

SCALING ML IN DIGITAL ADVERTISING

Claudia Perlich

Chief Scientist

@claudia_perlich

dstillery
dstillery

Programmatic Advertising

The screenshot shows the homepage of The New York Times on a desktop browser. Several programmatic advertisements are highlighted with red boxes:

- Top Left:** A banner for "THE ALL-NEW CADILLAC XTS" with the Cadillac logo.
- Top Right:** A smaller banner for "THE ALL-NEW CADILLAC XTS" showing a black car.
- Search Bar:** An advertisement for "ING DIRECT" is placed above the search input field.
- Bottom Right:** A large video player advertisement for "INTRODUCING THE ALL-NEW CADILLAC XTS" with a "CLICK FOR SOUND" button and a "LEARN MORE" link.

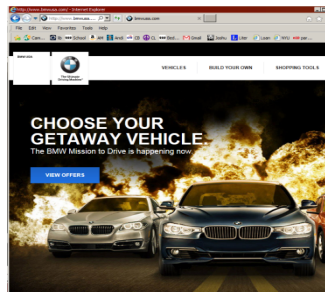
The main content area includes a navigation menu on the left, a featured article titled "Fiscal Crisis in States Will Last Beyond Slump, Report Warns", a "TimesCast" video player, and various other news sections like "Mormons' First Families Rally Behind Romney" and "Inside the Romney No. 2 Hunt: First, Do No Harm".

300 Million (US) consumer



Interacting with brand sites

Using Digital Devices



Measurement

Conversion

If we win an auction we serve ad of brand

100 ms to bid

Work with Brand



100 Billion bid requests per day

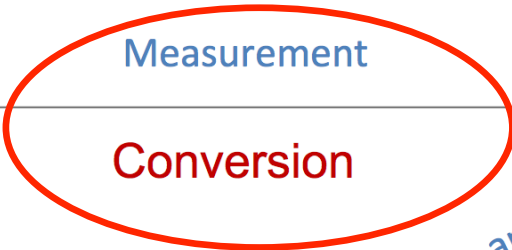
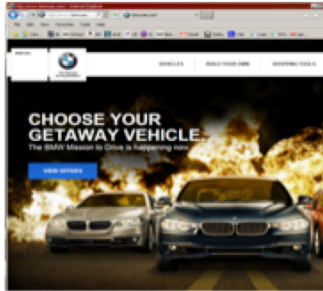
dstillery

300 Million (US) consumer



Interacting with brand sites

Using Digital Devices



If we win an auction we serve ad of brand

100 ms to bid

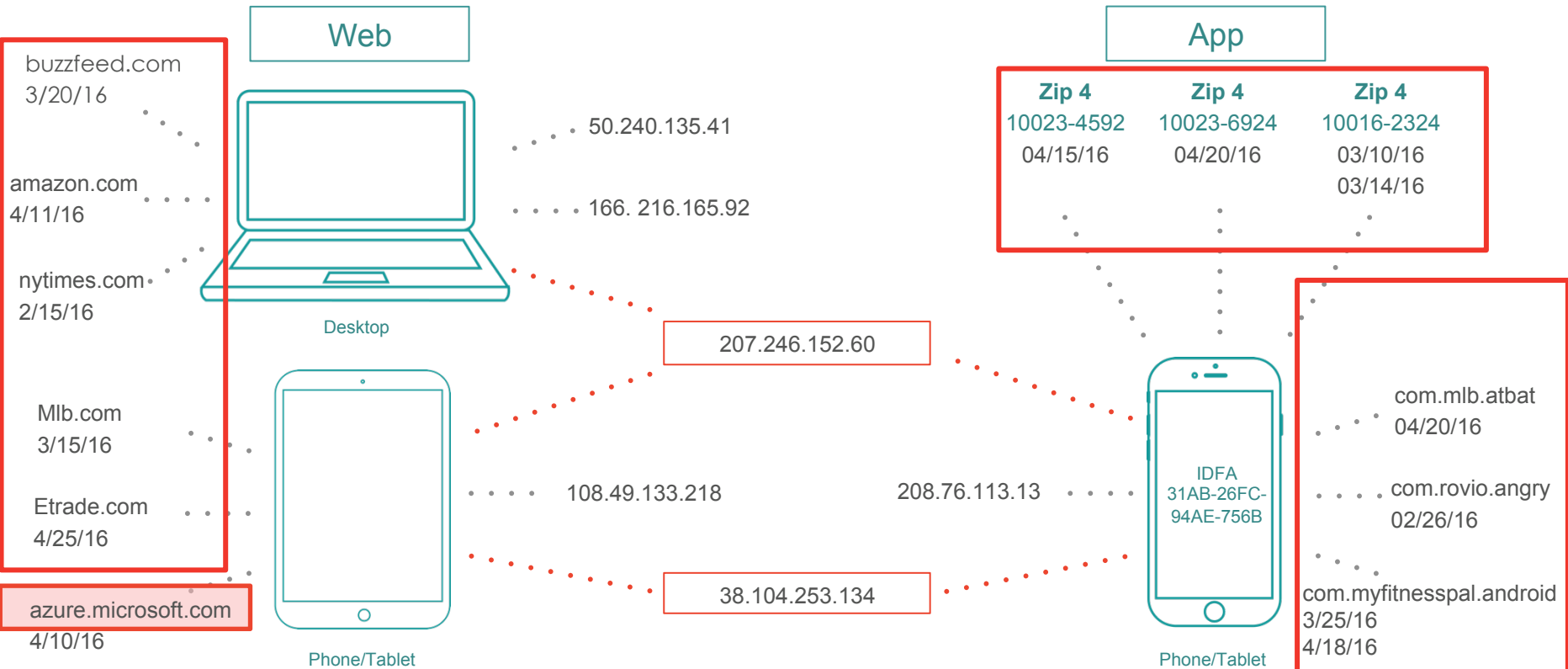
Work with Brand



100 Billion bid requests per day

dstillery

Predicting events on pretty much everything



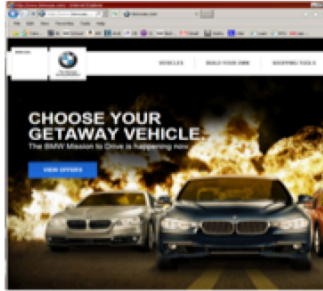
300 Million (US) consumer



SCALE?

Interacting with brand sites

Using Digital Devices



Measurement



Conversion

If we win an auction we serve ad of brand

100 ms to bid

Work with Brand



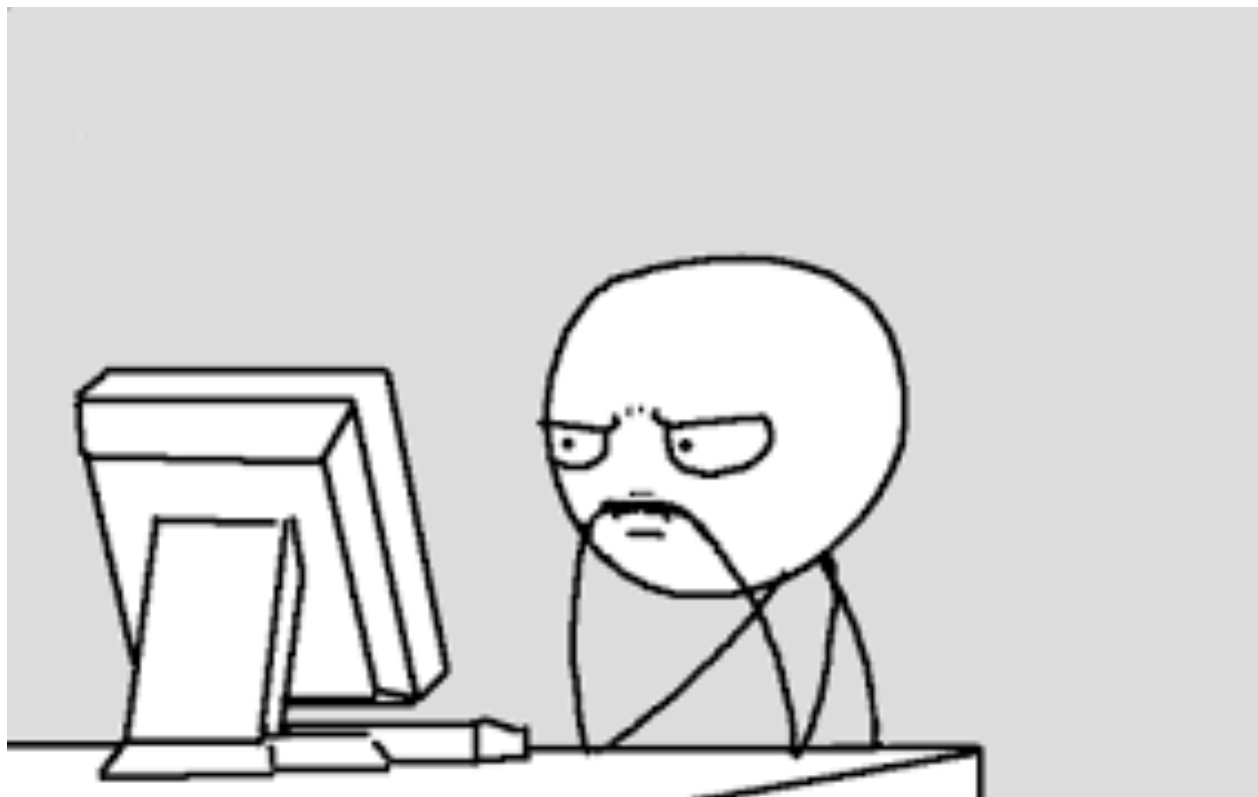
100 Billion bid requests per day

dstillery

Lets talk scaling? Most Limited resource?



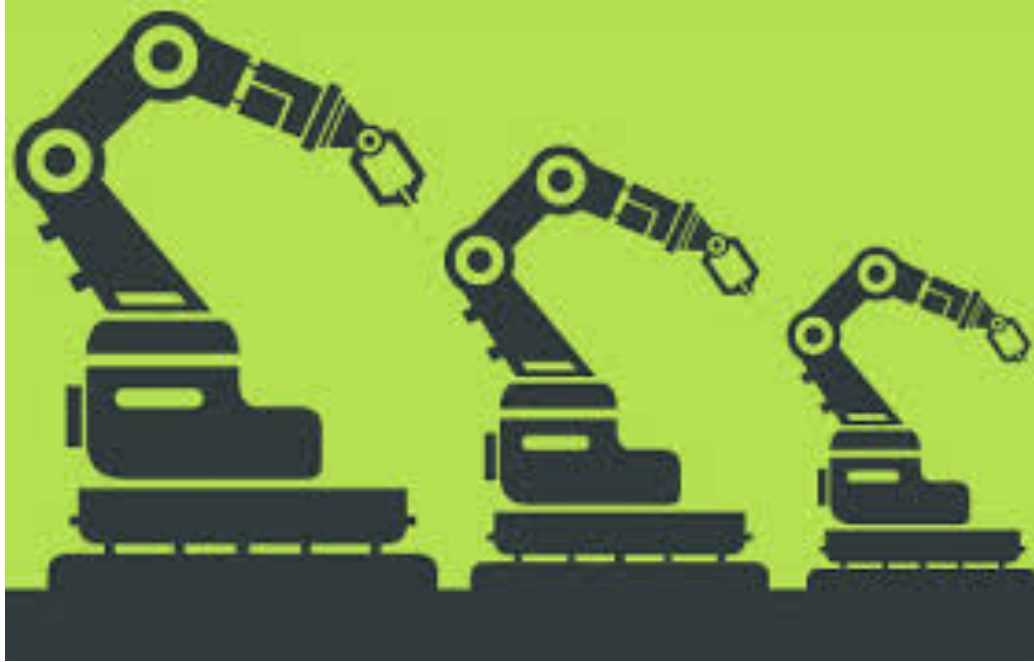
Personally: Waiting for your model estimation sucks!



Tip of the iceberg ...



World of Digital Advertising



Automation beats scaling ML

Scaling the Solution, not the ML ...

Build/manage thousands of models with limited DS resources ...

- All kinds of target/baserates:
various types of conversion, click, viewability, video completion, store visits, purchases
- Typically low predictability
- Create additional model for new target with a flip of a switch in hours
- Manage the feature space and automate deployment

Score ~200 Million examples daily against 10K models

- Less than 0.005 ms each

Economic efficiency

DS Dashboard ...

317

Campaigns on Optimus (out of 432 active campaigns; 249 auto set prices; 83 auto create targetgroups)

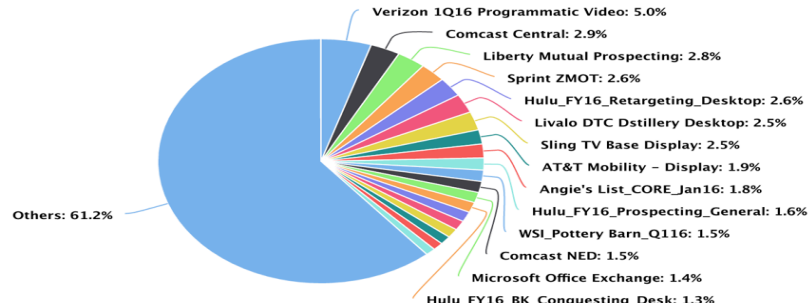
3116

SGD models were built in the past 24 hours

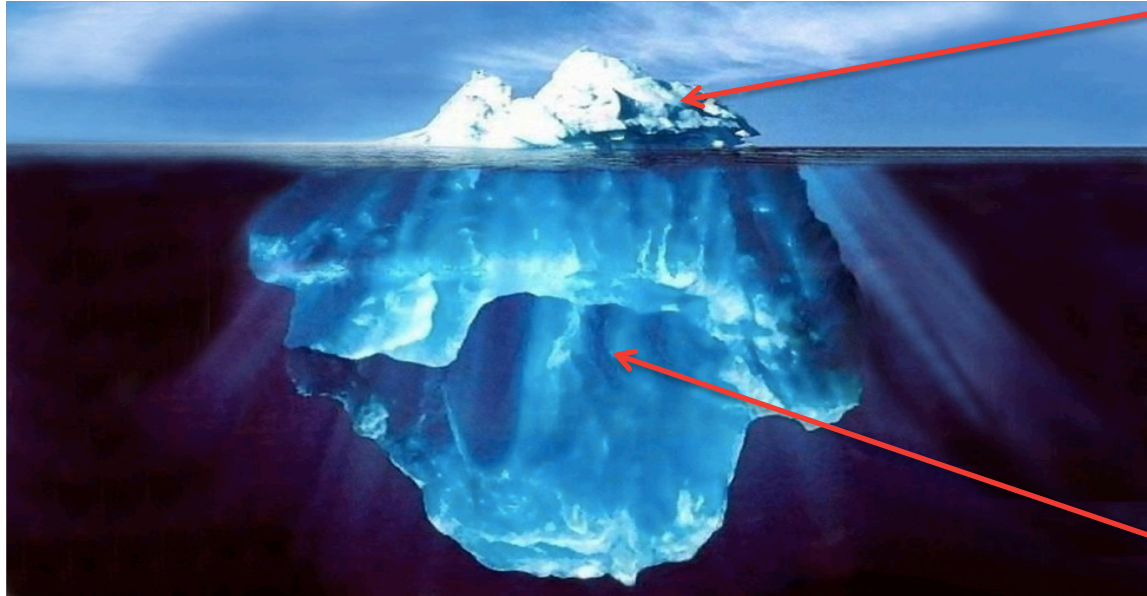
Projects:

