Data-Driven Product Innovation

Xin Fu
LinkedIn

Hernán Asorey
Salesforce

KDD2015

21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining
10 - 13 August 2015, Hilton, Sydney

WHY?

Data Science Fuels Product Innovation

Example of Data Guiding Product Dev

$$\text{Queries} = \text{Unique Searchers} * \frac{\text{Sessions}}{\text{Unique Searchers}} * \frac{\text{Queries}}{\text{Session}}$$

How Did Data Science Help?

- Metric to define product "true north"
- Tools to help understanding
- A/B experiments to learn and iterate

Data Science's Role in a Product Org

- Model 1: Data Science as an ***Owner***
- Model 2: Data Science as a ***Service***
- Model 3: Data Science as a ***Partner***

Data Science as an **Owner**

- Operates in a "hacky way" (early stage companies)
- Key steps in business rely heavily on data
  - Recommender
  - Relevance
  - Matching
  - Scoring
- Mostly back-end, relatively more stand-alone features

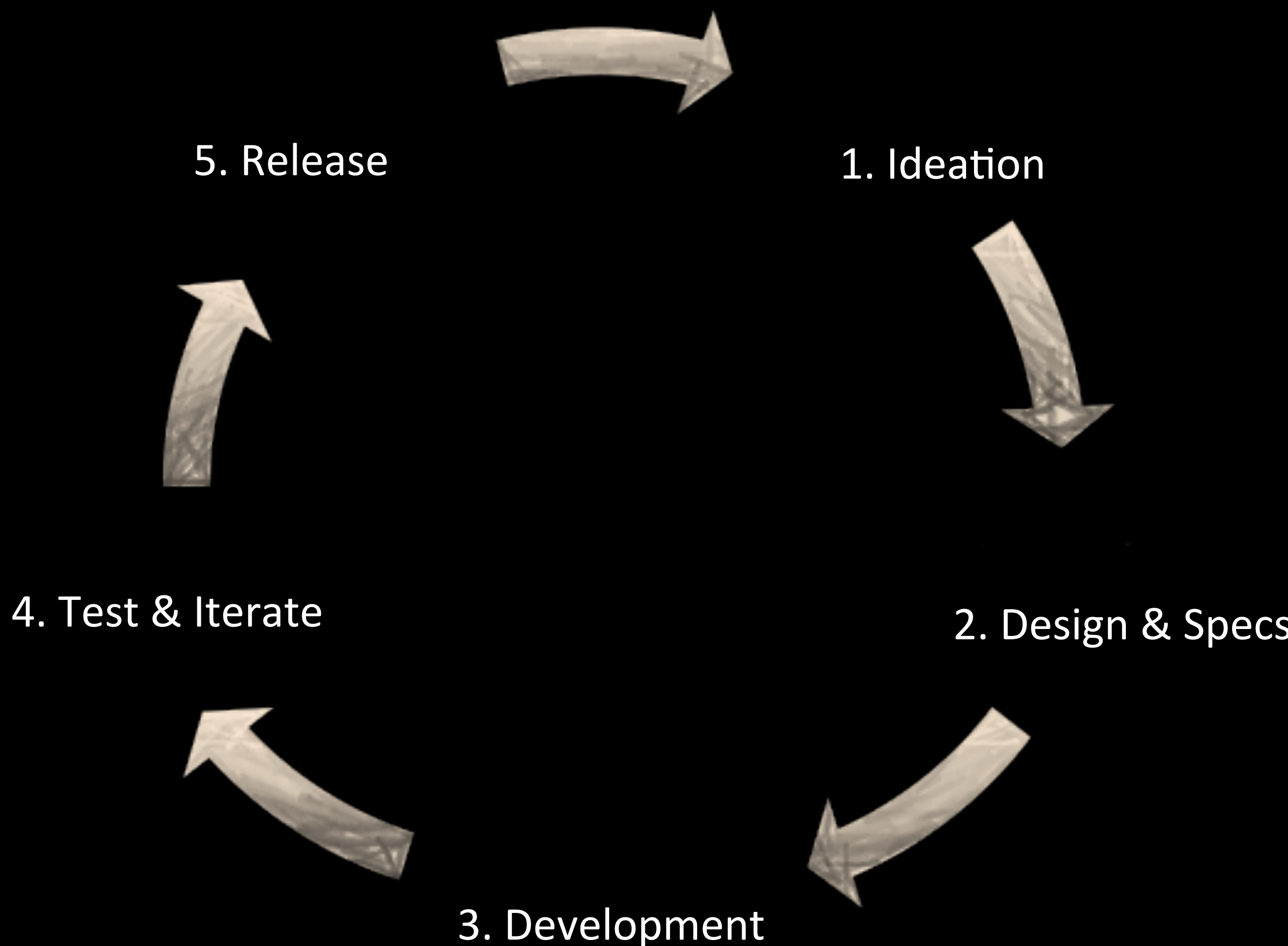
Data Science as a **Service**

- Engagement is “on-demand”, project-based
- Examples:
  - some of the BI roles
  - strategy roles (consultancy)
  - data API to product
  - modeling for specific purposes: propensity to {x}, where x: {buy, attrite, convert, etc.}

