Applied ML in Marketing Tech @ The Oracle Data Cloud

Alex Sadovsky
Senior Director of Data Science

ORACLE
DATA CLOUD

grapeshot

AddThis™

bluekai

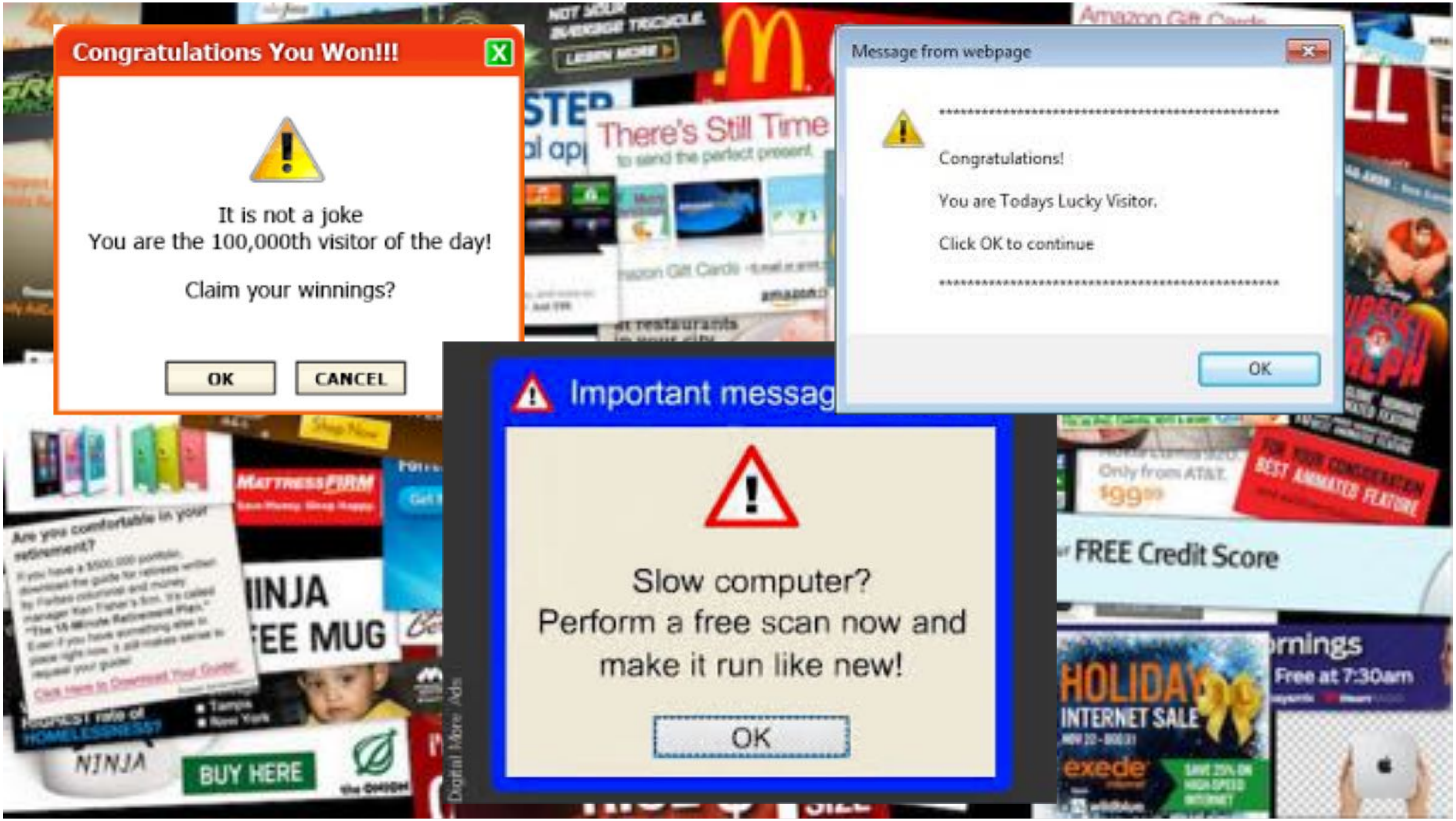
ORACLE®
DATA CLOUD

crosswise

Datalogix

MOAT

Why so many companies?



Congratulations You Won!!!
It is not a joke
You are the 100,000th visitor of the day!
Claim your winnings?
OK
CANCEL
Message from webpage
Congratulations!
You are Todays Lucky Visitor.
Click OK to continue
OK
Important messag
Slow computer?
Perform a free scan now and
make it run like new!
OK
There's Still Time
FREE Credit Score
HOLIDAY
INTERNET SALE
BUY HERE

ORACLE

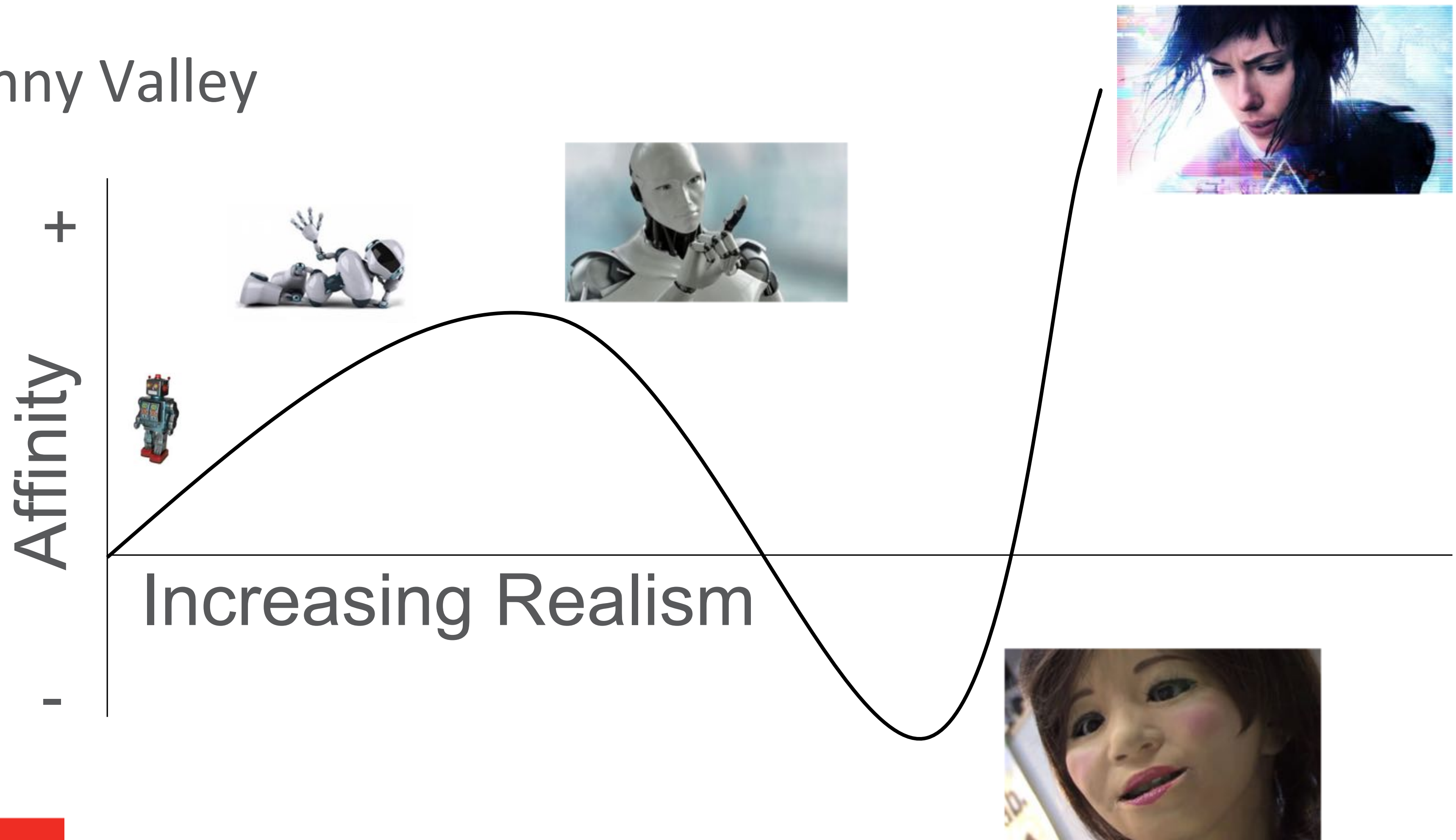
DATA CLOUD

Our mission:

Help advertisers connect with the right customer, personalize every interaction, and measure the effectiveness of each engagement.

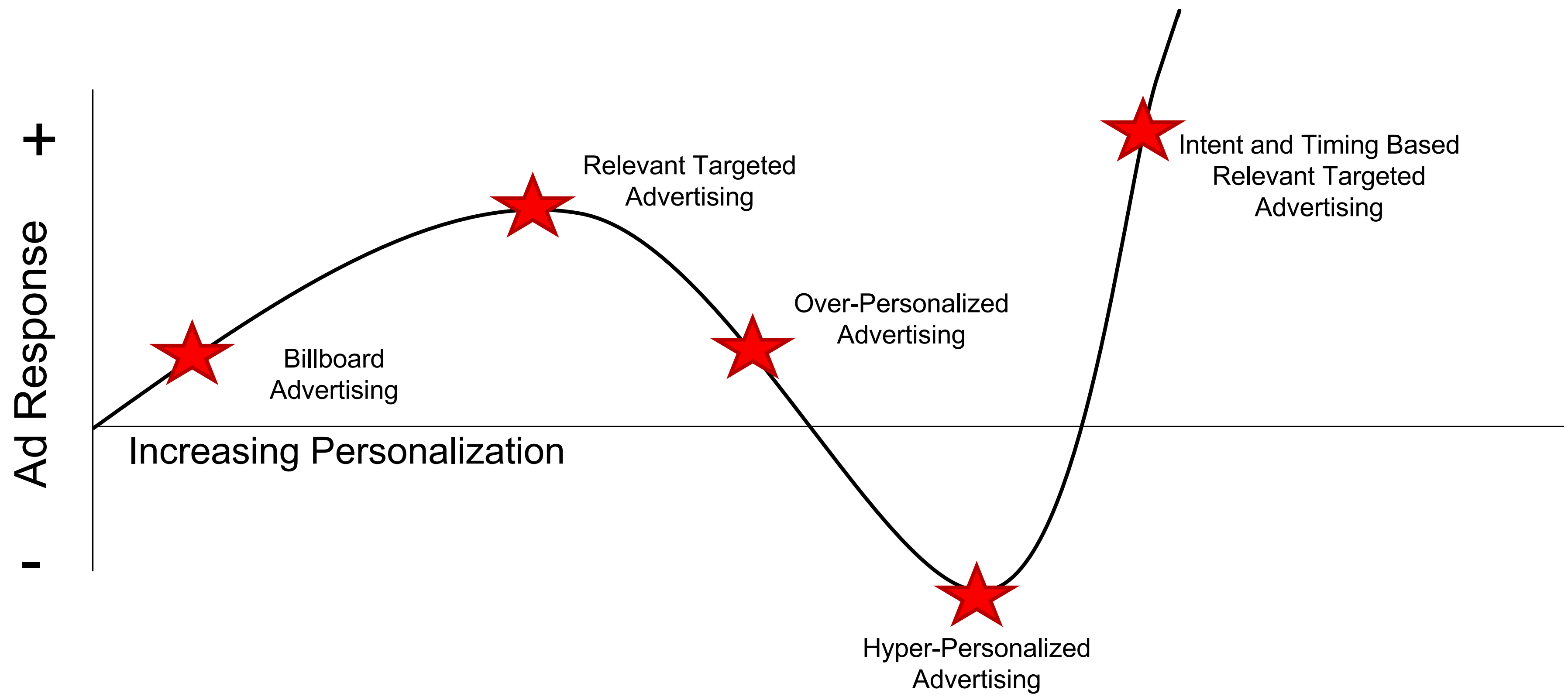
ORACLE

Uncanny Valley



ORACLE
DATA CLOUD

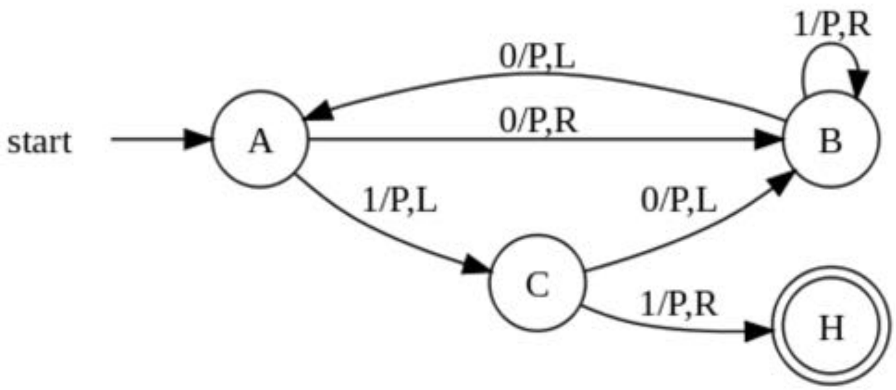
Uncanny Marketing Valley

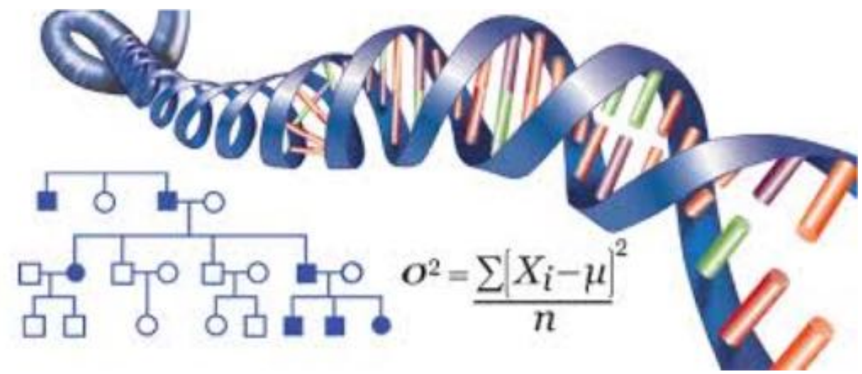


ORACLE DATA CLOUD

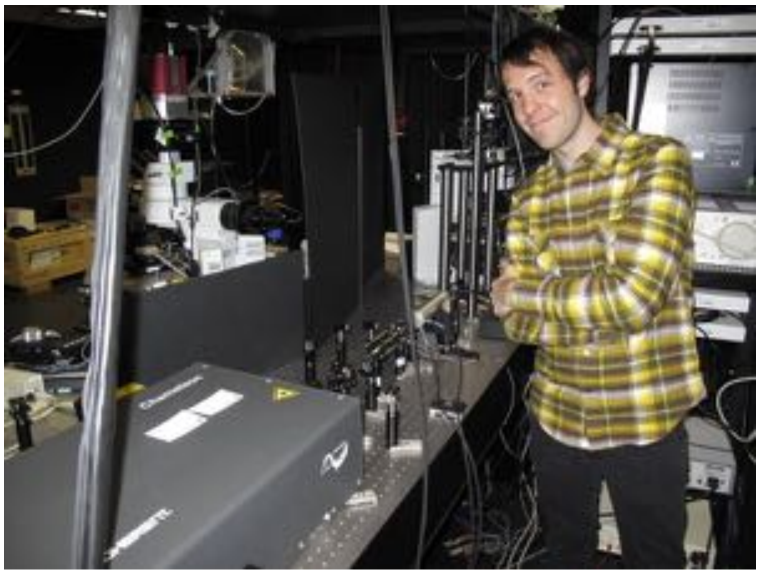
My path to Data Science: Non-formalized training

