

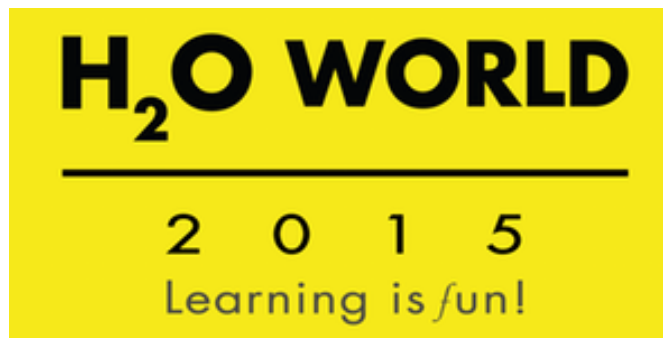


# Advanced Analytics at Macys.com

---

November 11, 2015

*Macys.com Advanced Analytics*



- Big data analytics and data scientists
- Challenges and solutions of big data predictive modeling
- Macy's Advanced Analytics Team
- Our analytics projects
- Personalized site recommendations
- Response propensity models
- Best practices of analysts and modeling

## Two type of data scientists

---

- DATA scientist” for Big Data infrastructure  
Data collections, data processing, pattern recognition,  
data access, data presentations  
Infrastructure and tool builders
- “data SCIENTIST” for Big Data domain problems  
Bid data domain solutions, data driven insights,  
what, how and why  
The science and art of use of big data and tools
- Big Data is not about data, it is about Big Analysis and Solutions
- Division of expertise is inevitable
- Focus, domain knowledge, specialization

# Modeling in the Big Data era

---

## ■ Challenges:

- ✓ Modeling needs to scale
- ✓ Timeliness of models
- ✓ It takes time to integrate
- ✓ Test and Experimentation

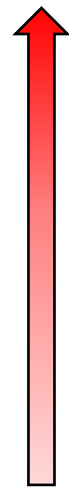
## ■ Solutions:

- ✓ Big data warehouse solutions
- ✓ Separation of concerns
- ✓ Scalable modeling tools
- ✓ Best practices in modeling



# Separation of concerns

Data  
Science



- Solution complexity
- Data complexity
- Variability of requirements
- Standard data mining algorithms
- Availability
- Reliability, scalability, latency
- CPU, help and disk IO issues



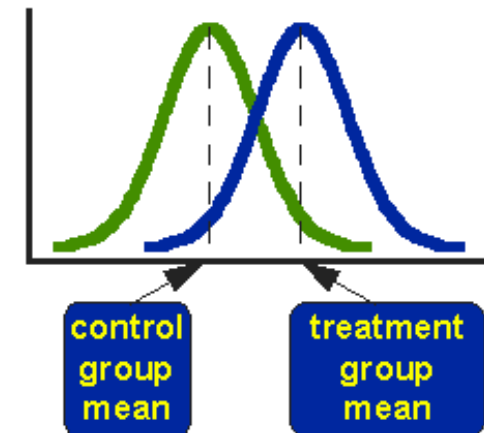
Platform  
Engineering



# Test and Experimentation

---

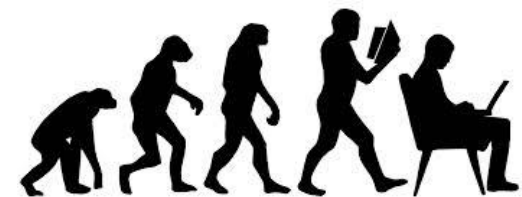
- Customer response behavior is complex
- Theory may or may not be right
- New layouts, new models, new messages
- Split traffic tests
- Find the winners, and gain learning
- Often there are test design problems and understanding their implications



## It takes time to integrate

---

- Make sure the right data are collected
- Measurement and attribution
- Start conversations about model based decisions
- Teams need to think in model metrics
- Organization needs to adapt
- Accumulate assets of creatives, best practices



