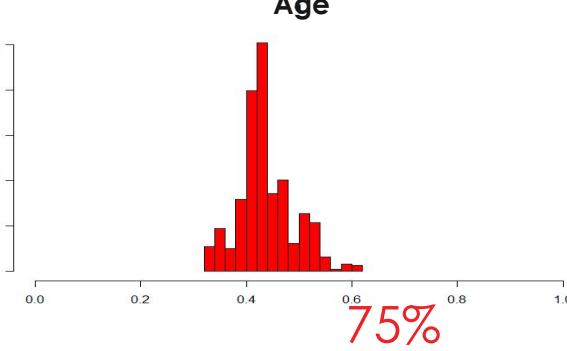


Predict gender base on all likes: a lot of signal ...

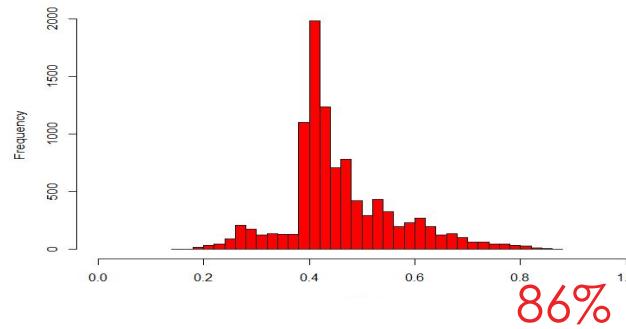


Progression: from age to all 'likes'

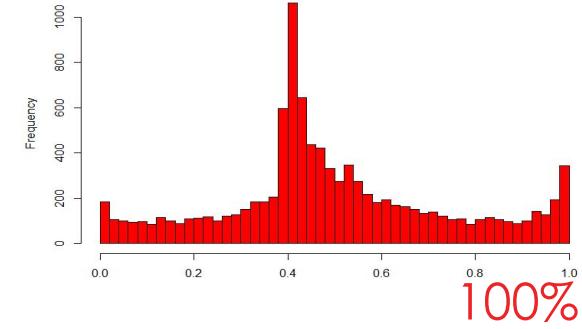
Age



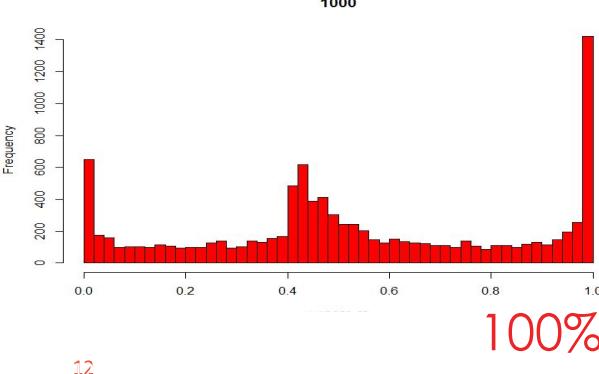
10



100

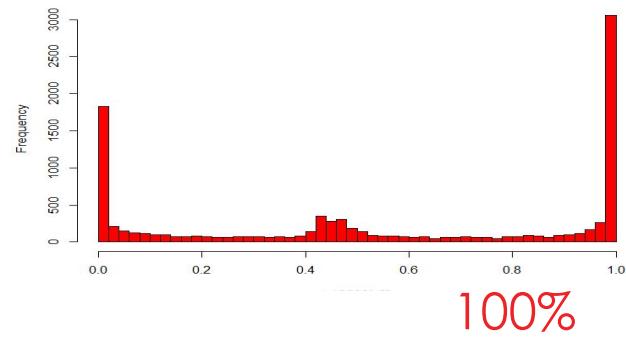


1000

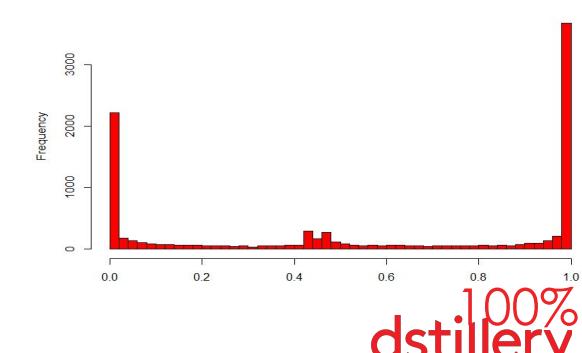


12

10000



100000



100%
dstillery

But what happens if your problem is a mixture of both?



Random



Deterministic

Witness a spike in human predictability ..



Death of
free will?

Predicting digital events based on browsing histories





SGD Model (Model ID: 1728493)

Marketer	Microsoft (Empowering Media SF) [6000]
Offer	Microsoft Azure Evergreen [7599]
Offer Level	40: Free Trial [19945]
Model Created At	2016-03-06 12:36:39
Model Event End Time	2016-03-06 17:40:00
# Positive	217
# Negatives	2,170
# Features	362,059
Parent Model	1725531
HDFS Path	hdfs://hdfs2/user/sgd/models/19945/OL19945_FM2_BS20160306123636.model

Top Features with Highest Scores

Show 10 entries

Search:

Feature	Coef
[46303095]	3.117
blogs.msdn.com [33892607]	3.010
www.why-recycle.com www.technet.com [46550201]	2.932
www.lz95.org [35757707]	2.779
www.dorisandjillycook.com www.dotnetfunda.com www.edenprairiecenter.com [35375642]	2.508
www.onebox.com www.avanquest.com [35325508]	2.467
www.office.com [46380052]	2.410
[46305049]	2.297
www.verbasoftware.com [46446087]	2.294
www.rosebyrne.org www.socialapocalypse.com www.milfmovies.biz www.milwaukee-wisconsin-wi.com www.msexchange.org [35355270]	2.288

Logistic Regression
Stochastic gradient descent
Hashing
Streaming
L1 & L2 Penalties