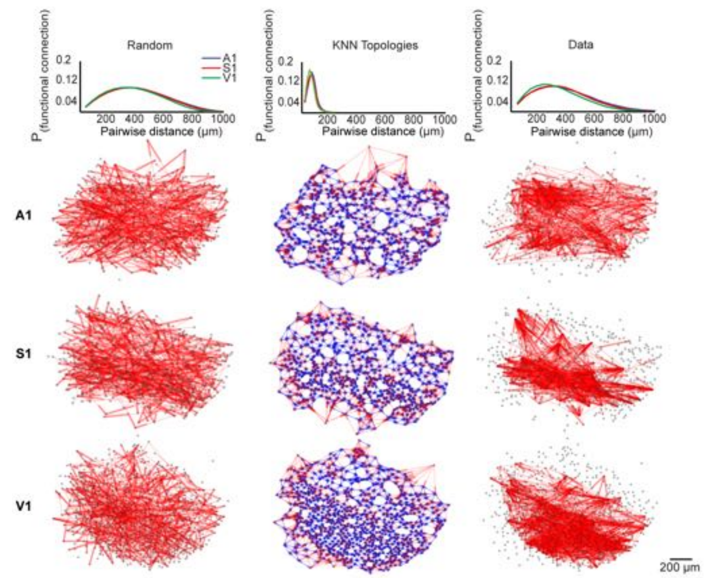
My path to Data Science: Formal education





My path to Data Science: Formal education





THE UNIVERSITY OF CHICAGO

ORACLE DATA CLOUD

What is Data Science?

Big Data Borat
@BigDataBorat
Follow

Data Science is statistics on a Mac.

7:32 AM - 27 Aug 2013

601 RETWEETS 257 FAVORITES

Zvi
@nivertech
Follow

"Data Scientist" is a Data Analyst who lives in California.

7:55 PM - 14 Mar 2012

141 RETWEETS 41 FAVORITES

Josh Wills
@josh_wills
Follow

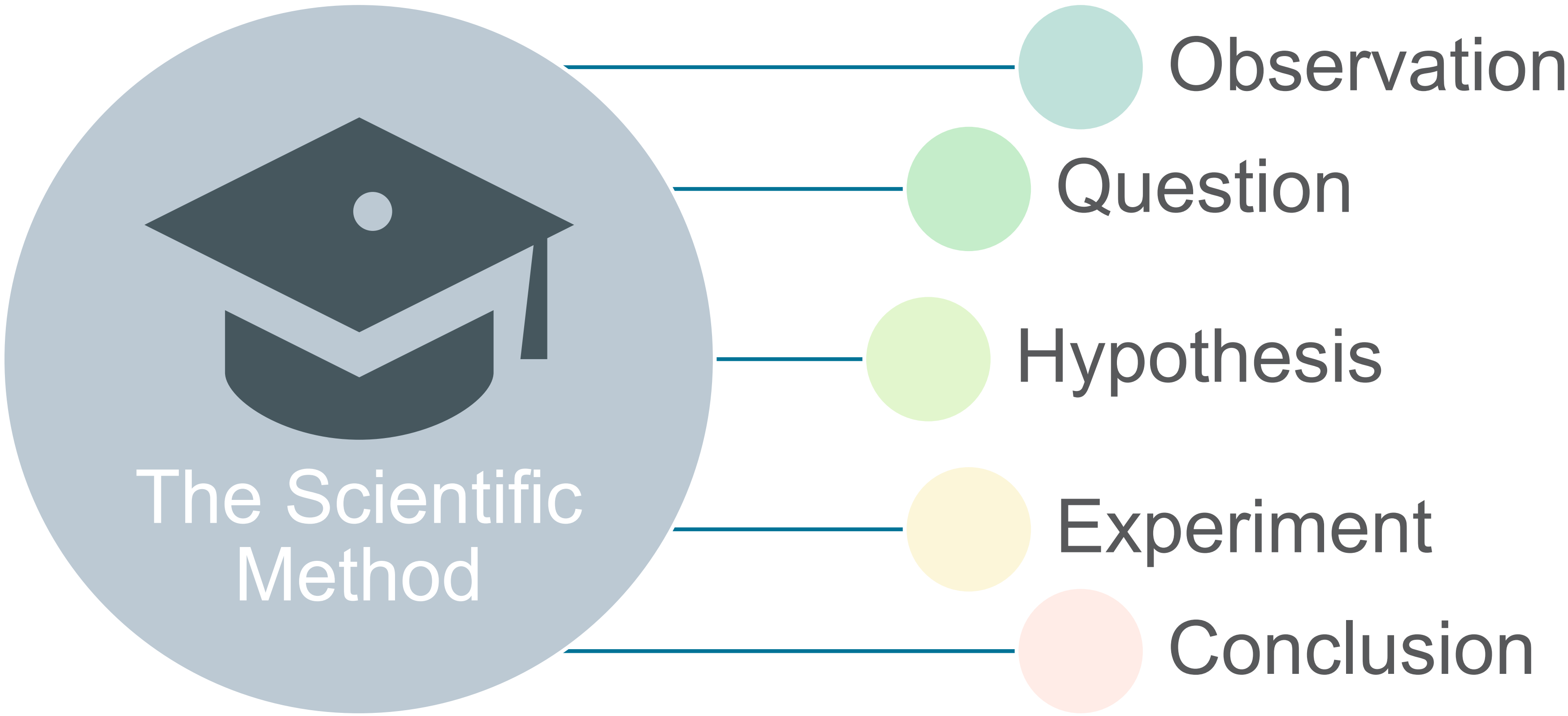
Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

10:55 AM - 3 May 2012

968 RETWEETS 457 FAVORITES

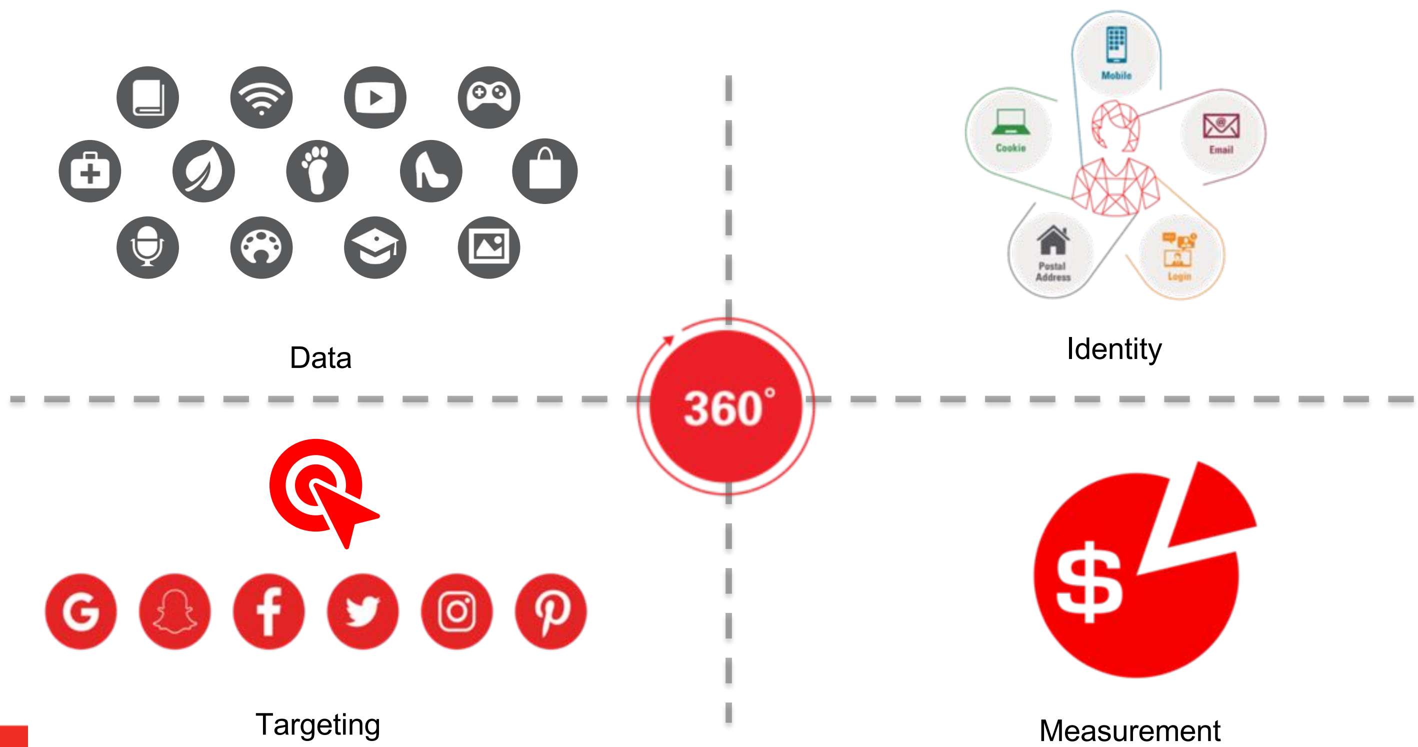
ORACLE
DATA CLOUD

What is Data Science at the Oracle Data Cloud?



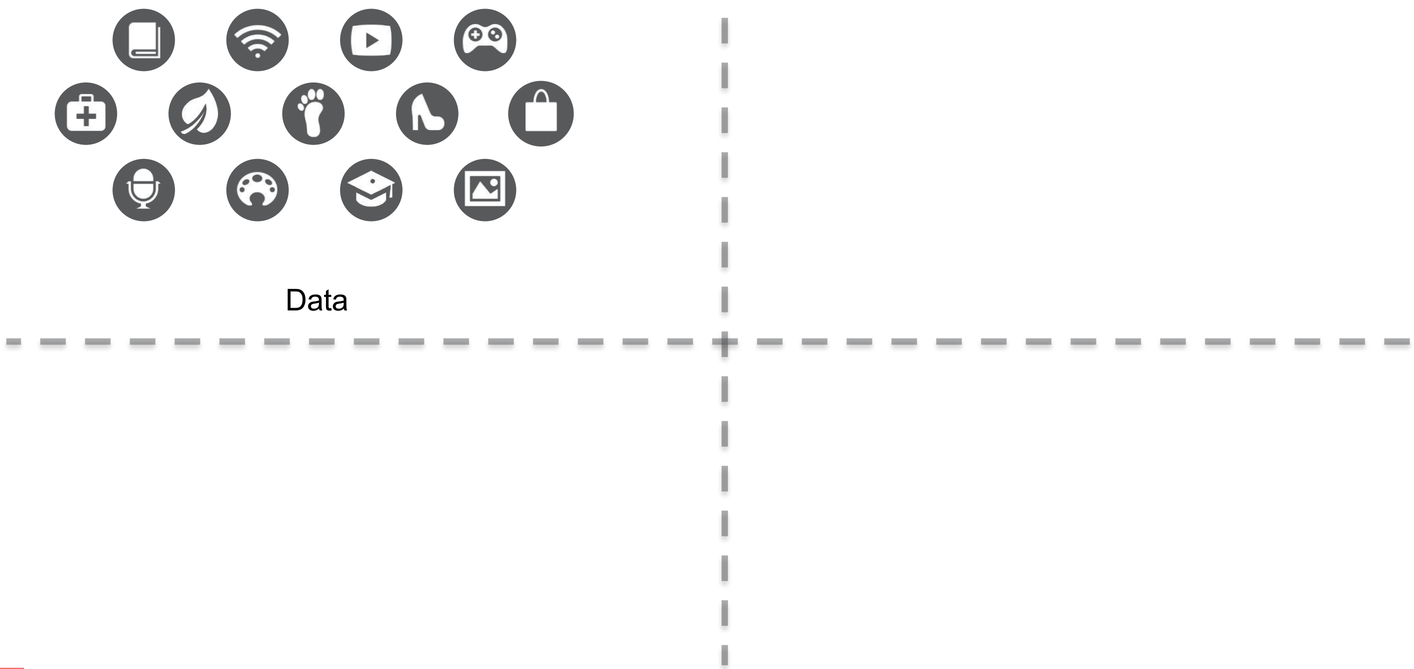
ORACLE DATA CLOUD

Data Science at Oracle Data Cloud



ORACLE DATA CLOUD

Data



To market efficiently, the Oracle Data Cloud uses a full view of consumer behavior

