

# Predictability and other Predicaments in Machine Learning

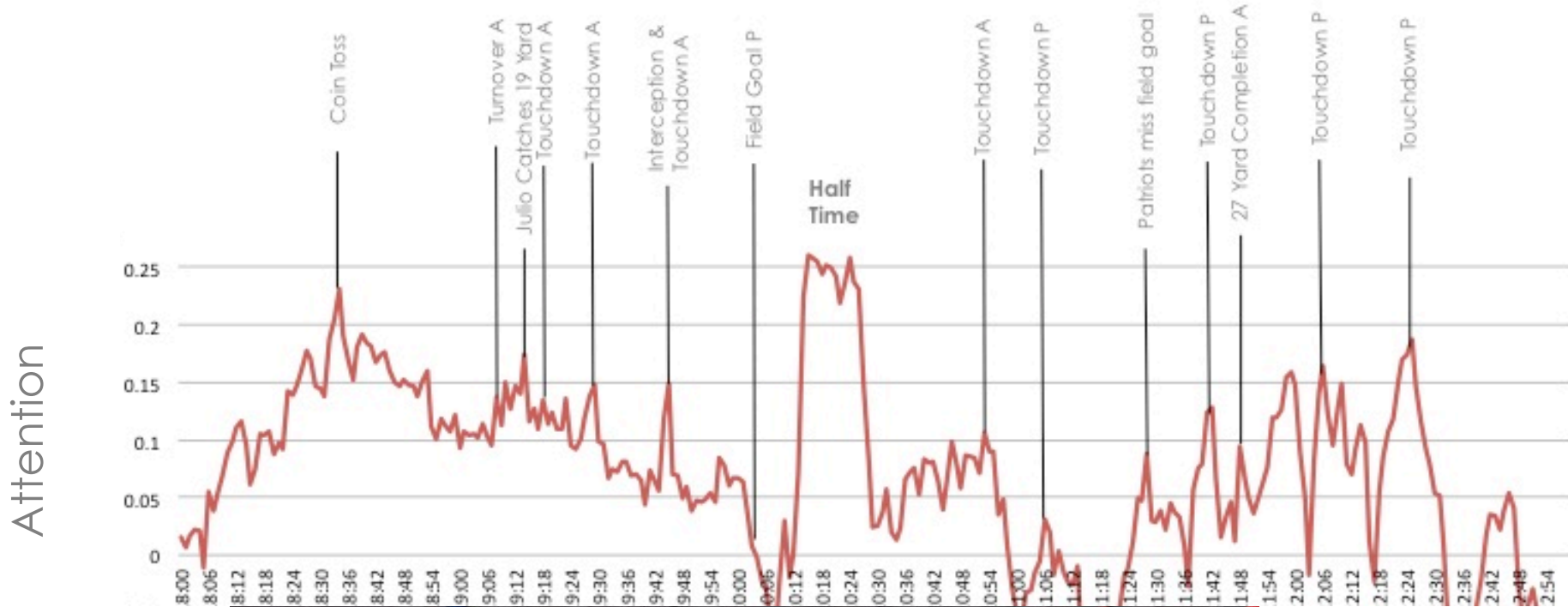
Claudia Perlich

Chief Scientist

@claudia\_perlich

dstillery

# The other day during the Super Bowl: What did people pay most attention to?



# Programmatic Advertising

The screenshot shows the homepage of The New York Times on Tuesday, July 17, 2012. Several programmatic advertisements are highlighted with red boxes:

- Top Left:** "THE ALL-NEW CADILLAC XTS" logo.
- Top Right:** A small image of a black Cadillac XTS.
- Search Bar:** The "ING DIRECT" logo is displayed in the search bar.
- Bottom Right:** A large video player for "INTRODUCING THE ALL-NEW CADILLAC XTS" with a "CLICK FOR SOUND" button and a "LEARN MORE" link.

The main content area features several articles:

- Fiscal Crisis in States Will Last Beyond Slump, Report Warns** by MARY WILLIAMS WALSH and MICHAEL COOPER. Summary: Long after the economy rebounds, states will face financial problems that include rising health care costs and underfunded pensions, a task force of budget experts said.
- Unity Government in Israel Disbanding Over Draft Dispute** by JODI RUDOREN and RICK GLADSTONE. Summary: The head of the centrist Kadima Party said it had "no choice" but to pull out because of differences over a proposed universal draft.
- Cautious on Growth, Bernanke Offers No Hint of New Action** by BINYAMIN APPELBAUM. Summary: Ben S. Bernanke, the
- TimesCast** video player featuring Evan Vuori.
- OPINION** section with articles like "Islamic School for Girls" and "Brooks: Capitalism Debate".
- MARKETS** section showing S.&P. 500, Dow, and Nasdaq indices.
- GET QUOTES** section with "My Portfolios" and "Go" buttons.
- CAMPAIGN 2012** section with articles like "Mormons' First Families Rally Behind Romney" and "Inside the Romney No. 2 Hunt: First, Do No Harm".

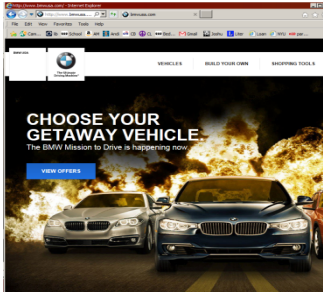
# 300 Million consumer



Interacting with one of our customer brand sites

Using Digital Devices

**conversion**

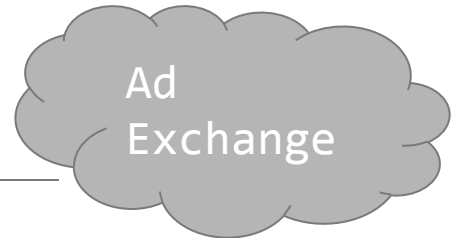


If we win an auction we serve an ad

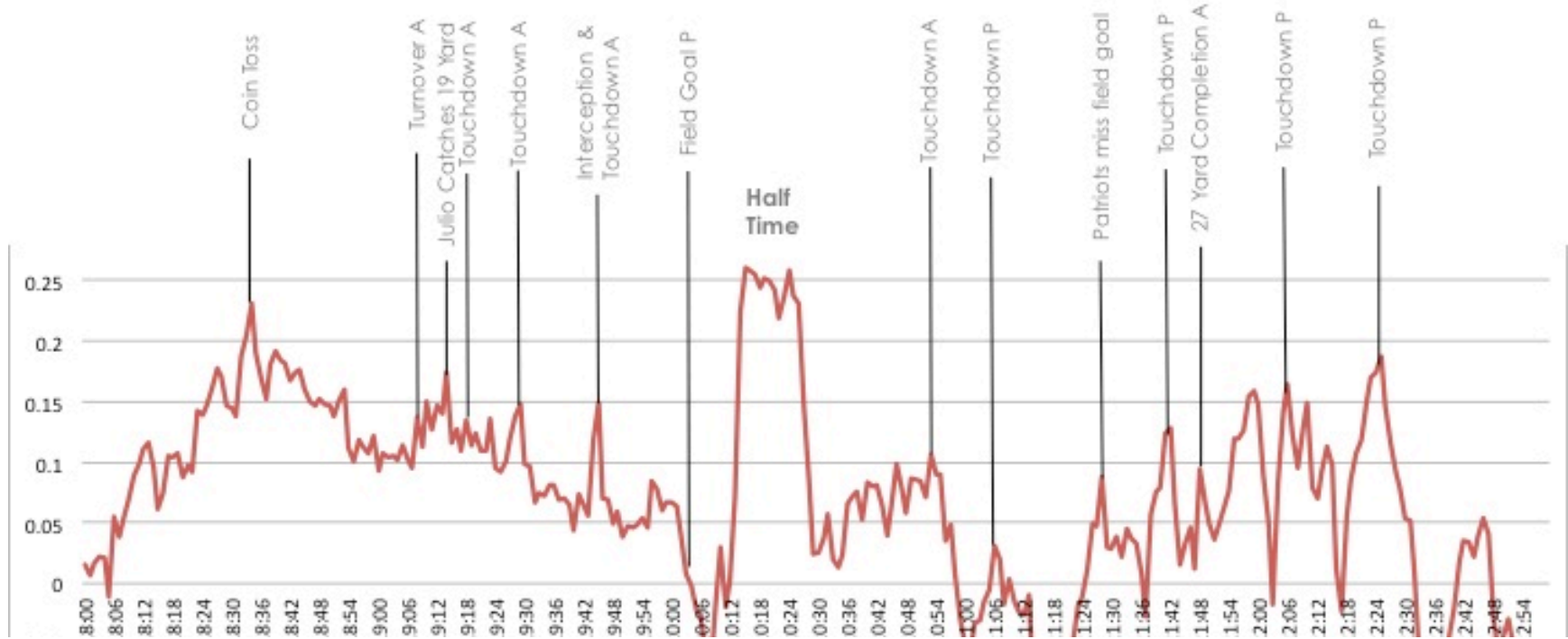
**100 ms response time**



**100 Billion bid requests per day**



# The other day during the Super Bowl: What did people pay most attention to?



# To whom should we show an AZURE ad?



# People who clicks on the AZURE ad?



# Visit the AZURE homepage after an ad?

← → ↻ 🏠 **Microsoft Corporation [US]** https://azure.microsoft.com/en-us/?cdn=disable ☆ ☰

📁 Apps 📁 OLD ★ Bookmarks 📧 Mail 📅 Cal 🦊 Dash 🐞 Jira 📄 Open 📄 open2 🐦 Wiki 🦋 Cellar 🦋 Top Shelf 📷 Pulse 🌿 expense >>

Microsoft Azure SALES 1-800-867-1389 ▼ MY ACCOUNT PORTAL Search 🔍

Why Azure Products Documentation Pricing Partners Blog Resources Support **FREE TRIAL** >

# The cloud for modern business

Move faster  
Save money  
Integrate on-premises apps and data

**Why Azure?** >

**57% of Fortune 500 companies already use Azure.**  
Hear their stories ▶

▣ ▣ ▣ ▣



# Sign up and use AZURE after an ad?

Sign up and deploy your first cloud solution in under 5 minutes

Try for free >

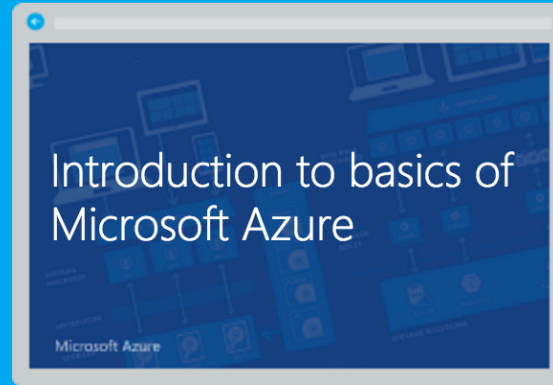
## Get Started



Watch three minute videos that teach you how to quickly get started with Azure.

[Get started ▶](#)

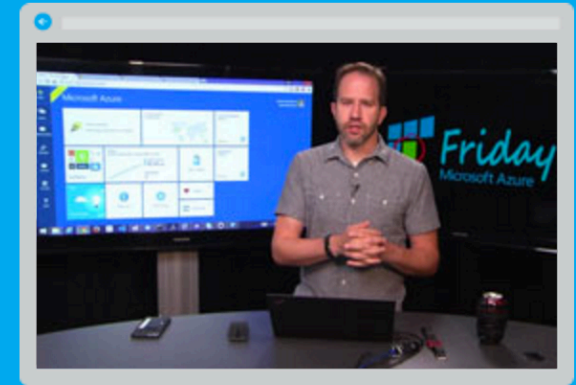
## Free Webinars



Sign up and watch live online presentations about the latest features in Azure.