# Scalable modeling tools

---

- Out of sample testing, cross validation
- Fast and scalable modeling algorithms
- Model comparisons and selections, model management tools
- Automated model optimization tools
- Penalize models being unnecessarily too large
- Ensemble models
- Robust models, handling missing variables, and outliers
- Convenient model building environment
- Graphical tools
- Model deployment tools

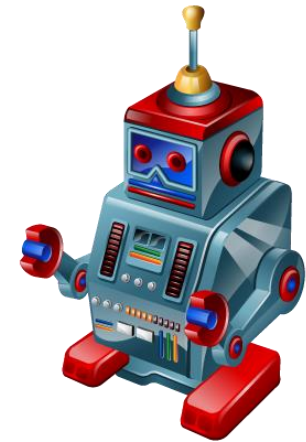




# Best practice big data modeling

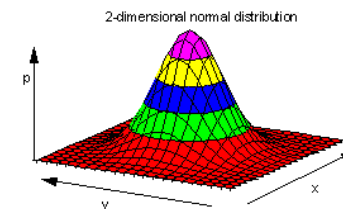
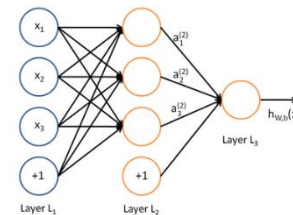
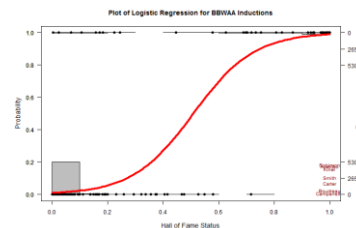
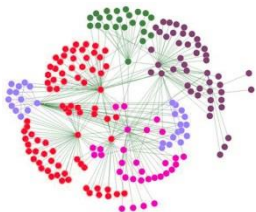
---

- Understand how the data are collected, what data can and cannot be collected
- Balance cost of collecting data and optimize modeling
- Model performance depends on quality of data
- Use automated, robust model building solutions
- Use feedback loop to test hypotheses
- Do simulations to see if changes are reasonable
- Good ideas are not necessarily complicated
- Focus on domain knowledge, not just data mining tools



# Macy.com's Advanced Analytics

- We are at the frontiers of Big Data science
- We have predictive modeling, experimental design and data science teams
- Our team members have very strong background in
  - ✓ Quantitative fields, math, stat, physics, bioinformatics, decision sciences, and computer science
  - ✓ We collaborate with systems and IT teams internally as well as 3<sup>rd</sup> party vendors like SAS Research, IBM Research, ...
- We use a wide range of tools Hadoop, SAS, SAP, H2O, R, Mahout, Spark, and others
- We are data scientists with keen focus on domain problems



# Customer acquisition and retention

---

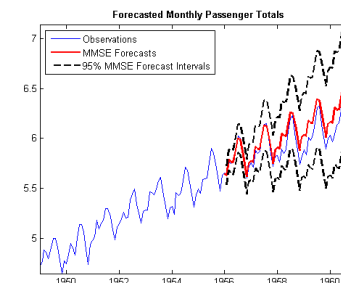
- Targeting the right message to the right customer at the right time
- Build predictive models of purchase behavior and identify drivers
- Customer retention modeling and identify drivers
- Site recommendation algorithms
- Most work is in batch mode, expanding slowly into real time
- Rapid-prototyping and testing of algorithms and policies
- Output of the team's work support other marketing teams to identify, and reach best customers



## Some other projects

---

- Data organization or data munging
  - ✓ Data collections, individual and event level, 360 degrees, ...
  - ✓ Segmentation of customers
  - ✓ Customer value
  - ✓ Multiple channel attribution
- Experimentation platform
  - ✓ Both for site layout as well as contents and recommendations
- Forecast and optimization
  - ✓ Prediction, simulation, and search and optimize
- Big data refinement and scalability
  - ✓ Find new data sources, more efficient ways of accessing data, and organizing and processing data