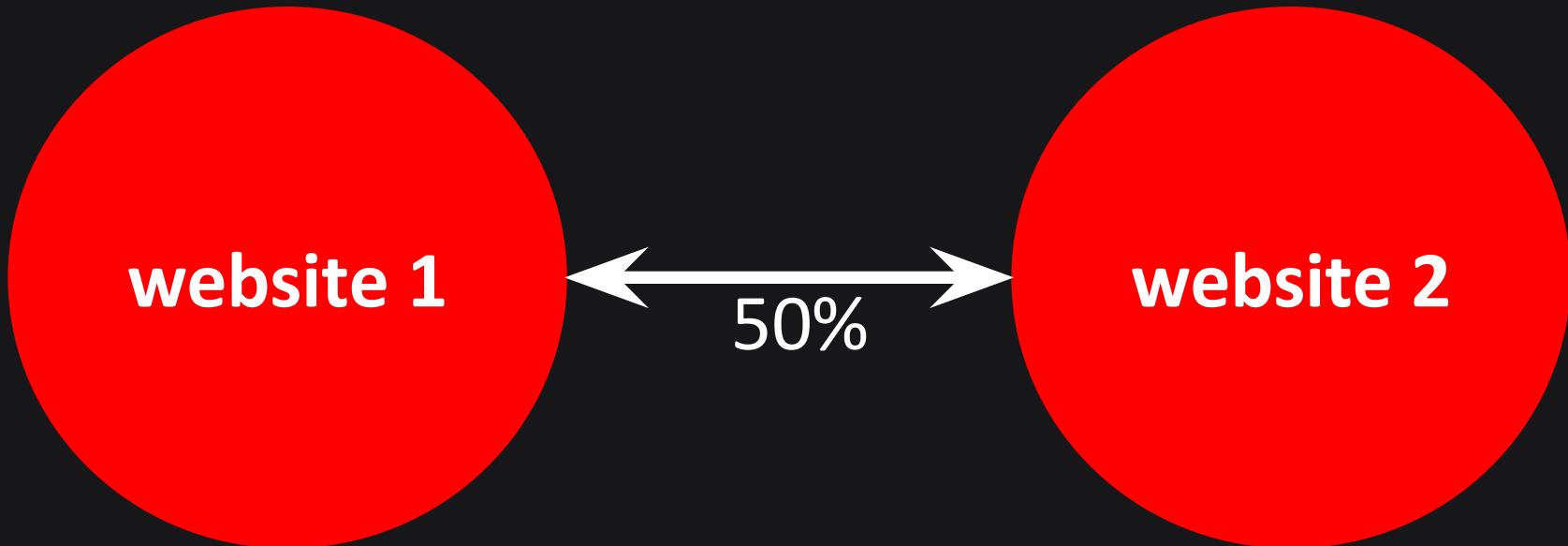
Sheraton®
HOTELS & RESORTS

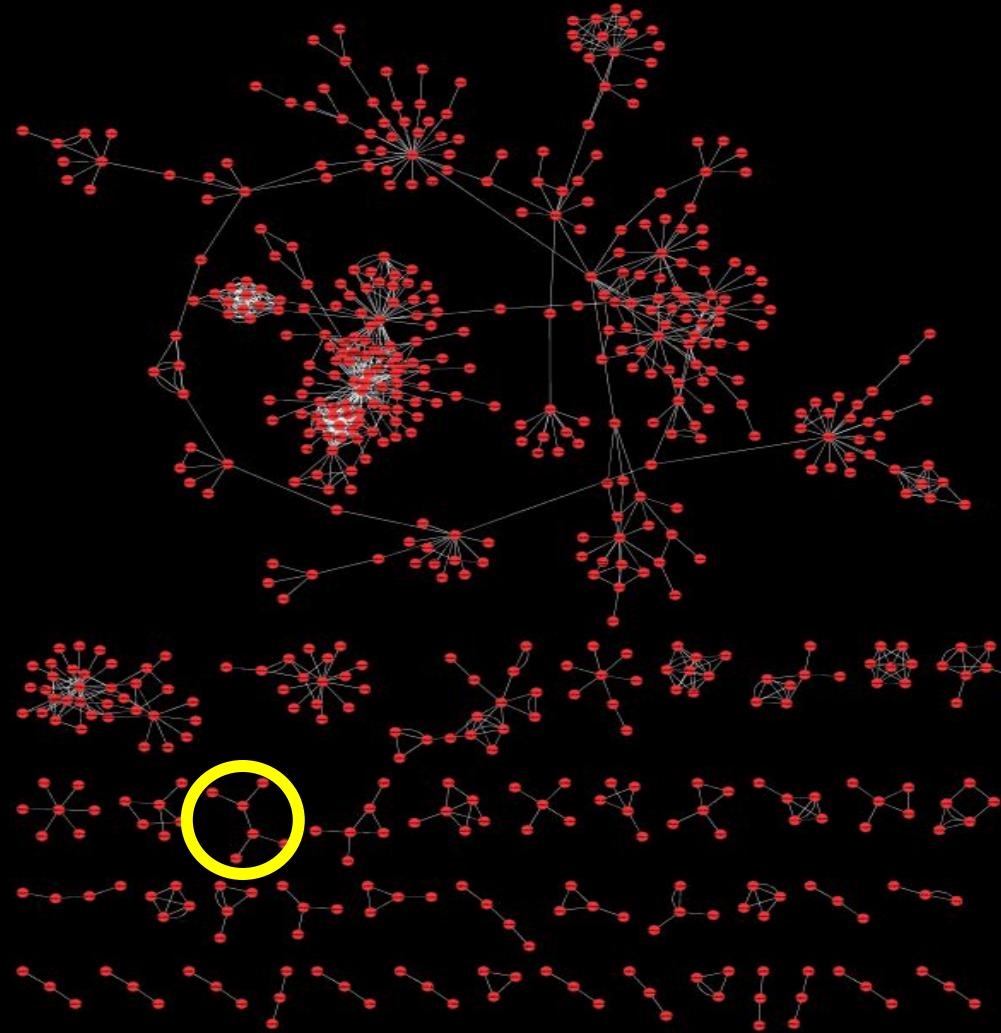


URL's that are very predictive for more than 10 brands

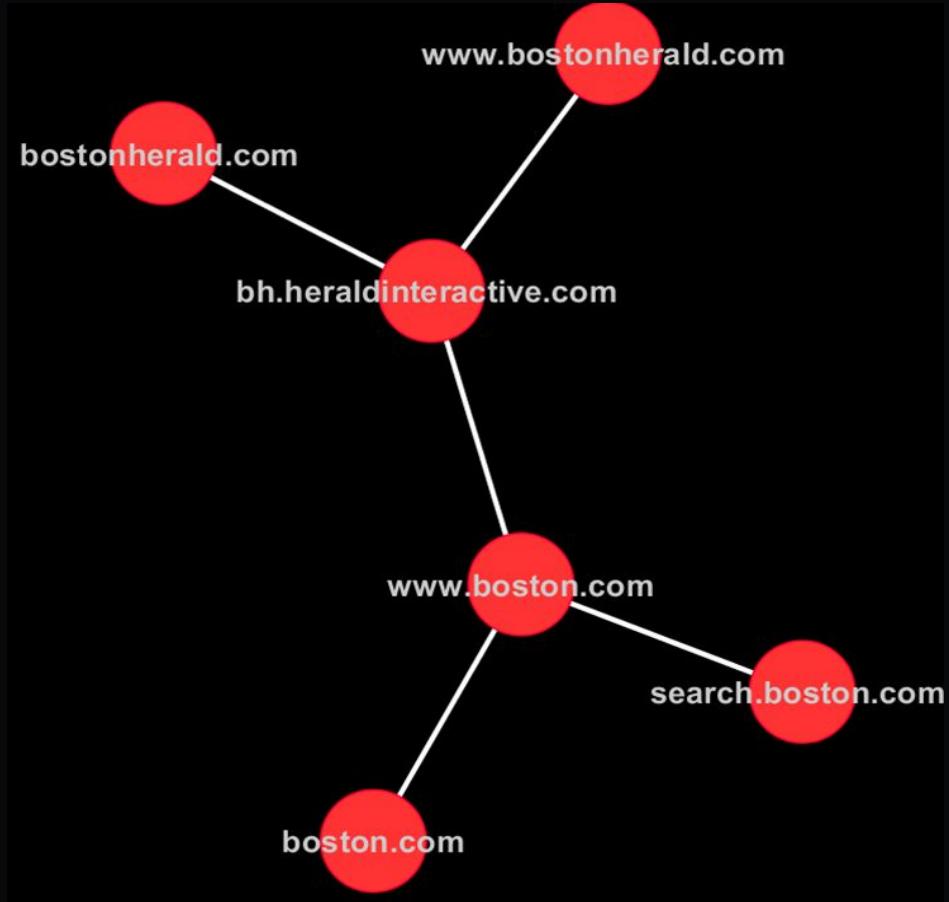
www.womenshealthbase.com
www.filmannex.com
www.ffog.net
www.drugsnews.org
www.menshealthbase.com
www.dailyfreshies.com
www.hark.com
www.gossipcenter.com
www.articletrunk.com
www.411answers.com
www.dailyrx.com
www.all-allergies.com
www.knowvehicles.com
www.chinaflix.com
www.parentingnewsstories.com
www.wrestlingnewz.com
www.gourmandia.com

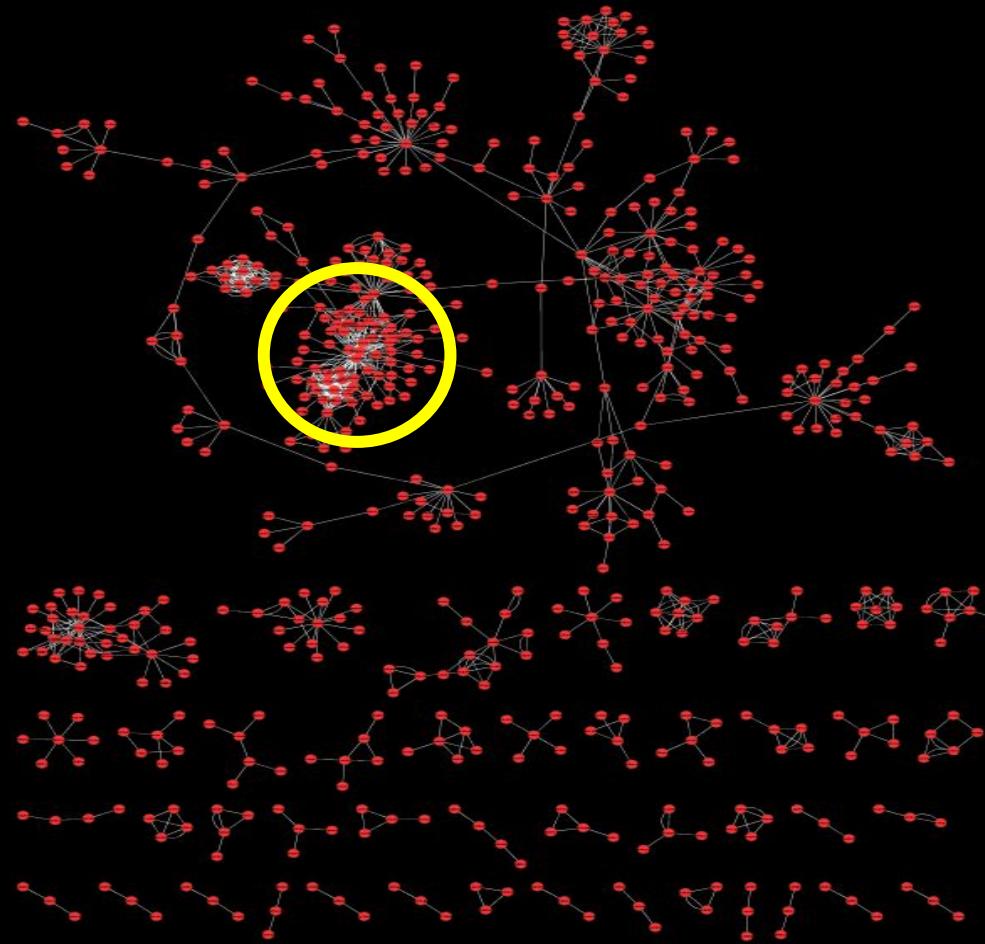
Traffic overlap of cookies from Bid Request