Who they ARE

Working mom,
Millennial

Where they GO

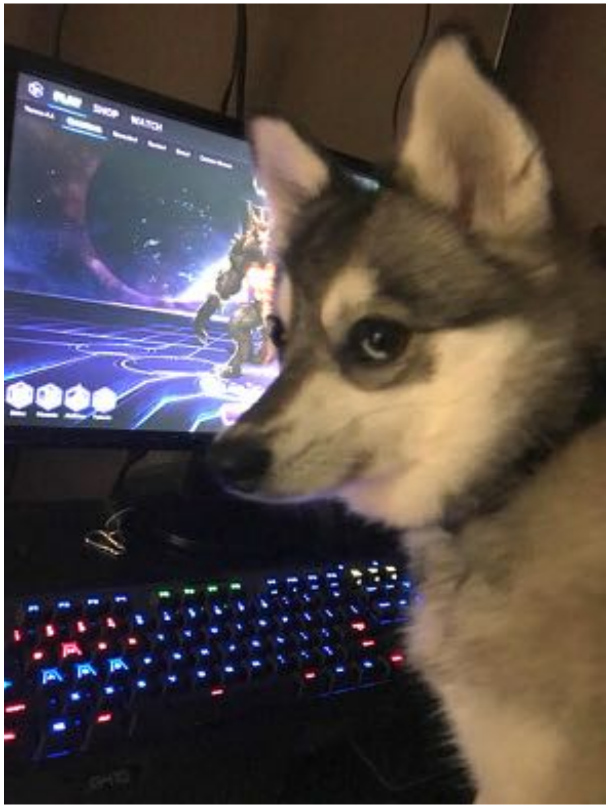
Shops at Big Box stores
Went to South by Southwest

What they DO

Searches for baking recipes
Visits hiking blogs

ORACLE
DATA CLOUD

On the internet no one knows you're a dog



ORACLE
DATA CLOUD

Aspirational vs IRL

Car search on the internet

Car ownership in real life

To market efficiently, the Oracle Data Cloud uses a full view of consumer behavior

What they **BUY**

Mac and Cheese,
Salty Snacks,
Tablet Computers,
Videogames

Who they **ARE**

Working mom,
Millennial

Where they **GO**

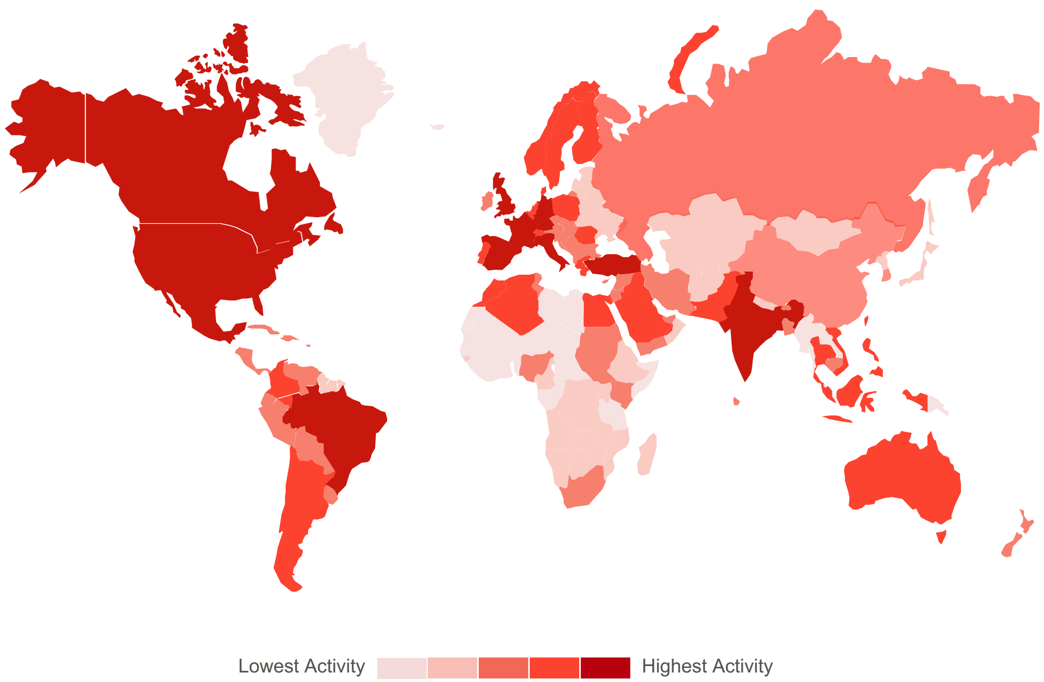
Shops at Big Box stores
Went to South by Southwest

What they **DO**

Searches for baking recipes
Visits hiking blogs

ORACLE
DATA CLOUD

Our data is huge in scale



Digital
Over 5 Billion global cookie IDs

Over 1 Billion mobile device IDs

Over 15 Million domains

1.1 Trillion page views

Analog
115M US House Holds

$3 Trillion in annual observed consumer spending

ORACLE DATA CLOUD

And exists at an **extremely detailed** level

Married: Yes

Kids: 2

Education level:
College grad

Online sites:
- Kayak travel – Italy
- Travel - High-end hotels
- Cars - cross-over vehicle

Location:
San Diego, CA

Hobby: Photography

Buys high-end women's apparel

Drives a Honda CRV

Avid home cook

Organic product purchaser

Buys children's apparel

Favorite wine is Pinot Noir

Shops at luxury retailers

Datalogix data retail data

hanna Andersson

Q: What differentiates ODC from others with these data?

A: Our specificity in cataloging every transaction

Transaction Data

Household ID:	#12345ab6789
Date:	December 2, 2013
Client Name:	Hanna Anderson
Client SKU:	#HT32966
Amount Paid:	$50.00

Product Data

Client Name:	Hanna Andersson		
Client SKU:	#HT32966		
Product Description:	Our Nordic-inspired crewneck sweaters are tightly knit from 100% combed cotton yarns. This is quality to hand down; you won't find a single loose thread inside the smoothly finished knit, so there's no catching on fingers. Washable. Certified by Öko-Tex Standard 100. Imported.		
Product Categories:	Sweater	Wool Clothing	Children's Clothing

How do we do it?

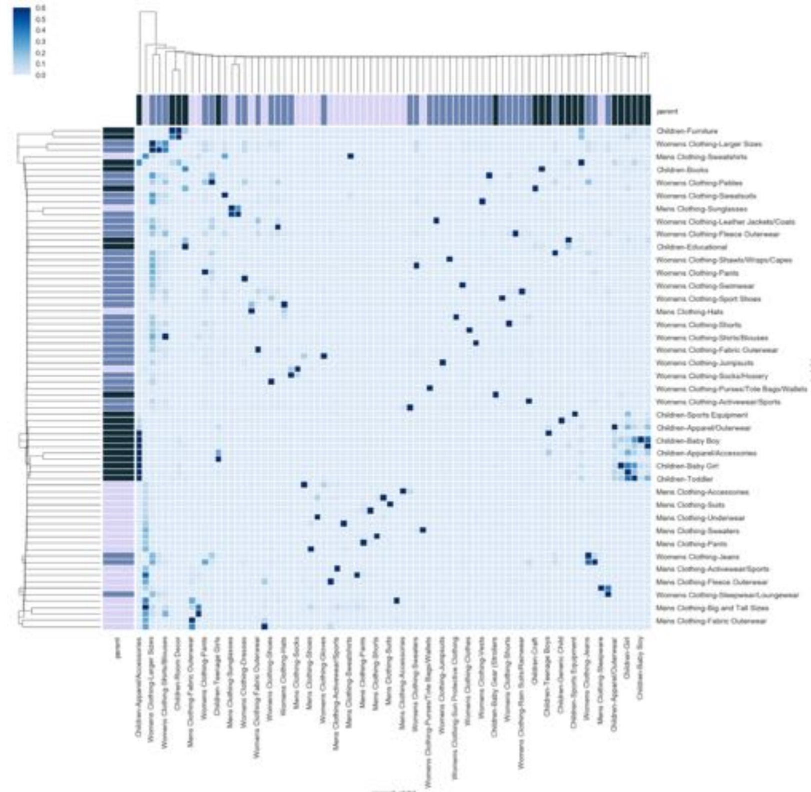
Very Deep Convolutional Networks for Text Classification

Alexis Conneau, Holger Schwenk, Loic Barrault, Yann Lecun

(Submitted on 6 Jun 2016 (v1), last revised 27 Jan 2017 (this version, v2))

The dominant approach for many NLP tasks are recurrent neural networks, in particular LSTMs, and convolutional neural networks. However, these architectures are rather shallow in comparison to the deep convolutional networks which have pushed the state-of-the-art in computer vision. We present a new architecture (VDCNN) for text processing which operates directly at the character level and uses only small convolutions and pooling operations. We are able to show that the performance of this model increases with depth: using up to 29 convolutional layers, we report improvements over the state-of-the-art on several public text classification tasks. To the best of our knowledge, this is the first time that very deep convolutional nets have been applied to text processing.

Comments: 10 pages, EACL 2017, camera-ready
Subjects: **Computation and Language (cs.CL)**; Machine Learning (cs.LG); Neural and Evolutionary Computing (cs.NE)
Cite as: arXiv:1606.01781 [cs.CL]
(or arXiv:1606.01781v2 [cs.CL] for this version)



ORACLE
DATA CLOUD

AddThis data

. . .Derived from javascript-based marketing tools on +15M domains

ORACLE
DATA CLOUD

Geolocation Data – Accuracy problem or opportunity?

Marketers Must Question Location Data Before They Use It for Attribution

Mobile Data Has A Quality Control Problem

by Allison Schiff // Monday, March 5th, 2018 – 9:00 am

Until We Fix Location Data Inaccuracy, We Will Never Close the Mobile Spending Gap

by AdExchanger // Friday, July 28th, 2017 – 1:05 am

GPS isn't the answer when it comes to location-based marketing

With GPS still only accurate to 10 meters and failing to account for height, marketers need to start figuring out novel ways of implementing their location-based marketing strategies. PoweredLocal's Gary Tramer explains.

How an internet mapping glitch turned a random Kansas farm into a digital hell

Report: Only 1% of exchange location data useful for offline attribution

Study also found that average accuracy of exchange-derived locations is 'over 4 New York City blocks.'

ORACLE
DATA CLOUD

Why does accurate location data matter?

Testing location data: 30+ Provider RFP

Each location provider was challenged to send us their "best" list possible for buyers at each of the three chains and we would compare to our data in the follow scenarios:

Better than random

How likely was the location data to outperform random buyers?

Location based machine learning

Can location data power a model built to find shopper look-a-likes?

Too much Information

How well could a location provider describe our own employees' day to day lives?

We found that 5 data providers are drastically ahead of the pack

Better than random

Question: When is it the busiest time to head to the grocery store?

Answer: Saturday, followed by Sunday, followed by any weekday before a holiday or major sporting event

Question: Who is most likely to make a purchase at a store?

Answer: Someone who has purchased there before

We wanted our random tests to reflect this!

Better than random: Creating random

Joe Grocery

Location Data

Thursday	Friday	Saturday	Sunday
*			*

Purchase Data

Thursday	Friday	Saturday	Sunday
*			*

Jane Random

Joe's Location Data

Thursday	Friday	Saturday	Sunday
*			*

Purchase Data

Thursday	Friday	Saturday	Sunday
	*		*

Data driven view of a consumer with location data

Married: Yes

Kids: 2

Multiple elementary school visits

Education level:
College grad

Online sites:
- Kayak travel – Italy
- Travel - High-end hotels
- Cars - cross-over vehicle