[Sign up and attend ▶](#)

## Azure Friday



Watch the weekly 10 minute video series with Scott Hanselman.

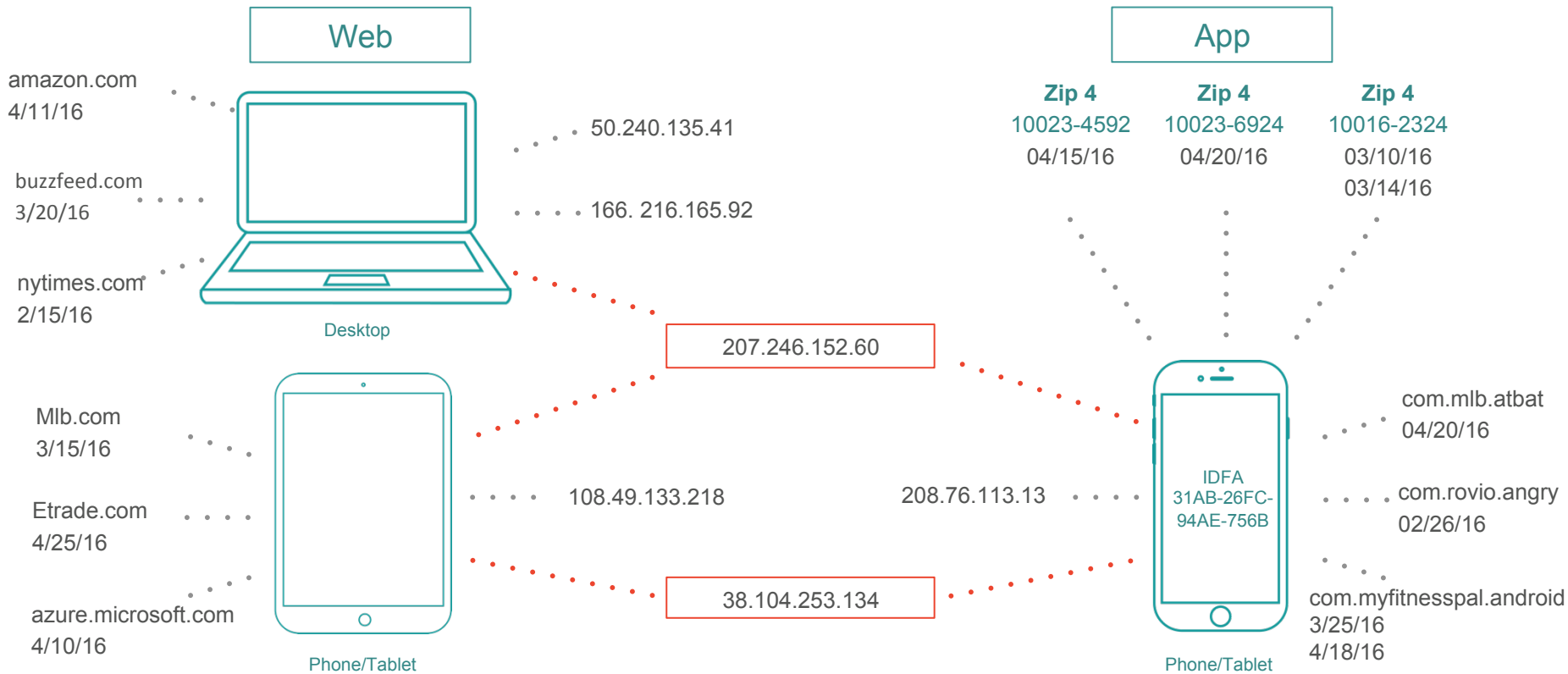
[Watch this week's show ▶](#)

## Who should I (Claudia) target?

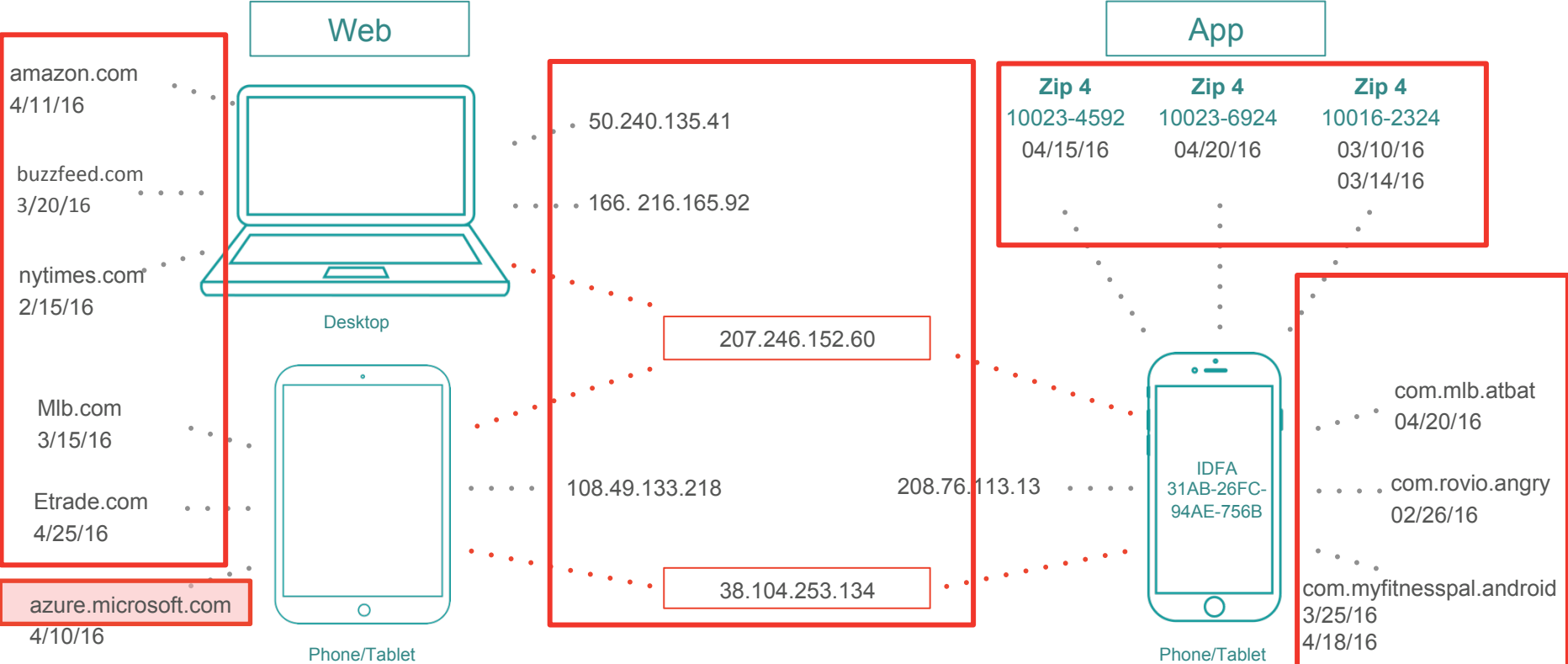


I will try to look good in your chosen metric

# Dstillery Data: 100 Billion events per day across devices



# Predicting events on pretty much everything





# Dstillery Automated Predictive Modeling

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

Margin of yesterday (20160306)

Revenue of yesterday (20160306)

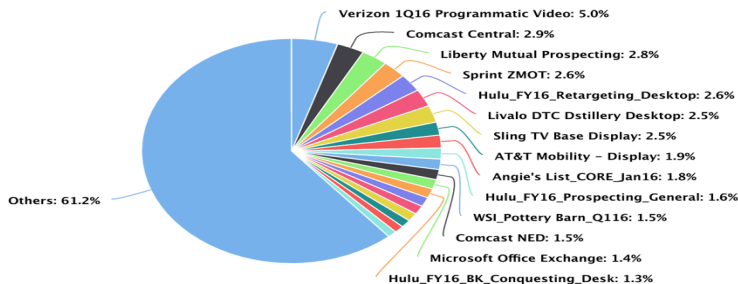
61M

Number of impressions yesterday (20160306)

## 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



# 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://hns2/user/sgd/models/19945/OL19945_FM2_BS20160306123636.model

Logistic Regression  
 Stochastic gradient descent  
 Hashing  
 Streaming  
 L1 & L2 Penalties

## Top Features with Highest Scores

Show 10 entries

Search:

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

Showing 1 to 10 of 30 entries

Previous **1** 2 3 Next

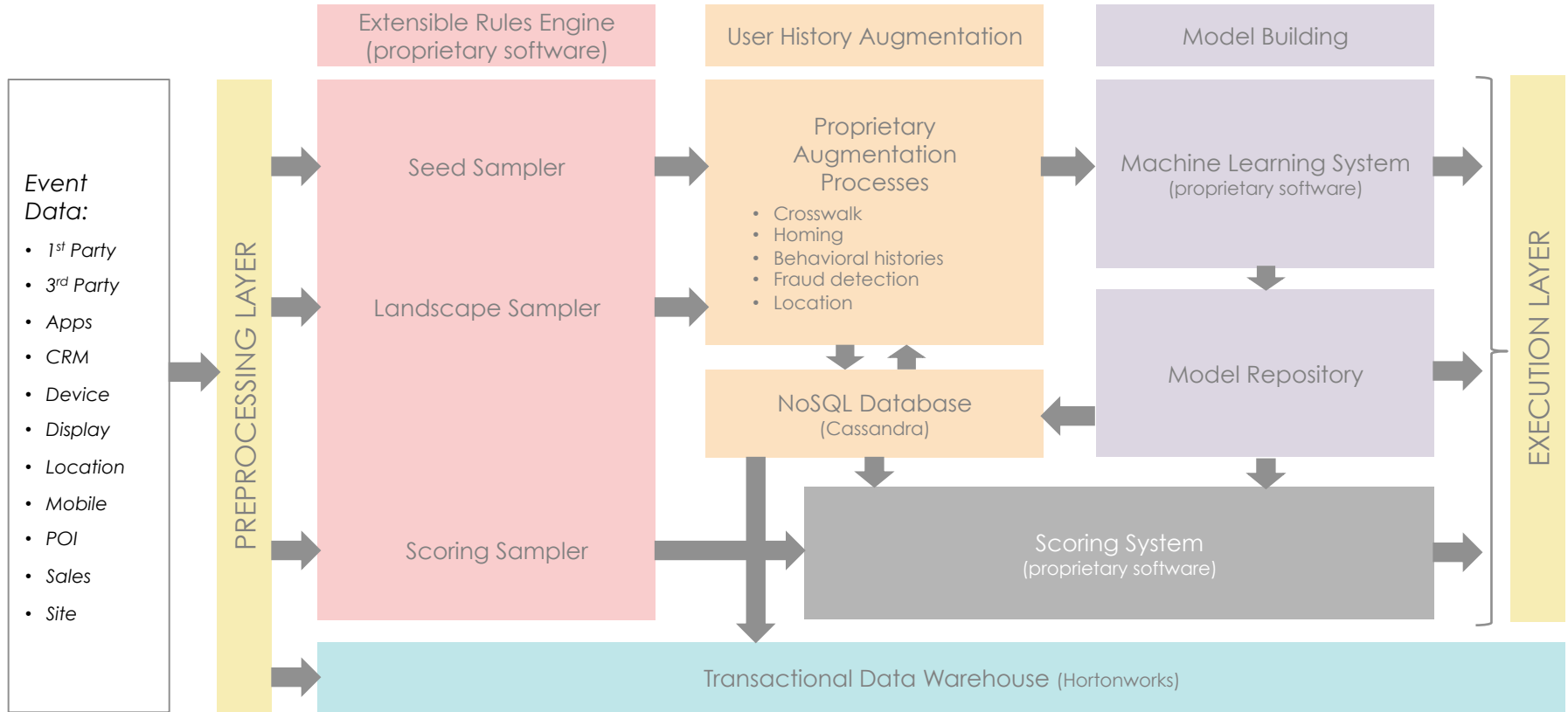
## Top Features with Lowest Scores

Show 10 entries

Search:

Feature	Coef
<a href="#">www.coolrunning.com</a> [46417350]	-4.078
<a href="#">www.intraxiz.com</a> <a href="#">www.mirabaiceiba.com</a> <a href="#">www.cdphp.com</a> [46491135]	-2.165

# Technology Profile of an Automated ML Infrastructure



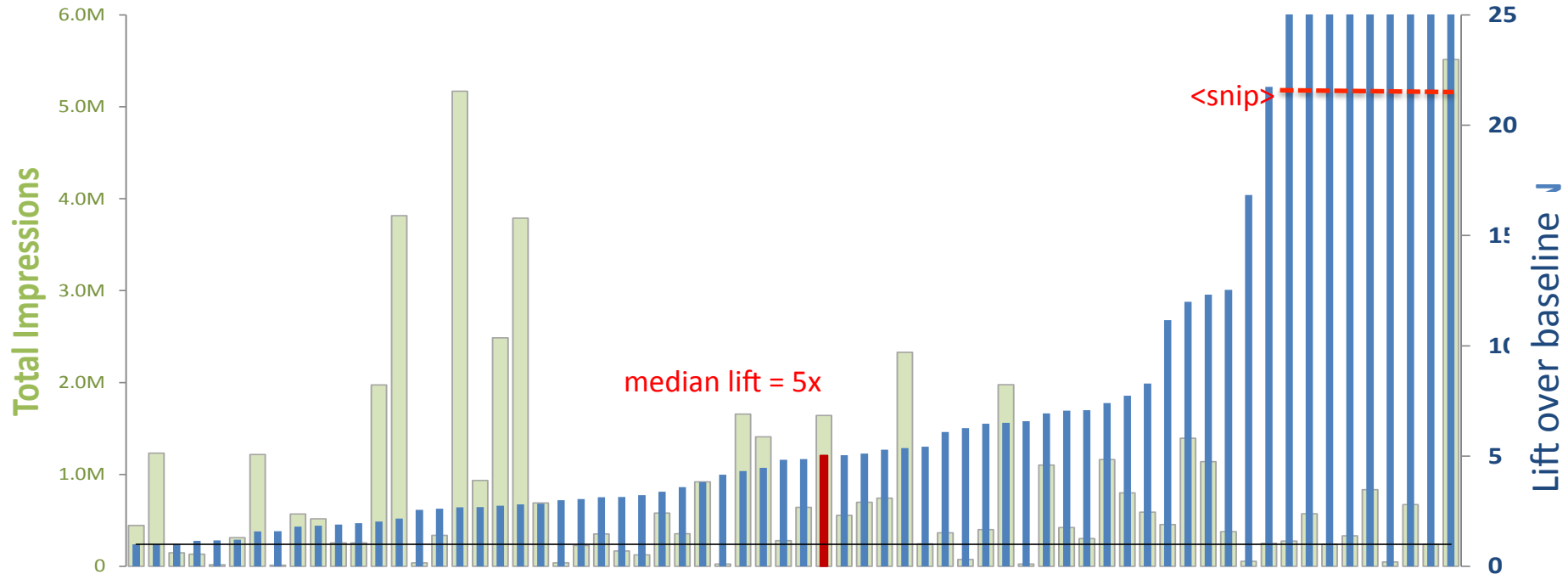




id	112
desc	SGD RON Sampler
process	init-request;populate-cookies-from-req;refresh-ssc(adh,vstcnt,lbh,sglst,browser_prop);feature-map(2);score(display,5:6:11:14:17:18,false,-1,false,true);cfp-to-kafka(cfpBaseline)
mode	ON
destination	
precondition	none
sampler_interval	86400
spec	{"isQ":"return (event.eventType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



# Lift over random for online display ad targeting



Beware what you ask for ...