- CFP Sampler Specs
- Data Nexus
- DS Crosswalk
- Inventory Stats
- Optimus Campaigns
- Place Still Tag Index
- Seg Avails
- Sport
- Segrank Campaigns
- Exchange Specific Stats
- Optimus Decider Audit Trail

Top campaigns by daily revenue



Automating Machine Learning

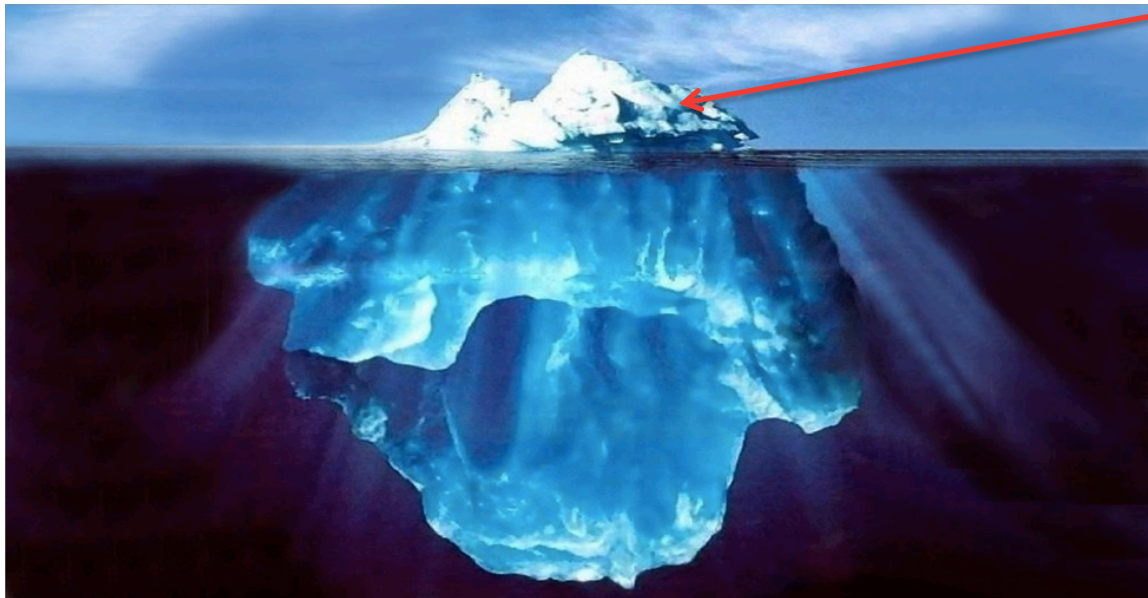


Strategic ML Choices ...


Infrastructure

Automating Machine Learning

Strategic ML Choices ...



ML Strategy: robustness & interpretation beats peak performance

- Feature engineering
 - Sparse binary or hostnames: easy to code all kinds of data
 - ~100 Million (possible) features
 - Stay 'close' to the raw data
 - Hashing to ensure consistent feature space
 - One MUST fit all ...
 - Algorithm
 - Logistic regression (and some Naïve Bayes)
 - Stochastic gradient descent
 - Penalties & Heuristics for Learning Rate Estimate
 - Constant refresh every ~3 days
 - Stacking & Transfer Learning & Sampling
- 



SGD Model (Model ID: 1728493)

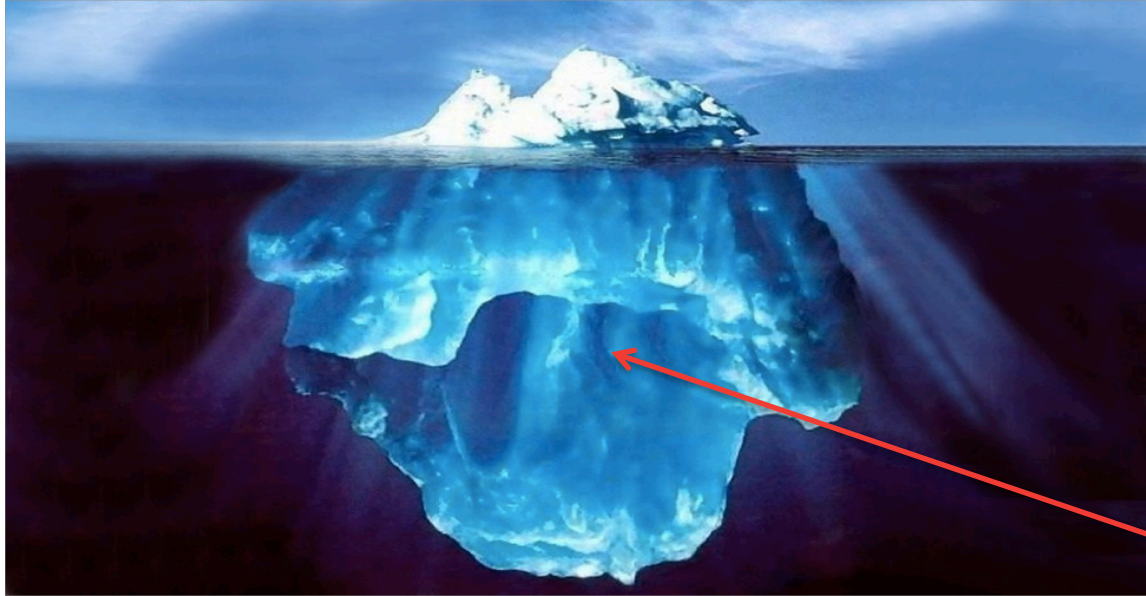
Marketer	Microsoft (Empowering Media SF) [6000]
Offer	Microsoft Azure Evergreen [7599]
Offer Level	40: Free Trial [19945]
Model Created At	2016-03-06 12:36:39
Model Event End Time	2016-03-06 17:40:00
# Positives	217
# Negatives	2,170
# Features	362,059
Parent Model	1725331
HDFS Path	hdfs://hdfs2/user/sgd/models/19945/OL19945_FM2_BS20160306123636.model

Top Features with Highest Scores

Show entriesSearch:

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

Automating Machine Learning



Infrastructure

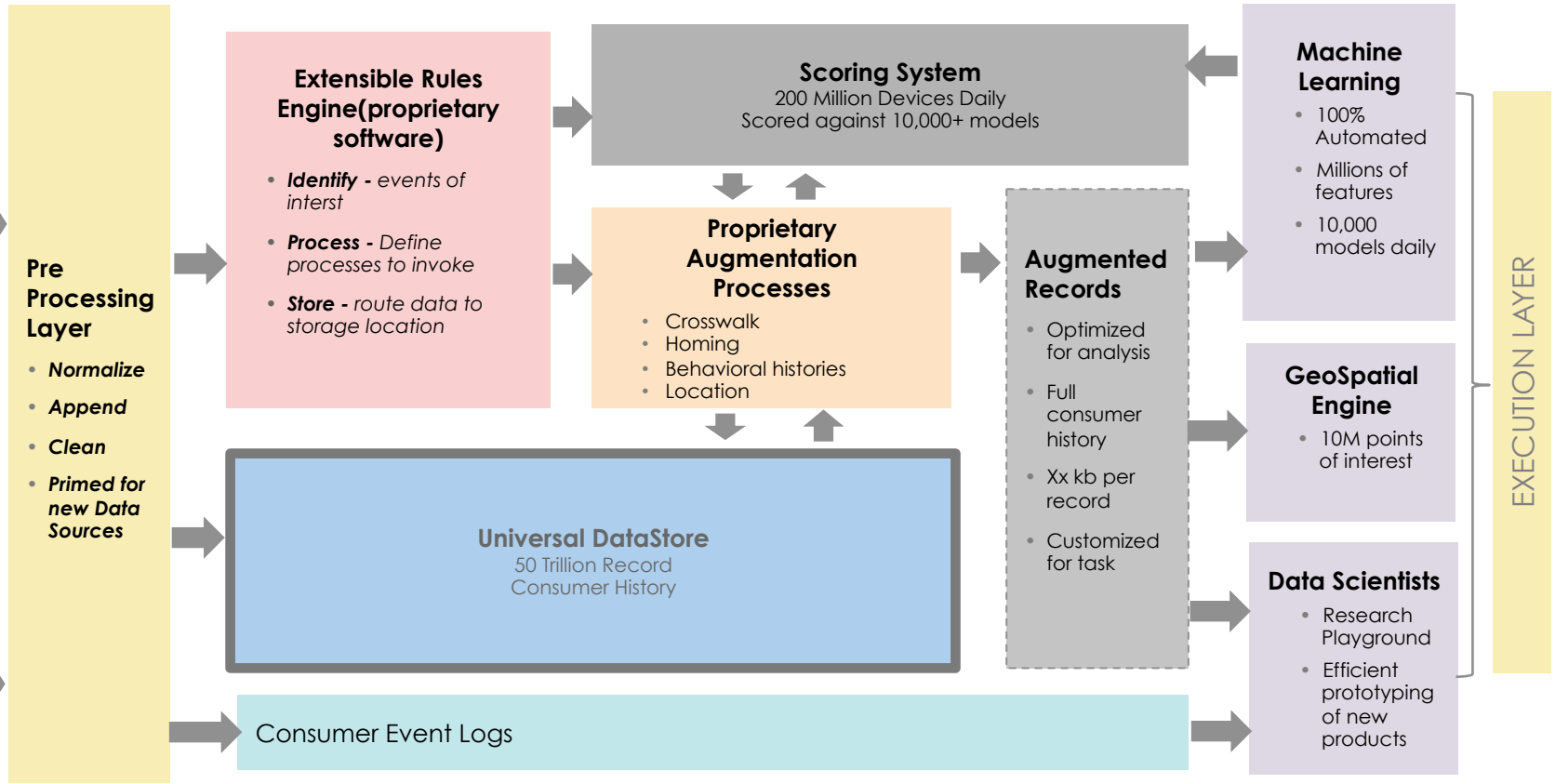
Consumer Sensors Pipe
Data Into ..

Streaming Data Infrastructure Unlocking Access to 50 Trillion record Consumer
History..

Empowering
Scientists & Machines

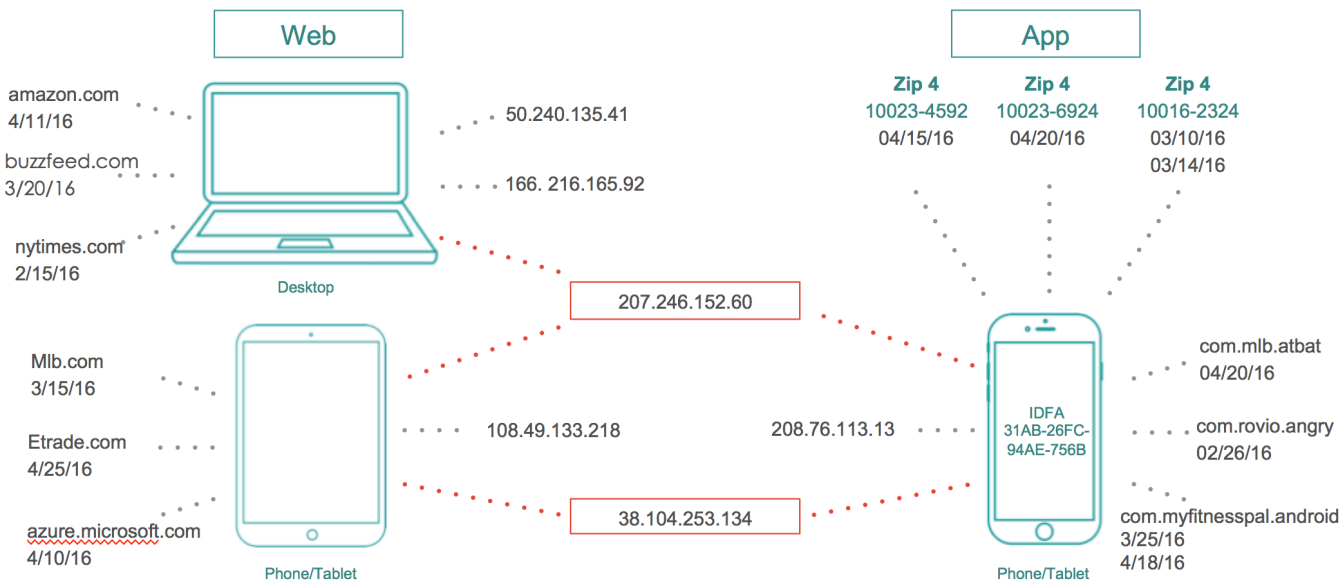
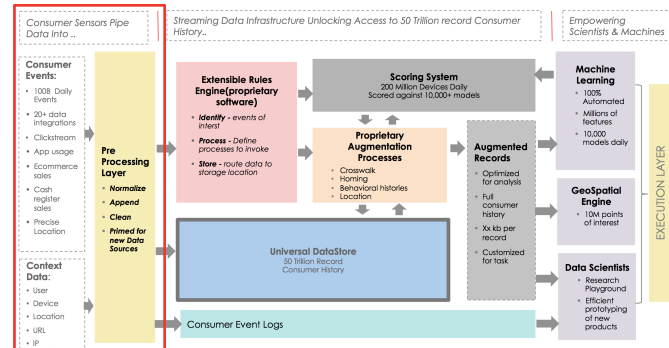
- Consumer Events:**
- 100B Daily Events
 - 20+ data integrations
 - Clickstream
 - App usage
 - Ecommerce sales
 - Cash register sales
 - Precise Location

- Context Data:**
- User
 - Device
 - Location
 - URL
 - IP



Consumer Events Data

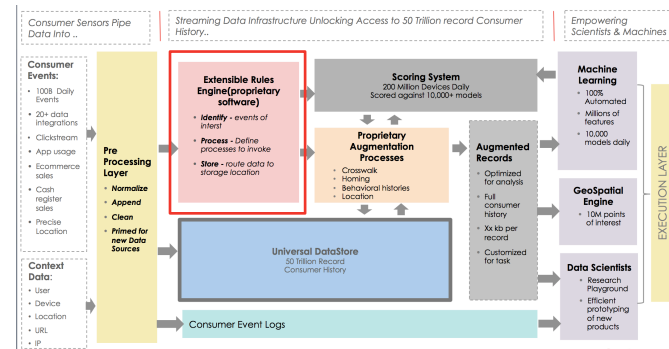
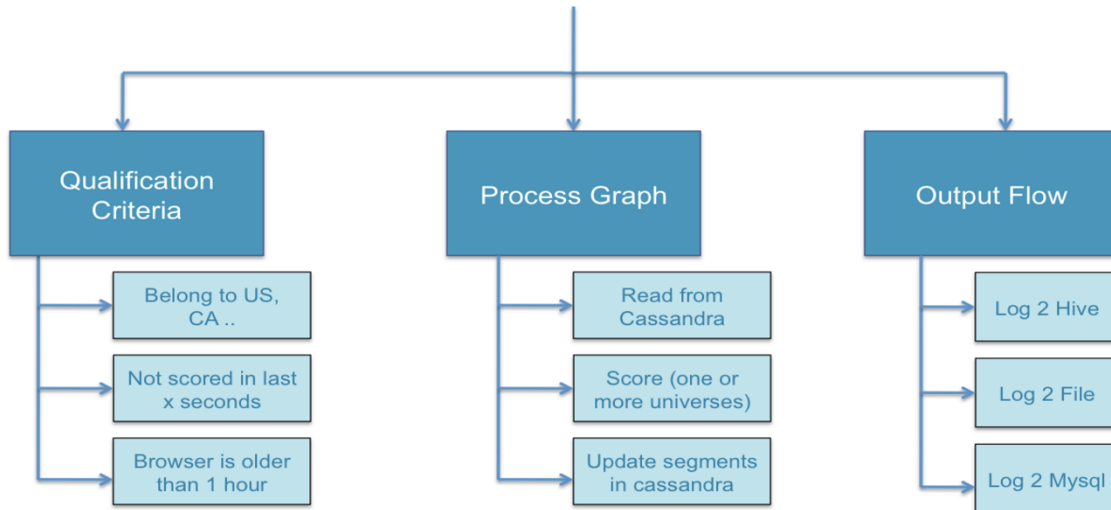
- Cleaning & consistent format
- Bot detection
- Appending contextual information