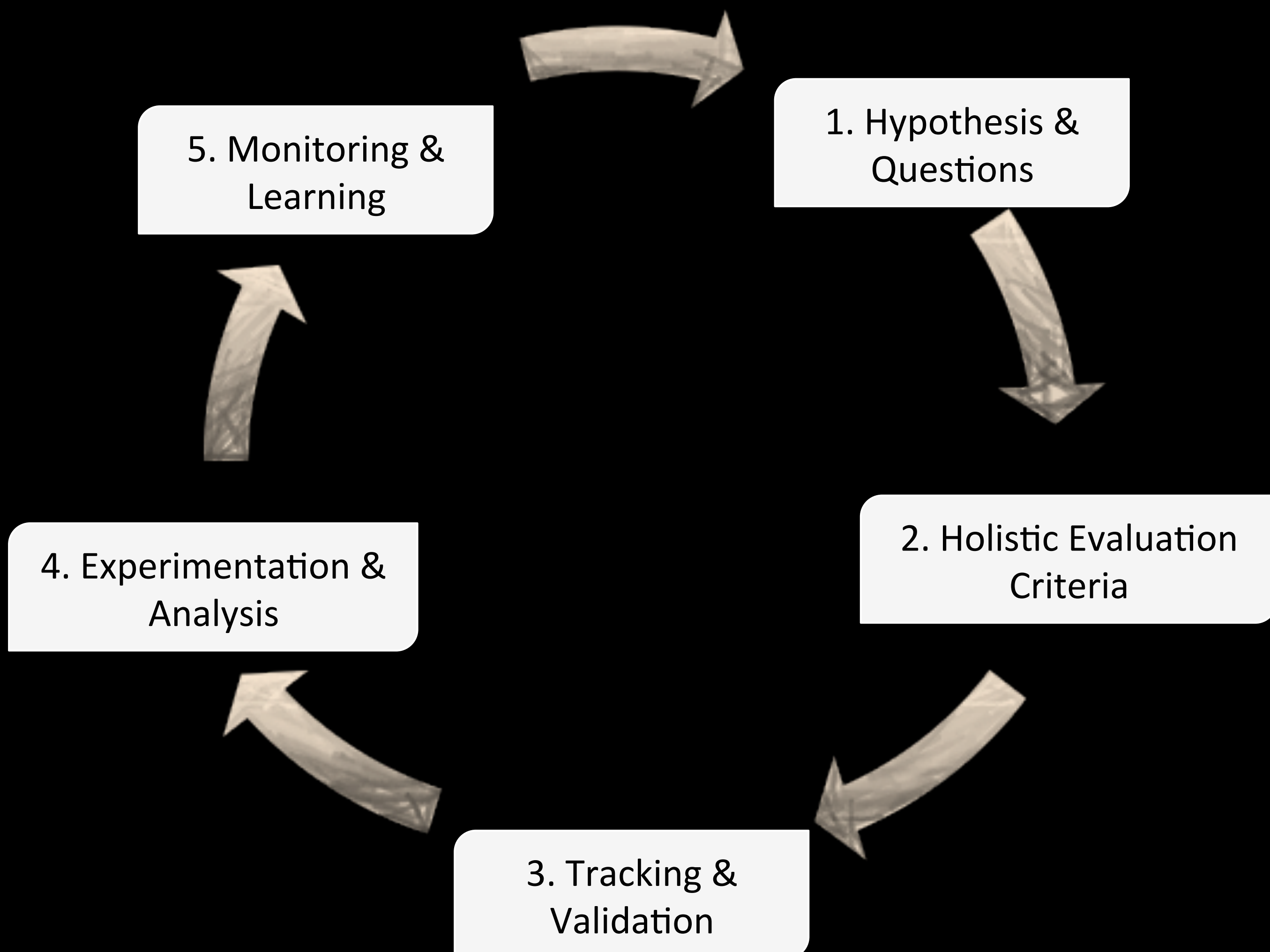
Data Science as a **Partner**

- Plays active role in every stage of ***Product Life Cycle***
- Shares the ultimate goal of product success
- Often requires an embedded engagement model