Visited a Honda dealership two days ago

DMA:
San Diego

Traveled to NY last week

Hobby: Photography

Buys high-end women's apparel

Data from stores w/o loyalty programs

Drives a Honda CRV

Gets aftermarket service

Avid home cook

Shops at local small markets

Organic product purchaser

Buys children's apparel

Favorite wine is Pinot Noir

Frequently goes to smaller liquor stores

Bargain Shopper

Big box shop frequenter

ORACLE
DATA CLOUD

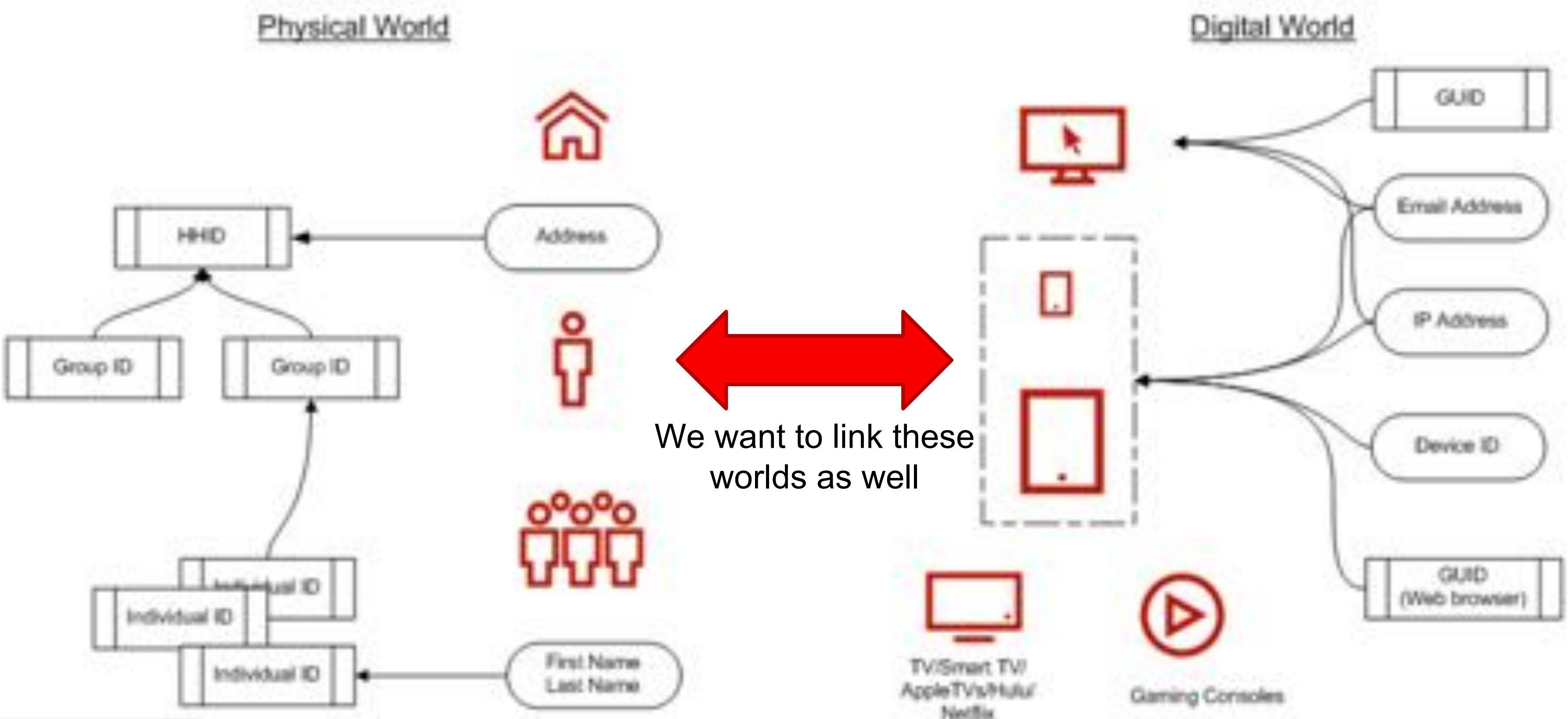
Identity

Data

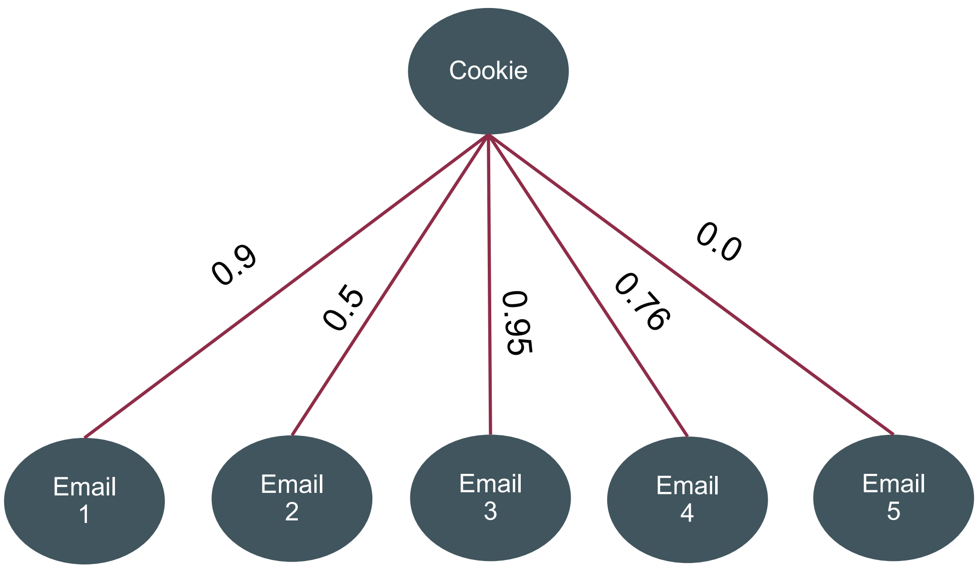
Identity

ORACLE DATA CLOUD

Where does data come from?



The problem: defining linkage

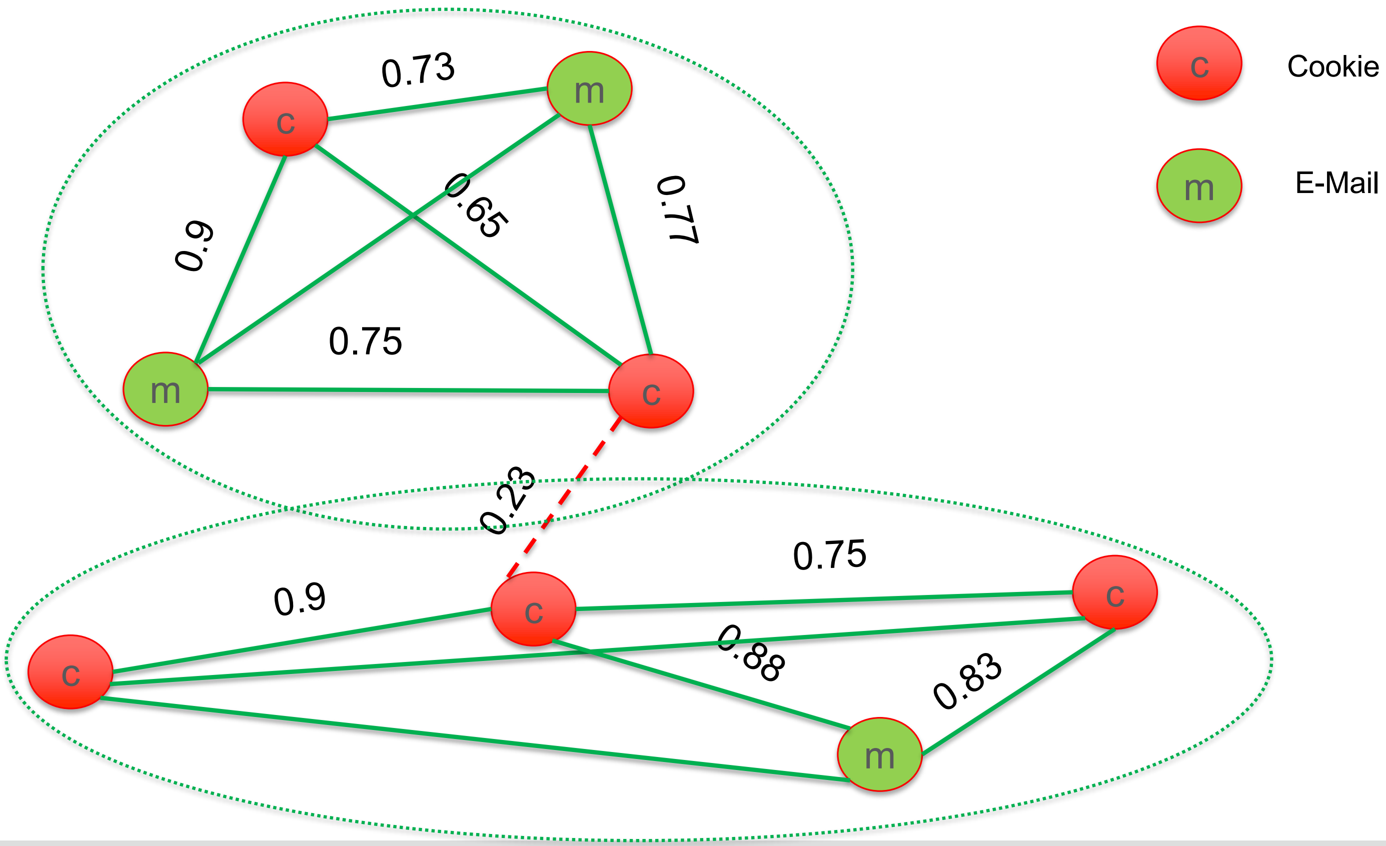


We use machine learning to generate probabilities for each candidate

- Natural language processing of email address
- Supervised learning on known linkages based upon relevant features

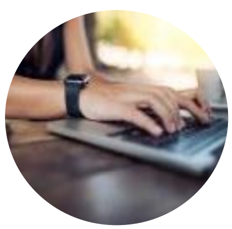
ORACLE DATA CLOUD

The problem: grouping linkage types



ORACLE DATA CLOUD

Why does a strong ID Graph matter for marketing?

You reach		instead of		**Wasted media spend**
Your target ends up in	**Control**	instead of	**Test**	**Inaccurate results**
You can only target via		instead of		**Limited cross-channel reach and disjointed consumer experience**

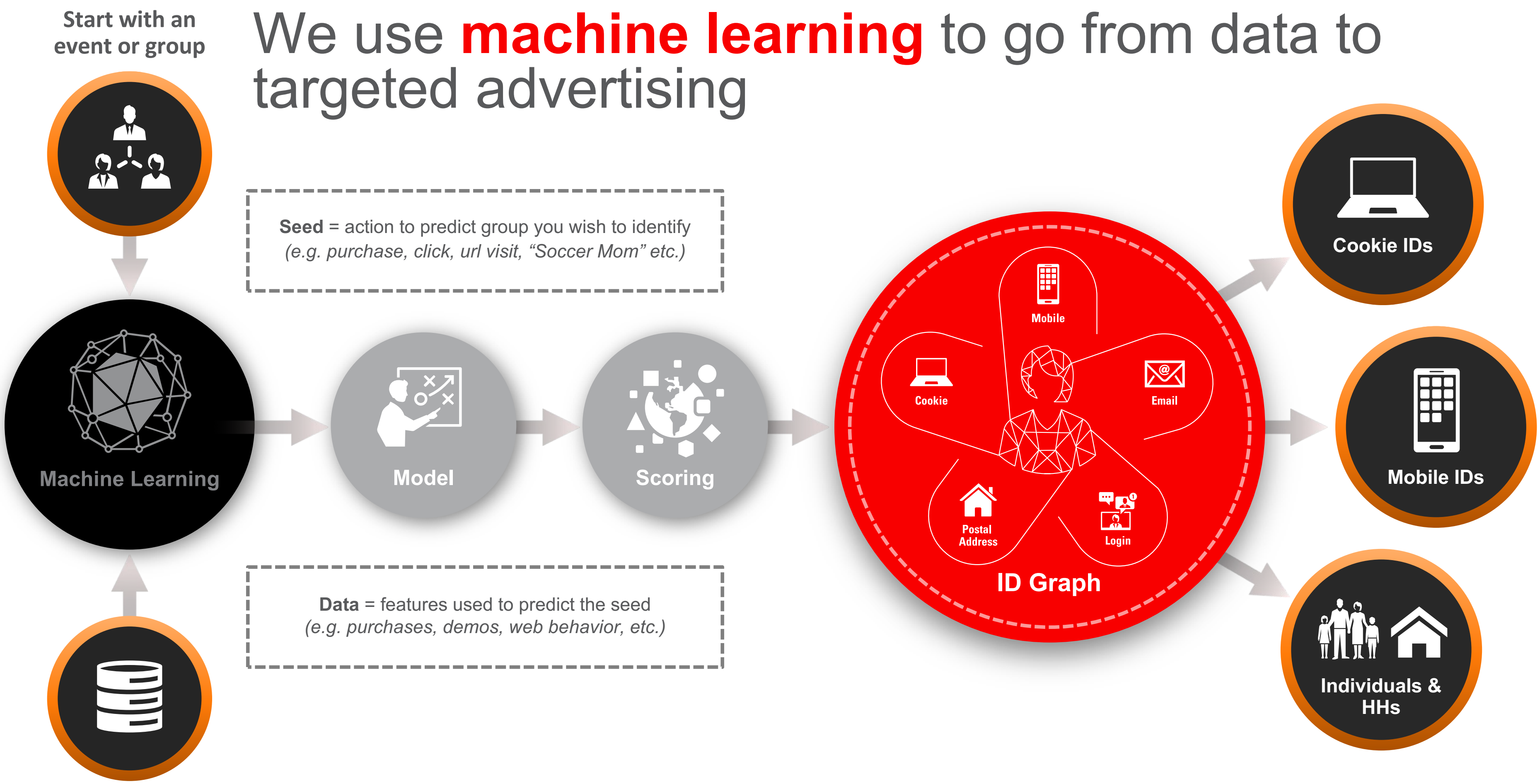
Targeting

Data

Identity

Targeting

We use **machine learning** to go from data to targeted advertising



ORACLE DATA CLOUD

We do this on a stable,
self built, machine learning
Platform: M360

Every month we create:

10,000+ models

Each with terabytes of data…

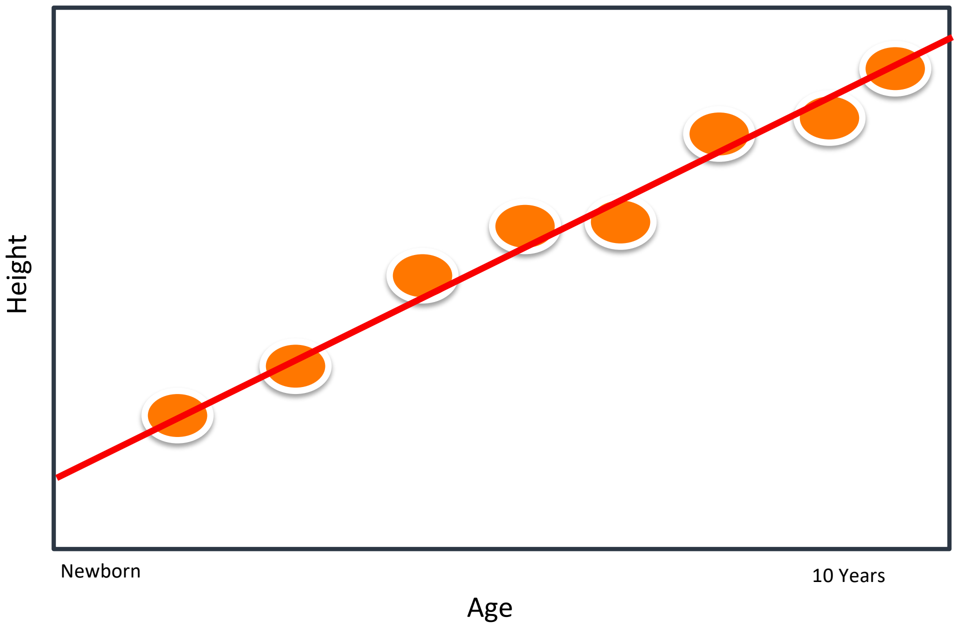
… and best in class machine learning accuracy