# 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

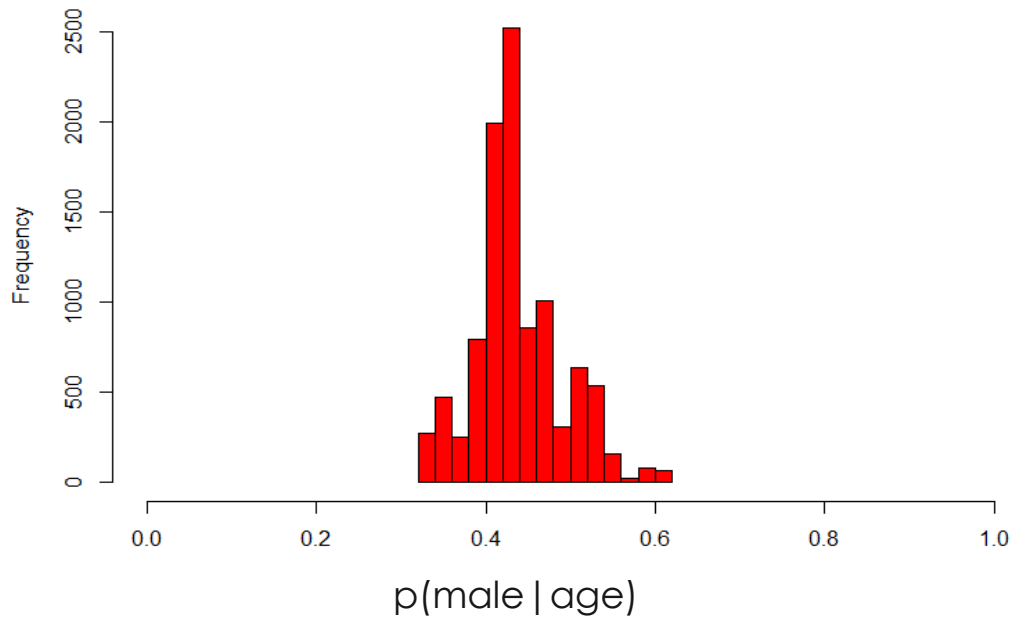
```
00001544469ae9b408869a463a1dd77a 100198443380917
00001544469ae9b408869a463a1dd77a 100248613695
00001544469ae9b408869a463a1dd77a 10050726267
00001544469ae9b408869a463a1dd77a 101021248409
00001544469ae9b408869a463a1dd77a 101054236602446
00001544469ae9b408869a463a1dd77a 101425333232551
00001544469ae9b408869a463a1dd77a 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 ...





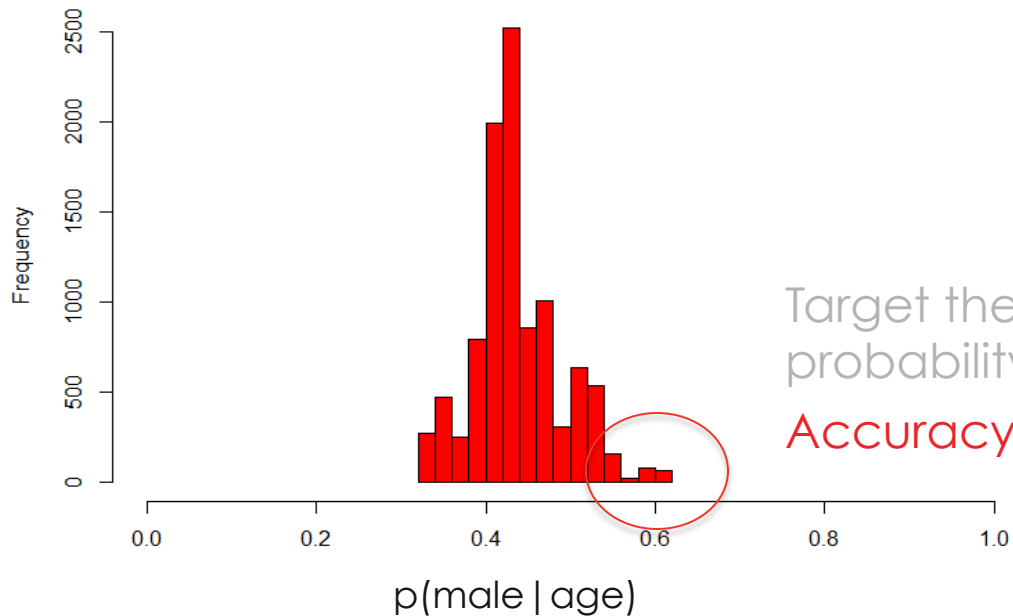
# Gender based on age: very little signal ...



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%

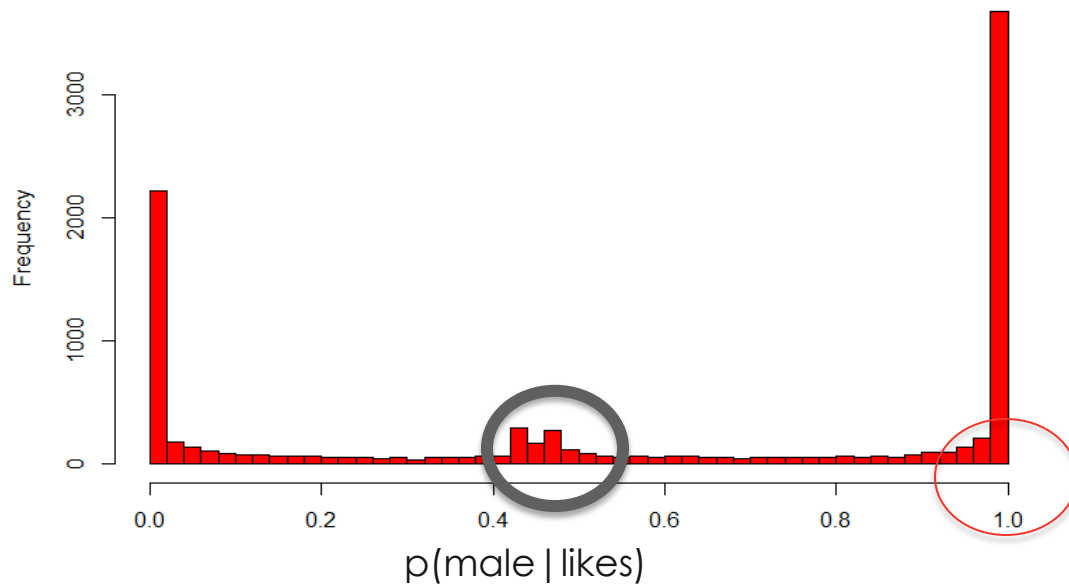
From small to bigger data ...



## Take 2: Gender based all your likes



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

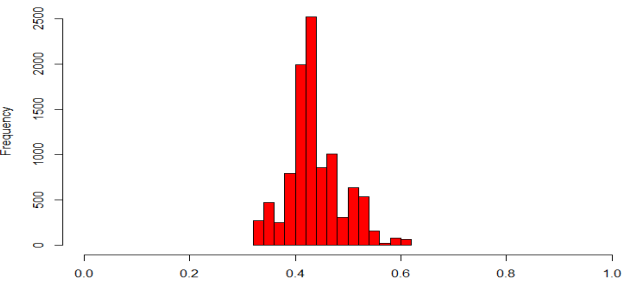


Target top 1%  
Accuracy: 100%

Overall Accuracy: 83%

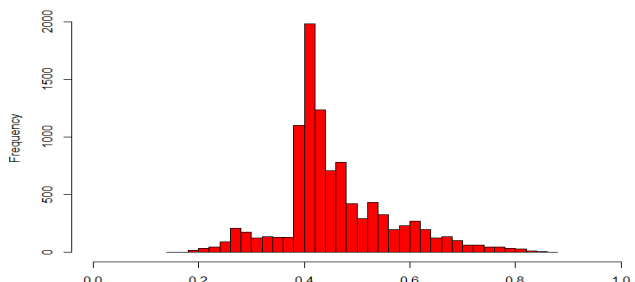
# Progression: from age to all 'likes'

Age



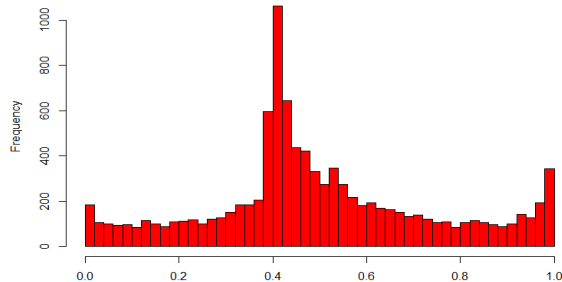
75%

10



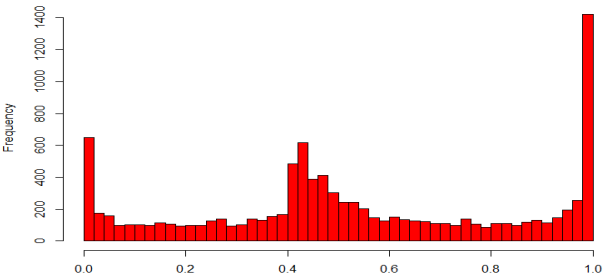
86%

100



100%

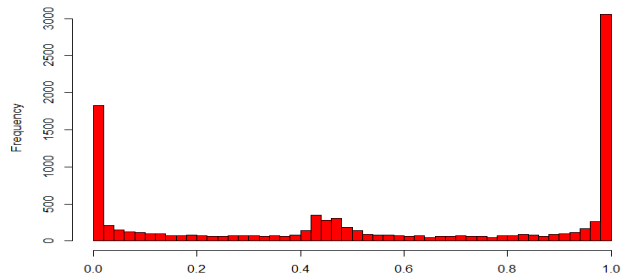
1000



100%

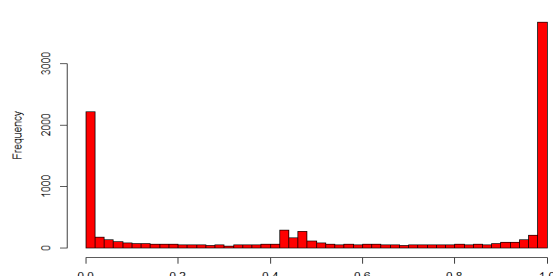
30

10000



100%

100000



100%  
dstillery

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