Product Life Cycle

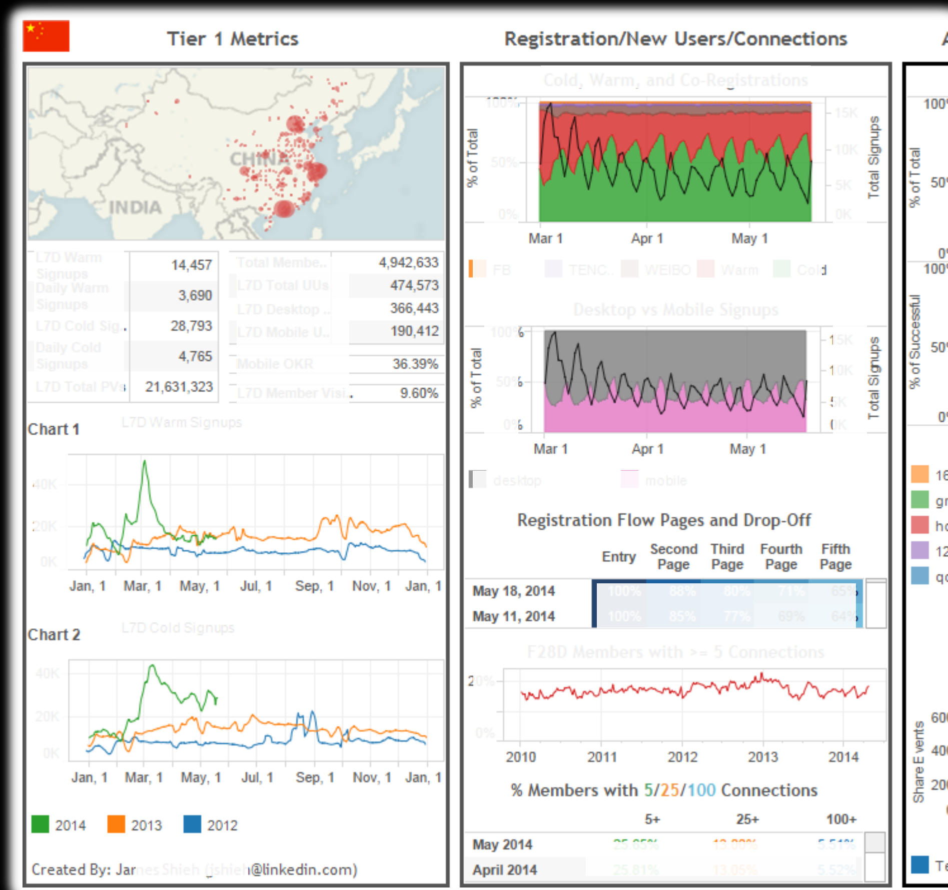


Data-Driven Product Innovation Framework

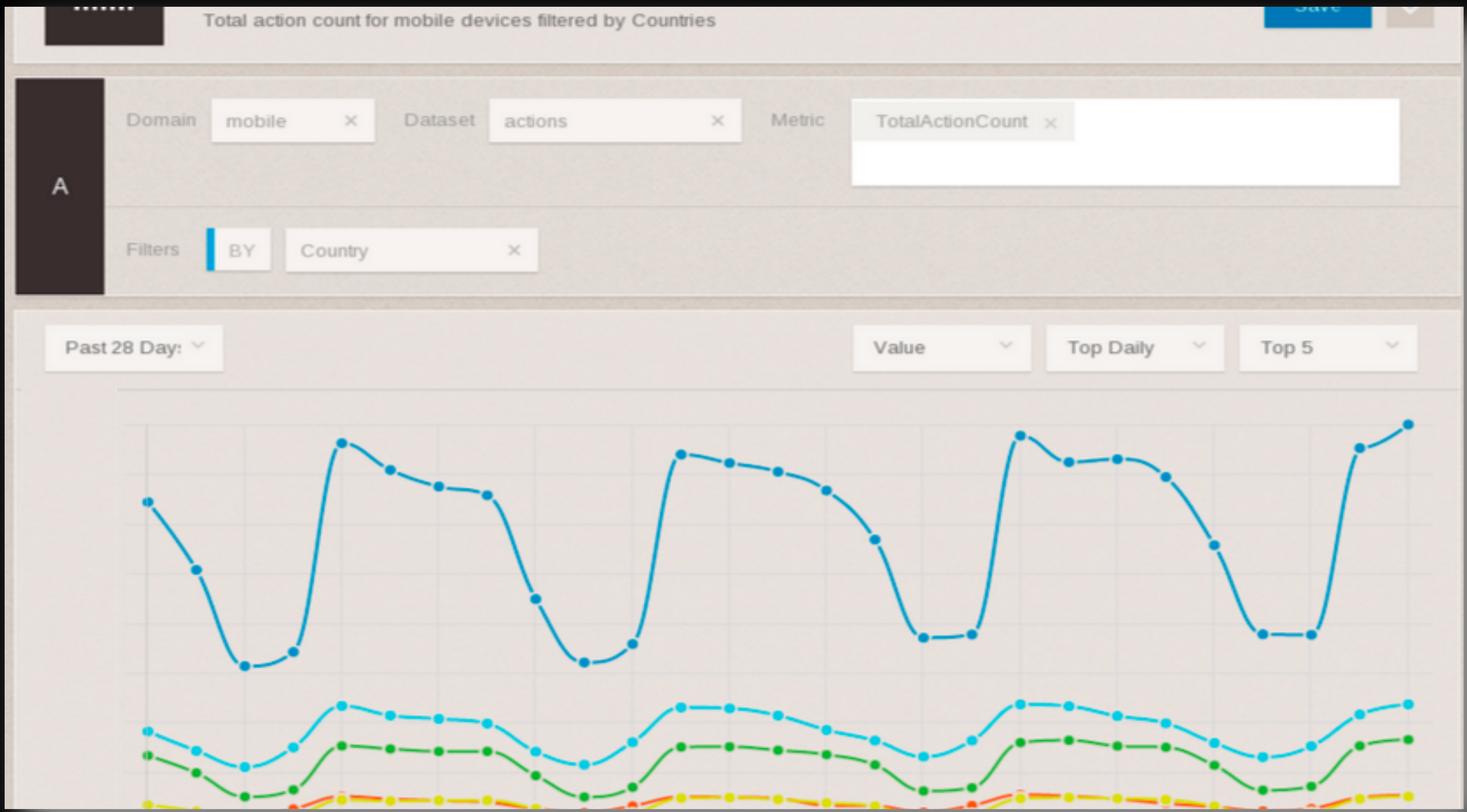


5. Monitoring & Learning

- Start from good reliable reporting of key product metrics
