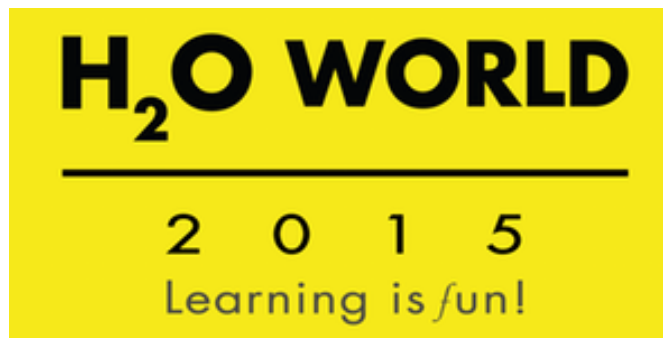# Advanced Analytics at Macys.com

November 11, 2015

Macys.com Advanced Analytics

Agenda

- 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

- 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

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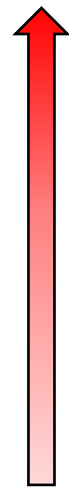- ✓ It takes time to integrate
- ✓ Test and Experimentation

■ **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
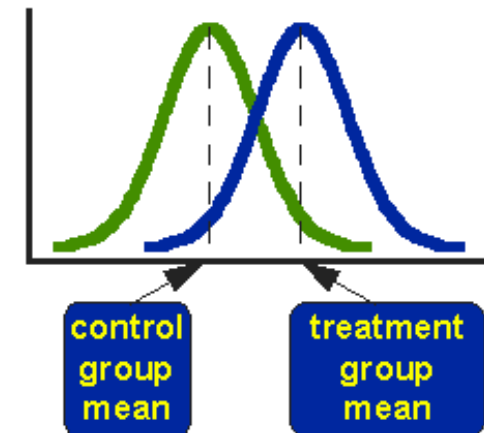- Availability
- Reliability, scalability, latency
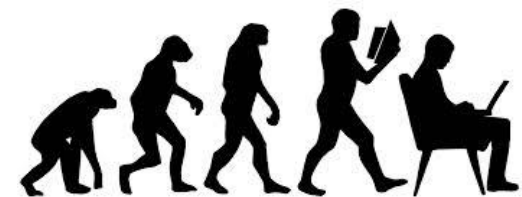- CPU, help and disk IO issues

Platform
Engineering

- 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

- 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
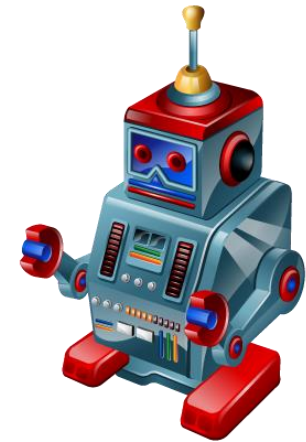
- We are at the frontiers of Big Data science
- We have predictive modeling, experimental design and data science teams
- Our team members have very strong background in
- ✓ Quantitative fields, math, stat, physics, bioinformatics, decision sciences, and computer science
- ✓ We collaborate with systems and IT teams internally as well as 3rd 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
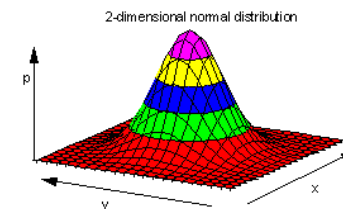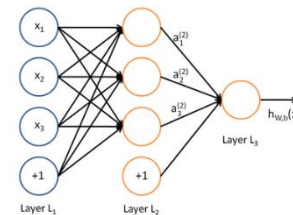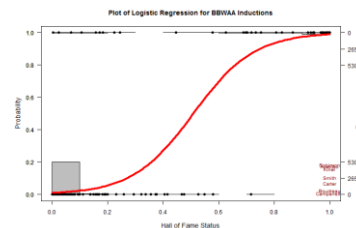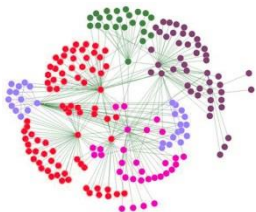
Customer acquisition and retention

- Targeting the <u>right</u> message to the <u>right</u> customer at the <u>right</u> 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

Forecasted Monthly Passenger Totals
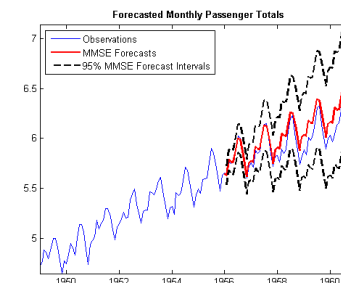
Macys.com's real time site personalization