# Macys.com's real time site personalization



★ Add to Bag x ★ Add to Bag x

www1.macys.com/bag/atbpage/?upcd["32429537"]]=1&source=PDPA2B&prodSelectionInfo=cmexplore=-\_569\_-1\_-SINGLE%20IT

the magic of macy's


Search or enter web ID

my bag (2)

FOR THE HOME BED & BATH WOMEN MEN JUNIORS KIDS BEAUTY SHOES HANDBAGS & ACCESSORIES JEWELRY & WATCHES SALE

mother's day gift guide wishlist deals & promotions gift cards wedding registry

1 item added to your bag >



Tommy Hilfiger Tan Linen with Lapel Piping Sport Coat

Color: Light Grey  
Size: 36R

Reg. \$295.00  
Sale \$99.99  
You just saved \$195.01

item(s) in bag: 2


subtotal: \$199.98


you saved: \$390.02


< keep shopping

checkout >

you might also like...

 Michael Michael Kors Light...  
Reg. \$295.00  
Was \$144.99  
Sale \$119.99  
★★★★★ (1)

 Polo Ralph Lauren Shirt...  
Orig. \$98.00  
Now \$79.99  
★★★★★ (2)

 Denim & Supply by Ralph...  
Reg. \$49.00  
Sale \$43.99

shop blazers & sport coats

shop casual button-down shirts

shop shoes

# Customer segmentation

**Demographic**

**Socio-economic**

**Behavioral**

**Values and styles**

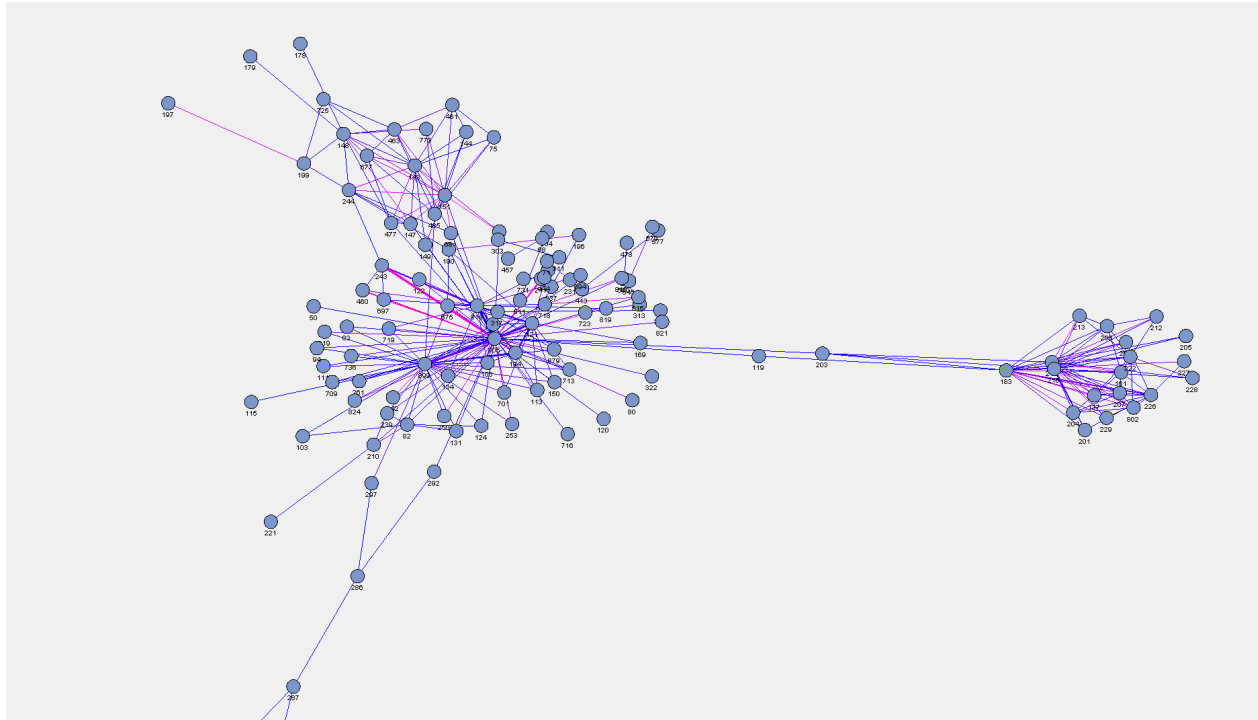
**Channels**

**Brand**





# Product social network



**Demographic**

**Style**

**Size**

**Brand**

**Price range**

**Season**



# Text mining for similarity

Augment data

Differentiation


Generalization






# Who gets which email?



 **CC - Private Sale & Clearance: extra 25% off new favorites!**  
Macy's Men to: daqing.zhao 05/14/2014 05:26 AM [Hide Details](#)

From: Macy's Men <shop@email.macys.com>  
To: daqing.zhao@macys.com,  
Please respond to Macy's Men  
<shop-ctg1aecdabw6papr4nlzjxgyhahcq@email.macys.com>

---

 Discover essential styles just for you at our Private Sale + extra 25% off! promo code: **FORYOU** [shop men](#)  