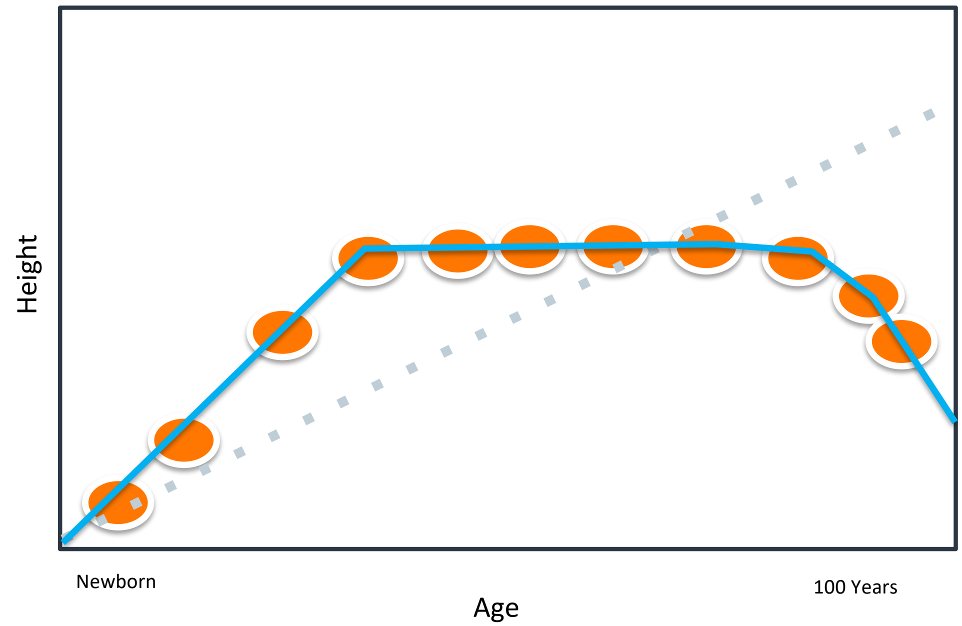
ORACLE
DATA CLOUD

The key to doing this successfully is letting the data tell the story

The non-data driven way to model data:

- Manually pick a modeling method for your data
- Apply it





ORACLE DATA CLOUD

Letting the data tell the story

No Free Lunch Theorem

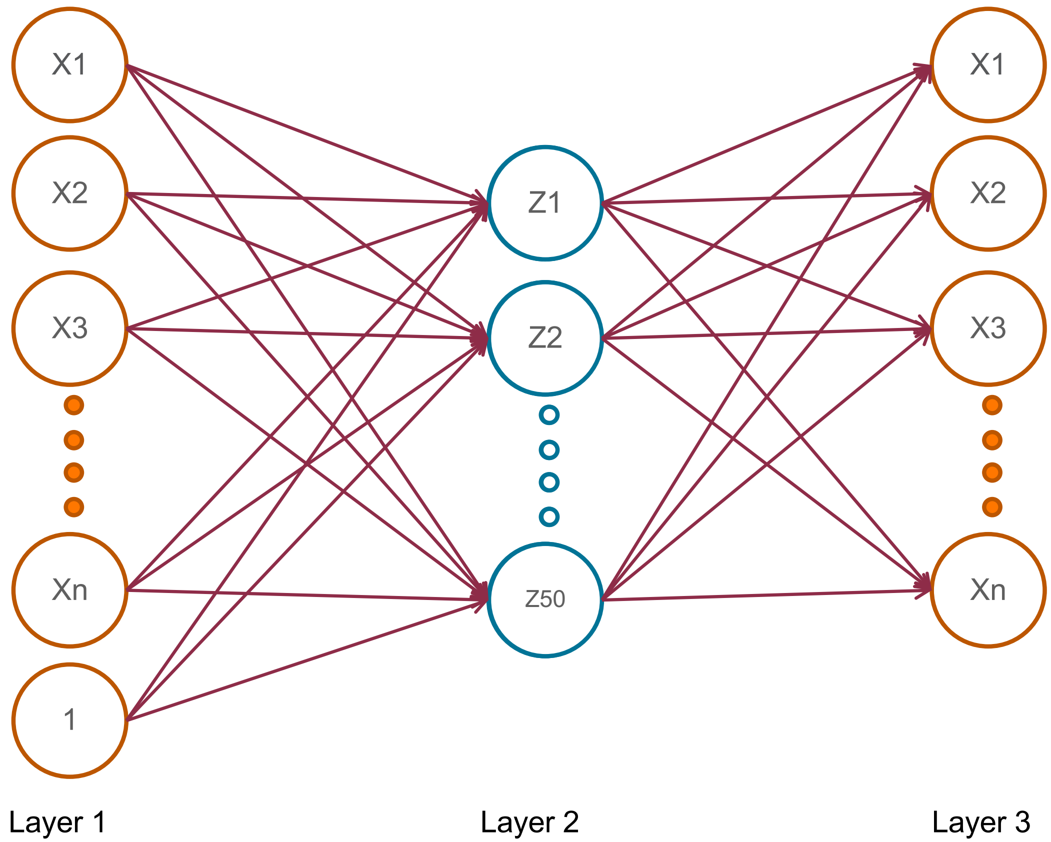
- For any model, any elevated performance for one type of problems is offset by poor performance for another type

So what can we do?

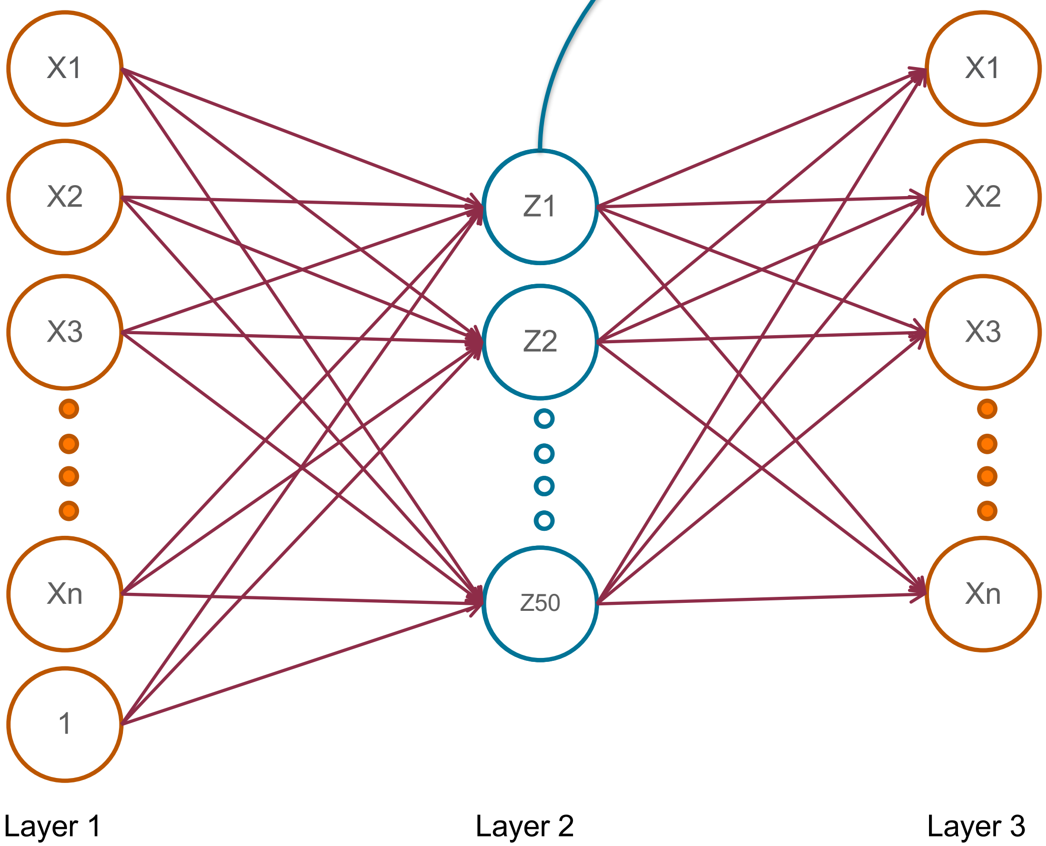
- Try multiple models with training data and pick the one or combination that works best with machine learning
- We can do it easily in under 50 lines of code!

GitHub https://github.com/sadovsky/no-free-lunch/blob/master/no_free_lunch.ipynb

Letting (some of) the data tell the story: Autoencoders

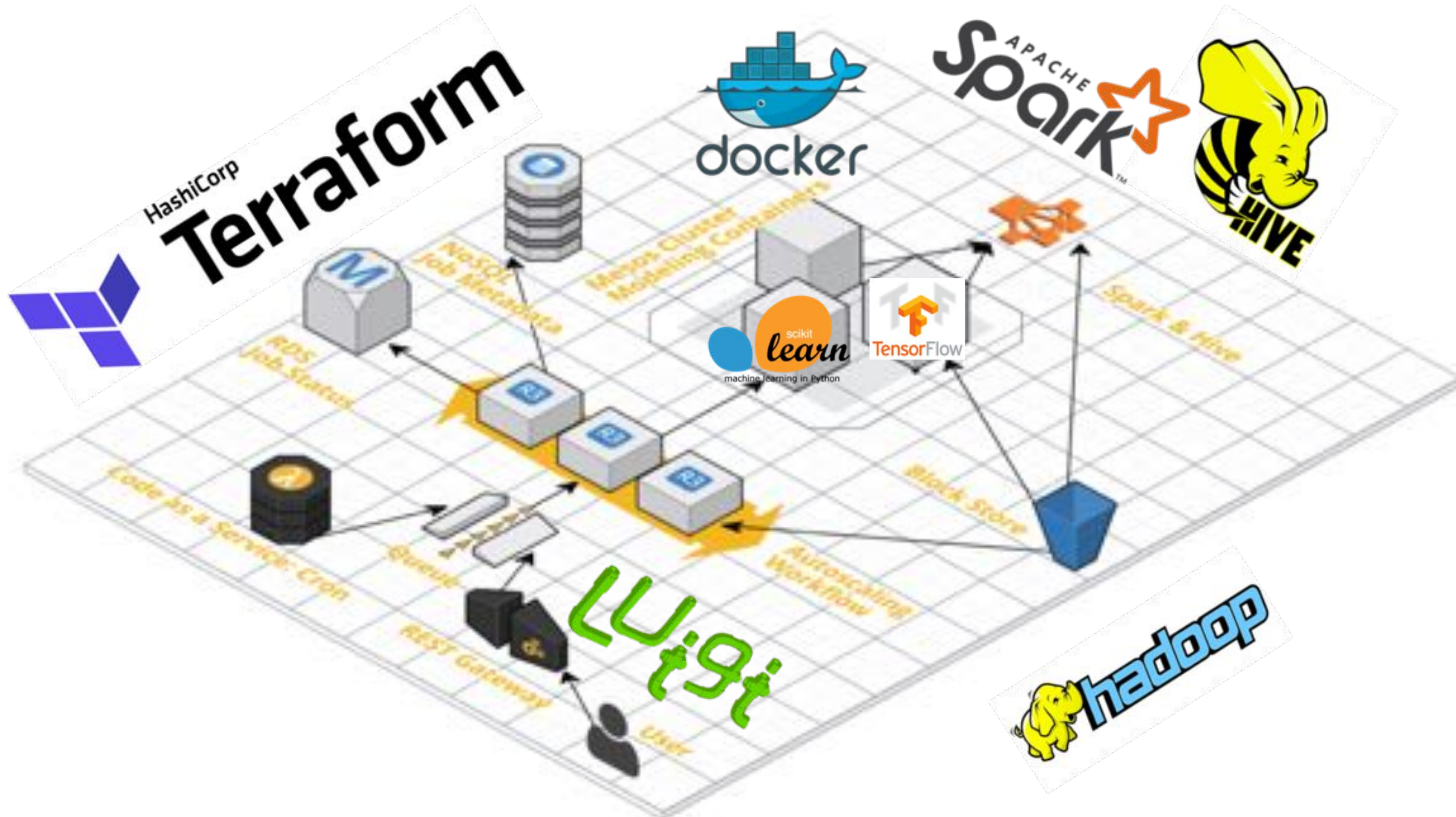


Autoencoder Features



Z1 Input	Weights
Occ Computer Programmer	0.81
Invest 401k	0.76
Children 16-18 male	0.73
Electronics hdtv	0.72
Occ Finance	0.71
Hobby Self Improvement	0.70
Physically Active	0.69
Own Dog	0.68

What do you do if you can't reduce dimensionality enough?



Why Spark/Hive/Hadoop?

Running multiple models: Hive solution

- Initial Solution: Data Science at the Command line

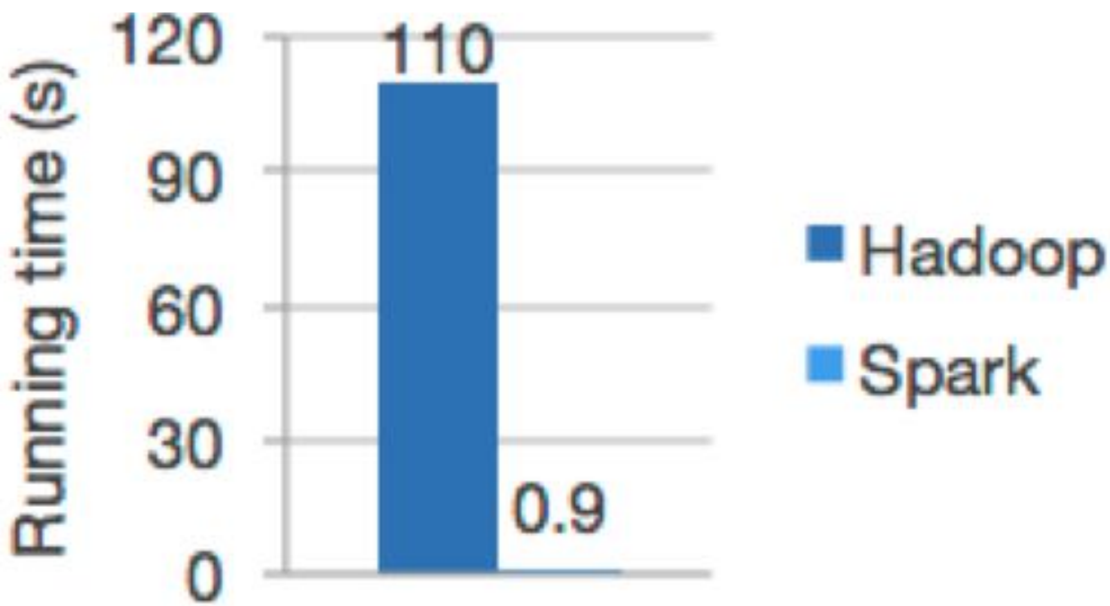
Bourne-again shell (BASH)
HIVE

Data → Model (In Hive) → Repeat

Takes about a week to run. Uses a ton of computer resources.