Sampler

- Extensible Rules Engine

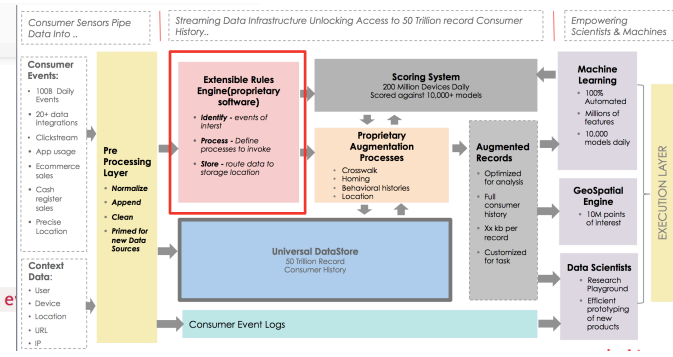
Samplers / Filters



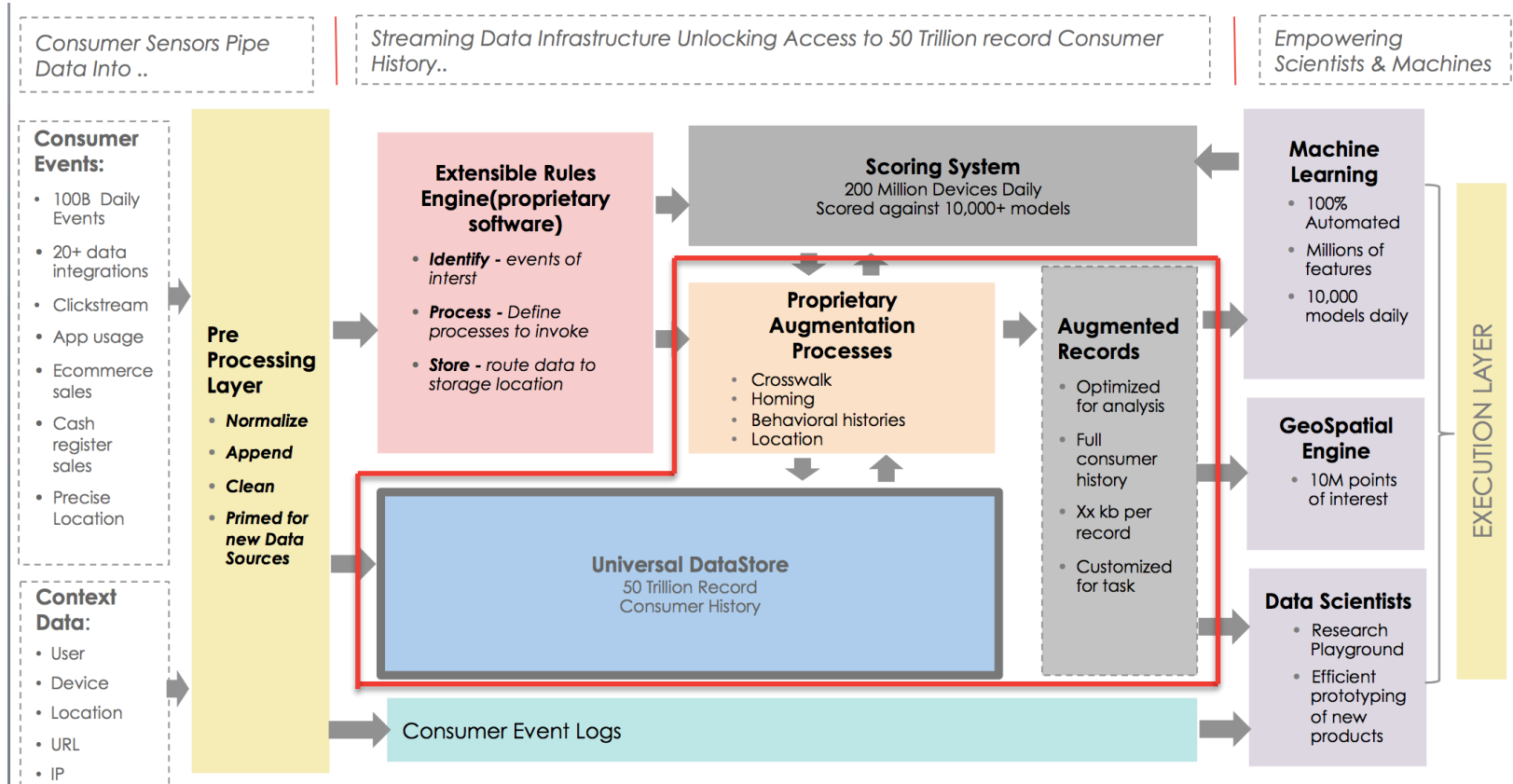
id 112
 desc SGD RON Sampler
 process init-request;populate-cookies-from-req;refresh-ssc(adh,vstcnt,lbh,sglst,browser_prop);feature-map(2);score(display,5:6:11:14:17:18,false,-1,false,true);cfp-to-kafka(cfpBaseline)
 mode ON
 destination
 precondition none
 sampler_interval 86400
 spec {"isQ":"return (event.eventType == 6 && event.audienceType == 72 && (event.getCountry() == \"US\" || e \\\"CA\\\" || event.getCountry() == \"GB\\\" || event.getCountry() == \"ZZ\\\" || event.getCountry() == \"unknown\\\"));\",\"onQ\":\"\",\"onR\":\"\"}
 type groovyTemplate
 version 3

id 113
 desc Fracture INS
 process init-request;populate-cookies-from-req;refresh-ssc(adh,vstcnt,lbh,sglst,browser_prop);feature-map(2);cfp-to-kafka(cfpBaseline);
 mode ON
 destination
 precondition none
 sampler_interval 0
 spec {"isQ":"return (event.getBrowserId() % 1000 == 0) && event.getBrowserId() > 999999999 && event.environmentType != \"APP\\\";\",\"onQ\":\"\",\"onR\":\"\"}
 type groovyTemplate
 version 2

id 114
 desc Bid Sampler 0.125%
 process init-request;populate-cookies-from-req;refresh-ssc(adh,vstcnt,lbh,sglst,browser_prop);cfp-to-kafka(cfpBaseline)
 mode ON
 destination
 precondition none
 sampler_interval 0
 spec {"isQ":"return ((new Random()).nextInt(800) == 1 && event.environmentType != \"APP\\\" && event.eventType == 10 && (event.getCountry() == \"US\\\" || event.getCountry() == \"CA\\\" || event.getCountry() == \"GB\\\"));\",\"onQ\":\"\",\"onR\":\"\"}
 type groovyTemplate
 version 2



Event Data Enrichment




Augmented Event Data (Json)

EventData	Data describing current event. Time, IP, etc.
GeoData	Info derived from precise physical location.
DeviceData	All data regarding the event device, including historical observations.
VisitData	List of historical visits by this device.
ScoresData	Thousands of propensity scores from the Dstillery machine learning system.
LinkedData	List of linked devices as well as list of linked visits.
SegmentData	Segments historically stored against device, as well as segments derived from all of the above augmented features.

 Paste JSON here

```
{
  "featureid":99,"offerlevelid":28568,"score":-0.28205407},
  {"featureid":100,"offerlevelid":28568,"score":-0.020146718},
  {"featureid":102,"offerlevelid":28568,"score":-0.025014956},
  {"featureid":9,"offerlevelid":100,"score":58.0},
  {"featureid":10,"offerlevelid":100,"score":62.0},
  {"featureid":14,"offerlevelid":100,"score":173.0},
  {"featureid":15,"offerlevelid":100,"score":2.9827585},
  {"featureid":23,"offerlevelid":100,"score":1.0},
  {"featureid":49,"offerlevelid":100,"score":5.1532917}},
  "sgdattrscout":1,"sgdattrslst":
  [{"featuremapperid":2,"sparseentriescount":62,"sparseentrieslist":
  [{"id":46275894,"value":1.0},{"id":46425750,"value":1.0},
  {"id":46392250,"value":1.0},{"id":33561858,"value":1.0},
  {"id":48004636,"value":1.0},{"id":46366453,"value":1.0},
  {"id":46570000,"value":1.0}, {"id":46160000,"value":1.0}, {"id":47001004,"value":1.0}, {"id":46474001,"value":1.0}],
  "appauditid":"zZJonVgkyYLTgAAABLC"
}
```

Edit your JSON directly in the textarea above. The JSON editor will get updated when you leave the field.

 ... or call a JSONP endpoint

URL:
display JSON

Example: <http://example.com/info.json?callback=?> # do not forget to put the actual callback name in the field on the left

JSON editor

Expand all *Legend:* array object string number boolean null Remove item by deleting a property name.

- eventattr

{"actiontakerinfocount":0,"actiontakerinfo": [],"appauditid":"zZJonVgkyYLTgAAABLC"
- actiontakerinfocount

0
- + actiontakerinfo

[]

DS Dashboard ...

317

Campaigns on Optimus (out of 432 active campaigns; 249 auto set prices; 83 auto create targetgroups)

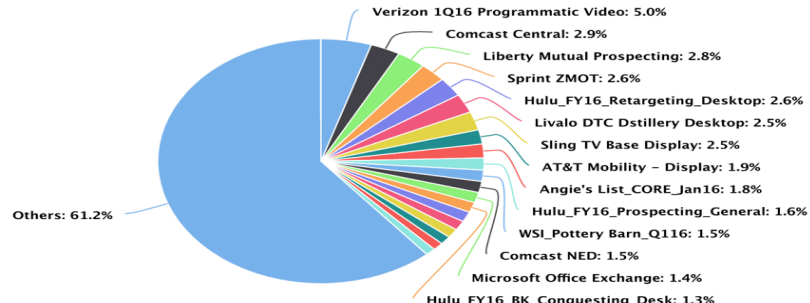
3116

SGD models were built in the past 24 hours

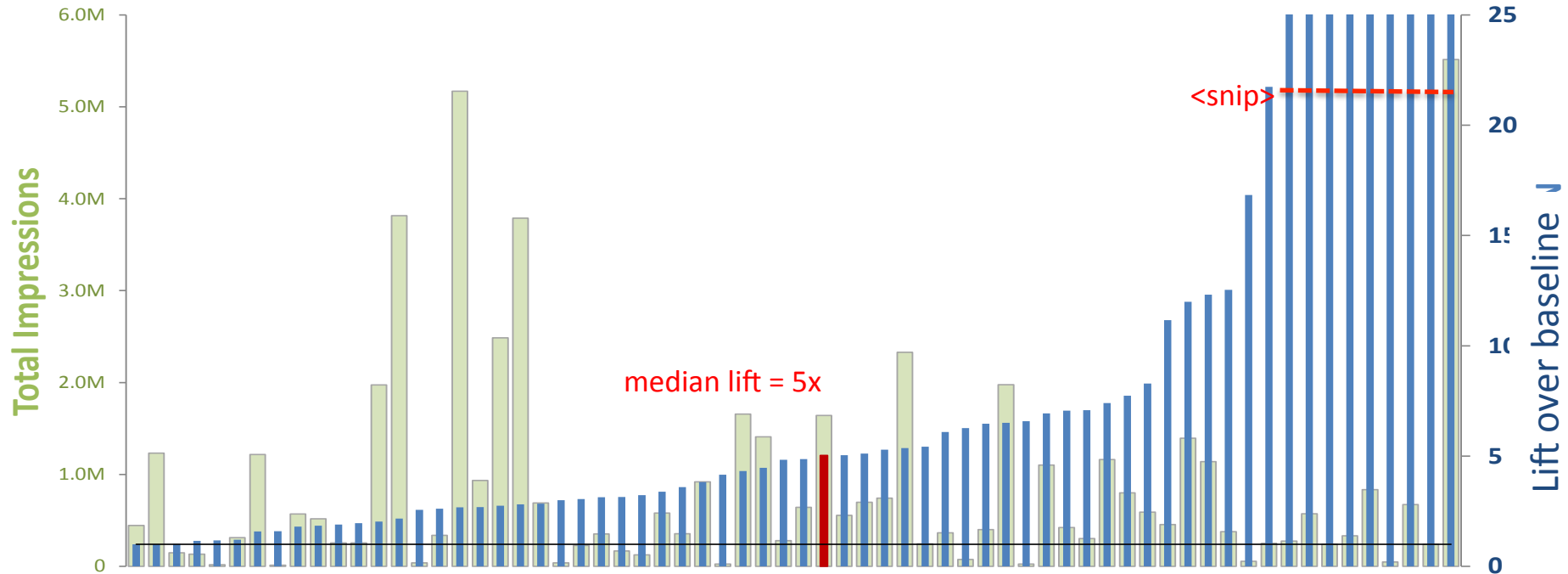
Projects:

- CFP Sampler Specs
- Data Nexus
- DS Crosswalk
- Inventory Stats
- Optimus Campaigns
- Place Still Tag Index
- Seg Avails
- Sport
- Segrank Campaigns
- Exchange Specific Stats
- Optimus Decider Audit Trail

Top campaigns by daily revenue

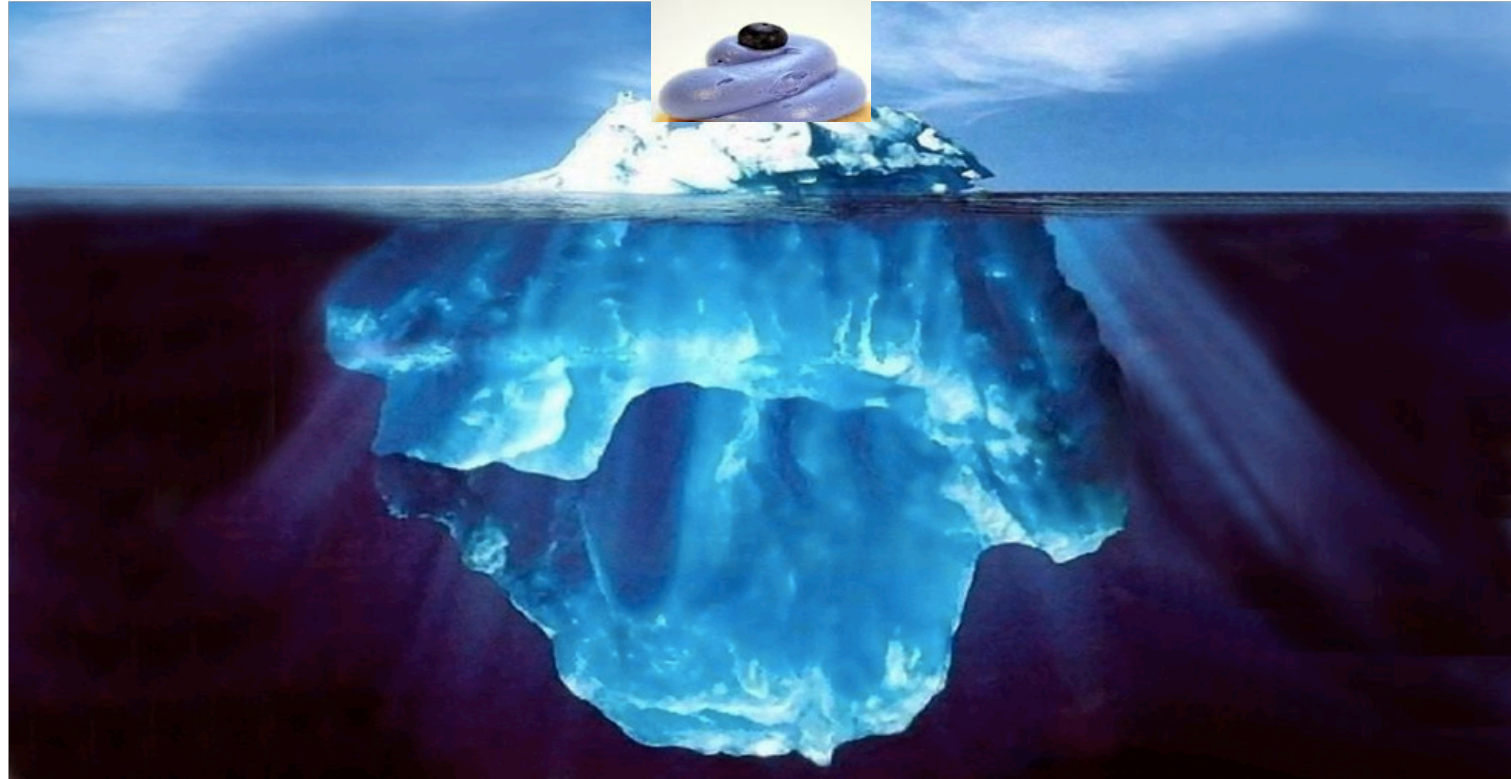


Lift over random for online display ad targeting



THE PREDICTABILITY PREDICAMENT: BEWARE WHAT YOU ASK FOR!