Can't see the images in this email? [Click here](#)

FOR THE HOME BED & BATH WOMEN MEN JUNIORS KIDS BEAUTY SHOES HANDBAGS & ACCESSORIES JEWELRY & WATCHES SALE

[the gift guide](#) stores deals & promotions gift cards wedding registry

A PRIVATE OFFER—JUST FOR YOU!

**EXTRA 25% OFF**

your men's clothing & shoe purchase of \$100 or more.  
promo code: **FORYOU** [exclusions & details](#)

**FREE SHIPPING & NEW! RETURNS EVERY DAY!**

Free Shipping with \$99 purchase.  
Free Returns by mail or in-store.  
U.S. only, other exclusions apply.


**ENDS SUNDAY!**

**MEN'S PRIVATE SALE & CLEARANCE**

SHOP ONLINE & IN-STORE MAY 14-18.

[shop now](#) also shop: [big & tall](#) [guys](#)

[get savings pass](#) [find a store](#)



| SEE HOW MUCH YOU CAN SAVE:   | Regular Price | Sale Price | Final Cost | Total Savings   |
|--|---------------|------------|------------|-----------------|
| <b><u>SUITS &amp; SUIT SEPARATES</u></b>                               |               |            |            |                 |
| Select Suit Separates from Michael Michael Kors & Lauren Ralph Lauren. | \$550         | \$329.99   | \$247.50   | <b>\$302.50</b> |
| Select Tommy Hilfinger Suit Separates                                  | \$550         | \$274.99   | \$206.24   | <b>\$343.76</b> |

 Discover essential styles just for you at our Private Sale + extra 25% off! promo code: **FORYOU** [shop big & tall](#)  
Can't see the images in this email? [Click here](#)

FOR THE HOME BED & BATH WOMEN MEN JUNIORS KIDS BEAUTY SHOES HANDBAGS & ACCESSORIES JEWELRY & WATCHES SALE

[the gift guide](#) stores deals & promotions gift cards wedding registry

 **FREE SHIPPING & NEW! FREE RETURNS**  
Free Shipping with \$99 purchase. Free Returns by mail or in-store, excludes furniture & mattresses. U.S. only, other exclusions apply.