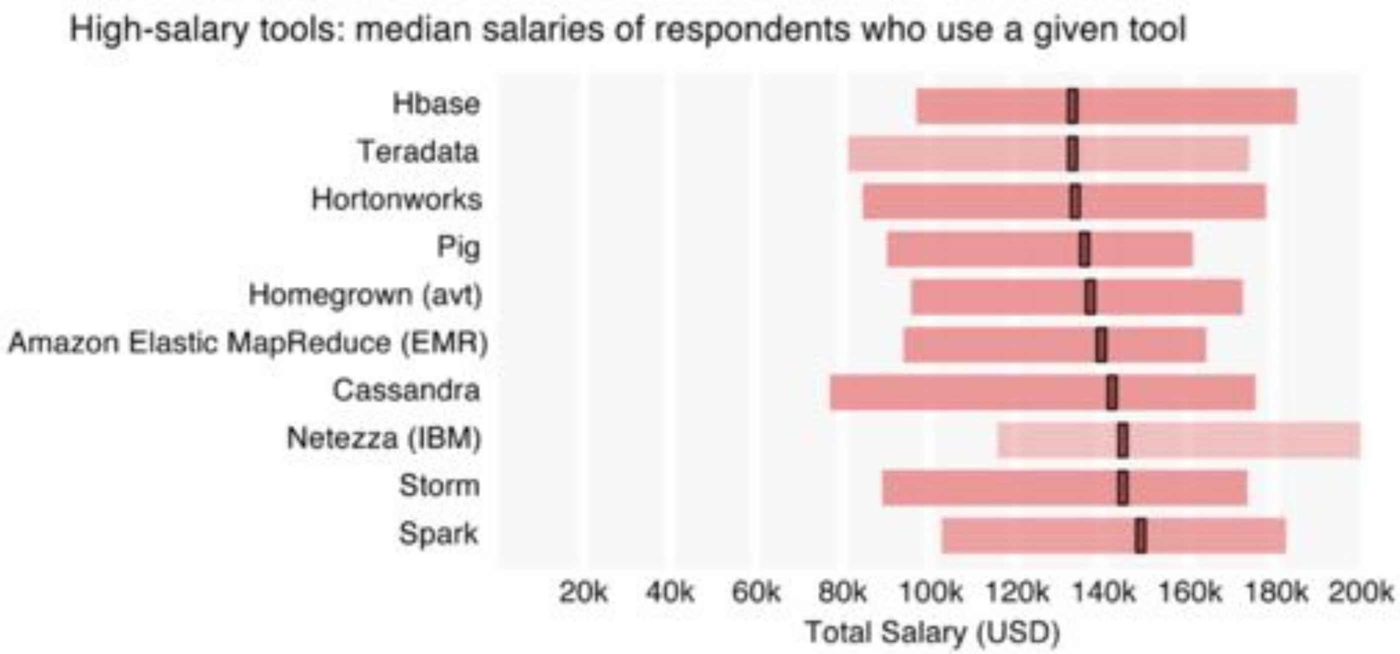
Moving from Hive to Spark

Why Spark?



ORACLE DATA CLOUD

Why else Spark?



http://www.oreilly.com/data/free/2014-data-science-salary-survey.csp

Running multiple models: Spark solution

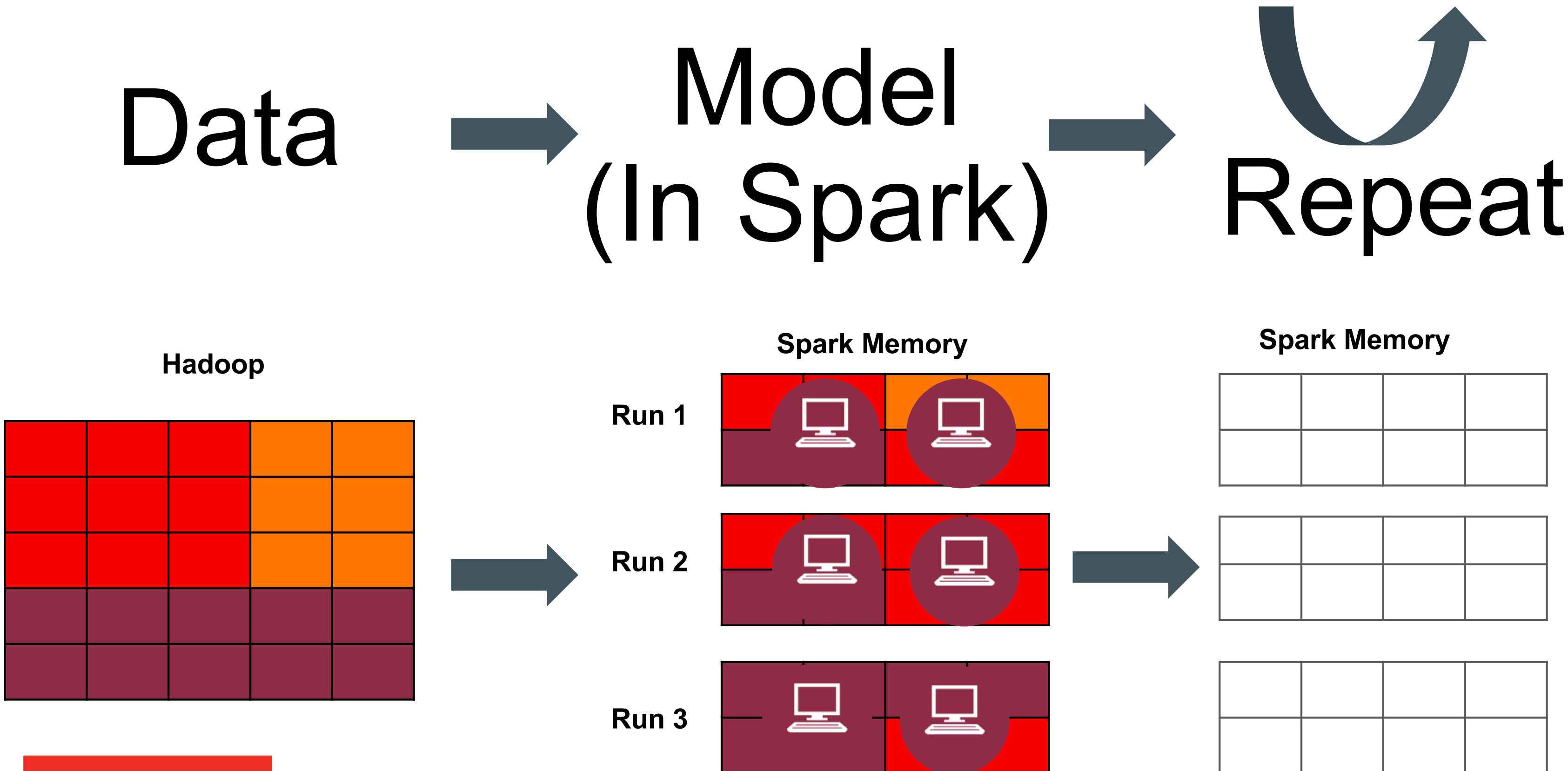
- Novel Solution

All spark

Data → Model (In Spark) → Repeat

Still takes forever to run.

Running multiple models: Spark solution

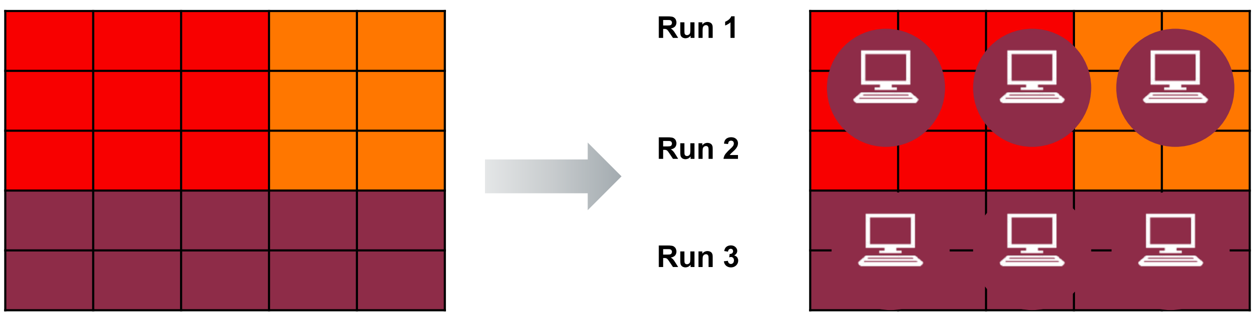


Running multiple models: Spark solution w/caching

Data → Model (In Spark) → Repeat Modeling

Hadoop

Spark Memory Cached

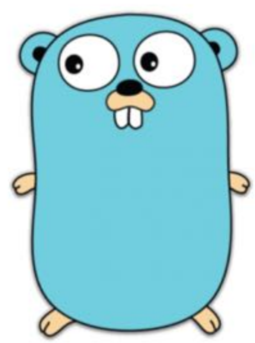


1 Week Process -> 11 Hours

- More computers but way shorter duration
- Faster and Cheaper

Python is great, but not always the answer

FORTRAN



Go

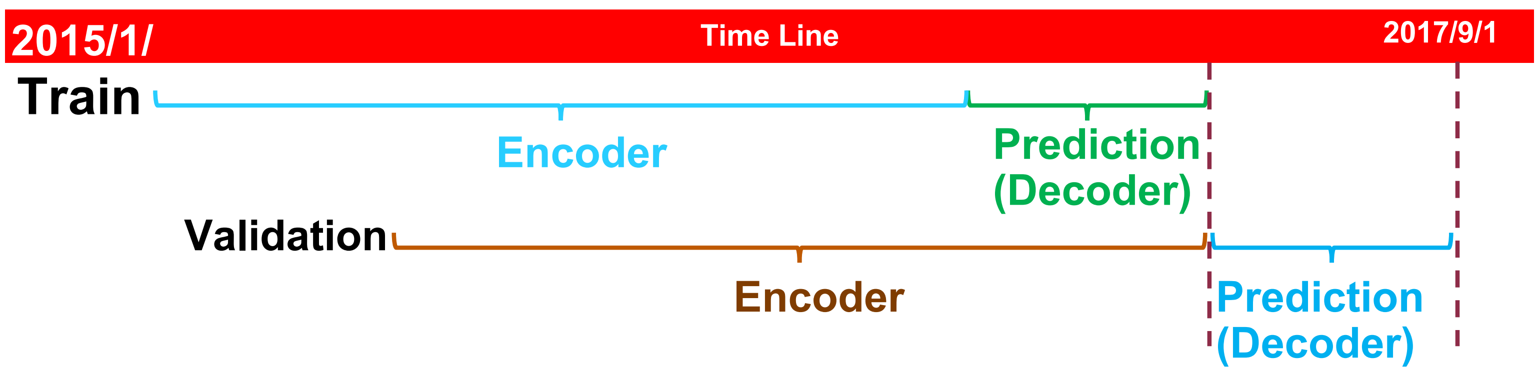
https://github.com/sadovsky/binomial_sim

Targeting beyond buy-alike/look-alike modeling

- Sales cycle / Temporal models
- Contextual targeting

ORACLE DATA CLOUD

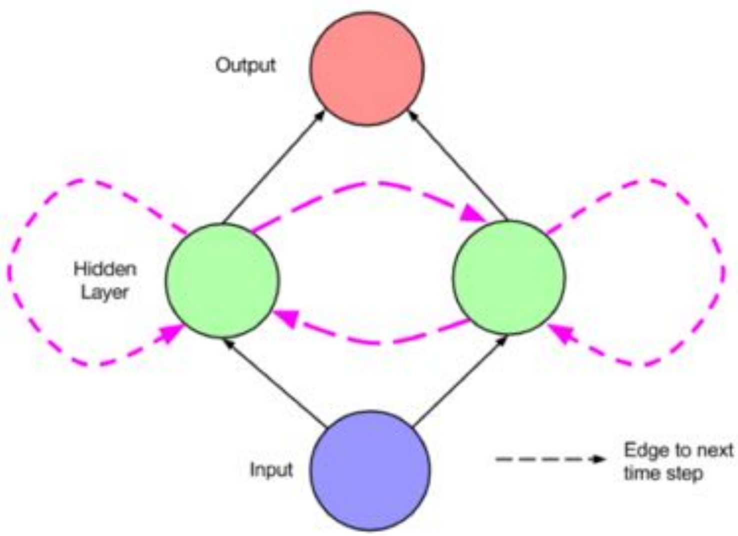
Predictive HH Purchase Forecasting



Work inspired by:

Laptev, Nikolay et al. “Time-series Extreme Event Forecasting with Neural Networks at Uber.” (2017).
Zhu, Lingxue and Nikolay Laptev. “Deep and Confident Prediction for Time Series at Uber.” *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*(2017): 103-110.