“Icing” on top of the Ice Berg?



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

Pizza for
Dinner?



Sexual
Orientation



Random

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:

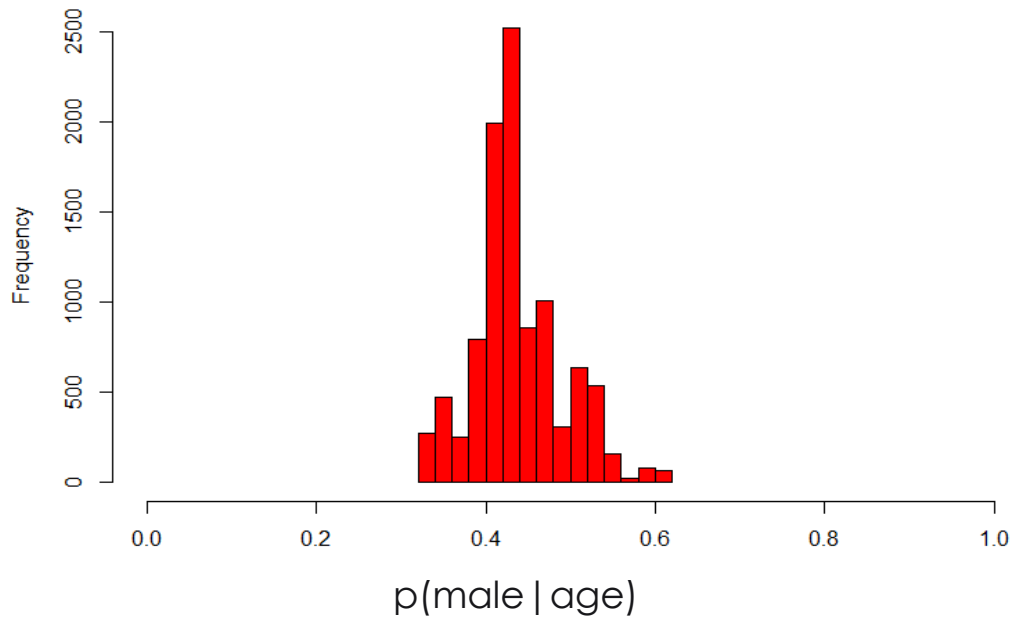
Logistic regression on sparse representation

```
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 ...



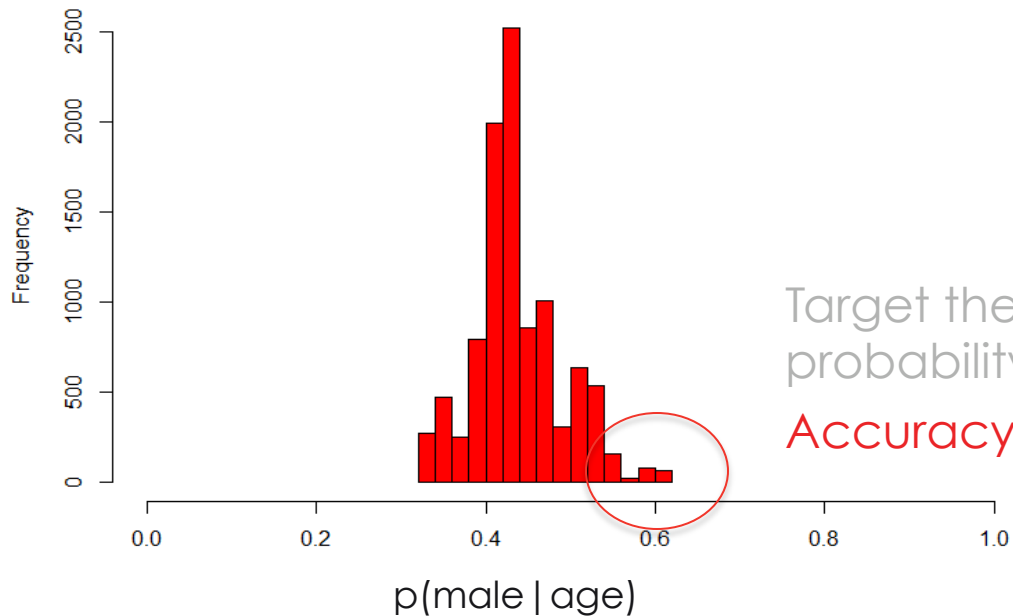
Gender based on age: very little signal ...



Overall Accuracy: 60%

AUC: 58%

Gender based on age: very little signal ...



Target the 1% with highest probability:

Accuracy: 75%

Overall Accuracy: 60%

AUC: 58%

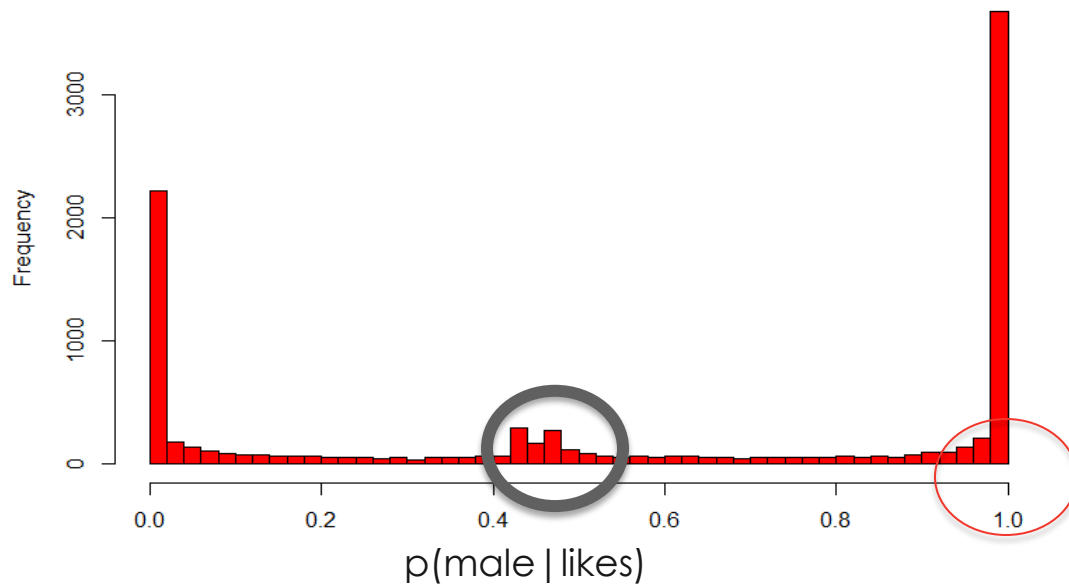
From small to bigger data ...



Take 2: Gender based all your likes



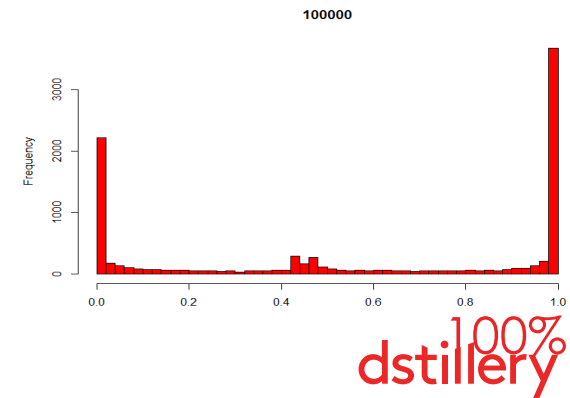
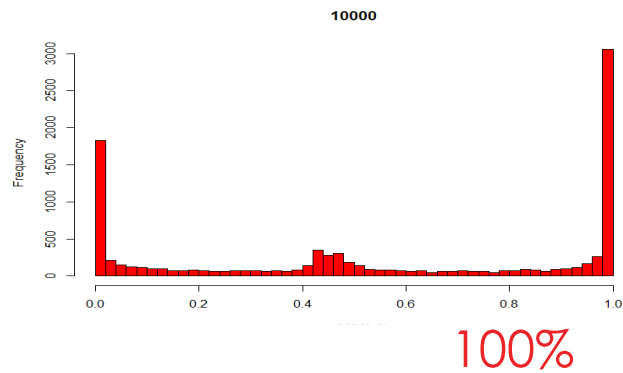
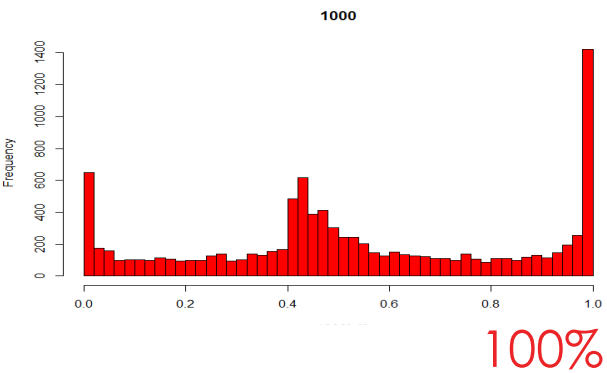
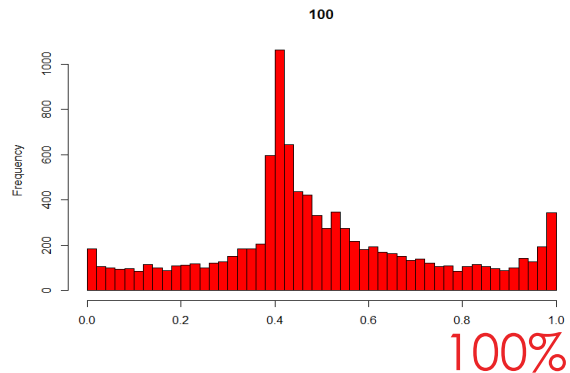
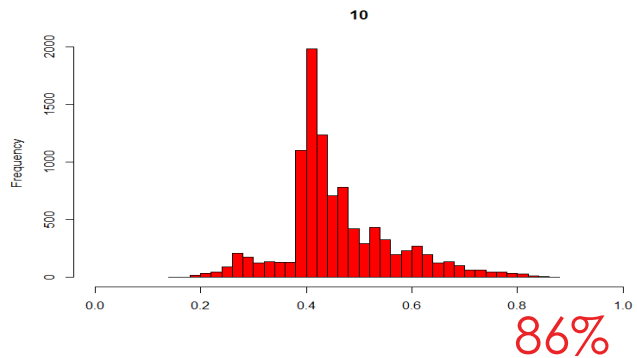
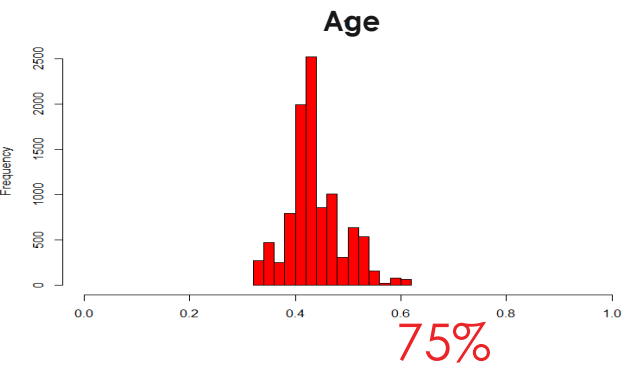
Predict gender base on all likes: a lot of signal ...



Target top 1%
Accuracy: 100%

Overall Accuracy: 83%

Progression: from age to all 'likes'



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



Random

Deterministic

DS Dashboard ...

317

Campaigns on Optimus (out of 432 active campaigns; 249 auto set prices; 83 auto create targetgroups)

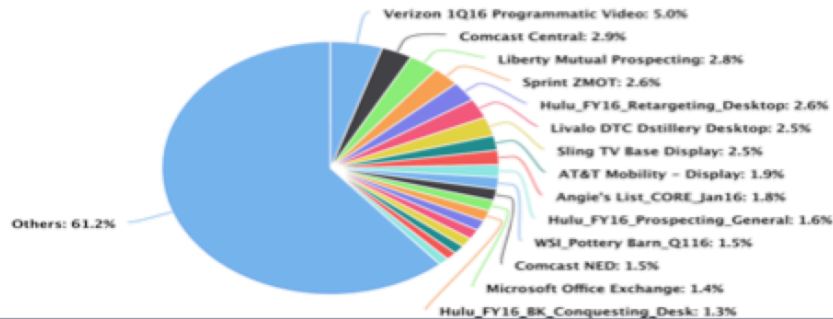
3116

SGD models were built in the past 24 hours

Projects:

- CFP Sampler Specs
- Data Nexus
- DS Crosswalk
- Inventory Stats
- Optimus Campaigns
- Place Still Tag Index
- Seg Avails
- Sport
- Segrank Campaigns
- Exchange Specific Stats
- Optimus Decider Audit Trail

Top campaigns by daily revenue



Witness a spike in human predictability ..



Death of
free will?



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

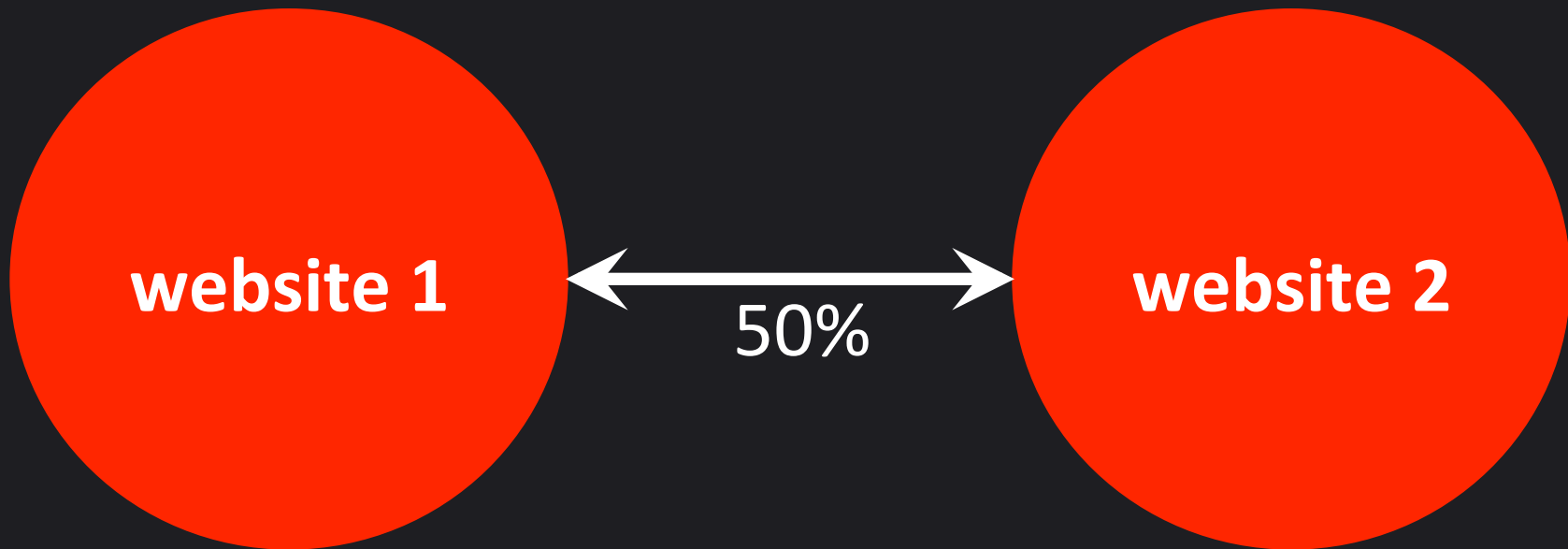
www.chinaflix.com

www.parentingnewsstories.com

www.wrestlingnewz.com

www.gourmandia.com

Traffic overlap of cookies from Bid Request



What website am I?

Web Site	Percent Common Users
www.youtube.com	55%
www.weather.com	28%