**ENDS SUNDAY: A SPECIAL OFFER—JUST FOR YOU!**

**BIG & TALL PRIVATE SALE & CLEARANCE**

**EXTRA 25% OFF**

your men's clothing & shoe purchase of \$100 or more. promo code: **FORYOU**  
[exclusions & details](#)

**SHOP ONLINE & IN-STORE MAY 14-18.**

[shop all big & tall styles](#) [shop all big & tall shoes](#) also shop: [men](#) [guys](#)

[get savings pass](#) [find a store](#)



**FINAL COST**  
**\$33.74-\$44.99**  
WITH EXCLUSIVE SAVINGS  
SALE \$44.99-\$69.99 Reg. \$70-\$125  
select dress & casual pants from Michael Michael Kors & Lauren Ralph Lauren



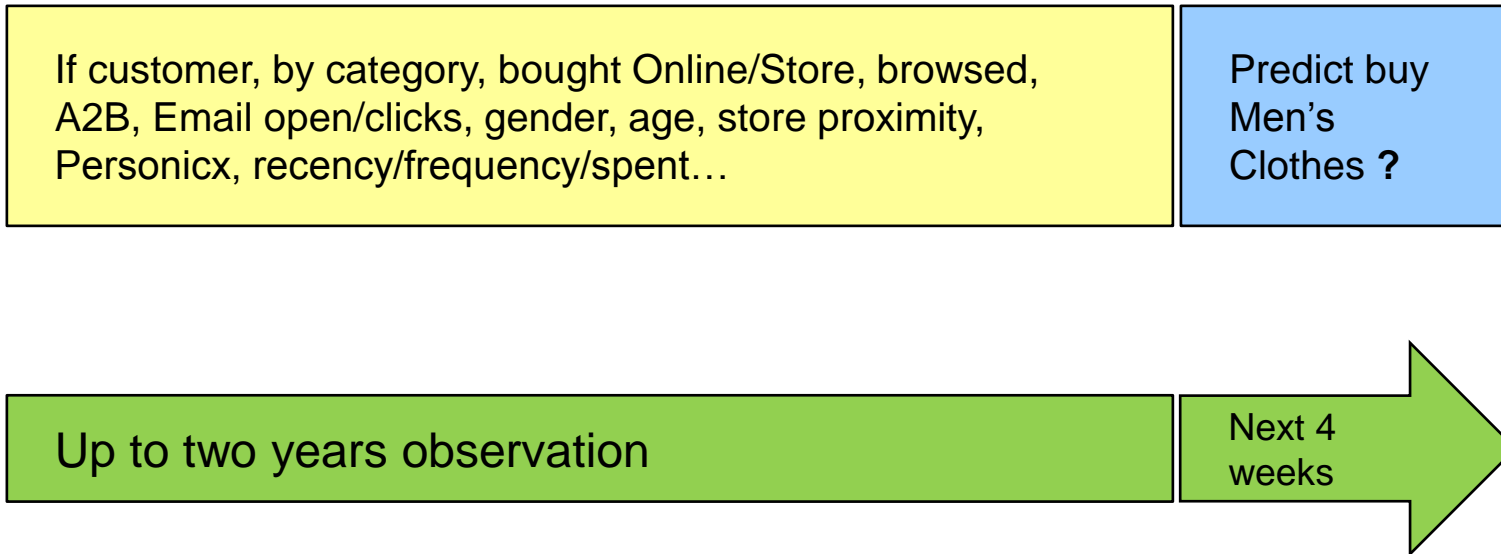
**FINAL COST**  
**\$247.49**  
WITH EXCLUSIVE SAVINGS  
SALE \$329.99 Reg. \$550  
select suit separates from Lauren Ralph Lauren



**FINAL COST**  
**\$23.99-\$41.24**  
WITH EXCLUSIVE SAVINGS  
SALE \$31.99-\$54.99 Reg. \$52.50-\$74.50  
select dress shirts from Tommy Hilfinger & more