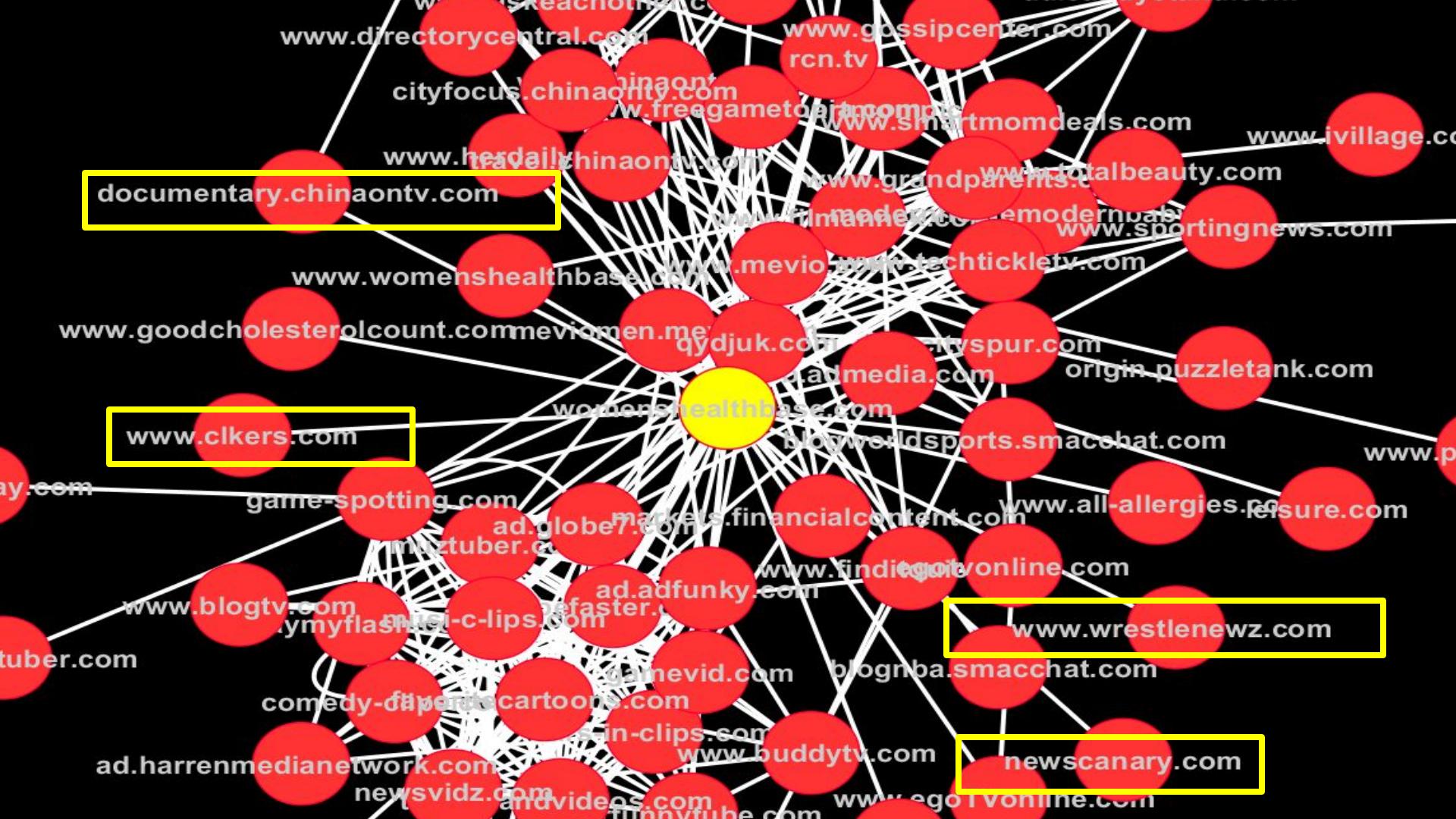


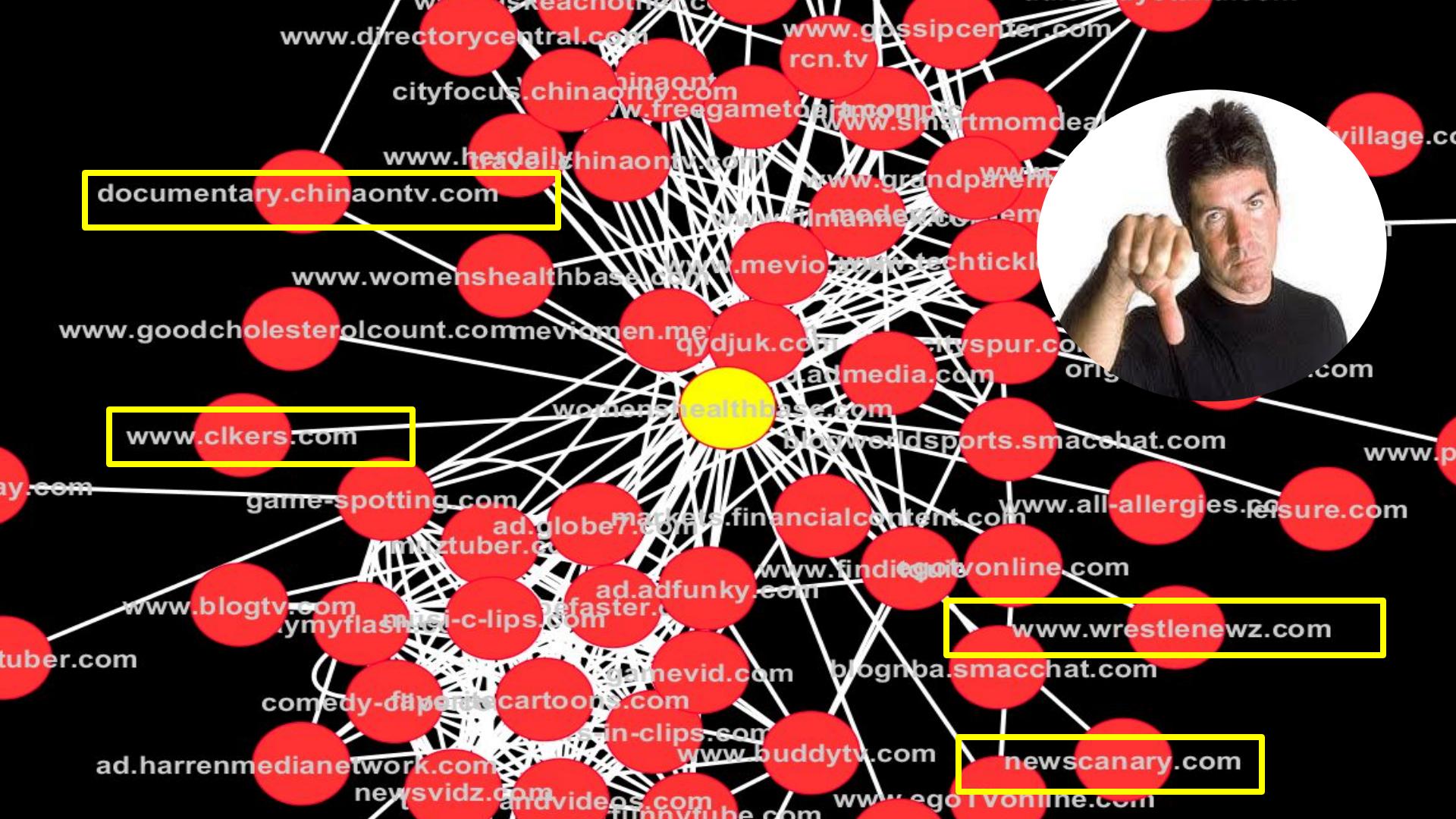


Boston Herald

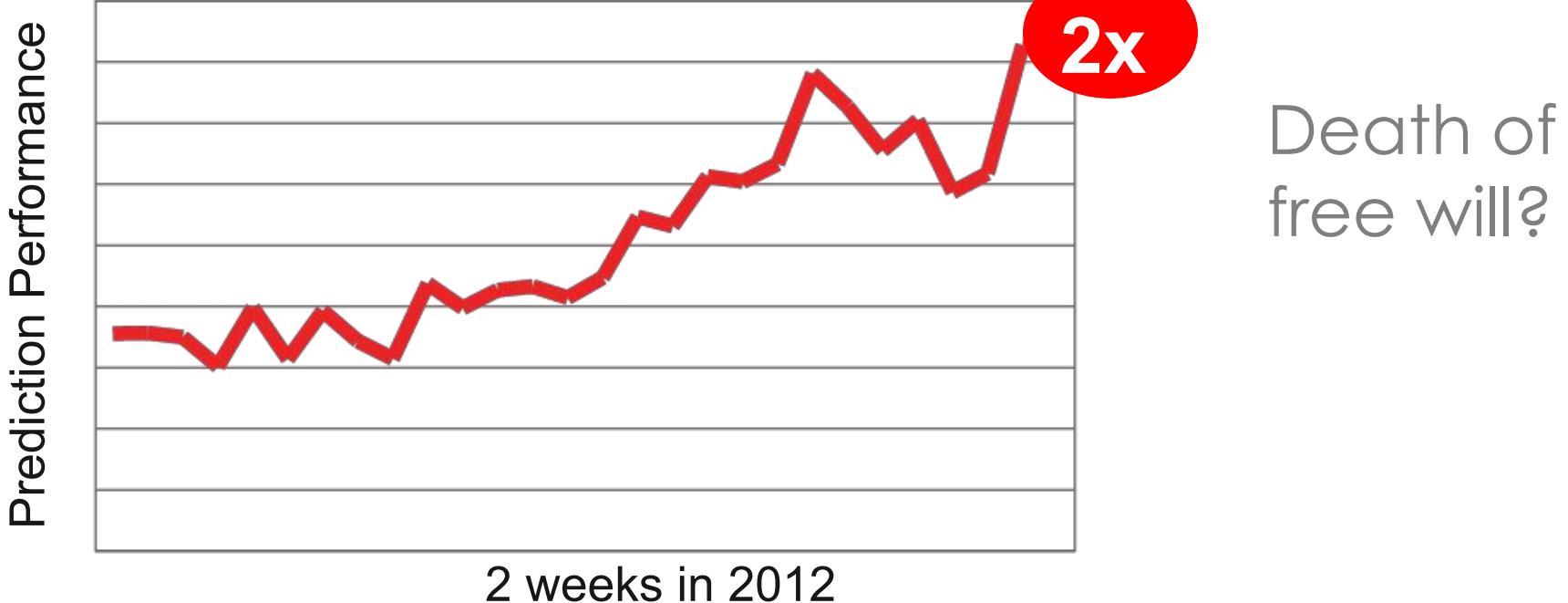




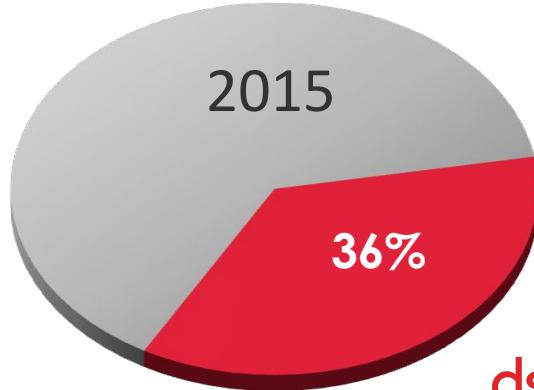
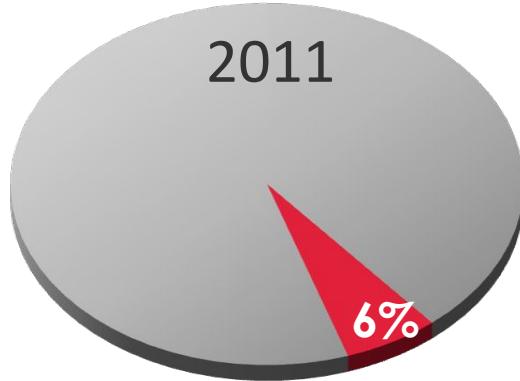




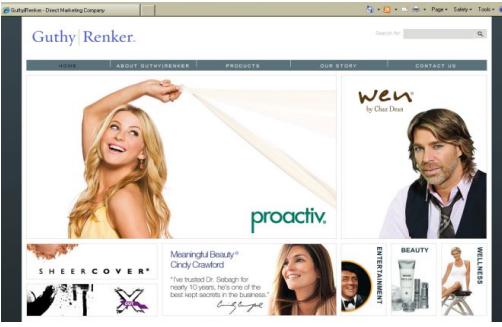
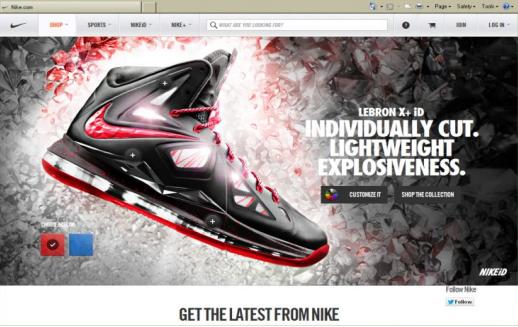
Witness a spike in human predictability ..