Customer segmentation

Demographic

Socio-economic

Behavioral

Values and styles
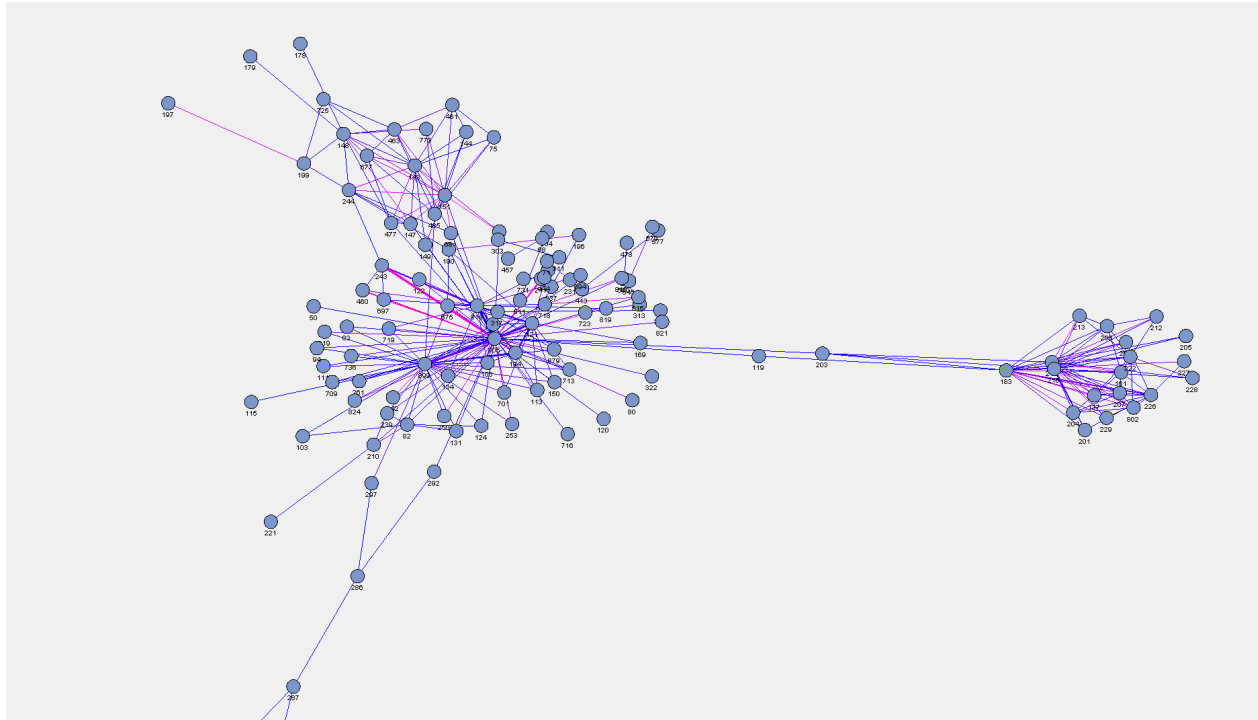
Channels

Brand

Product social network

Demographic

Style

Size

Brand

Price range

Season

Text mining for similarity

Augment data

Differentiation

Generalization

Who gets which email?

Propensity Models

If customer, by category, bought Online/Store, browsed, A2B, Email open/clicks, gender, age, store proximity, Personicx, recency/frequency/spent…

Predict buy Men's Clothes **?**

Up to two years observation

Next 4 weeks

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
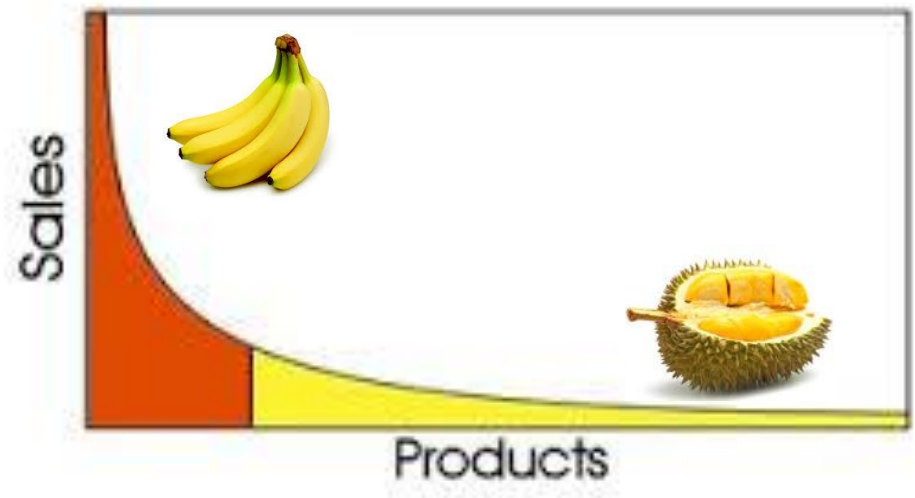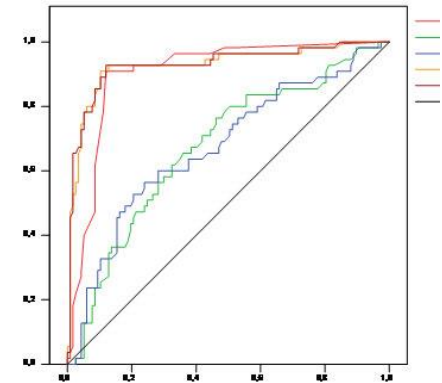
- H2O Demo: 100MM rows and 50 cols, logistic regression on a 16 nodes cluster, finishes in 11 sec

- 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