Random

Deterministic

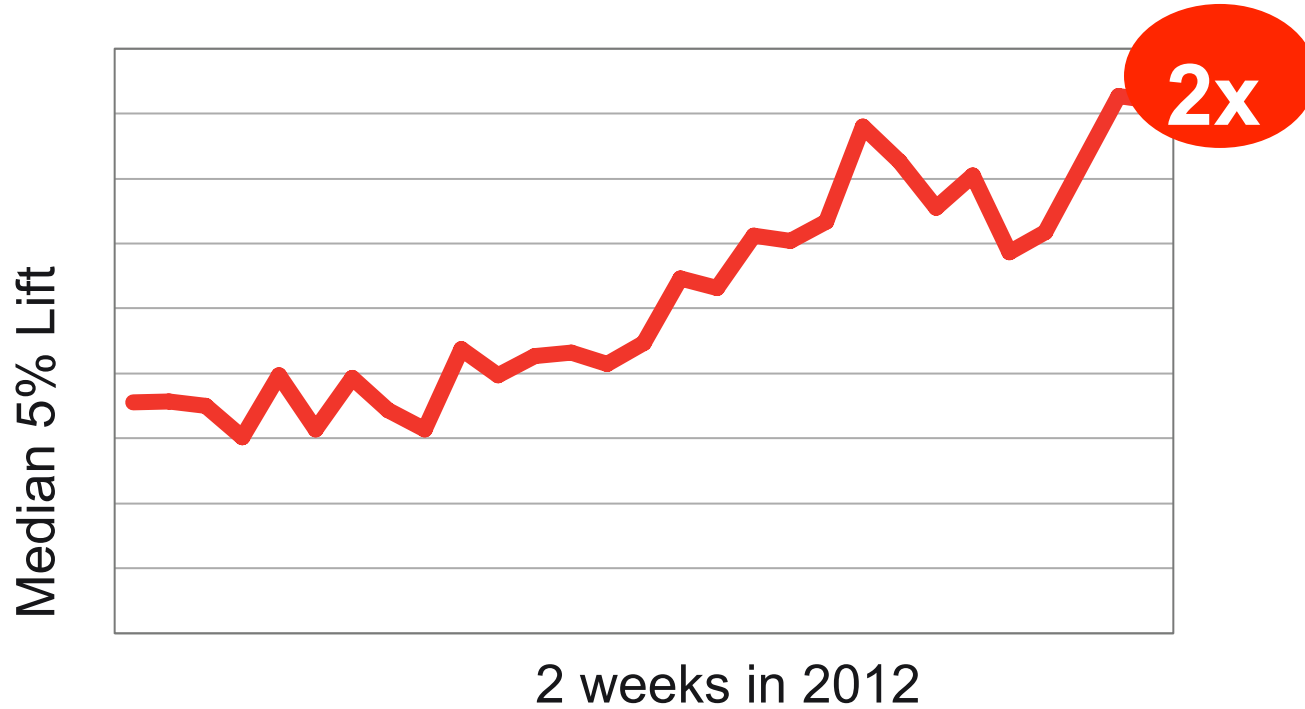
“Models tend to go where the signal is”



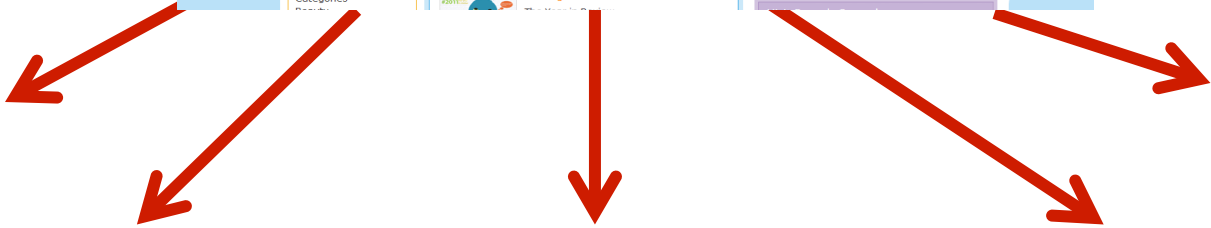
Certainty Vortex



# Witness a spike in human predictability ..



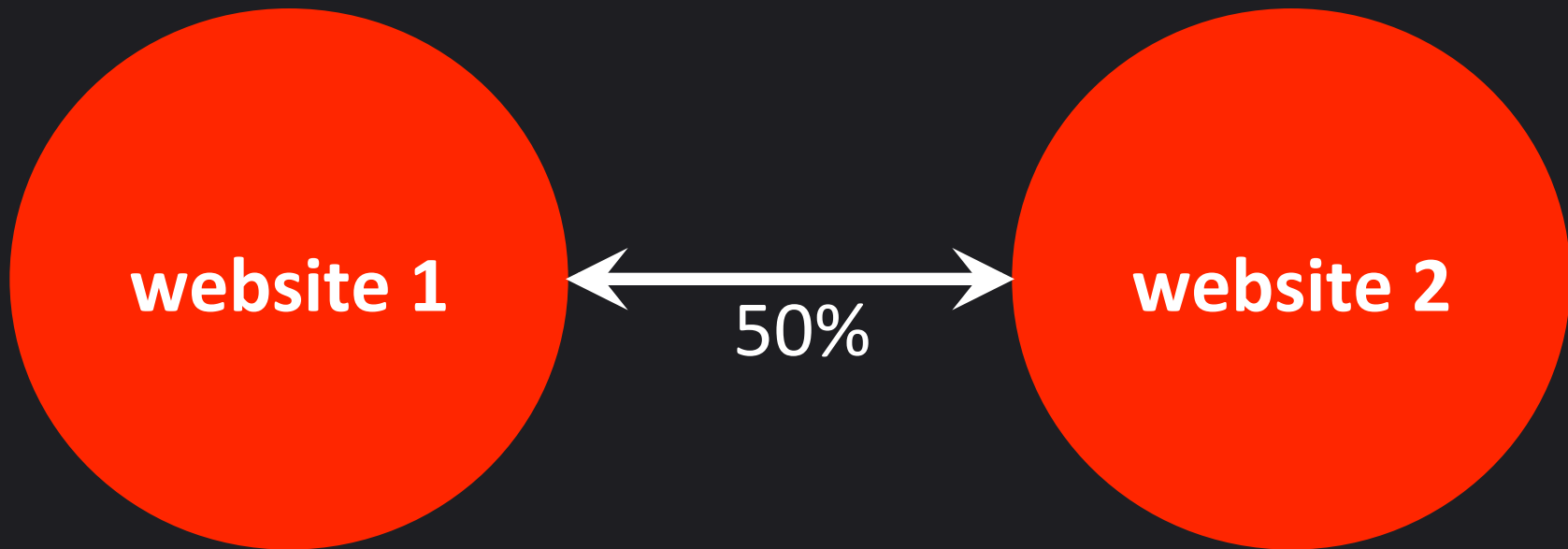
Death of  
free will?

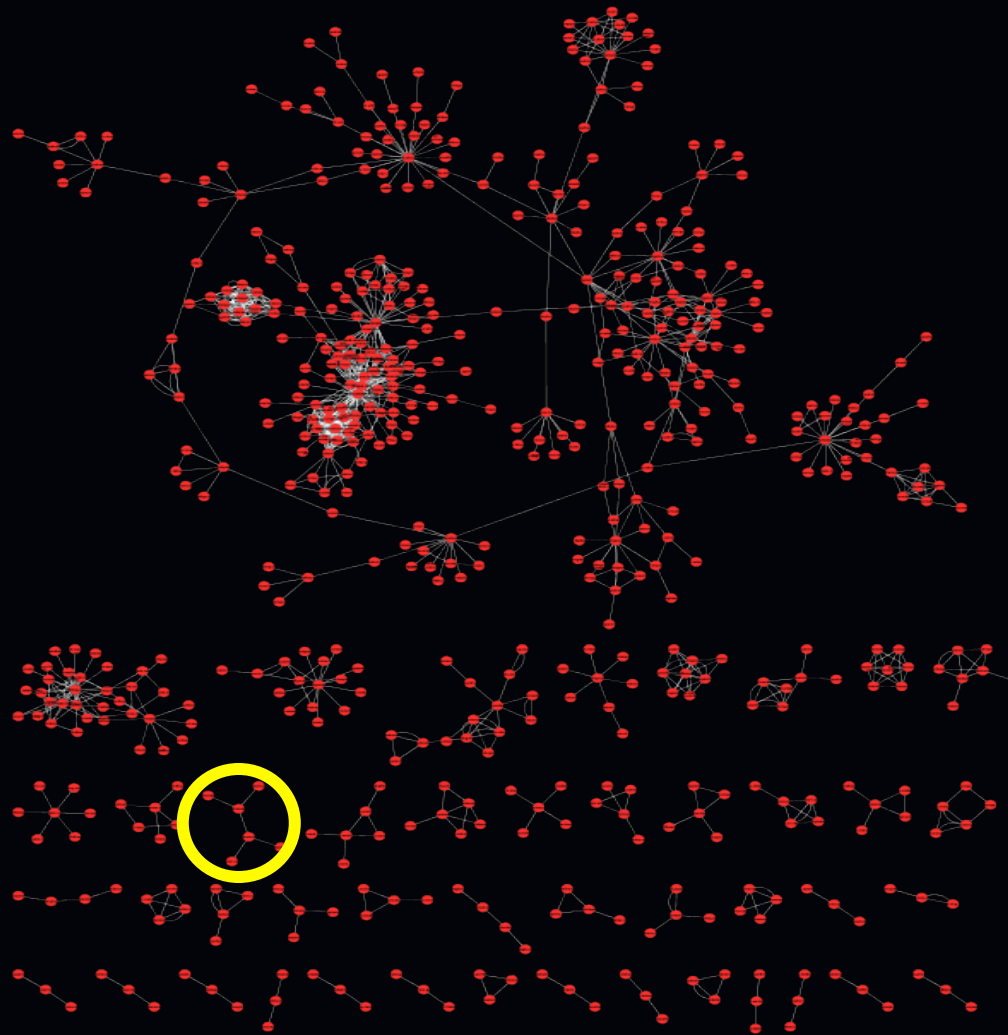


# URL's that are very predictive for more than 10 brands

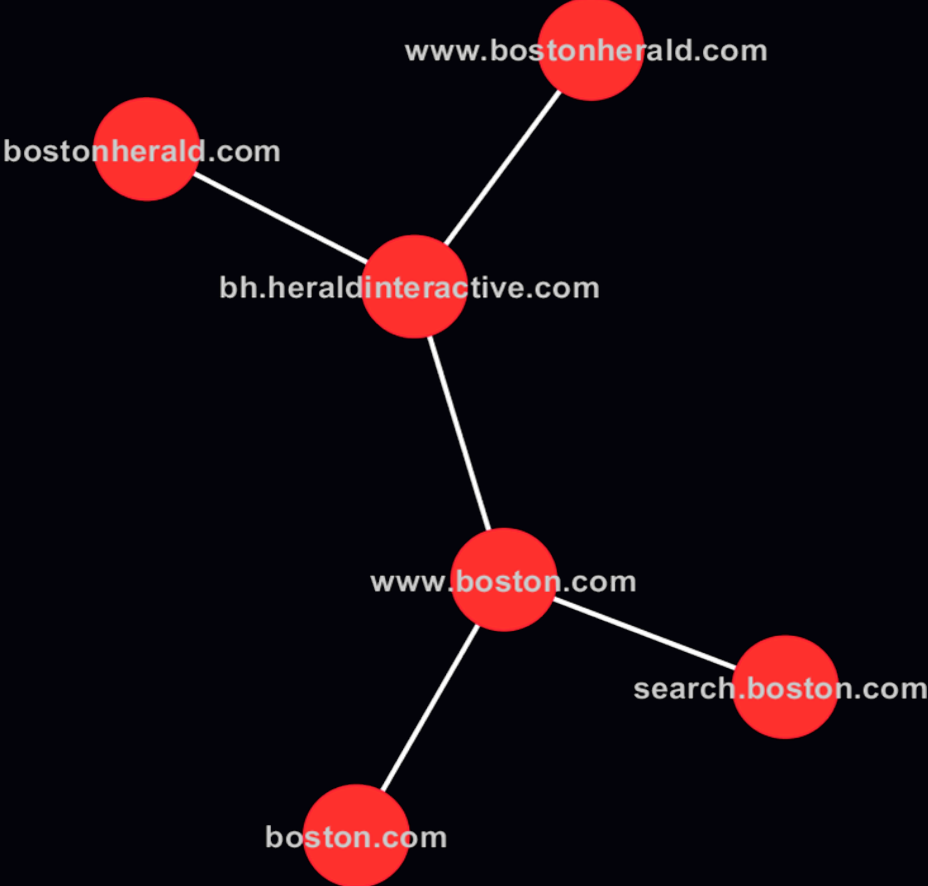
[www.womenshealthbase.com](http://www.womenshealthbase.com)  
[www.filmannex.com](http://www.filmannex.com)  
[www.ffog.net](http://www.ffog.net)  
[www.drugsnews.org](http://www.drugsnews.org)  
[www.menshealthbase.com](http://www.menshealthbase.com)  
[www.dailyfreshies.com](http://www.dailyfreshies.com)  
[www.hark.com](http://www.hark.com)  
[www.gossipcenter.com](http://www.gossipcenter.com)  
[www.articletrunk.com](http://www.articletrunk.com)  
[www.411answers.com](http://www.411answers.com)  
[www.dailyrx.com](http://www.dailyrx.com)  
[www.all-allergies.com](http://www.all-allergies.com)  
[www.knowvehicles.com](http://www.knowvehicles.com)  
[www.chinaflix.com](http://www.chinaflix.com)  
[www.parentingnewsstories.com](http://www.parentingnewsstories.com)  
[www.wrestlingnewz.com](http://www.wrestlingnewz.com)  
[www.gourmandia.com](http://www.gourmandia.com)