Meet 'Non-Human traffic'

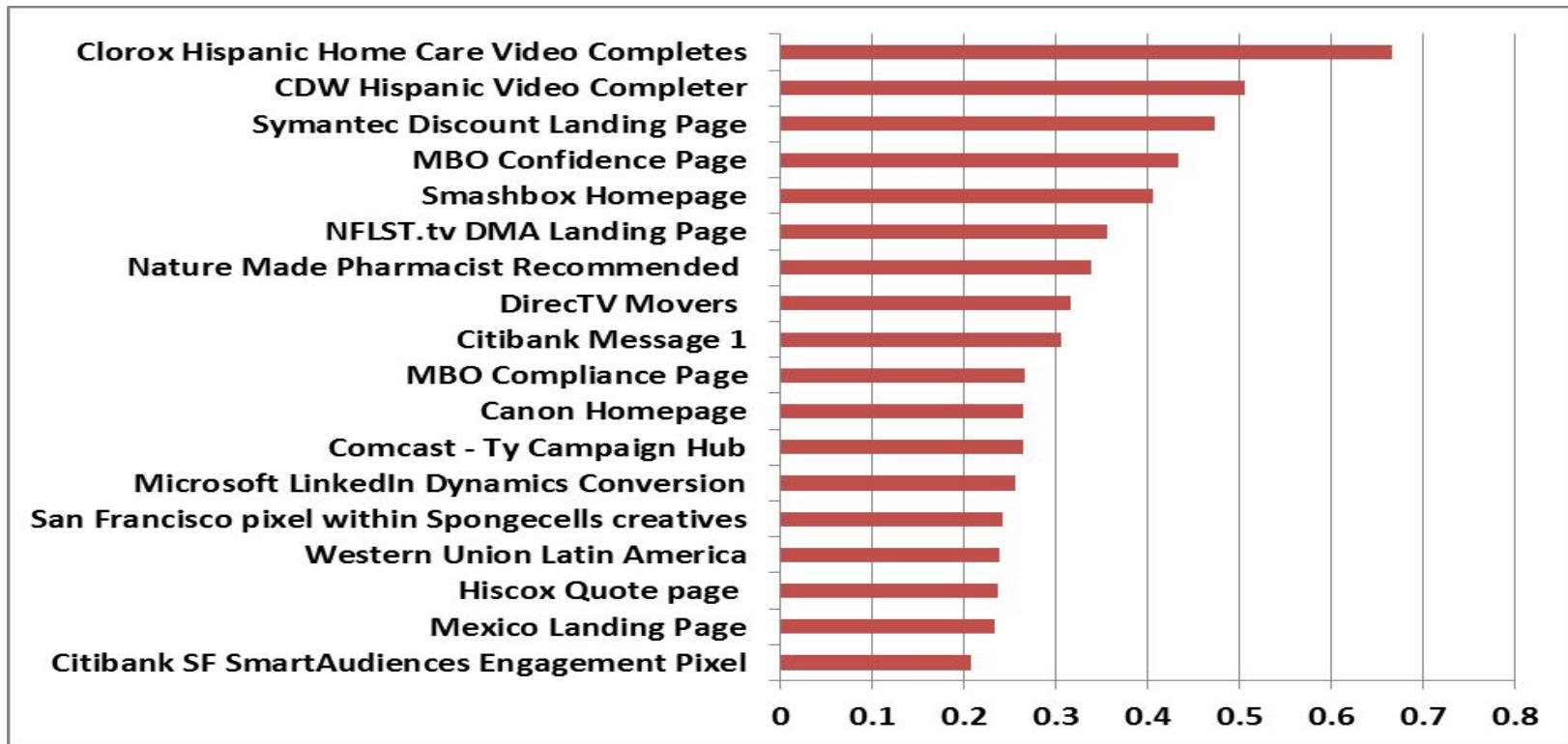


Bots are executing conversion events

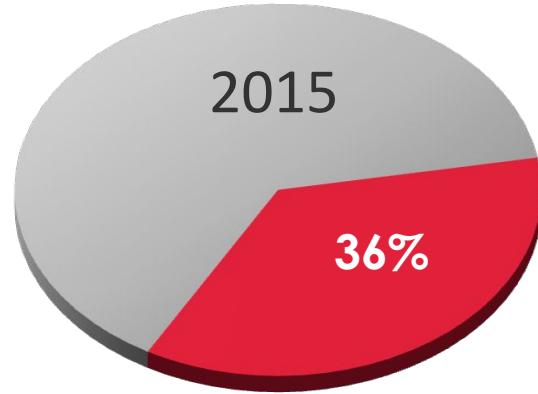


- 'Cookie Stuffing' increases the value of the ad for retargeting
- Messing up Web analytics ...
- Messes up my models because a bot is easier to predict than a human