www.foxnews.com	23%
www.imdb.com	22%
www.ehow.com	22%
www.latimes.com	20%
www.washingtonpost.com	19%
wiki.answers.com	18%
abcnews.go.com	18%
www.forbes.com	17%
www.dailymail.co.uk	14%
www.examiner.com	13%
premiumweathersite1.com	11%
www.politico.com	11%
www.nydailynews.com	11%
www.usatoday.com	11%
www.slate.com	10%
content.usatoday.com	10%
www.nypost.com	9%
www.npr.org	9%

Answer: nytimes.com

HOME PAGE TODAY'S PAPER VIDEO MOST POPULAR U.S. Edition ▼

Subscribe: Digital / Home Delivery | Log In | Register Now

blomingdales

The New York Times

Saturday, September 29, 2012 Last Update: 2:41 PM ET

SEE WHAT'S ON ROLLOVER TO EXPAND

Search Orange Savings Account™

Follow Us f t | Subscribe to Home Delivery | Personalize Your Weather

nexus⁷

The new tablet from Google. From \$199

Watch video

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
ARTS
STYLE
OPINION

Autos
Blogs
Books
Cartoons
Classifieds
Crosswords

Supreme Court Faces Crucial Rulings in Coming Term

By ADAM LIPTAK 12:29 PM ET

The justices, starting on Monday, are expected to take up same-sex marriage, the Voting Rights Act of 1965 and affirmative action in higher education.

Kurds to Pursue More Autonomy if Syrian Regime Falls

By TIM ARANGO

Some in the long divided and oppressed ethnic group are

ARTHUR OCHS SULZBERGER, 1926 - 2012



OPINION »
EDITORIAL APPRECIATION
A Publisher and Projectionist
A personal reminiscence of Arthur Ochs Sulzberger, the man who led The New York Times for 34 years.

- Blow: 40 Days of Night
- Nocera: College Rankings
- Editorial: Overhaul Special Education
- Glanville: Team as Family
- Op-Ed: The N.F.L. Lockout and the Modern Economy

BUSINESS DAY »
COMMON SENSE
Are You Better Off? That Depends on Who You Are
The people who should be most receptive to Mitt Romney's message are less likely to vote for him.

SPORTS »

No platform fees.
Trade free for 60 days.



What website am I?

Web Site	Percent Common Users
www.filmannex.com	96%
www.ffog.net	93%
www.drugsnews.org	93%
menshealthbase.com	89%
www.womenshealthbase.com	85%