We need our networks to remember



ORACLE DATA CLOUD

LSTMs

Input gate learns when to let activation into the internal state

Output gate learns when to let value out of the internal state

When both gates are closed, the activation is trapped in the memory cell, neither growing nor shrinking, nor affecting the output at intermediate time steps

Constant error carousel enables the gradient to propagate back across many time steps, neither exploding nor vanishing

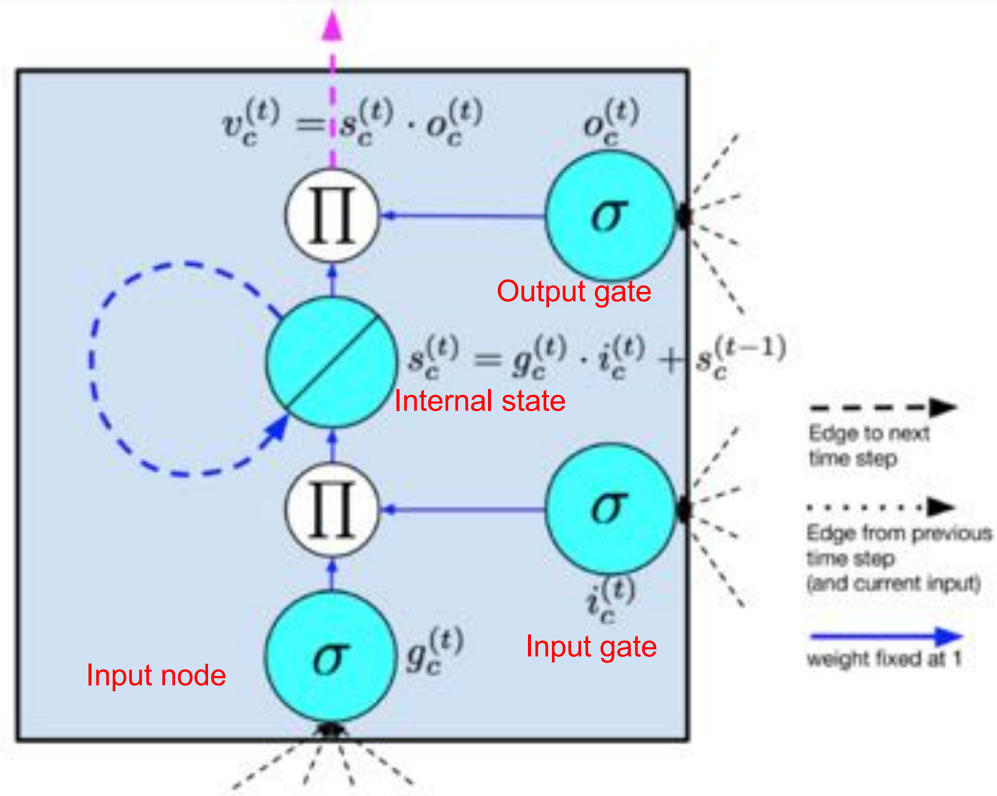
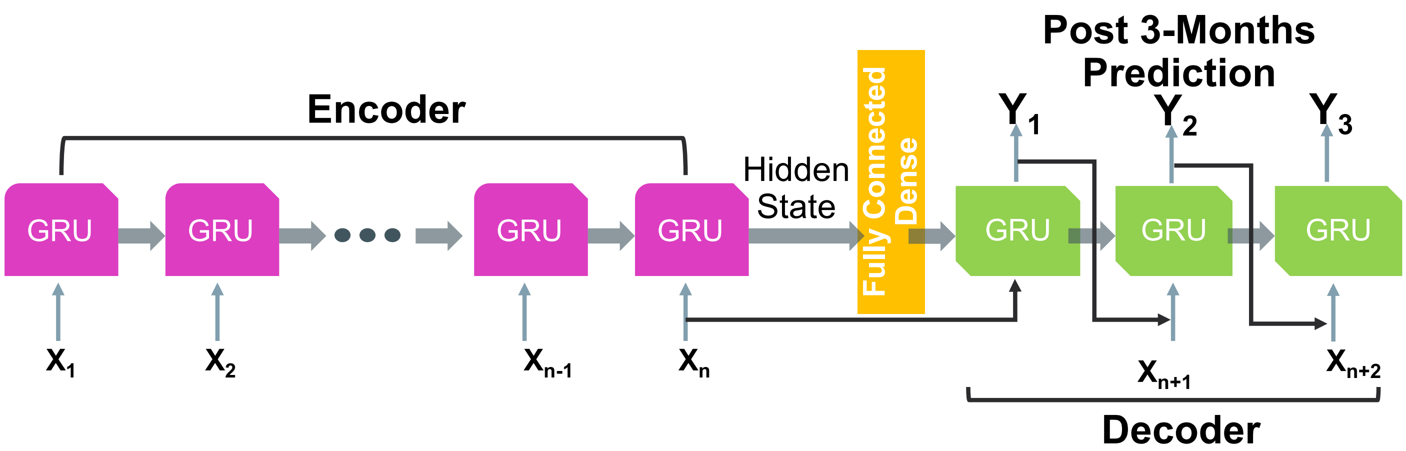


Figure 9: One LSTM memory cell as proposed by Hochreiter and Schmidhuber [1997]. The self-connected node is the internal state s. The diagonal line indicates that it is linear, i.e. the identity link function is applied. The blue dashed line is the recurrent edge, which has fixed unit weight. Nodes marked Π output the product of their inputs. All edges into and from Π nodes also have fixed unit weight.

Predictive HH Purchase Forecasting

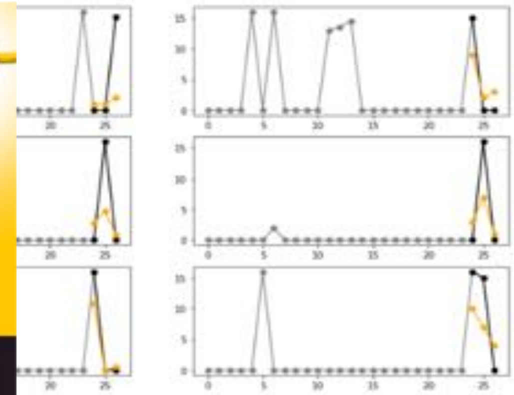


- Teacher-Forcing is used during model training
- X_n is monthly input features
- Y_n is monthly sales prediction
- n = 24, two-year worth of monthly sales data was used for encoder

ORACLE DATA CLOUD

Predictive HH Purchase Forecasting

Lagging Largest-Sale Month Accuracy is
For the households that have purchases ove
prediction period, the predicted largest-sale
within one-month distance to true largest-sal
The predicted largest-sale month has to pre
largest-sale month



- Households are ranked
three-month predictive
- Baseline: ranking buye
pre-period sales (pre-p
monthly sales data)

...racy	Baseline
...9%	0.28%
...0%	1.31%
...1%	4.93%

ORACLE DATA CLOUD

Grapeshot: Contextual brand safety

11-year-old charged with driving drunk

REUTERS

NEWS ALERTS

Get an alert when there are new stories about:

- Orange Beach, Alabama
- Perdido Key, Florida
- Chevrolet Monte Carlo

Add Selected Alerts

• More Alerts

ELSEWHERE ON THE WEB

CNN.COM
Explosion rocks besieged mosque in Pakistani capital

ABC NEWS
'Out of the Blue': Do Aliens Exist?

THE CHRISTIAN SCIENCE MONITOR
William Bratton: Lauded chief of troubled LAPD

Fri Jul 6, 3:23 PM ET

MIAMI (Reuters) - An 11-year-old girl was charged with drunken driving after leading police on a chase at speeds of up to 100 mph that ended when she flipped the car in an Alabama beach town.

A video camera in the police car captured the look of surprise on the officer's face when he approached the wrecked car and got a look at the motorist.

ADVERTISEMENT

The Mobile Press-Register newspaper said the patrolman saw the Chevrolet Monte Carlo speeding and flashed his lights to signal the driver to stop. Instead, the car sped faster, traveling at up to 100 mph (160 kph) before sideswiping another vehicle and flipping over in the Gulf Coast town of Orange Beach, Alabama, on Tuesday night.

The young driver, who lived nearby in Perdido Key, Florida, was treated at a hospital for scrapes and bruises and released to relatives. Police also charged her with speeding, leaving the scene of an accident and reckless endangerment.

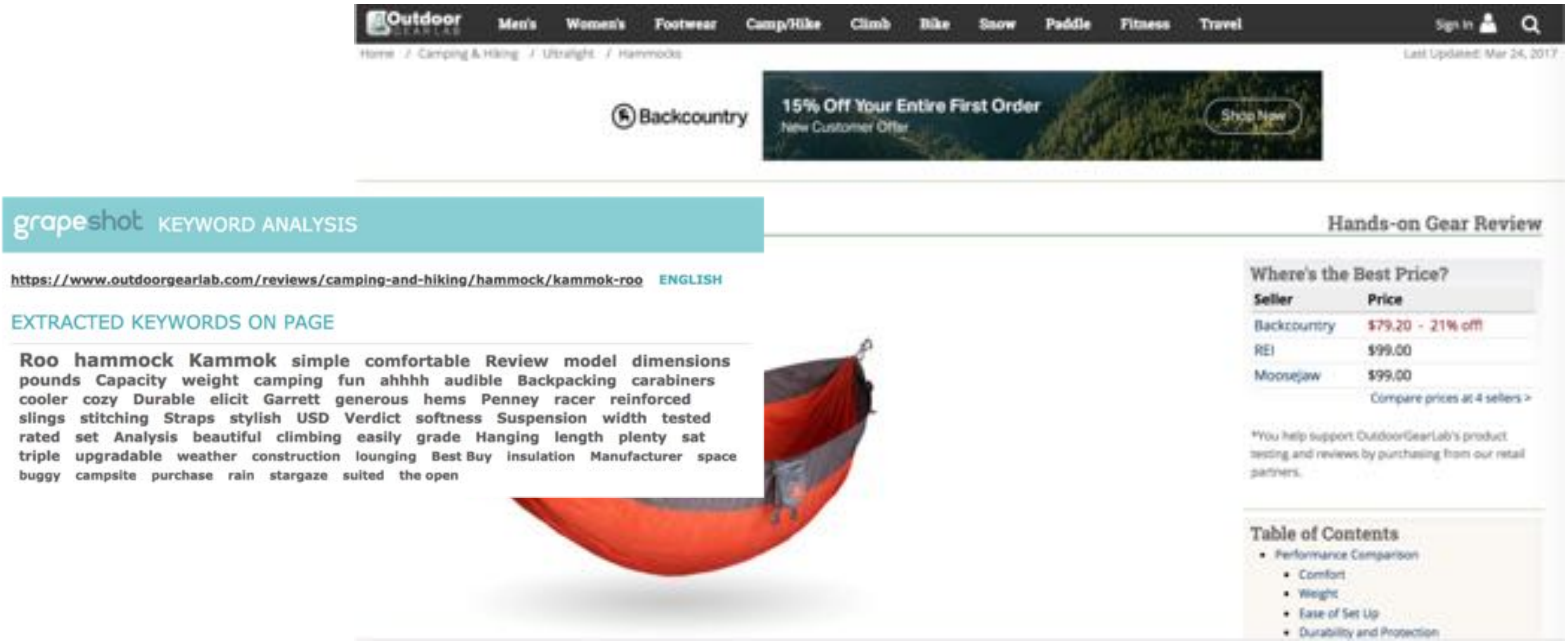
The car belonged to a relative and police were still trying to find out where she got the alcohol. There was none in the vehicle but her blood alcohol level was over the limit for adult motorists, police told the newspaper.

Email Story IM Story Printable View

ORACLE DATA CLOUD

Grapeshot Part Deux

Grapeshot Part Three



ORACLE DATA CLOUD

Measurement

Data

Identity

Targeting

Measurement

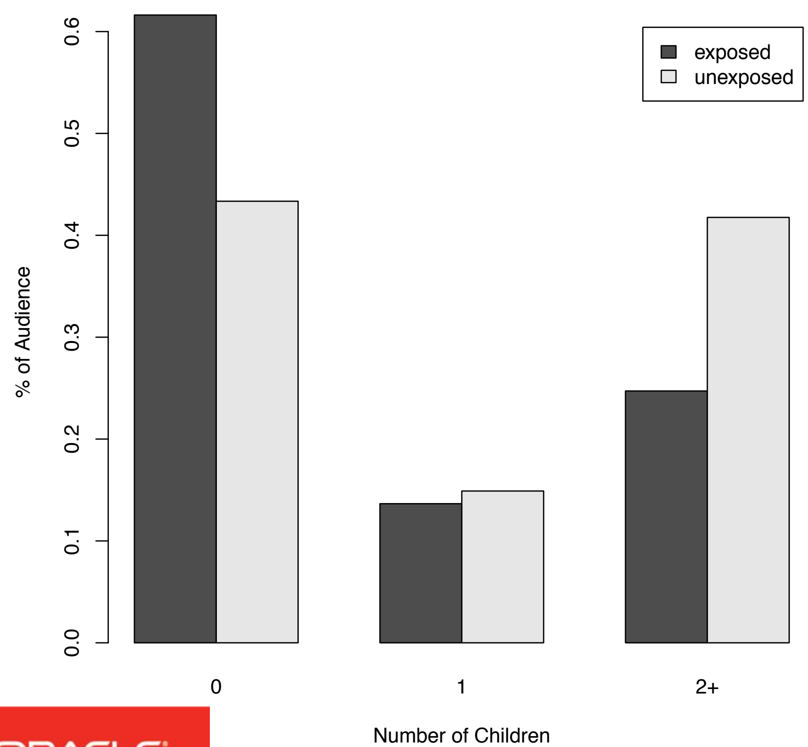
Did our targeting ***cause*** consumers to go out and buy?

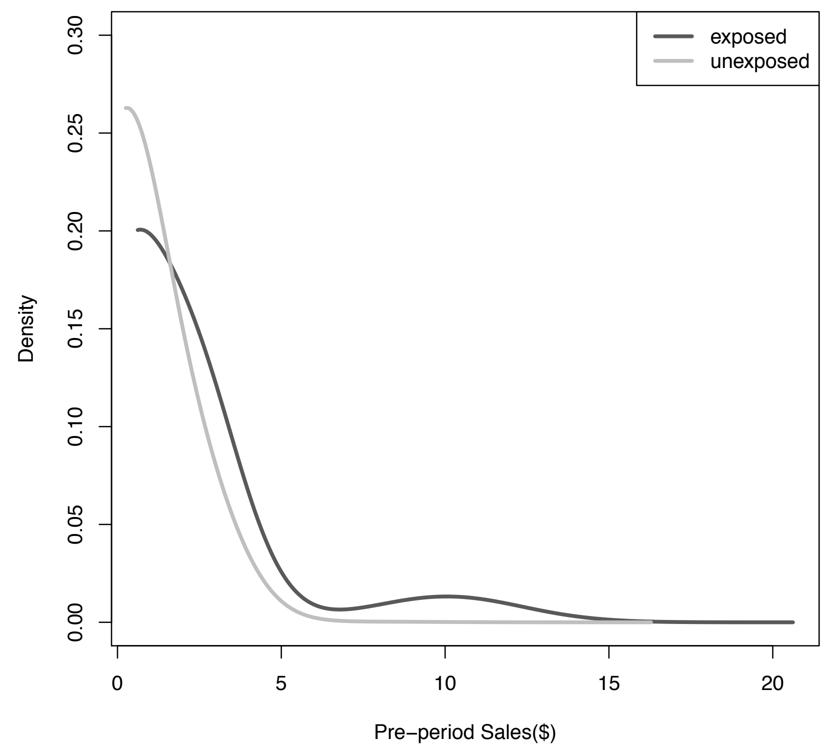
Ideally, we want to do a careful test/control experiment…

… but we can't actually do this!