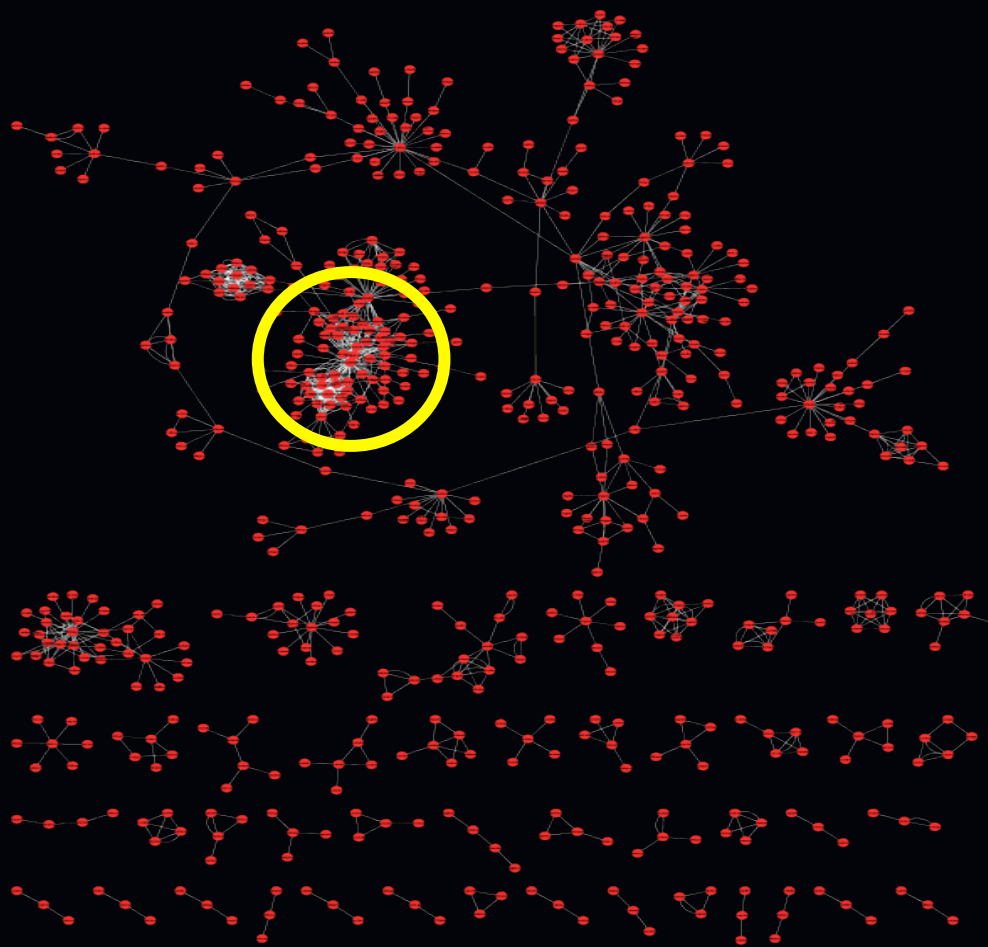
## Traffic overlap of cookies from Bid Request

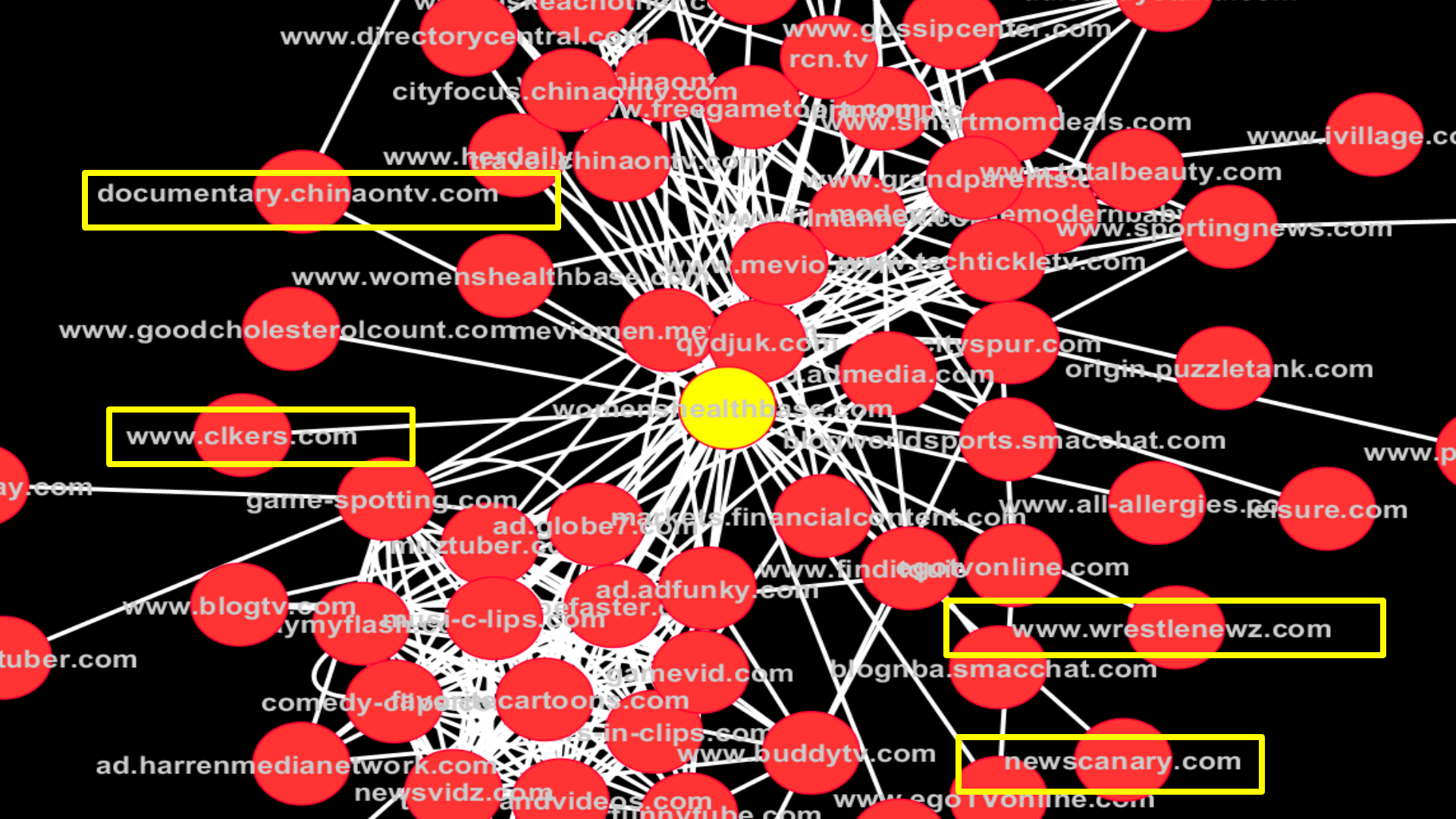




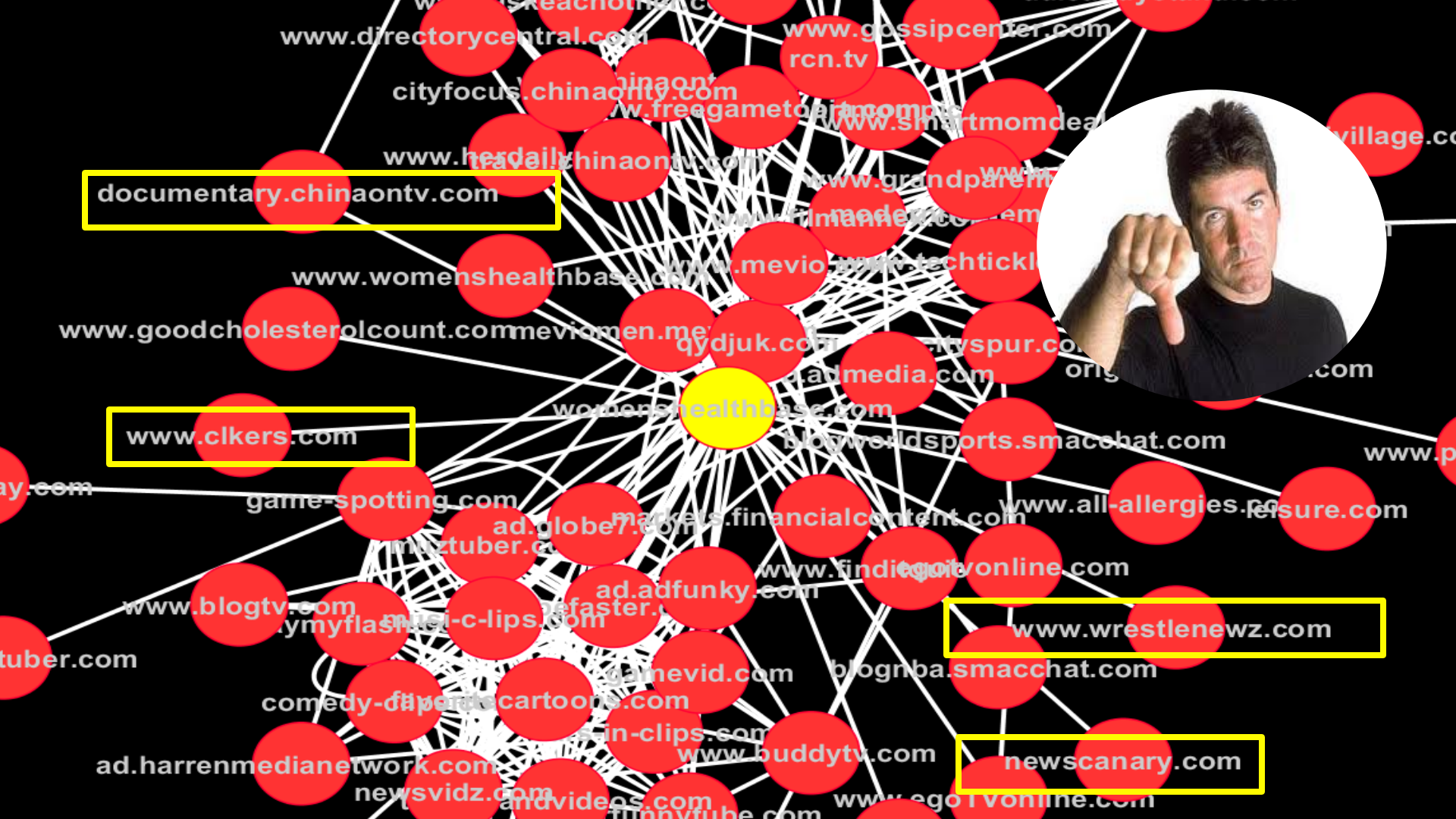
# Boston Herald



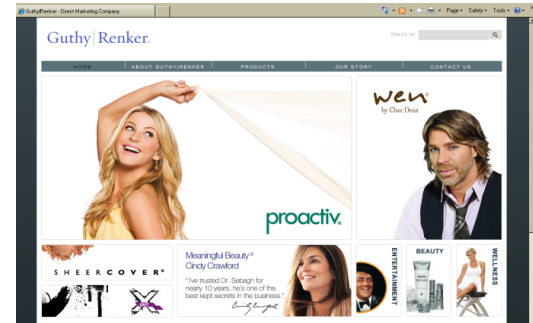
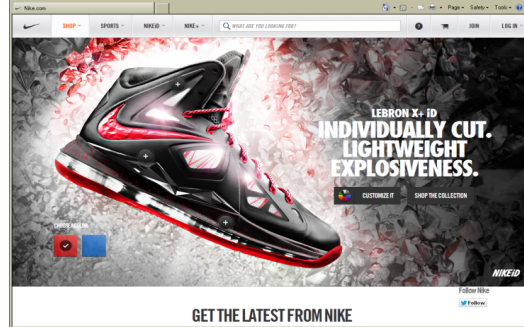
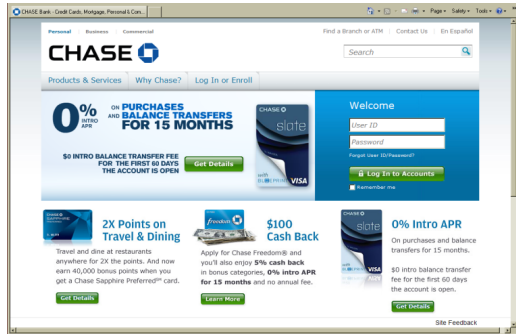






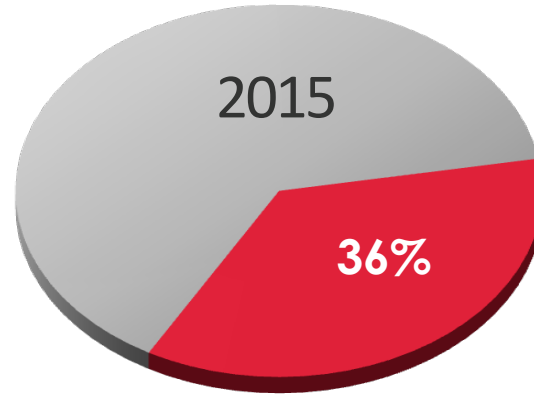


# Bots are executing conversion events



- '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

# Two populations surfing the we: Bot vs. Human

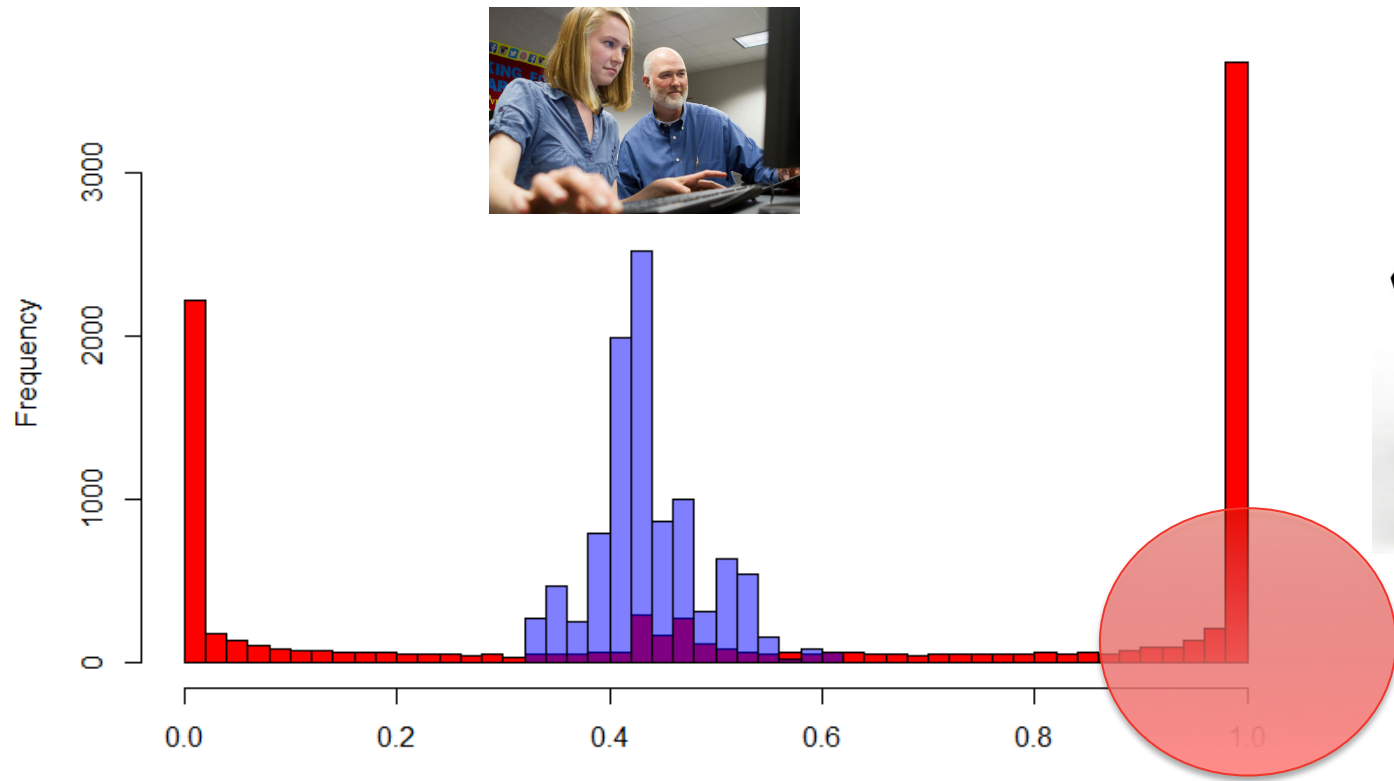




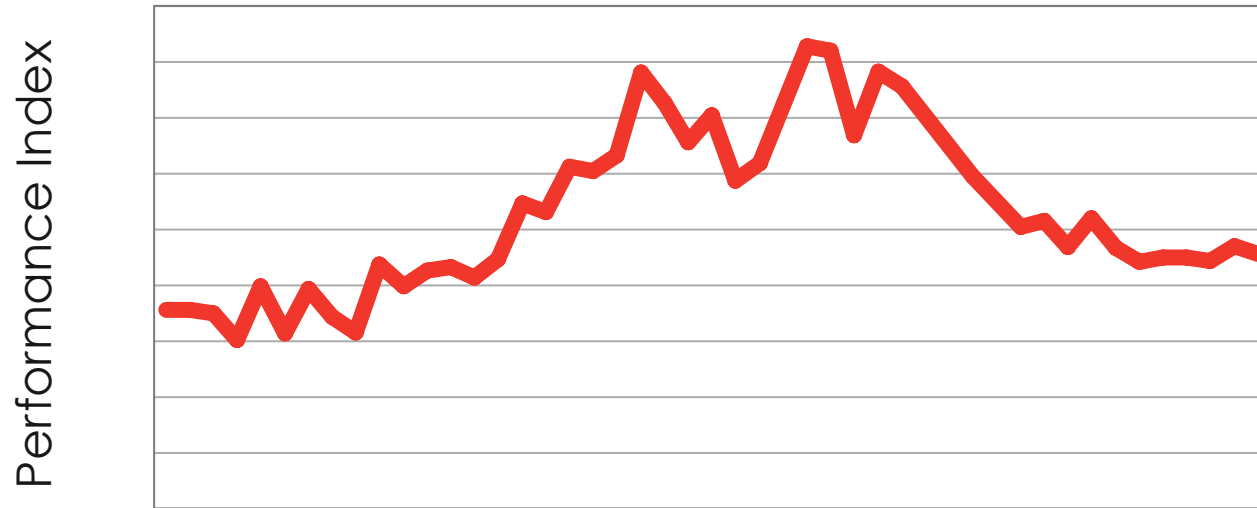
Random

Deterministic

# Bot activity has more signal – higher predictions



# Eliminate labels generated by bots ....

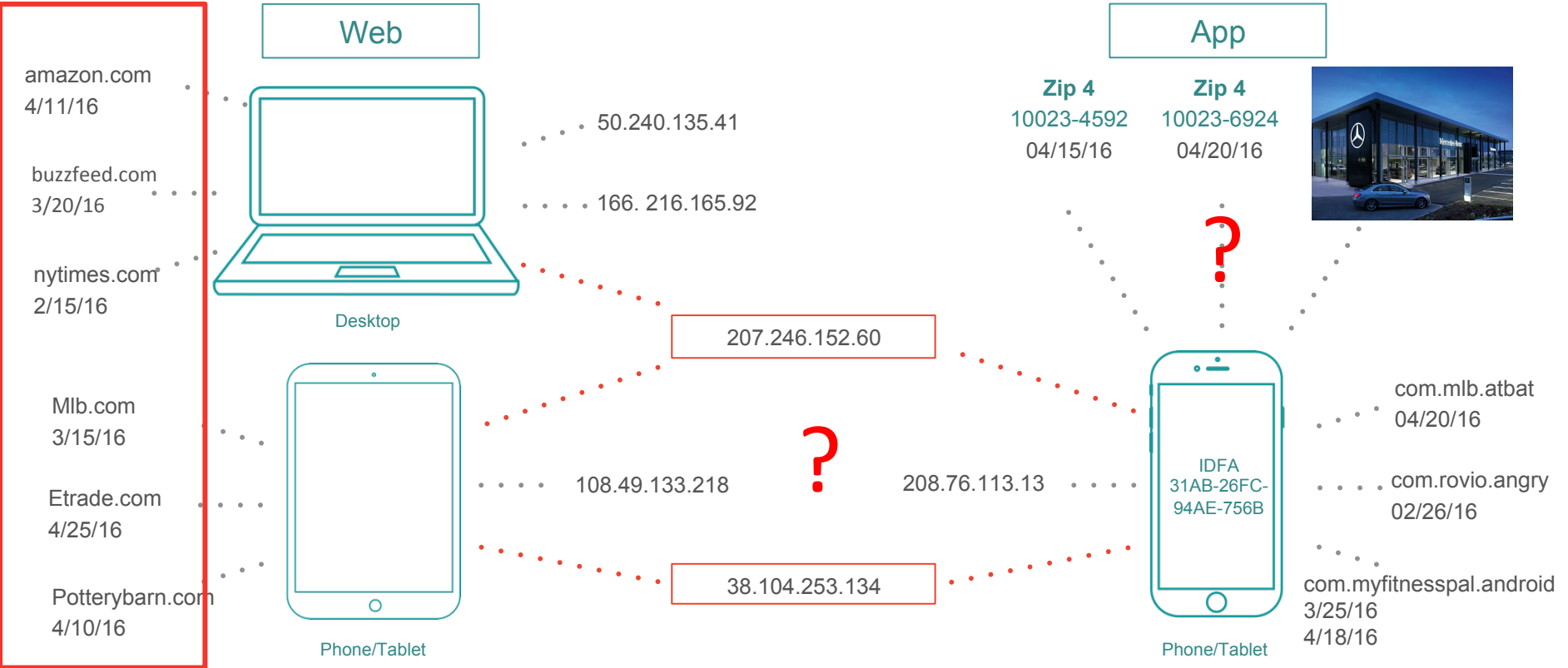


3 more weeks in spring 2012

# Finding good audiences for luxury cars? Predict dealership visits?

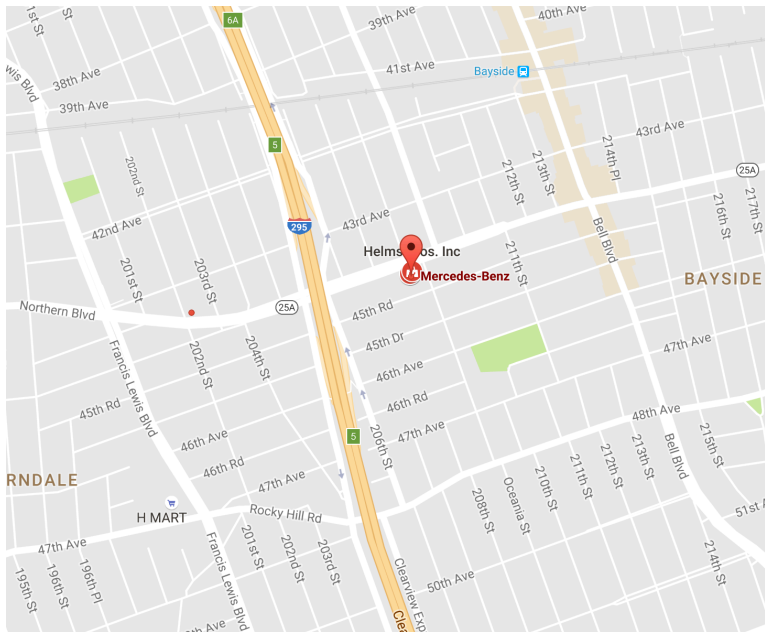


# Browsing history of your neighbor who hacked your WIFI?





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

# 'In the market' signal



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

# Real Estate



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

# Fitness ...



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

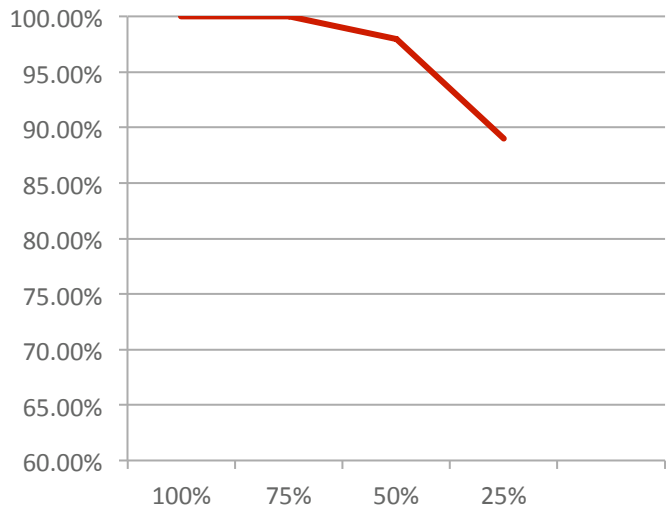


