Analytics for Decision Making

smart people don’t make blind bets
About me

• SDM ’04, AA MS ’06
• Founder and Analytics and Innovation lead at DataXu
  – In 2007, with Prof Ed Crawley, Bill Simmons, launched the company, out of room in building 33
• I’m also a mech eng, from back in the day when you would draw blueprints by hand
  – Yes, I also know how to use a slide ruler
• Worked as consultant for a while (hence the ‘beautiful’ slides)
Founded in 2009; Headquartered in Boston, MA; 300+ employees, 14 offices globally, PetaByte scale dataset

Real-time decision system based on MIT research by Bill Simmons and Sandro Catanzaro (first used for NASA Mars Mission project)

Easy-to-use software platform enabling brands to use Big Marketing Data

5th fastest growth US company

Rated #1 algorithmic optimization technology by Forrester Research

Powering blue-chip brands: Ford, Lexus, General Mills, Wells Fargo, Discover

DataXu is all about applying data science to the art of marketing
Agenda

WHY ANALYTICS

CHALLENGES

OPPORTUNITY

© 2013 Sandro Catanzaro
What about Analytics

- Data is a means to an end: **insight for decision support**
- **Analytics job** is to **bridge** data and insight

- Bain surveys shows **significant advantage** from companies using data to create insight

- Analytics popularity is making it a mainstream discipline

Sources
How analytics differentiates winners – Bain Brief, Sep 2013
Just a trend?

• Declared the new black
  – Symptom of shark jumping?

• Analytics will stay relevant as long as it continues providing the right answer
  – Or the right answer more often than not
Analytics creates fact-based decision support

• Decision making process used to be driven by experience
  – “Wisdom/Intuition” – we have been using human brains to accumulate and process big data

• Data supported decisions outperform “old wisdom”
  – Find non-obvious answers
  – Provide an answer in record time
  – Answer can follows more closely dynamic environments
  – Can be methodical within a specific very large dataset

© 2013 Sandro Catanzaro
Marketing budgets have been driven by intuition-based decisions

**Challenge:**

71% of CMOs said they don’t have the tools needed to master data driven marketing*

The marketing budget allocation problem

Consumers migrating towards digital media are providing a wealth of under-utilized data

**Solution:**

Platform that automatically and in a closed feedback loop:

- Ingests large consumer datasets, including media exposure
- Derives insight about what is effective
- Decides in real time on message and context to talk with consumers

* IBM Global CMO Survey, 2011
A solution that spawn a new industry

Example Automated Analytics Platform

- 6 TB of data per day – processed by 1000s servers, part of a real time synchronized cluster
- 1M decisions per second, each answered in less than 40 milliseconds, across 3 continents
- Over 150 analysis, web-accessible, continuously and automatically refreshed
A solution that spawn a new industry

Example Automated Analytics Platform

Data driven decisioning provides an unfair advantage

- Identifies and executes under-valued opportunities
- Quickly reacts to market changes
- Learns without prejudice about the market

- 6 TB of data per day – processed by 1000s servers, part of a real time synchronized cluster
- 1M decisions per second, each answered in less than 40 milliseconds, across 3 continents
- Over 150 analysis, web-accessible, continuously and automatically refreshed
Agenda

WHY ANALYTICS

CHALLENGES

OPPORTUNITY
Valuable but not trivial

• While, data based decisions do provide best opportunity for success

• Yet, they incur in significant complexity to arrive to answers

• Discipline and commitment are required to meet this complexity challenge
A multifaceted set of issues conspire against success

<table>
<thead>
<tr>
<th>Input</th>
<th>Process</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dirty data</td>
<td>Scaling up the team</td>
<td>Unclear need</td>
</tr>
<tr>
<td>Too much data</td>
<td>New toolset dynamics</td>
<td>Unwillingness to pay</td>
</tr>
<tr>
<td>Far away data</td>
<td>Thinking takes time</td>
<td>Answer not simple enough</td>
</tr>
<tr>
<td>Too little data</td>
<td>R&amp;D type of risk</td>
<td>Output contradicts business</td>
</tr>
<tr>
<td>Legal implications</td>
<td>Market noise – new tools</td>
<td>Politics</td>
</tr>
<tr>
<td>Reputational risk</td>
<td>Threat of disruption from output to input</td>
<td>NIH – fear of change</td>
</tr>
<tr>
<td>Trust and proprietary data</td>
<td></td>
<td>Proving ROI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Little intellectual curiosity</td>
</tr>
</tbody>
</table>

**Technical stakeholders**
**Business – Domain stakeholders**

© 2013 Sandro Catanzaro
The breadth of the issues require a team that is uniquely qualified

- **Valuable insights** take into consideration business, domain, and analytics process...

- ...with just-good-enough technology

- The Analytics Manager main job is to staff a team that can master them all

© 2013 Sandro Catanzaro
The breadth of the issues require a team that is uniquely qualified:

- Valuable insights take into consideration business, domain and analytics process.
- Systems Thinking and Management skills are at the core of the solution.
- The best analytics leaders have industry experience, management skills and reasoning and analytical rigor.
- The Analytics Manager main job is to staff a team that can master them all.
Output issues

- Many problems are not clear
  - Business stakeholders don’t know what can be answered
  - Even worse, the problem is not acknowledged
- A solution is identified before the problem is defined
  - Big data as a panacea
- Non technical stakeholders not willing to invest
- Answer complexity
Process and Input issues

- Good analysts who are willing to see the big picture are scarce
- Tool selection blues
- Analysis “gold polishing”
- Intrinsic risk that the analysis may not work
- Too much data – just in case
- Data collection politics
Reporting vs Analytics

All you need is a “report writer software”

• Analytics = curated report
  – Curation: identifying what matters, editing out what doesn’t
  – Joining the dots

• Analytics is a Bridge to the right insight
  – Not a pile of needles to replace a pile of hay
  – Decisions are made with just THE ONE needle
  – Careful you all detail oriented nation!