Percent bot traffic on conversion metrics



Two populations surfing the we: Bot vs. Human



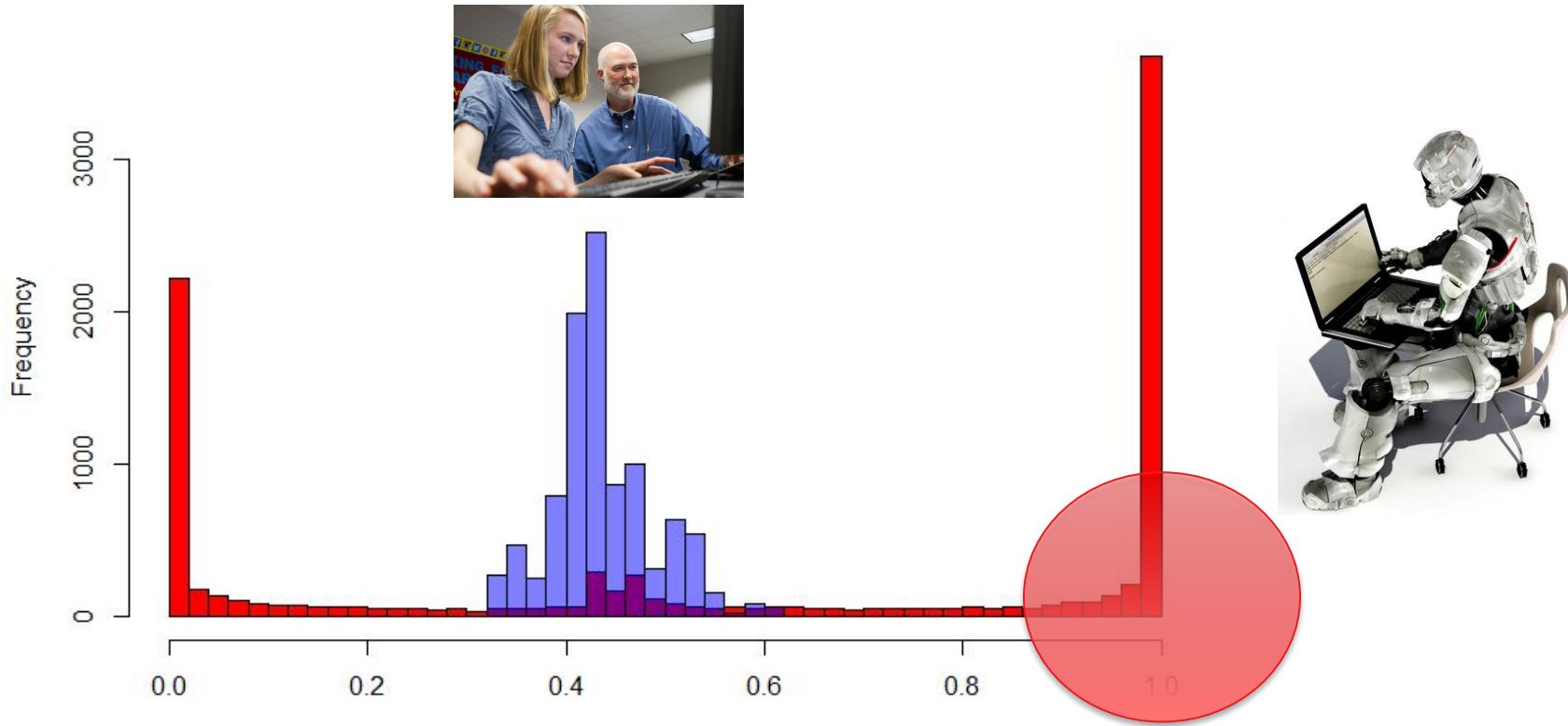


Random



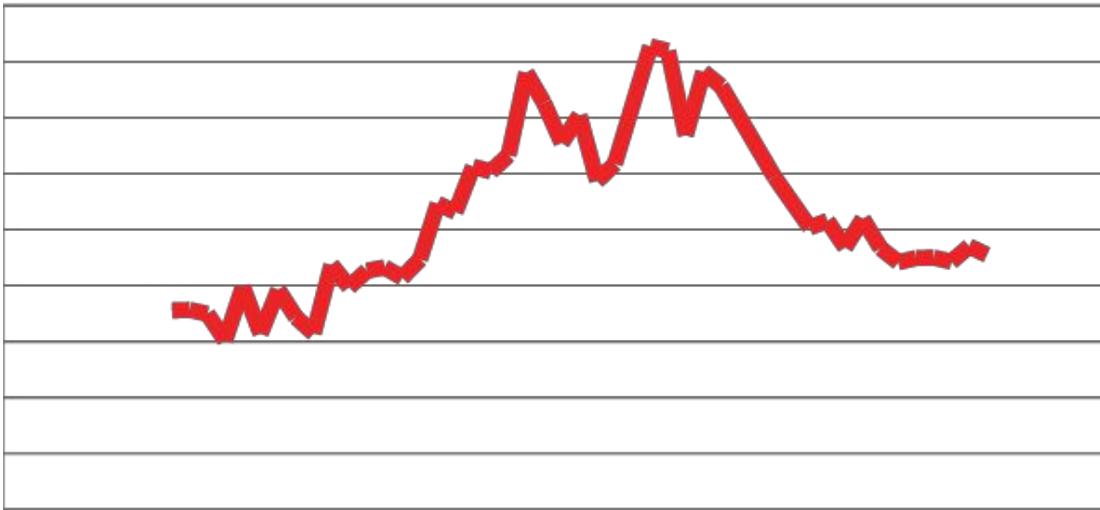
Deterministic

Bot activity has more signal – higher predictions



Eliminate labels generated by bots

Performance Index



3 more weeks in spring 2012

What about Clicks?

“Measure of consumers interest in the product”

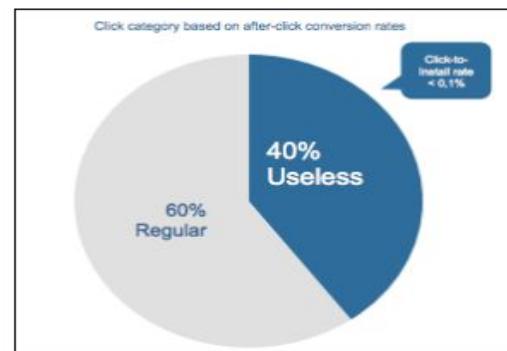


Report: 40% Of Clicks On Mobile Ads Are Fraudulent Or Accidental

Pamela Parker on September 5, 2012 at 5:42 pm

Mobile advertising may still be in its infancy, but if [a new study](#) by Trademob is correct, click fraud on mobile ads is already quite well developed. The mobile app marketing company conducted research that showed that 40% of mobile ad clicks are either accidental or fraudulent.

◀
F
n
IRE



Old days of the Click Metric ...

Strategy 1:
target women



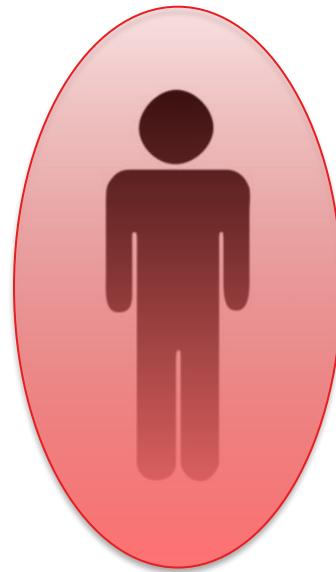
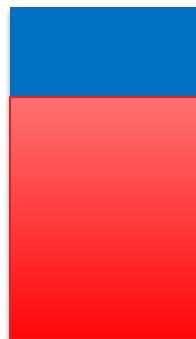
intentional