www.hark.com	85%
womenshealthbase.com	81%
www.articletrunk.com	81%
411answers.com	78%
www.dailyrx.com	74%
www.all-allergies.com	70%
knowvehicles.com	67%
modernbaby.com	63%
parentingnewsstories.com	59%
www.thefutoncritic.com	59%
www.gourmandia.com	59%
www.gossipcenter.com	59%
dailyfreshies.com	56%
www.greeting-cards.com	56%

Answer: hiphophavoc.com



choose from a large selection of *Android™* phones

Shop Now
free shipping

online deals

HipHopHavoc

THE HOME OF HIP HOP AND GANG BANGAZ

Search Hip Hop Audio:

- HOME
- NEWS
- MUSIC
- VIDEOS
- LIFESTYLES
- AUDIO CLIPS
- INTERVIEWS
- REVIEWS
- HHH RSS feed
- Audio Problems?

HIPHOPHAVOC TOP STORY



LIL WAYNE BREAKS ELVIS PRESLEY'S BILLBOARD RECORD

LOS ANGELES (Reuters) – Lil Wayne toppled Elvis Presley to become the new King of the Billboard Hot 100 singles chart on Thursday, with a

TOP NEW EXCLUSIVES

- DJ Drama ft Fabolous, T Pain, Yo Gotti - [GoIn Down](#)
 - Belly ft Juicy J - [1200](#)
 - J Cole ft Major Lazer - [Get Free ColeWorld](#)
 - J.Ford ft Peedi Crakk - [Favorite](#)
 - Mistah FAB ft Willie Joe - [Automatic](#)
- [See All Exclusives](#)

HIPHOPHAVOC E-MAIL

Login ID:

Password:

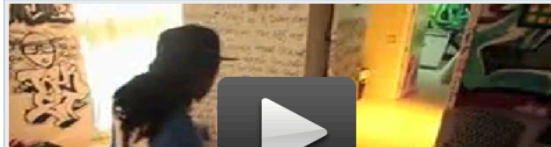
New users [sign up!](#)

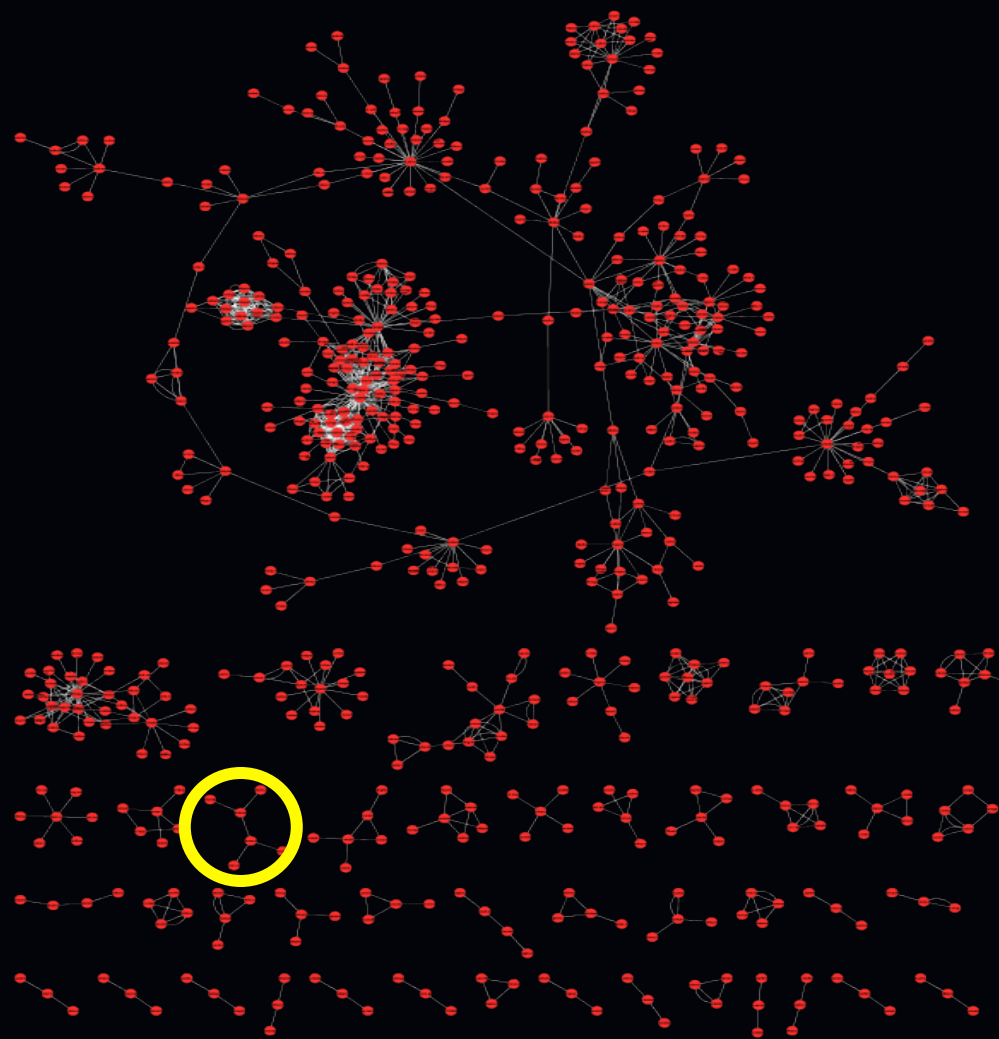
SEARCH SITE CONTENT

UNDERGROUND HITZ

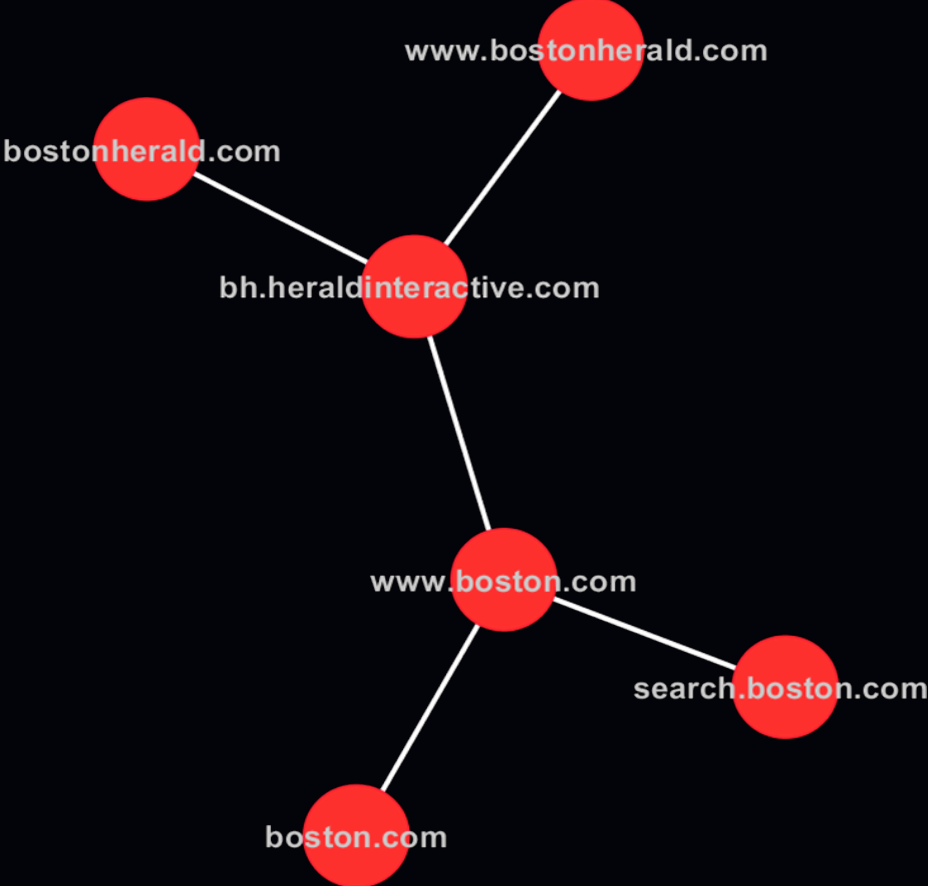
- Roc Marciano - [I Shot The King\(Prod by Roc Marciano\)](#)
 - Mikkey Halsted ft Twista - [Get Low\(Prod. By The Legendary Traxster\)](#)
 - Mistah F.A.B. - [Swimming Pools](#)
 - Chino XL ft. Sick Jacken, Immortal Technique, DV Alias Krist - [Arm Yourself](#)
 - Graft - [Bands A Make Her Dance\(G Mix\)](#)
- [See All UG Hitz](#)

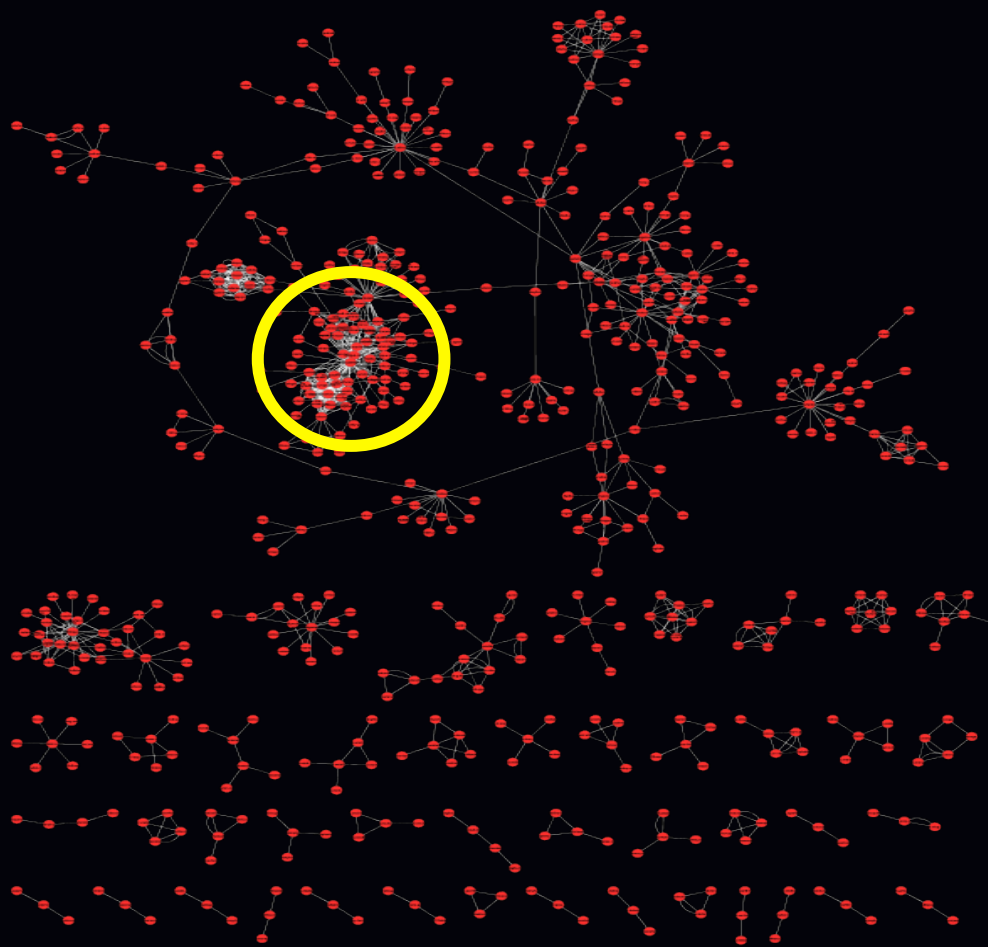
LIL WAYNE-SAME DAMN TUNE(DIRECTED BY DJ SCOOB DOO)

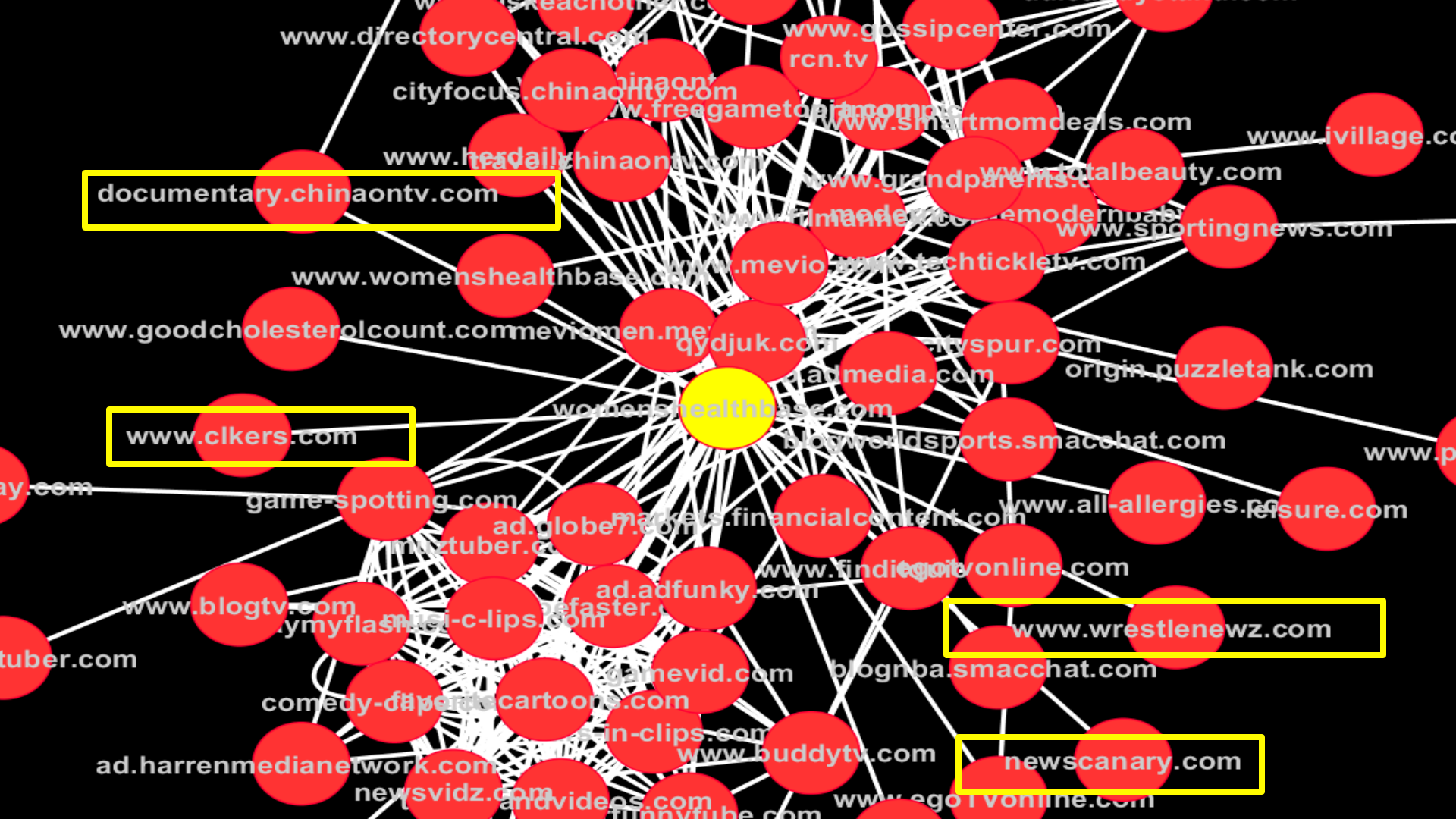


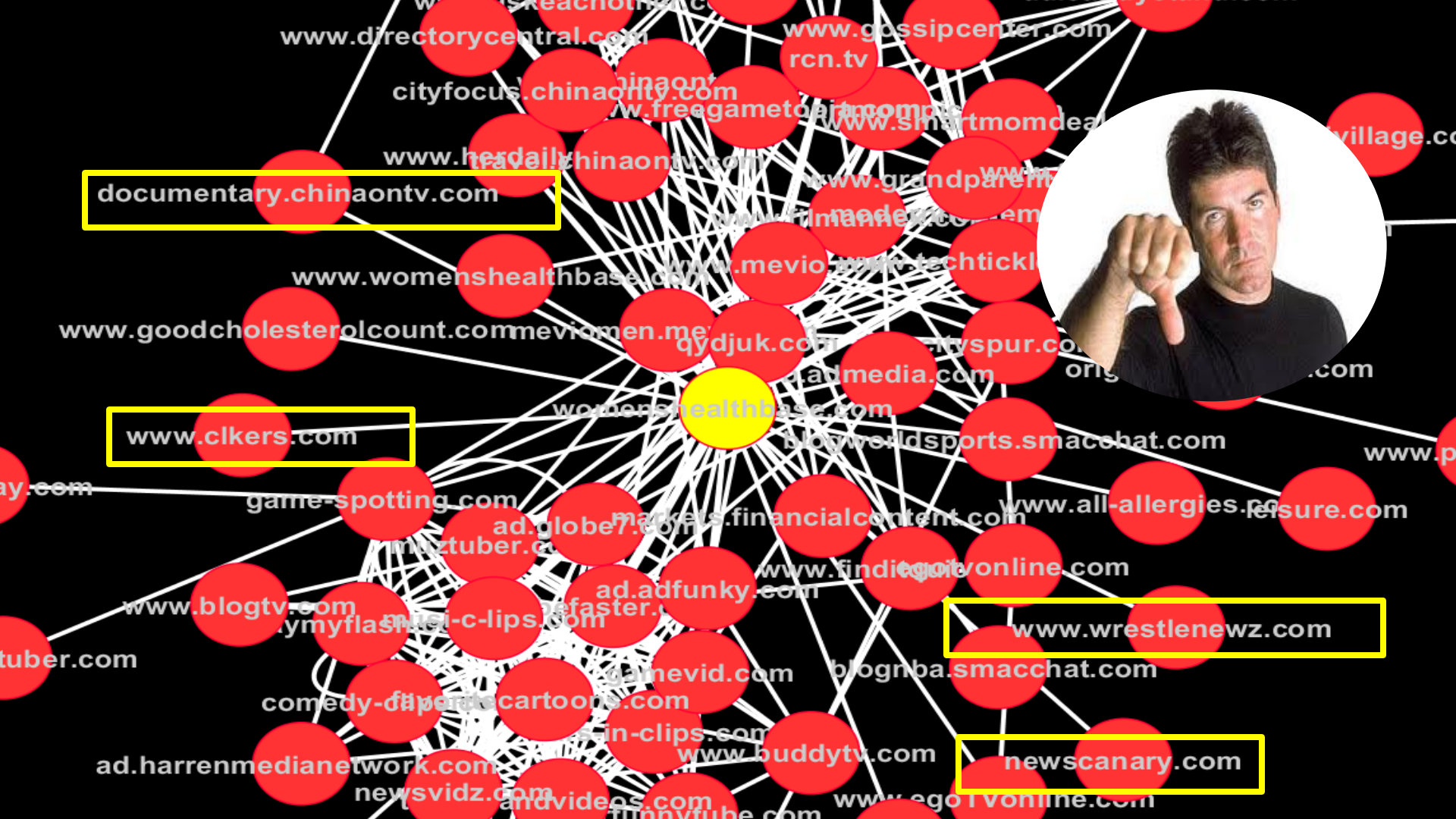


Boston Herald

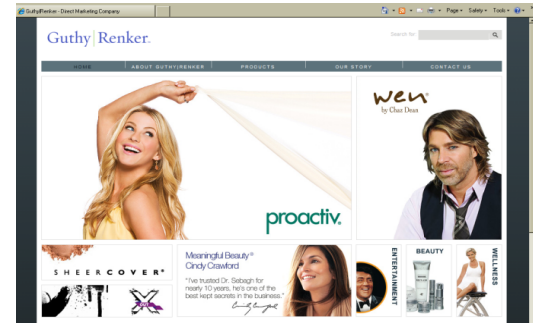
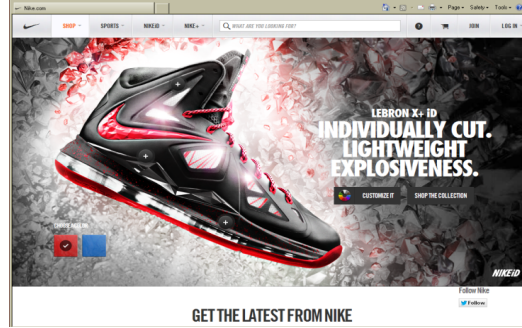
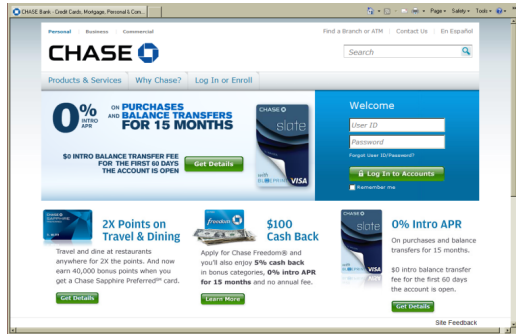






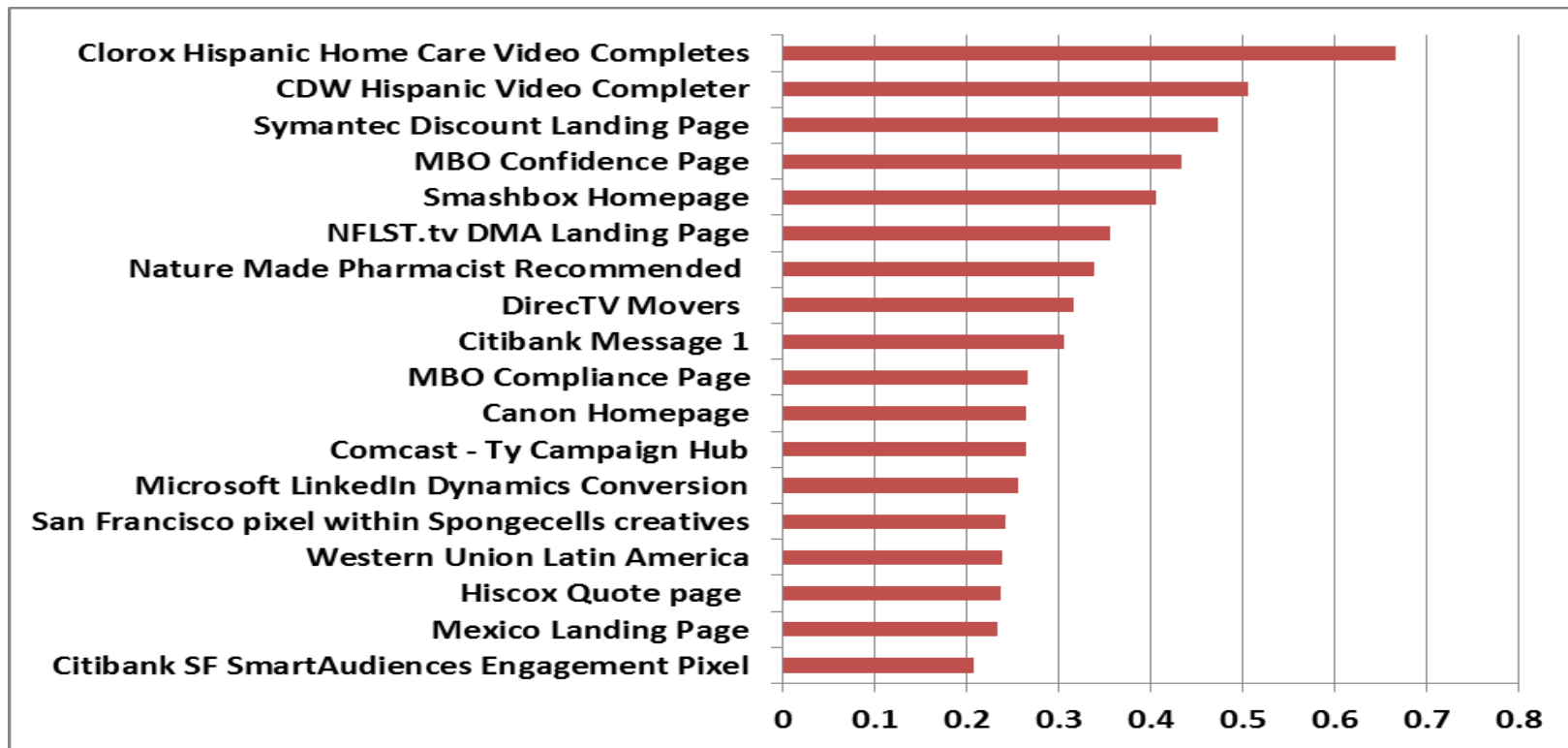