© 2013 Sandro Catanzaro
Agenda

WHY ANALYTICS

CHALLENGES

OPPORTUNITY
Analytics is a bridge

A feat of engineering
(art of tradeoffs with incomplete information)

• Goes to a specific point to another specific point
• Good enough foundations (data and question clarity)
• Can be used by most people who want to cross it
• Has explicit durability assumptions
Analytics is a system

Input
- Data
- Data stakeholders

Process
- Technology
- Selection of what matters
- QA
- Cleaning
- Model limits
- Insight
- Aggregates
- Error quantification

Output
- Presentation
- Insight recipients
- P&L owner

Easier to manage interface
Brittle interface

© 2013 Sandro Catanzaro
Types of Analytics

1. Research
   – Unclear need, unclear answer, high risk

2. Development
   – Clear answer and method, but some risks on implementation

3. Operations
   – Good experience with execution. Can be automated
   – QA for Automation becomes a type 1 problem, again

*My company does all three – I personally focus on innovation & analytics, thus #1*
FRAMEWORK
(to make magic work)
Where to start

Clear and present need “you are lucky”

Need is not crisp, but challenge is known “we just need Analytics”

Lack of awareness about challenge “Beware”

Crisp and clear statement of the “question to answer”

**good**

Can you tell us how sales correlates with factors A, B, C, D and E; using a quarter by quarter dataset, and data from 2010 to date

**bad**

Please do some attribution modeling

© 2013 Sandro Catanzaro
The pull model

*Starts from the end*

Dataset
- Collection (data sources, pick, parse cleanup)
- Processing (storage technology, aggregation, processing technology)
- Analysis output (analysis tools)

Is this truly important to you?
Are you willing to pay for it?

Start from the Business problem
Then ensure there is willingness to pay
Then proceed step by step backwards
The pull model

Starts from the end...

...You would cross the “bridge” 3 times

- **First time** from where you are, to the business stakeholder and problem
- **Second time** from the business problem, all the way back to the data, so that every piece “pulls” the next previous one
  - With business problem in hand, you will know the minimum analysis output that provides the answer
  - With the analysis in hand, you will know analysis tools needed to prepare, and also what is the processing and aggregations needed to support it
  - With the aggregations in hand, you will know the minimum dataset required
  - .. All the way to data collection
- **Third time** from the data, you will come back as you execute the analysis, step by step to finally arrive to the desired answer
Continuous iteration

• Share interim results: continue selling on value, continue selling on model belief
• May modify the “crisp question” – ensure you also reframe time, resources, cost
• Whether you’ll share the methodology depends on the audience
• Creates dialog and sense of shared ownership in the answer

Incorporate feedback

- Crisp question statement
- Mockup answer (Answer First)
- Back of the envelope answer
- Complete answer

Share output

(maybe methodology)

Nothing fancy here
Just risk mitigation / project management 101...

© 2013 Sandro Catanzaro
Best models: 70% belief - 30% math

• Simplify

• Share with all shareholders – get them to believe in the answer and model

• Simplify

• Acquire buy in – “prewire meetings”
  – Must have on board the P&L owner
  – Also recruit influencers, other stakeholders

• Simplify again
  – Review and clean up model
  – Remove unneeded items, keep just enough to prove your point
  – Maybe add a single cherry (non-requested-by-so-cool-to-know insight)

Be careful on adding accuracy at the cost of complexity
Scaling the team

• Best bet is for introducing geeks, into big picture thinking
  – Smarts beat expertise
  – “Spin Doctors” are useful, to a point

• Aim for “T shaped” people
  – Broad big picture understanding
  – Deep expertise in one area

• Provide context
  – Make sure team understands where does the analysis fit

• Be willing to let go
  – Understand costs of potential mistakes
  – Responsibility is never delegated

"You do as you're told, we pay as we please. You work like a slave, we punish at random. That, in a nutshell, is our corporate culture."
The Truth

- The value of Analytics team is guiding decision-making using fact-derived insight

- Not always possible
  - Not enough data or tools to arrive to conclusion
  - Business pressure to change the answer
  - The analysis is not ready in time (pesky bugs?)

- The only value of data, derived insight and platforms to create such is to be correct
  - Otherwise, not worth the paper they are written in

Insight that can’t be trusted is worse than flipping coins

© 2013 Sandro Catanzaro
Sandro’s Analytics principles

1. Start with the business/domain problem. Never with data
2. Don’t ever ignore humans – especially the one who pays
3. 70% belief – 30% math
4. Teach smart geeks to think big picture & let go
5. Simplify - weight complexity vs. accuracy
6. Quick and good enough wins the day
7. Model and iterate – rinse and repeat
8. Tackle risks from riskiest to least risky
9. Be flexible in scope, but take victory lap and re-budget
10. Never compromise with “the truth”
system design and management

Thank you!