# Propensity Models

---



**We are building an expanding family of models, at  
Category/Brand/Outfit...**

# Customer behavioral models

Insights and predictive models and decisions

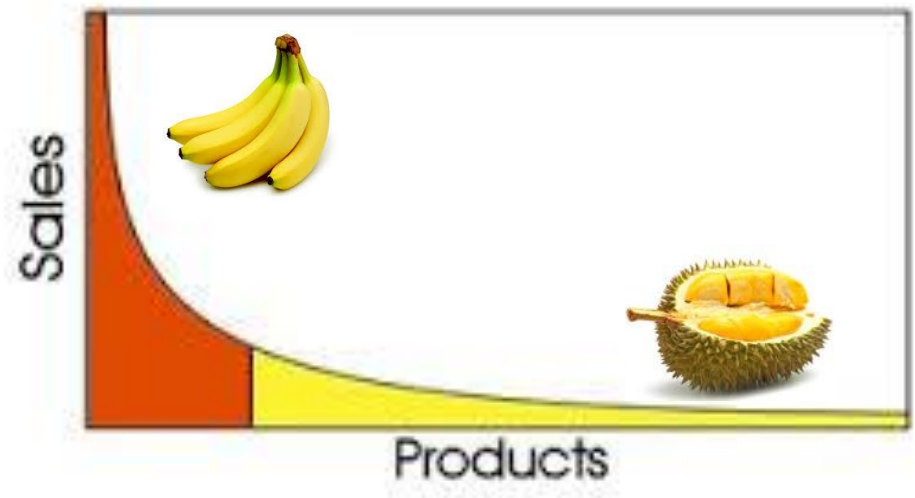
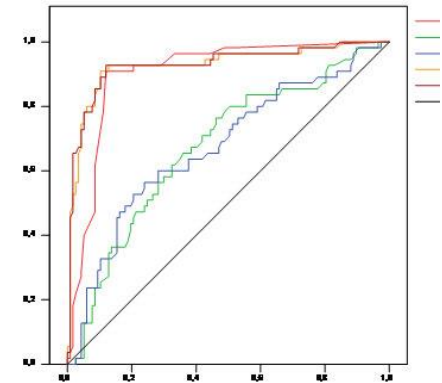
Popular and long tail products

Easy to model popular products

Hard to model long tail niche products

Balance trade offs between  
increasing differentiation  
and reducing noise

Model decisioning using  
business constraints  
and linear programming



## H2O Proof of Concept Project

---

- H2O Demo: 100MM rows and 50 cols, logistic regression on a 16 nodes cluster, finishes in 11 sec
- We provide challenging use case for H2O
- Research cluster of 15 machines
- Customer sample of 7 million, 4000+ columns,
- Test 100 models, R interface, model training in two days
- Model accuracy by ROC AUC distributions, robustness
- Score all customers in several hours



## Considerations of Recommendations

---

- Data sources
  - Data completeness, timeliness
  - Data organization
- Define similarity
  - Look alike in what way
- Cross sell and up sell
- Business rules, assortments
- Repetition
  - Freshness vs burn into memory
- Seasonality
- New product
- Cold start

## Considerations of Recommendations

---

- Intrinsic needs
- Character, belief, value
- Habits, belong
- Style, long term taste, theme
- Status
- “That’s me”
- Familiar, known
- Looking for value, convenience, deal, services
- Trust and service
- Easier to build a profile, preference matching
- Social needs, aspirations
- Trend, fashionable, hot, cool
- Friends, family, neighbors
- Be respected, accepted
- New, impulse
- Explore, learn, serendipity
- Risk taking, find out self
- Less familiar
- Looking for guidance, willing to invest and be influenced
- Trust, authority
- Acquisition modeling
- Hit and miss, nurture, branding

## Concluding thoughts

---

- Big data presents big opportunity and big challenges
- Data science is not about data, but domain solutions
- Modeling in Big Data era is different from traditional practices
- Resolution, robustness, turn around time, timeliness
- Organizations need to adapt to model based decisions
- Data are not clean until thoroughly analyzed
- Scalable and efficient modeling tools are essential