accidental

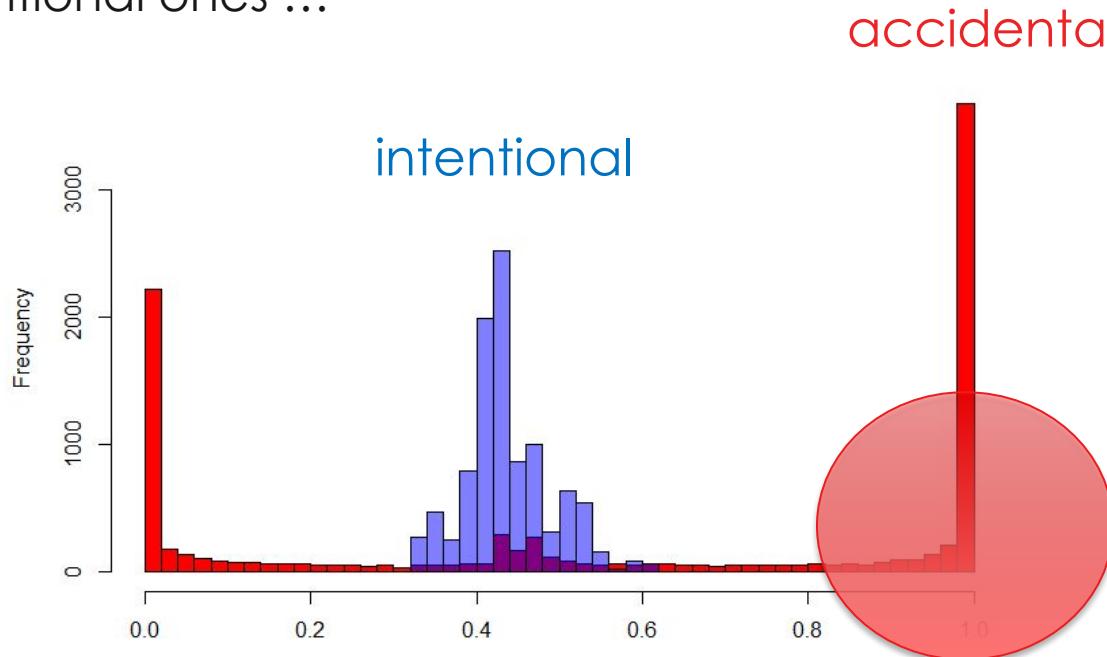


Strategy 2:
target men



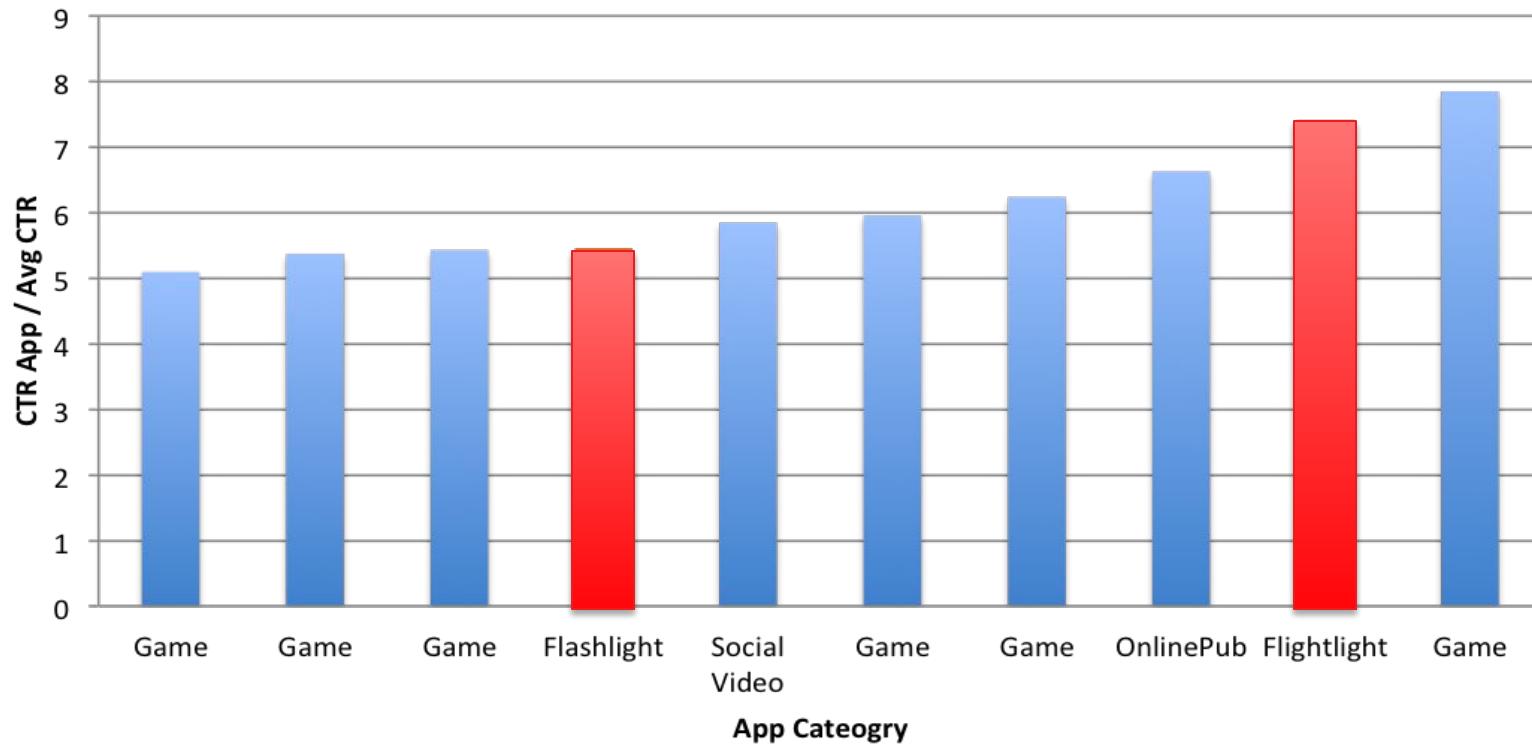
Accidental clicks are more predictable than intentional

Model learns to predict the accidental much more easily than the intentional ones ...



Accidents: People fumbling in the dark ...

Top 10 Apps by CTR



'Creative Optimization' Last night



38



Mahjong Village

4.5★

30 thousand



1 million

Play free mahjong solitaire game! Hundreds of unique levels!

Google Play

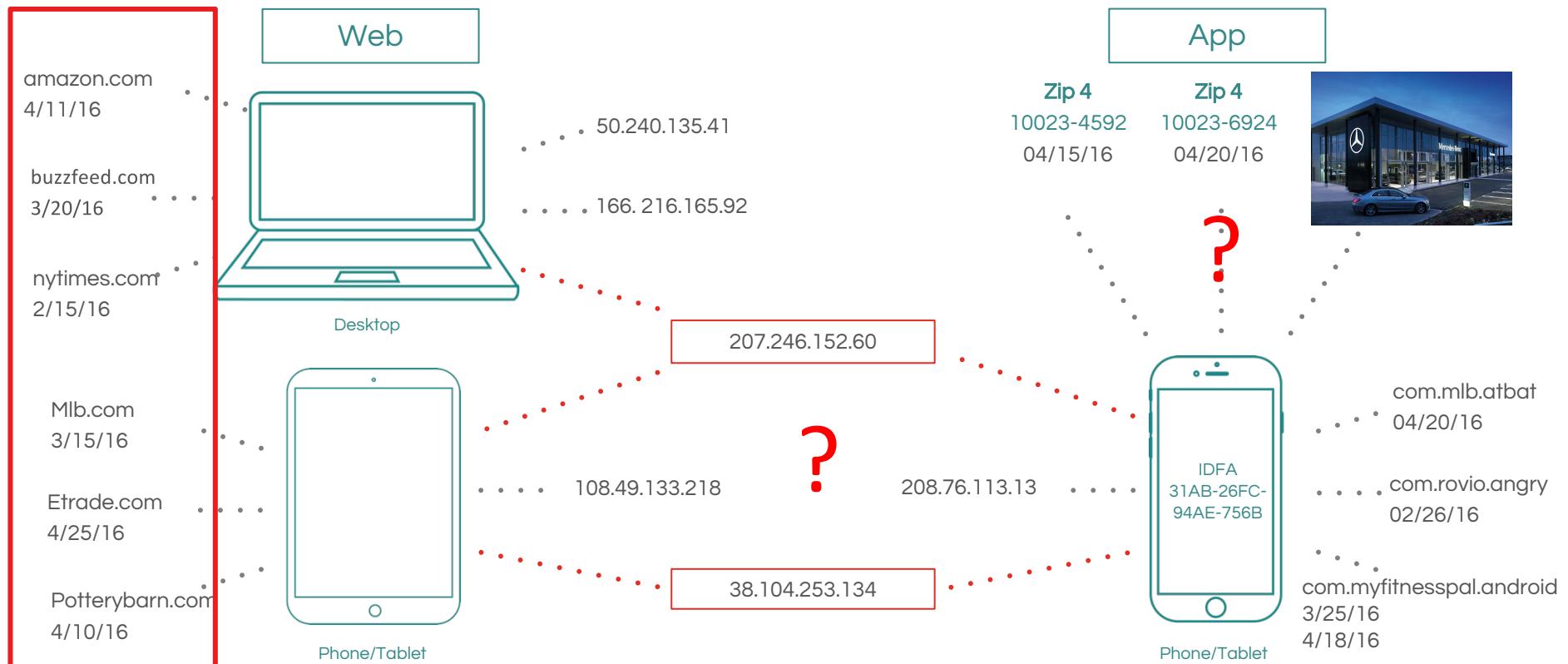


dstillery

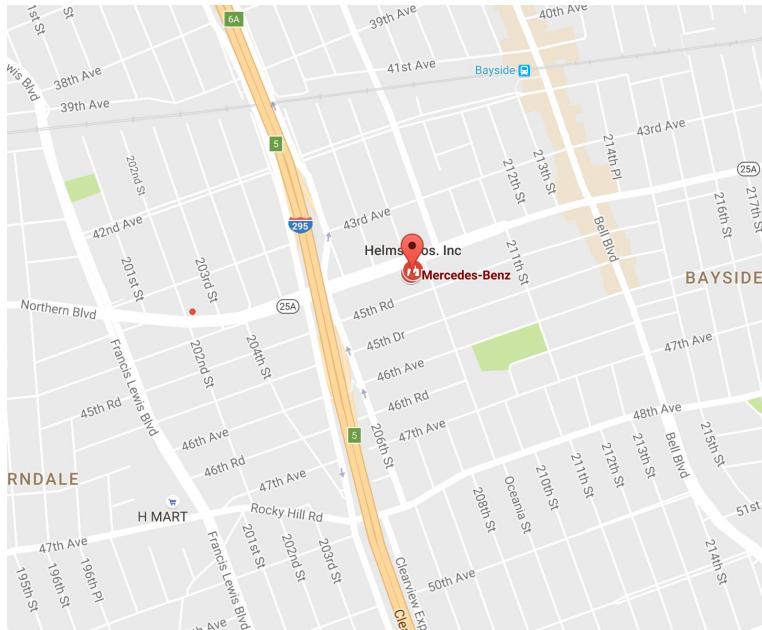
Finding good audiences for luxury cars? Predict dealership visits?