Bots are executing conversion events

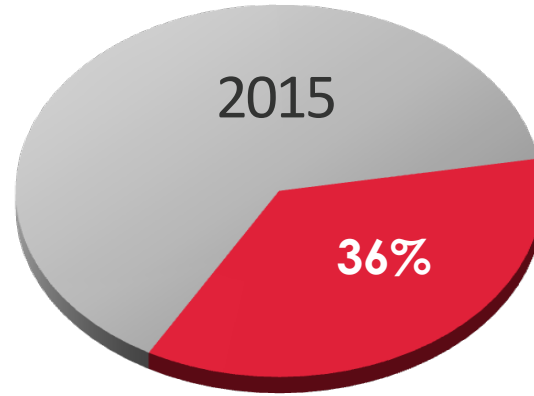


- 'Cookie Stuffing' increases the value of the ad for retargeting
- Messing up Web analytics ...
- Messes up my models because a bot is easier to predict than a human

Percent bot traffic on conversion metrics



Two populations surfing the we: Bot vs. Human

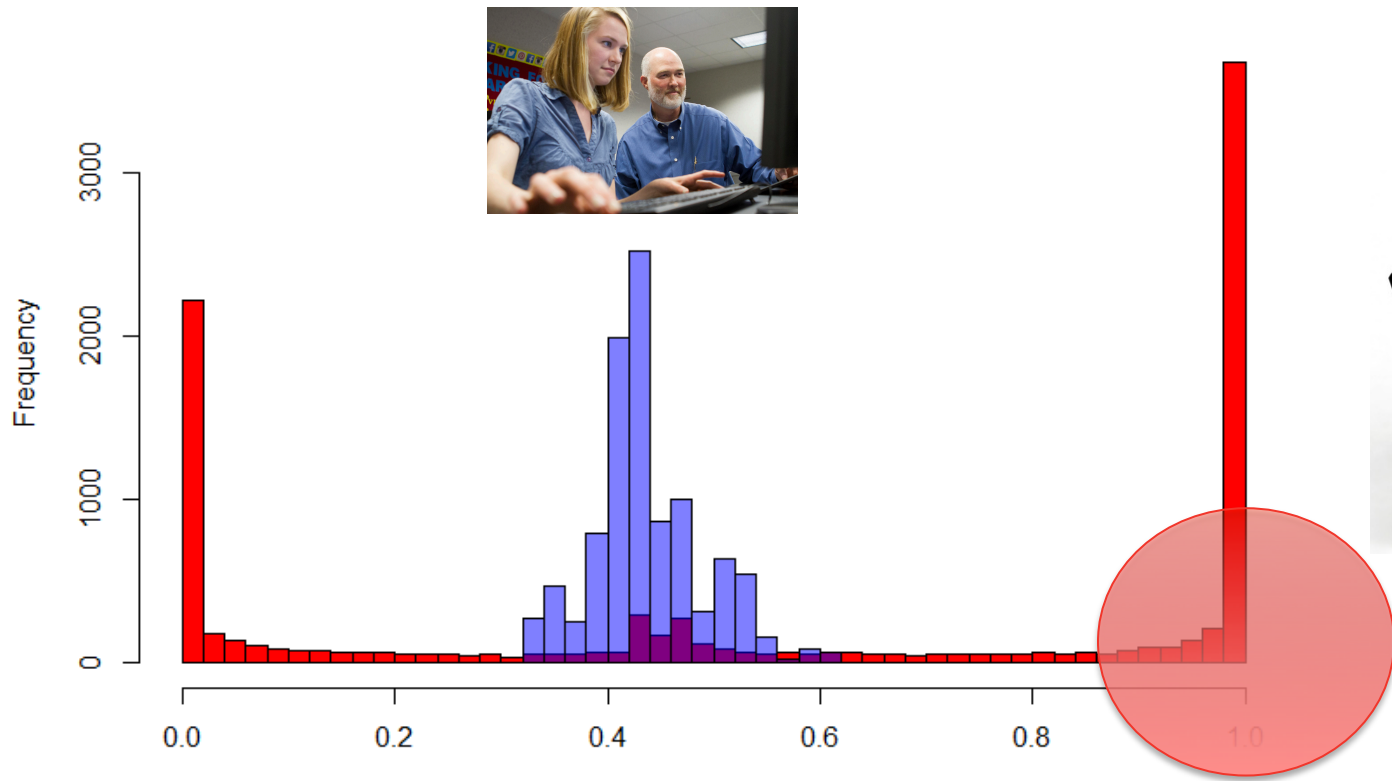




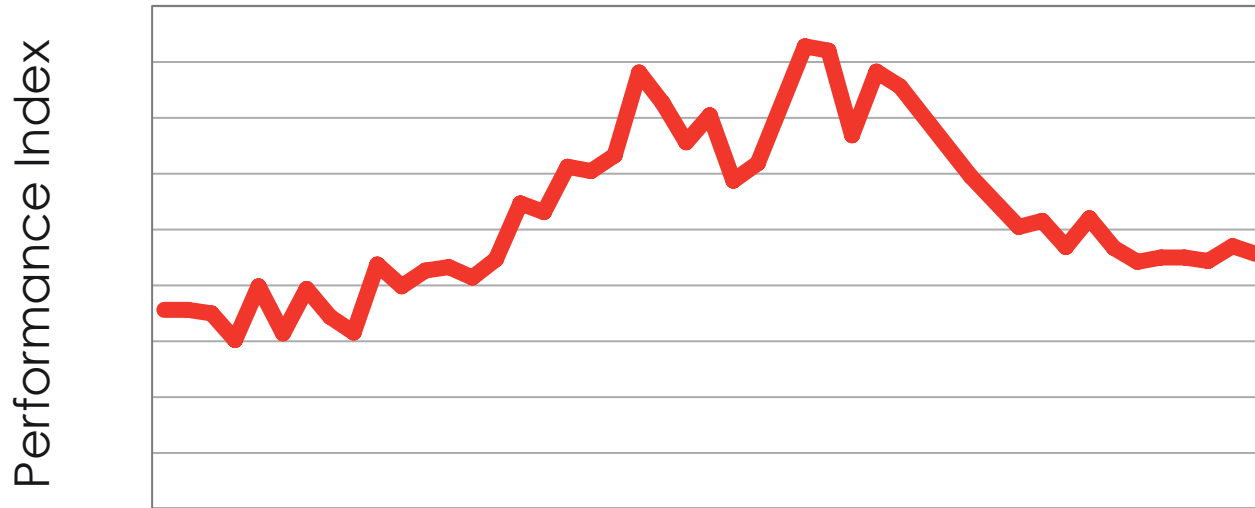
Random

Deterministic

Bot activity has more signal – higher predictions



Eliminate labels generated by bots



3 more weeks in spring 2012

What about Clicks?

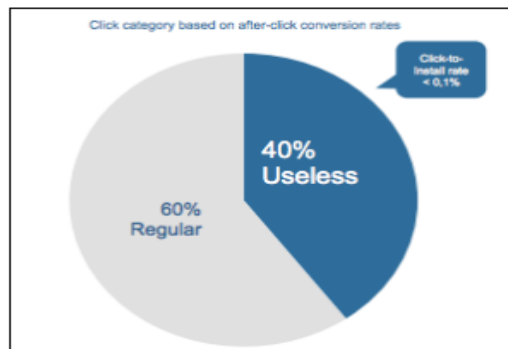
“Measure of consumers interest in the product”



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

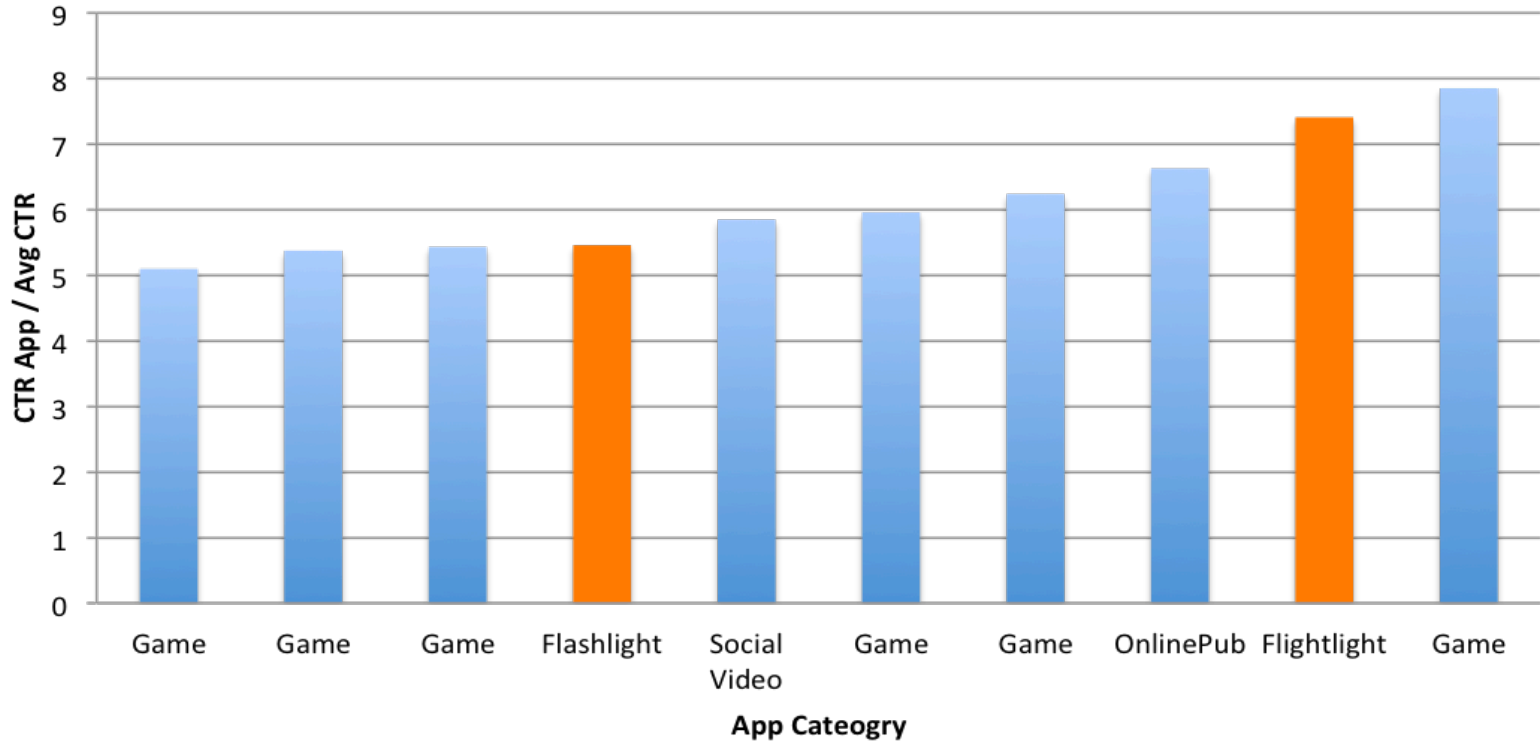
Pamela Parker on September 5, 2012 at 5:42 pm

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.

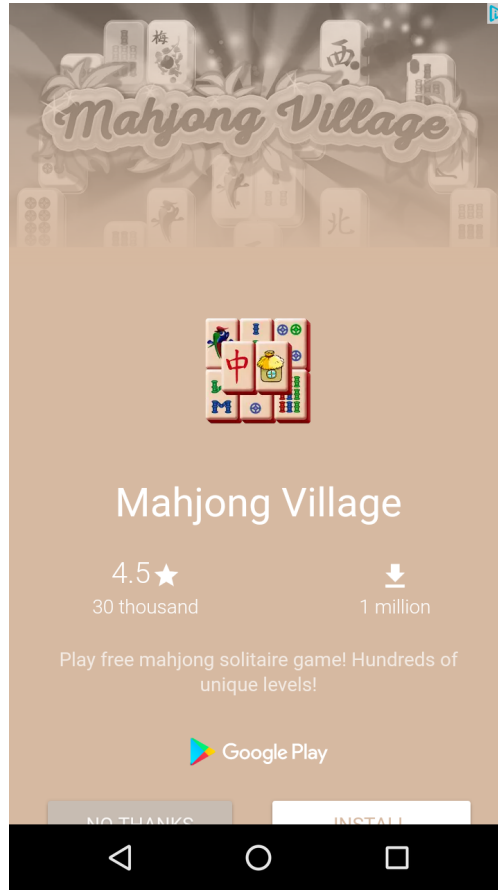


Model learned when people click accidentally: Fumbling in the dark ...

Top 10 Apps by CTR

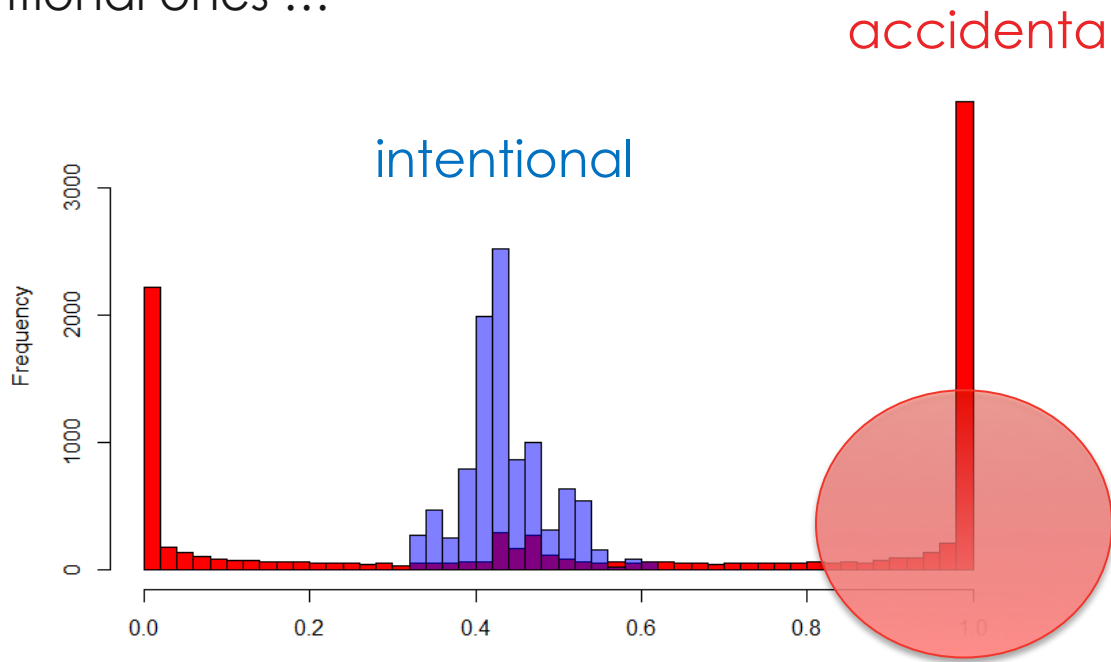


'Optimizing Clicks ...'



Accidental clicks are more predictable than intentional

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

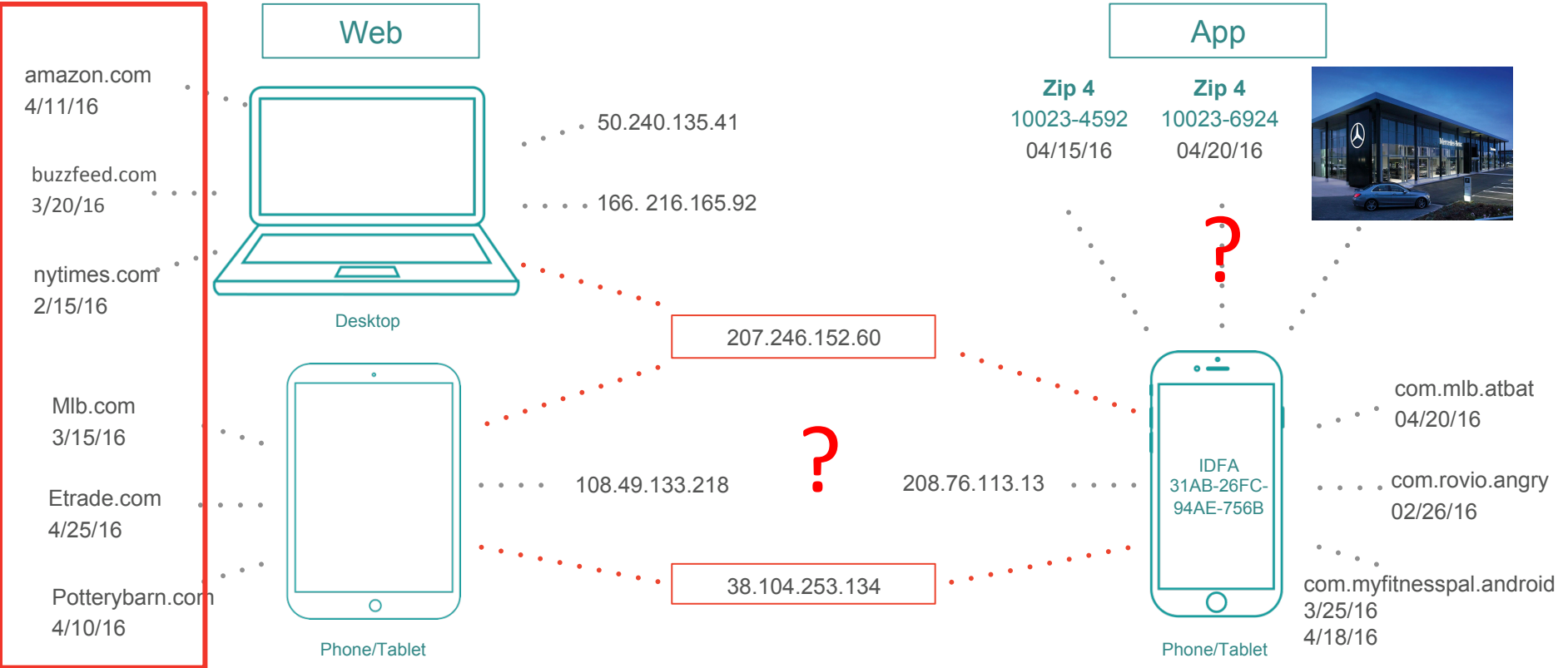


I will only target clicks that are likely to happen accidentally but not randomly ...

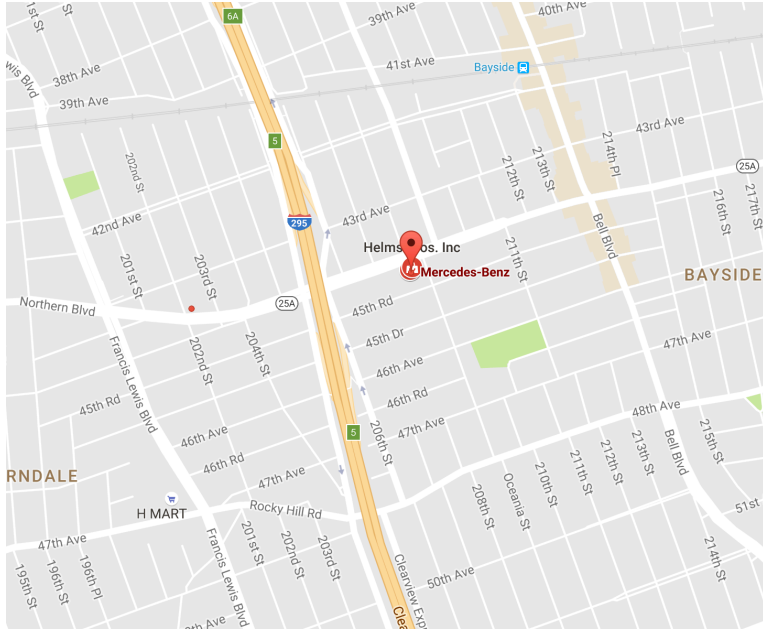
Finding good audiences for luxury cars? Predict dealership visits?



Browsing history of your neighbor who hacked your WIFI?



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

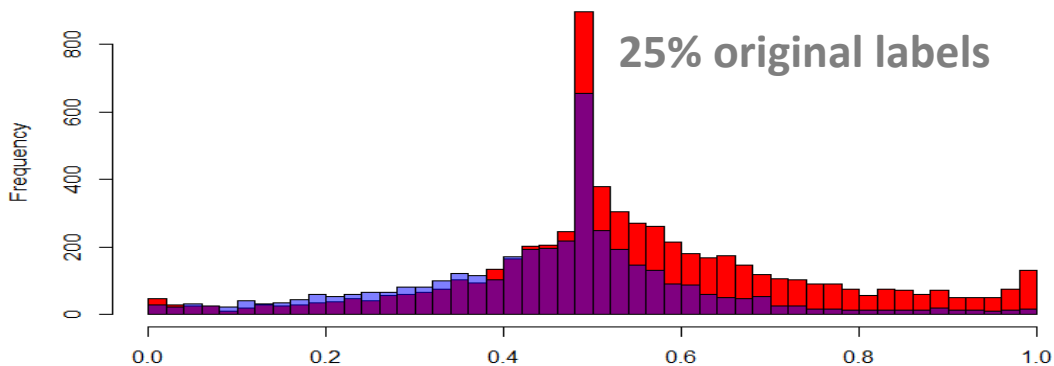
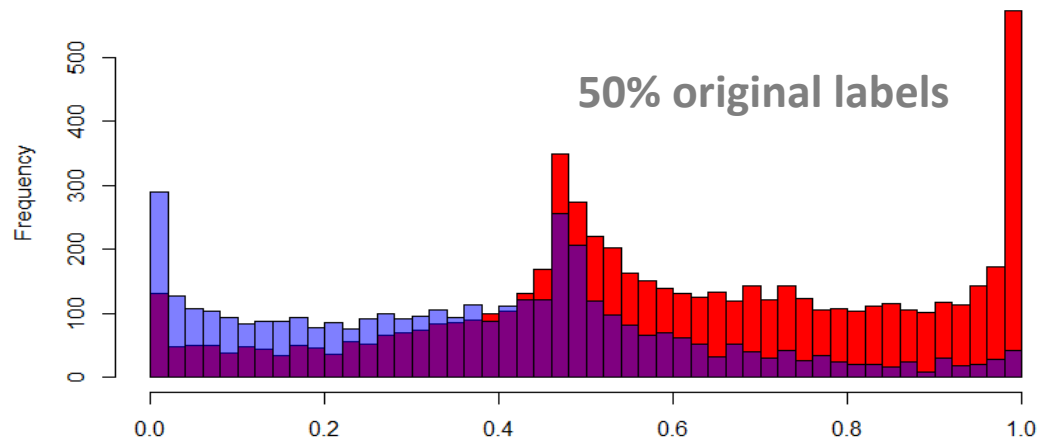
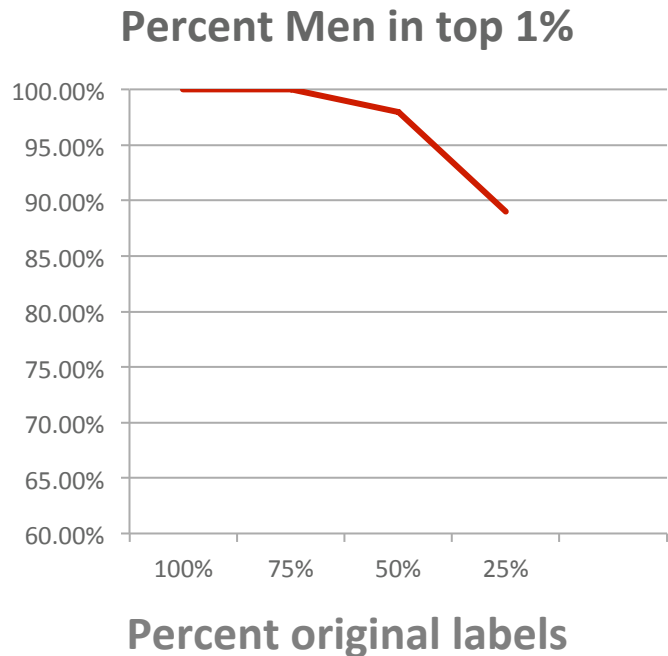
```
0000d41ed774823fca142945ec915c0,1,,,,,,,,,en_GB,,  
0000dee02d70cf8c0d79f96b6d1c59d,0,,,,,,,,,en_US,,  
0000f232abfe25a80156fe069395460,0,1992,20,2,2,,,,,,,,,19,-5  
0000f4ba0cff946b1c0e3b051287ede,0,1993,19,2,,,,,,,,,en_US,310,8  
0000130571654e3afaa62f4e9d2e4f63,0,,,2,2,,,,,,,,,en_US,193,7  
00001544469ae9b408869a463a1dd77a,1,1984,28,2,,,,,,,,,en_US,,-4
```