```
0000d41ed774823fca142945ec915c0,0,,,,,,,,,en_GB,,  
0000dee02d70cf8c0d79f96b6d1c59d,0,,,,,,,,,en_US,,  
0000f232abfe25a80156fe069395460,1,1992,20,2,2,,,,,,,,,19,-5  
0000f4ba0cff946b1c0e3b051287ede,0,1993,19,2,,,,,,,,,en_US,310,8  
0000130571654e3afaa62f4e9d2e4f63,1,,2,2,,,,,,,,,en_US,193,7  
00001544469ae9b408869a463a1dd77a,1,1984,28,2,,,,,,,,,en_US,,-4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

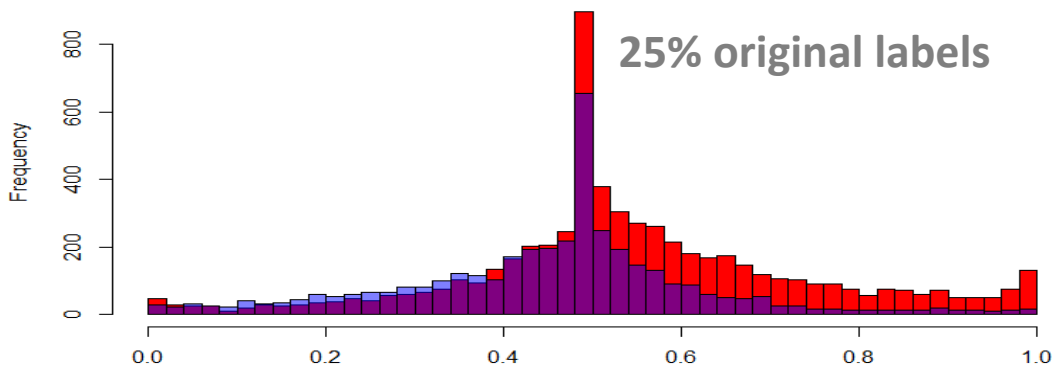
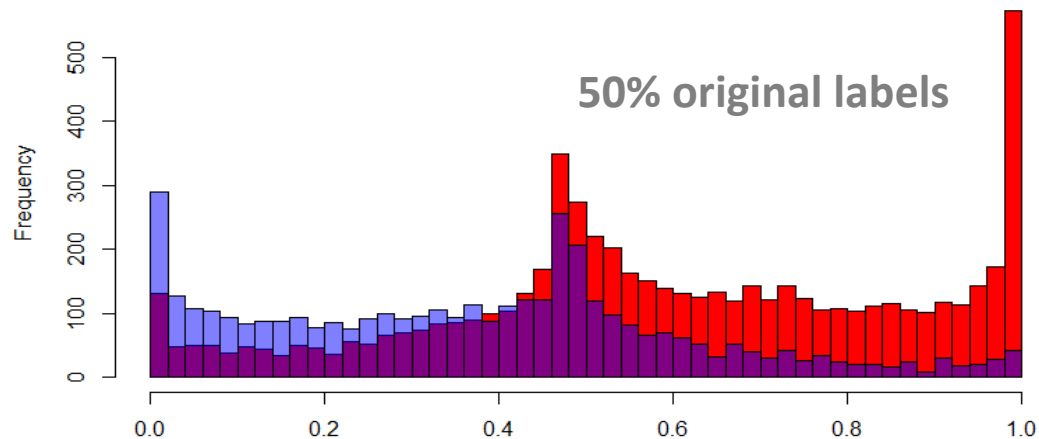


# Performance is surprisingly stable even under random noise

## Percent Men in top 1%



## Percent original labels



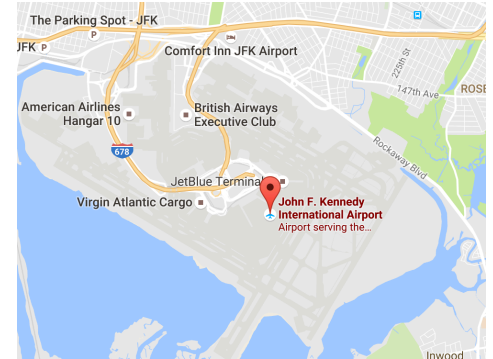
# Identify people who will go to Mercedes dealerships?