ORACLE DATA CLOUD

Problem: Exposed vs un-exposed is biased





Solution: Forensic Control

A **forensic control** weights the unexposed households to be equivalent to the exposed households

1. Stratification

Create smaller groups (strata) of households with similar characteristics (based on spend and variables related to exposure)

2. Propensity Model

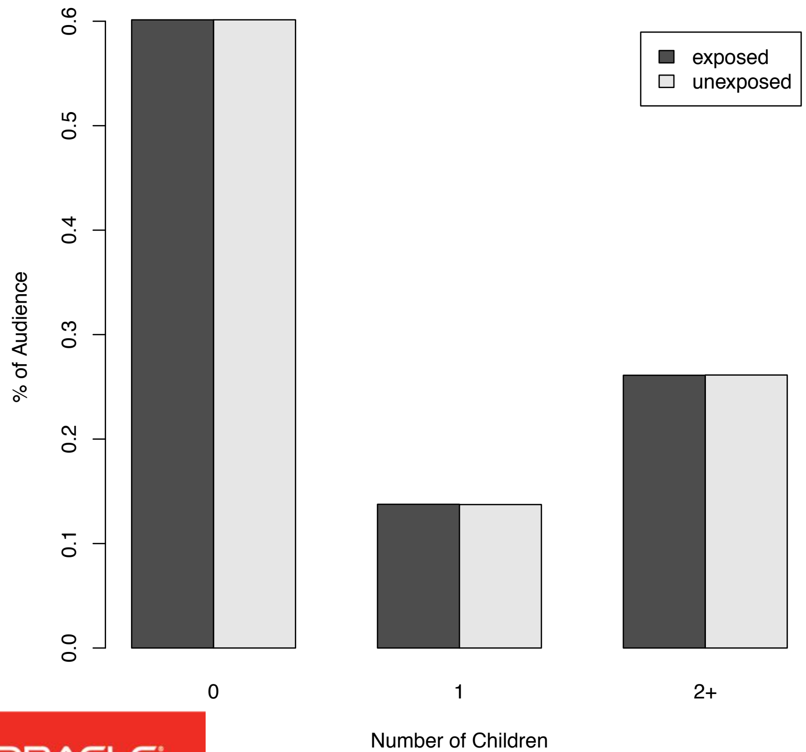
Understand what characteristics define the exposed group

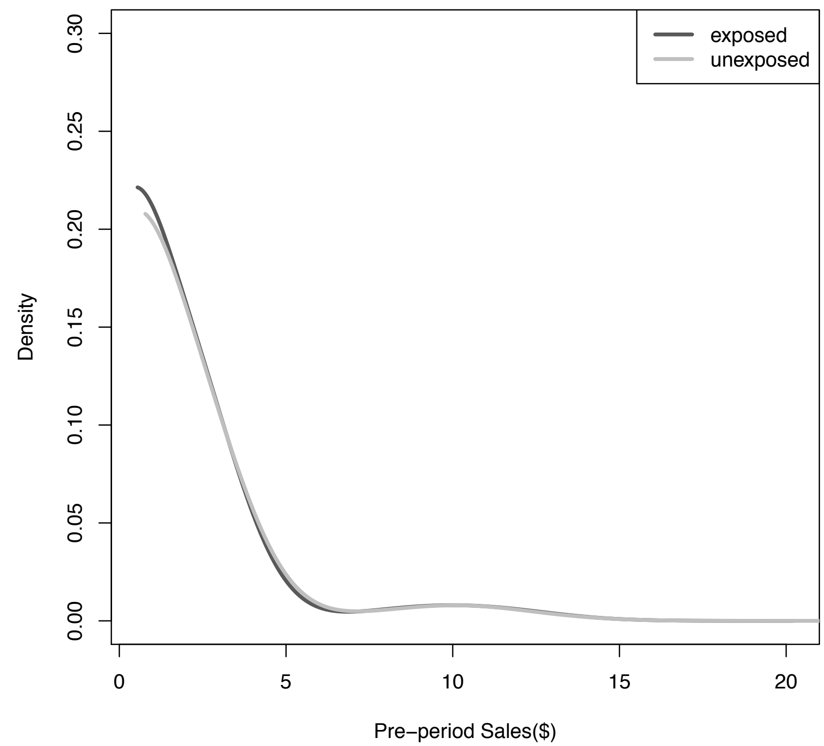
3. Weighting via entropy balancing

For every strata level, use entropy balancing to ensure equal sample size and equal mean, variance, skew for pre-period spend, and likelihood of exposure.

ORACLE DATA CLOUD

Exposed vs. control (after weighting)





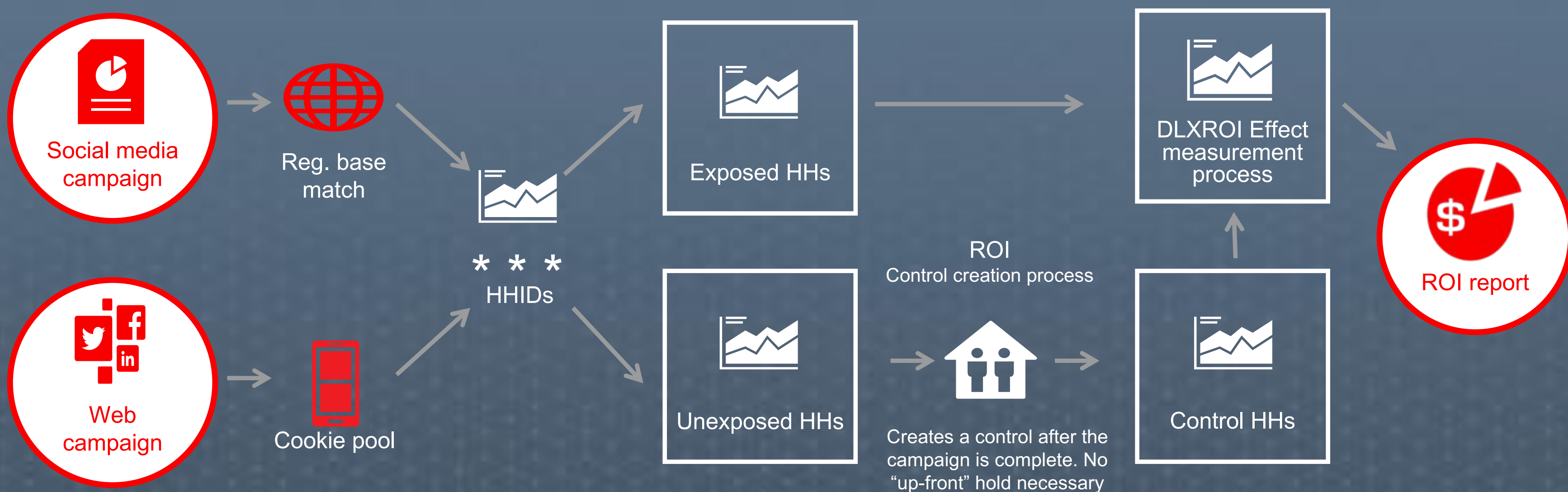
Lift estimation

1. Using only **control** households, we model and estimate post-period sales, given **ALL** available pre-period data.
2. Using this model, compute predicted post-period sales for **BOTH** the control households and the exposed households.
3. Use **the residuals** (observed sales – predicted sales) from step 2 to calculate differences between exposed and control households.
4. The observed **difference in the residuals** between control and exposed is the campaign lift!

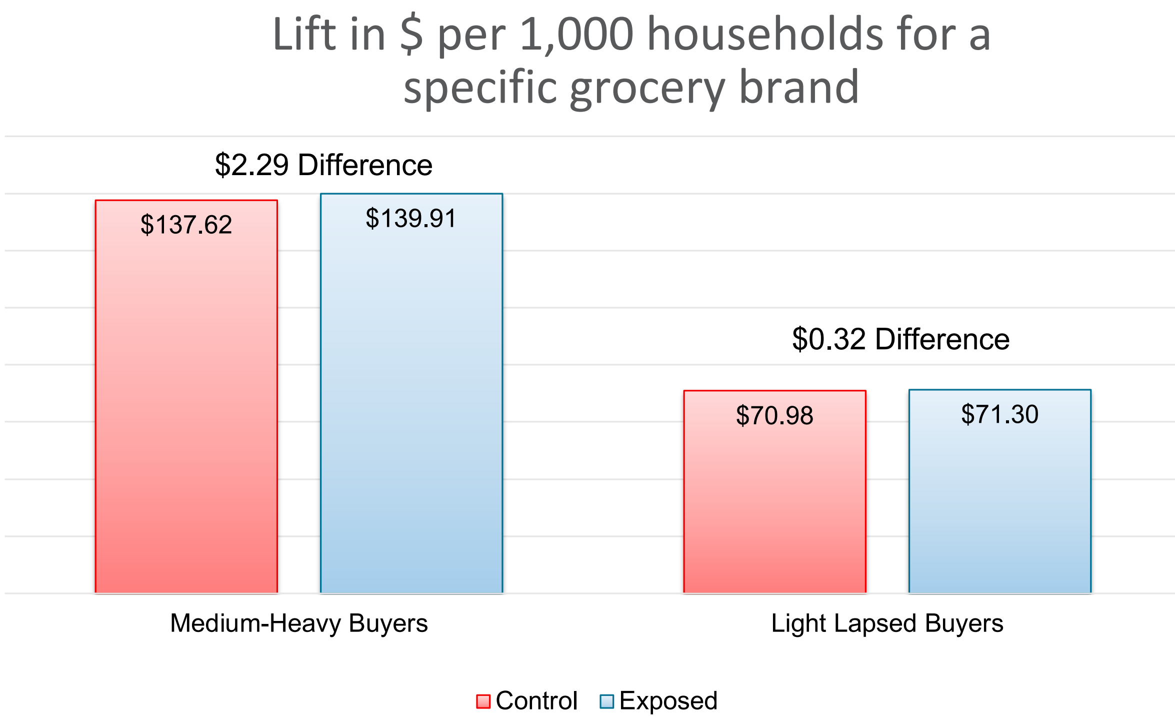
A/B testing, without needing an upfront holdout!

ORACLE DATA CLOUD

Measurement



Yes! Ads (can) have a causal effect on consumers!



Marketing can’t quite brainwash, but it can make an impact

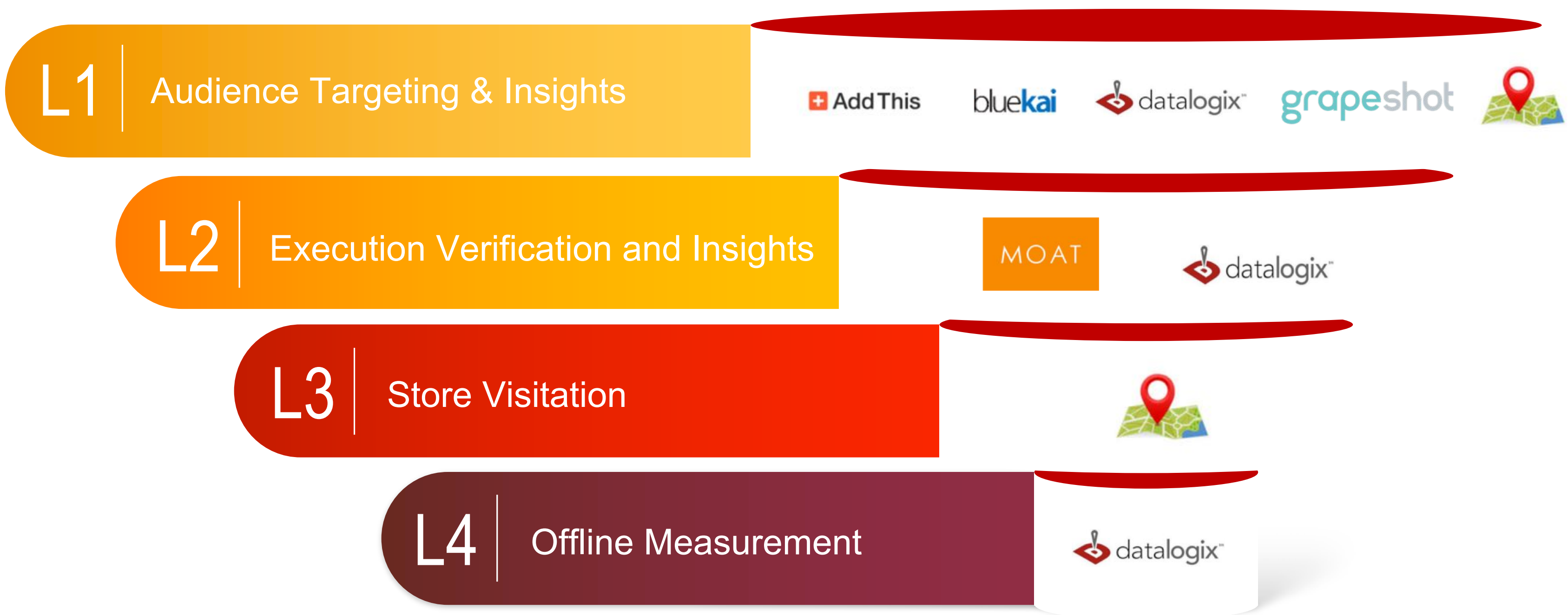
- The value lies in the scale of advertising audiences
- If 33 Million individuals see an ad, and every 1000 ads generate $2.29 of additional sales…

$75,570 of immediate, proven, incremental sales

ORACLE
DATA CLOUD

ORACLE
DATA CLOUD

A Better Serviced Marketing Funnel



Questions?

alex.sadovsky@oracle.com

ORACLE
DATA CLOUD