My actual data: 100 Billion daily events across devices ...



Potentially three populations in the location prediction



- **People who are indeed at dealership and their history**
- People who are somewhere close
- People who hacked into your WIFI

Identify people who will go to Mercedes dealerships



URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www.fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

'In the market' signal



URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www/fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

Real Estate



URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www/fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

Fitness ...



URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www.fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

How much randomness can a model absorb?

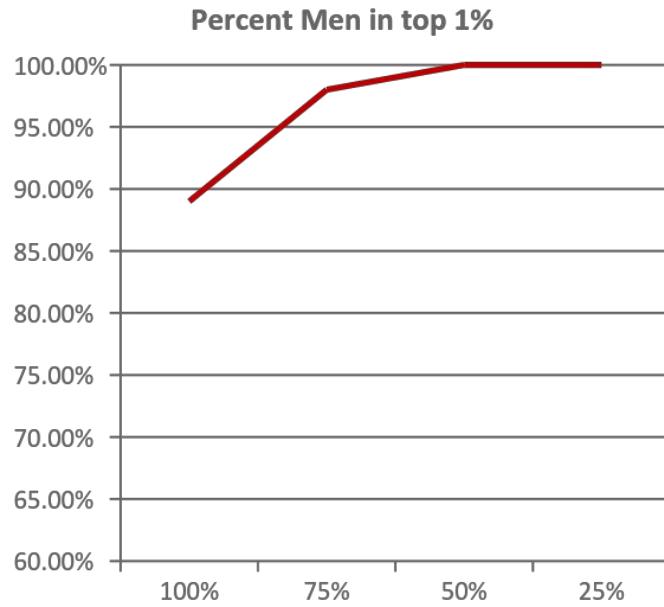
We will randomly switch the gender value for increasing percentages

```
00000d41ed774823fca142945ec915c0,1,,,...,en_GB,,  
00000dee02d70cf8c0d79f96b6d1c59d,0,,,...,en_US,,  
00000f232abfe25a80156fe069395460,0,1992,20,2,2,,,...,19,-5  
00000f4ba0cff946b1c0e3b051287ede,0,1993,19,2,,,...,en_US,310,8  
0000130571654e3afaa62f4e9d2e4f63,0,,,2,2,,,...,en_US,193,7  
00001544469ae9b408869a463a1dd77a,1,1984,28,2,,,...,en_US,,-4
```

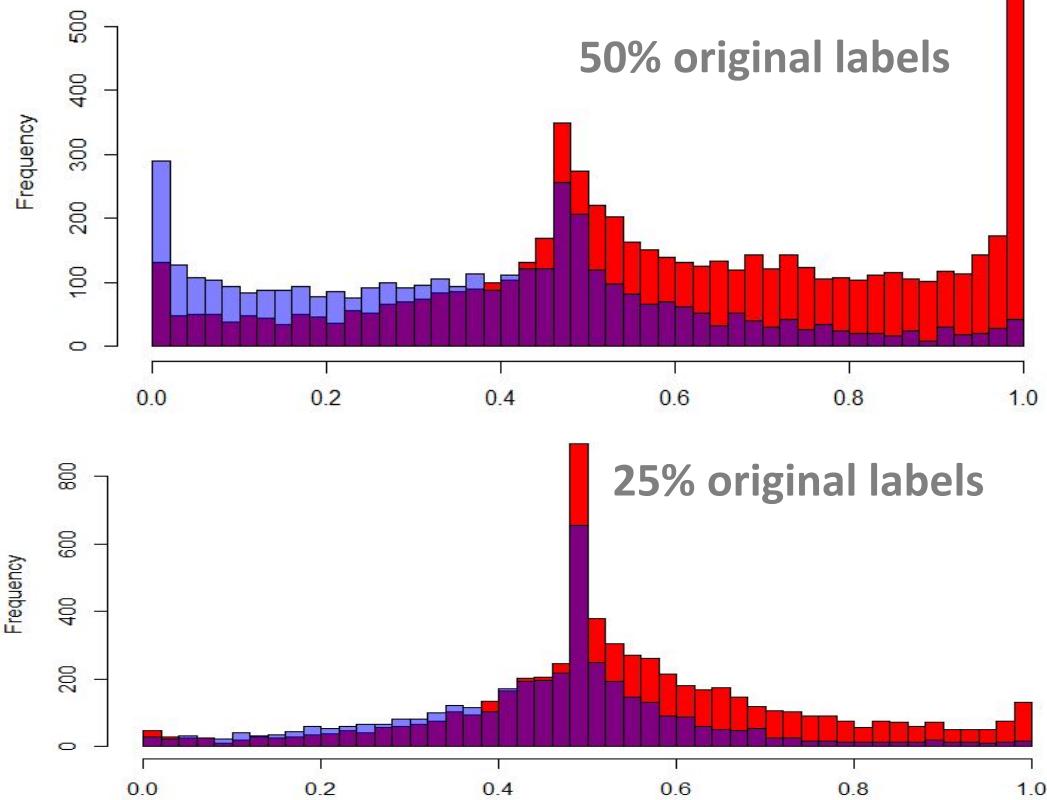


```
00000d41ed774823fca142945ec915c0,0,,,...,en_GB,,  
00000dee02d70cf8c0d79f96b6d1c59d,0,,,...,en_US,,  
00000f232abfe25a80156fe069395460,1,1992,20,2,2,,,...,19,-5  
00000f4ba0cff946b1c0e3b051287ede,0,1993,19,2,,,...,en_US,310,8  
0000130571654e3afaa62f4e9d2e4f63,1,,,2,2,,,...,en_US,193,7  
00001544469ae9b408869a463a1dd77a,1,1984,28,2,,,...,en_US,,-4
```

Performance is surprisingly stable even under random noise



Percent original labels

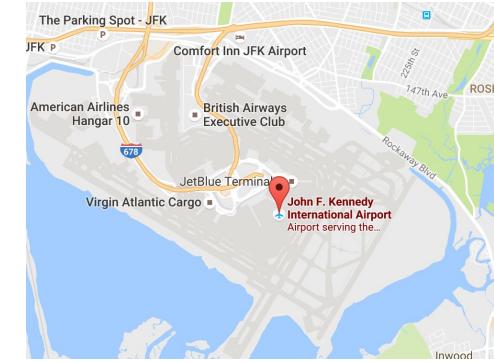


Identify people who will go to Mercedes dealerships?

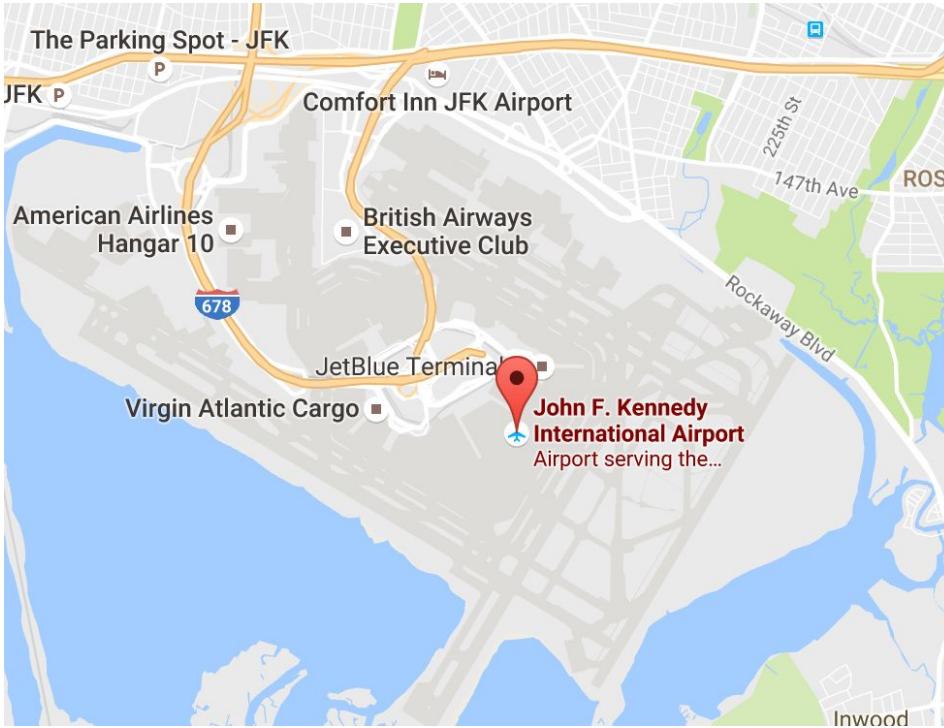


My predictions might be better than my 'ground truth'

Where do we find frequent traveler?