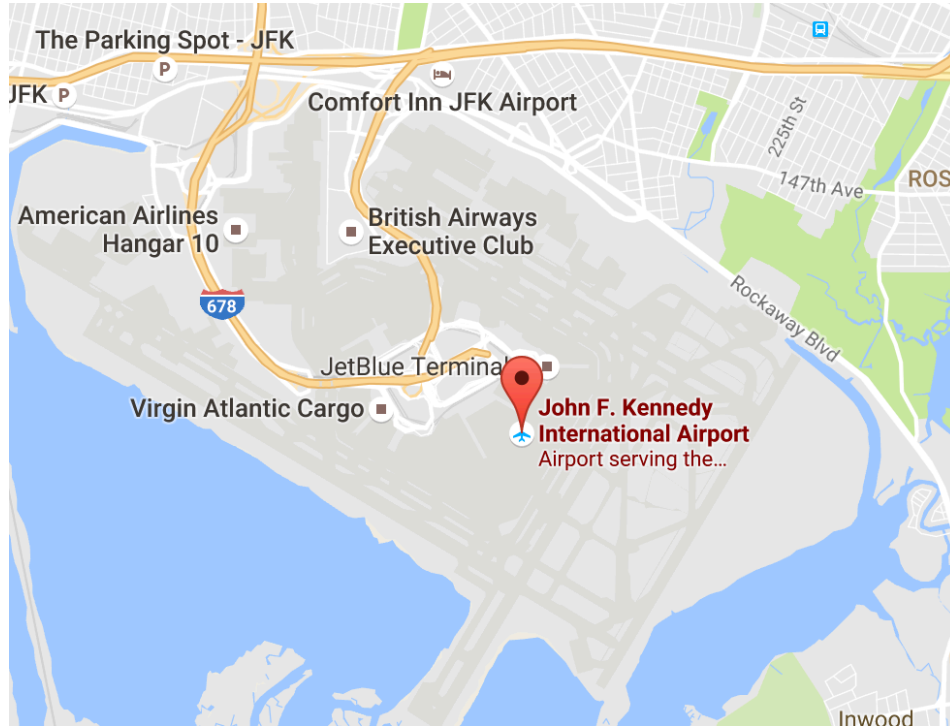
My predictions might be better than my 'ground truth'



# Where do we find frequent traveler?



# What do you think indicates people going to JFK?



## URL

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

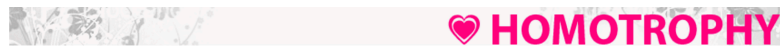
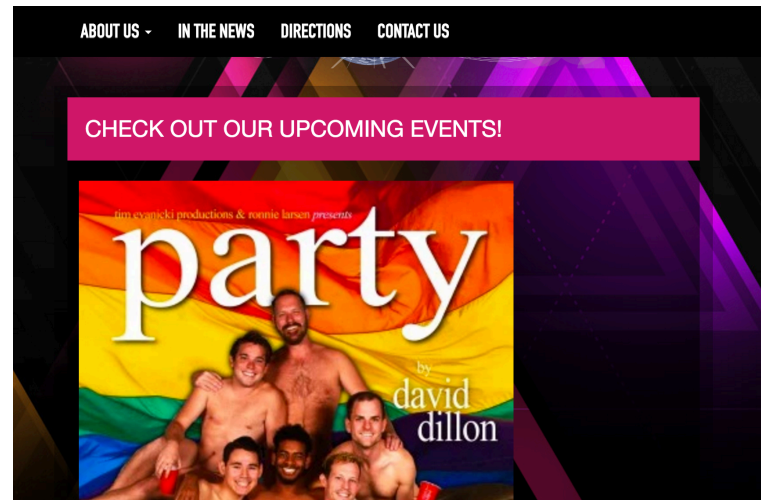
URL	Logistic Parameter
<a href="http://www.iglesiaelfaroinc.org">www.iglesiaelfaroinc.org</a>	2.38
<a href="http://www.jumpseatnews.com">www.jumpseatnews.com</a>	2.25
<a href="http://www.bluelineprop.com">www.bluelineprop.com</a>	2.21
<a href="http://www.ktxdtv.com">www.ktxdtv.com</a>	2.14
<a href="http://www.southjefffootball.org">www.southjefffootball.org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



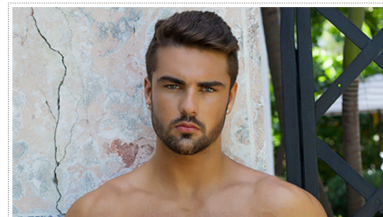
RECENT ARTICLES

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](#) [+1](#) [Pin it](#)



SUBSCRIBE

Posts (RSS)  
 Comments (RSS)

TROPHY : GET US

Become a FAN  
 Twitter

TROPHY::LISTS

VISIT LOVELY AD

INNduge Palm Springs



Swimsuits Allowed

# Predict who goes to JFK?



People who  
work there ...



Random

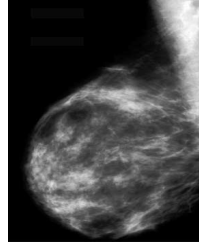
Deterministic

# Siemens: Breast Cancer Detection in Mammograms

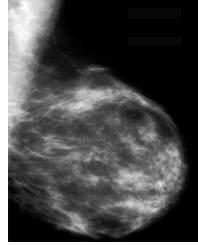


1712 Patients

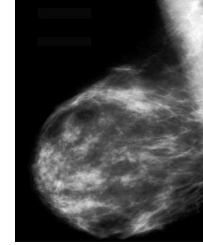
MLO



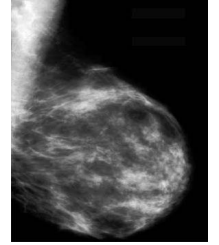
CC



MLO

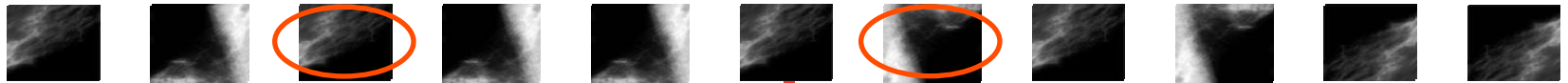


CC



6816 Images

Malignant



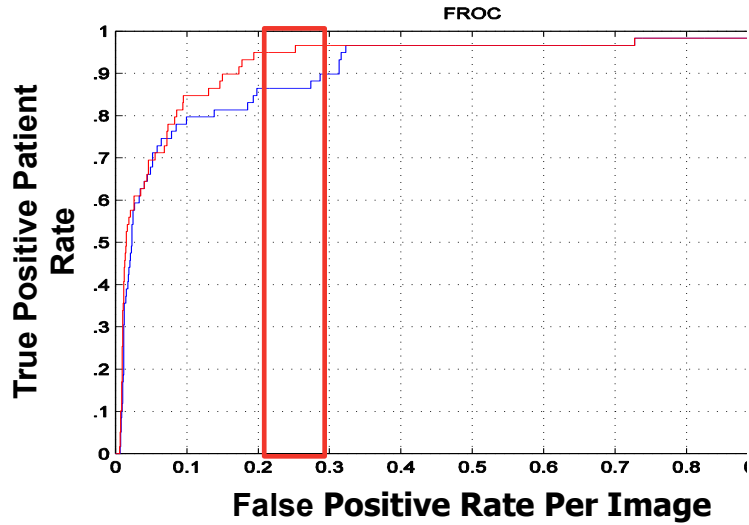
105,000 Candidates

$[x_1, x_2, \dots, x_{117}]$

Image feature vector



# Great Results!



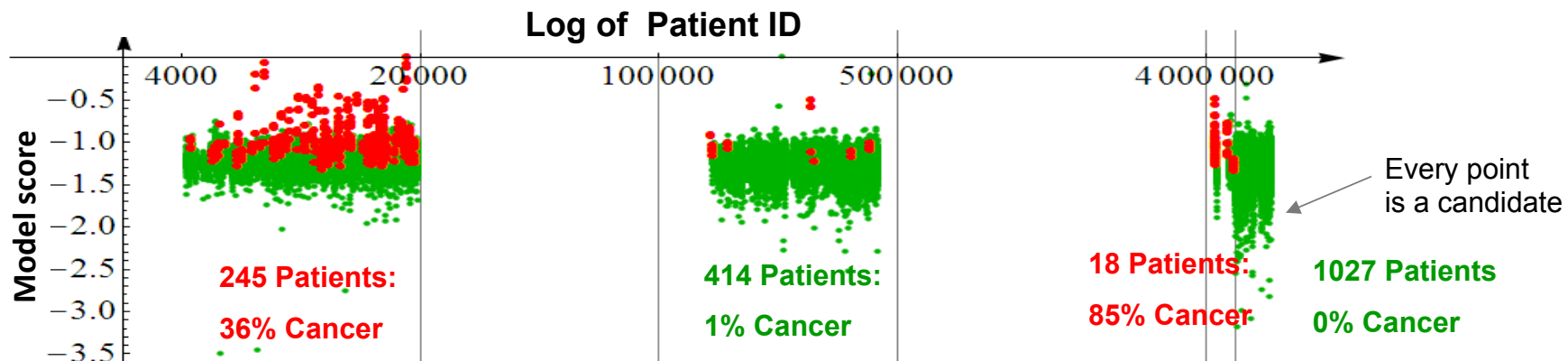
FROC: 0.093



Sent 60% of healthy women home

Except ...

# Something is strange about the Patient ID



Patient ID is extremely predictive

The model learned the implicit location of the fMRI...



Big Picture?

# Beware of (unintentional) discrimination based on predictability



## LinkedIn Jobs: Targeted Email Delivery is Here

 Parker Barrile February 28, 2011

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

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.

Ad related to latanya sweeney ⓘ

[Latanya Sweeney Truth](#)  
[www.instantcheckmate.com/](http://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/](http://www.instantcheckmate.com/)

[Latanya Sweeney](#)  
Public Records Found For: **Latanya Sweeney**. View Now.  
[www.publicrecords.com/](http://www.publicrecords.com/)

[La Tanya](#)  
Search for La Tanya Look Up Fast Results now!  
[www.ask.com/La+Tanya](http://www.ask.com/La+Tanya)

Claudia Perlich

Thank You!

dstillery