```
0000d41ed774823fca142945ec915c0,0,,,,,,,,,en_GB,,  
0000dee02d70cf8c0d79f96b6d1c59d,0,,,,,,,,,en_US,,  
0000f232abfe25a80156fe069395460,1,1992,20,2,2,,,,,,,,,19,-5  
0000f4ba0cff946b1c0e3b051287ede,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



Identify people who will go to Mercedes dealerships?

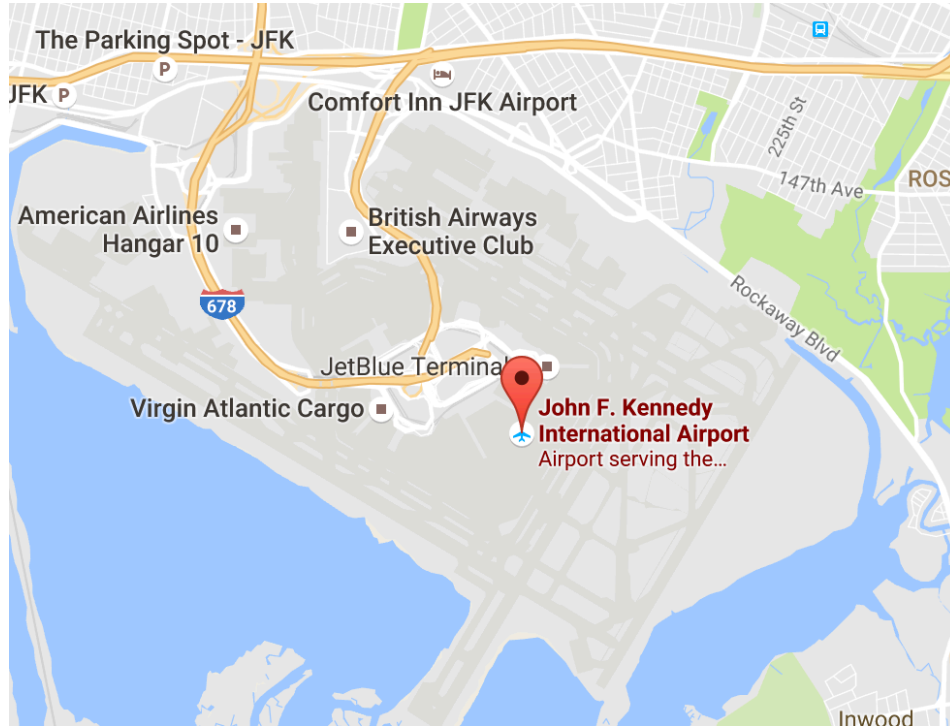


My predictions might be better than my 'ground truth'

Where do we find frequent traveler?



What do you think indicates people going to JFK?



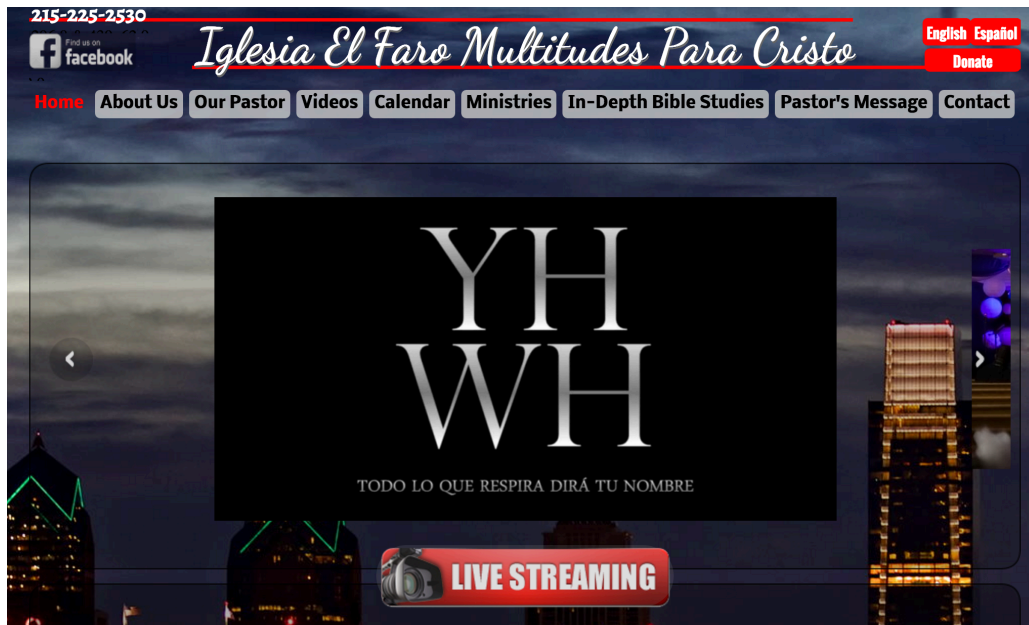
URL

www.iglesiaelfaroinc.org
www.jumpseatnews.com
www.bluelineprop.com
www.ktxdtv.com
www.southjefffootball.org
www.unitedafa.org
www.parliamenthouse.com
www.yunghova.com
www.interlinetravel.com
www.aclin.org
www.swissport.com
www.gcsanc.com
www.swacu.org
www.airlinepilotcentral.com
www.homotrophy.com
www.beggfuneralhome.net
www.tvathletics.org
www.2shopper.com
www.nextmagazine.com
www.dailyjocks.com
www.pullzone.com
www.diamondoffshore.com
www.myerspolaris.com
www.ryandeyer.com
www.okllo.com
www.ifihadtochoose.com
www.ivoirmixdj.com

Logistic Parameter

2.38
2.25
2.21
2.14
2.1
2.09
2.07
2.06
2.03
2.03
2.03
2.01
2.01
1.97
1.97
1.94
1.92
1.91
1.91
1.87
1.87
1.86
1.86
1.86
1.84
1.83
1.83

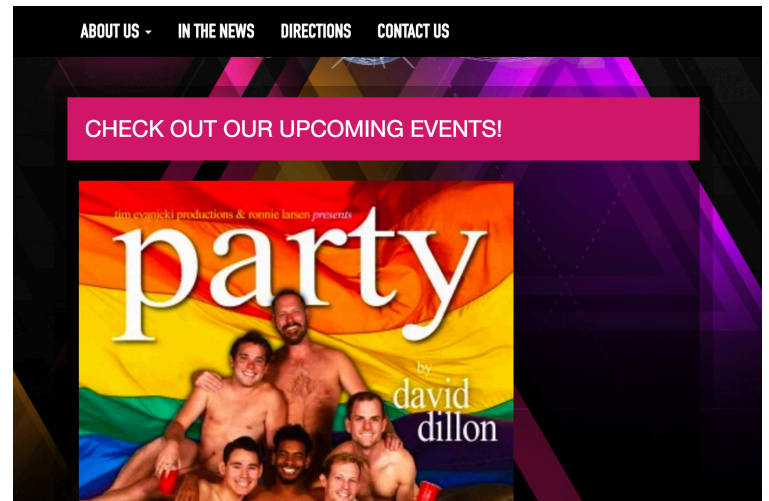
URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggfuneralhome.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



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.beggfuneralhome.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

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.beggfuneralhome.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

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.beggfuneralhome.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



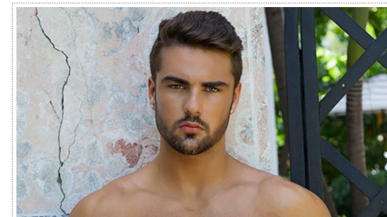
RECENT ARTICLES

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](#)
[StumbleUpon](#)
[Pin it](#)



SUBSCRIBE

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

TROPHY : GET US

[Become a FAN](#)
[Twitter](#)

TROPHY::LISTS

Predict who goes to JFK?



People who
work there ...



Random

Deterministic



Big Picture?

Thank You!

Claudia Perlich

Chief Scientist

@claudia_perlich

dstillery