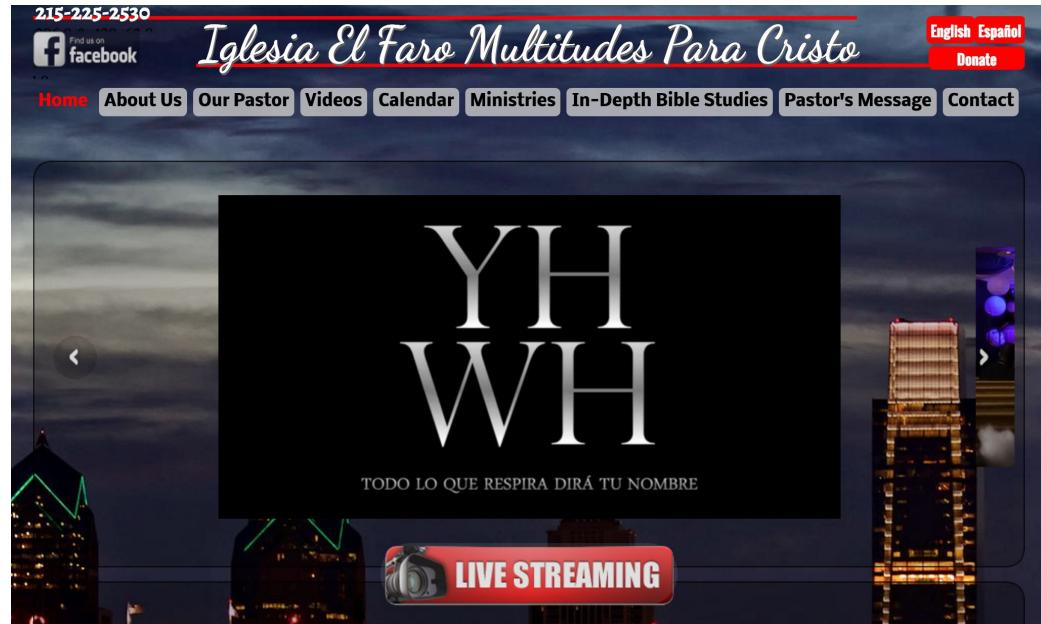


What do you think indicates people going to JFK?



URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotropy.com	1.97
www.beggsfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.oklo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.82

URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.begg'sfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.oklo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83



URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggstfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.oklo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83

The screenshot shows the homepage of the Association of Flight Attendants (AFA) website. At the top, there's a logo for 'AFA' and the text 'Association of Flight Attendants - CWA, AFL-CIO'. Below this is a banner with the text 'September 11 Remembered' and 'WE REMEMBER SEPTEMBER 11, 2001 SEPTEMBER 11, 2016'. A search bar is located at the top right. Below the banner, a large headline reads 'Show Our Pride' with the subtitle 'Show you have AFA PRIDE'. It encourages wearing the official AFA pin and wings everyday. To the right is a stylized graphic of wings. Below the main headline are several navigation links: 'Contract', 'Education', 'Reserve Preferencing', and 'Reports and Forms'. The 'Reports and Forms' link is specifically noted as 'LEC WORKSHEETS, SAFETY, HOTELS, CONTRACT, ETC.'.

The screenshot shows the homepage of Jumpseatnews.com. The header features the site's name 'Jumpseatnews.com' and the subtext 'United Airlines Flight Attendant News and Resource Center'. Below the header is a navigation menu with links for Home, Latest News, Crew Lounge, Unimatic, Discounts, Domiciles, Paperwork, About JSN, Membership, and Search. The main content area contains several news headlines: 'September 11 Remembered', 'Contract Vote Results', 'Complete Tentative Agreement Now Available', 'New Hires and Open Premium Positions', 'Tentative Agreement Approved by Joint MEC', 'AFA and United Agree to Terms for Joint Collective Bargaining Agreement', 'Early Boarding – Accommodating Special Needs Passengers', 'Crew Verification and Flight Attendant FAA Minimum', and a link to 'MORE NEWS...'. To the right of the news is a 'Quick Find' sidebar with links for 'Unimatic Access', 'Bid Packages', 'Flying Together', 'EmployeeRES', 'PED Supplement', 'Briefing Sheets', 'Uniforms', 'Travel and Safety', and 'Edit Your Pass Riders'. The sidebar also includes links for 'Help | Commands', 'Domestic | Transfers', 'Webmail', 'Pass Travel | Fare Calculator', 'Paychecks | Benefits | Discounts', 'Leave of Absence | eLearning | CQ', 'Tools | FAOM', 'Tailors', 'Medical | W2 Form', and 'Get Reports'.

URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghiova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggstfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.okllo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83



Football Homepage		Football Teams		Information		SJ Football Academy	
FRI, SEP 16	7:30 PM	FRI, SEP 16	9:30 PM	SAT, SEP 17	11:00 AM	SAT, SEP 17	11:00 AM
SJ18	--	SJ23	--	BC04	--	A01	--
WR06	--	SL06	--	SJ05	--	SJ01	--
WHEAT RIDGE, WR-B PROSPECT P		WESTMINSTER, SJ-B WOLF RUN P		LITTLETON, SJ-C EASTON PARK #		LITTLETON, SJ-A EASTON PARK #	



HEADLINES

Support Chatfield High School's Coach McGatlin
South Jeffco Family – A treasured member of our...

Dick's Sporting Goods Everyday Coupon
07/06/2016

Find us on Faceb

RECENT RESULTS

BF01
SJ01 - HOFFSCHNEIDER - I

[HOME](#) | [ABOUT](#) | [FOLLOW US ON TWITTER](#) | [CONTACT US](#)

[SUBSCRIBE](#)

- [+][Bloglines](#)
- [+][Feedster](#)
- [+][Google Reader](#)
- [+][My AOL](#)
- [+][My MSN](#)
- [+][My Yahoo](#)
- [+][Newsgrap](#)
- [+][Pluckit](#)
- [+][Rojo](#)

RECENT CHOICES

- [College Football or NFL Football?](#)
- [Raw Fish or Cooked Fish?](#)
- [Seinfeld or Arrested Development?](#)
- [Lloyd Christmas or Ace Ventura?](#)
- [Tetris or Solitaire?](#)

CATEGORIES

- [Art](#)
- [Books](#)
- [Film](#)

College Football
 NFL Football

Vote

Share / Save

College Football or NFL Football?

By The Serpent - Monday, September 14, 2009


OR


College Football

NFL Football

[« Previous Choice](#)

URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggstfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.oklo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83



HOMOTROPHY

RECENT ARTICLES

Page 1 of 1,039 | [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [Last](#)

Sebastian K. for All American Guys
SEPTEMBER 15TH, 2016 / HUNK, MODEL, PHOTOGRAPHY / NO COMMENTS

[Like](#) [Share](#) [Tweet](#) [Email](#) [Pin it](#)



SUBSCRIBE

Posts (RSS) [Comments \(RSS\)](#)

TROPHY : GET US

Become a FAN

Twitter

TROPHY::LISTS

Swimsuits Allowed



Predict who goes to JFK?



People who work there ...



Random

Deterministic



Big Picture?

Beware of (unintentional) discrimination based on predictability



LinkedIn Jobs: Targeted Email Delivery is Here

Parker Barile February 28, 2011

[in Share](#) 62 [Tweet](#) [Like](#) 0

LinkedIn just rolled out an exciting new feature that makes it even easier to reach qualified candidates automatically. Members who view jobs on LinkedIn will now receive targeted job recommendations automatically via email. Powered by LinkedIn's 'Jobs You May Be Interested In' system, the emails include jobs that are personalized for each recipient based on his or her professional profile.

LinkedIn. Jobs

Ad related to latanya sweeney ⓘ

[Latanya Sweeney Truth](#)

[www.instantcheckmate.com/](#)

Looking for Latanya Sweeney? Check Latanya Sweeney's Arrests.

Ads by Google

[Latanya Sweeney Arrested?](#)

1) Enter Name and State. 2) Access Full Background Checks Instantly.

[www.instantcheckmate.com/](#)

[Latanya Sweeney](#)

Public Records Found For: Latanya Sweeney. View Now.

[www.publicrecords.com/](#)

[La Tanya](#)

Search for La Tanya Look Up Fast Results now!

[www.ask.com/La+Tanya](#)

dstillery

Claudia Perlich

Thank You!

dstillery