- Particularly important for new partnership ("credibility projects")



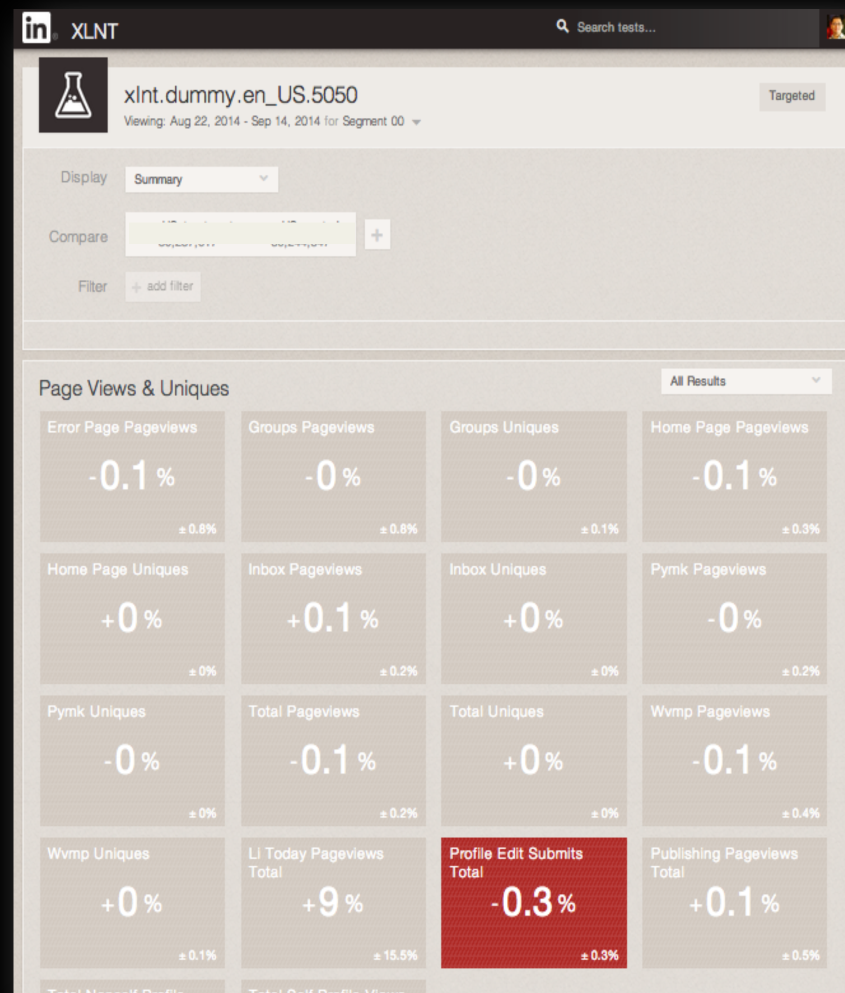
Partner with Data Infrastructure Team to Build Great Tools



4. Experimentation & Analysis

- A company-wide platform for A/B testing, ramping, and advanced targeting needs

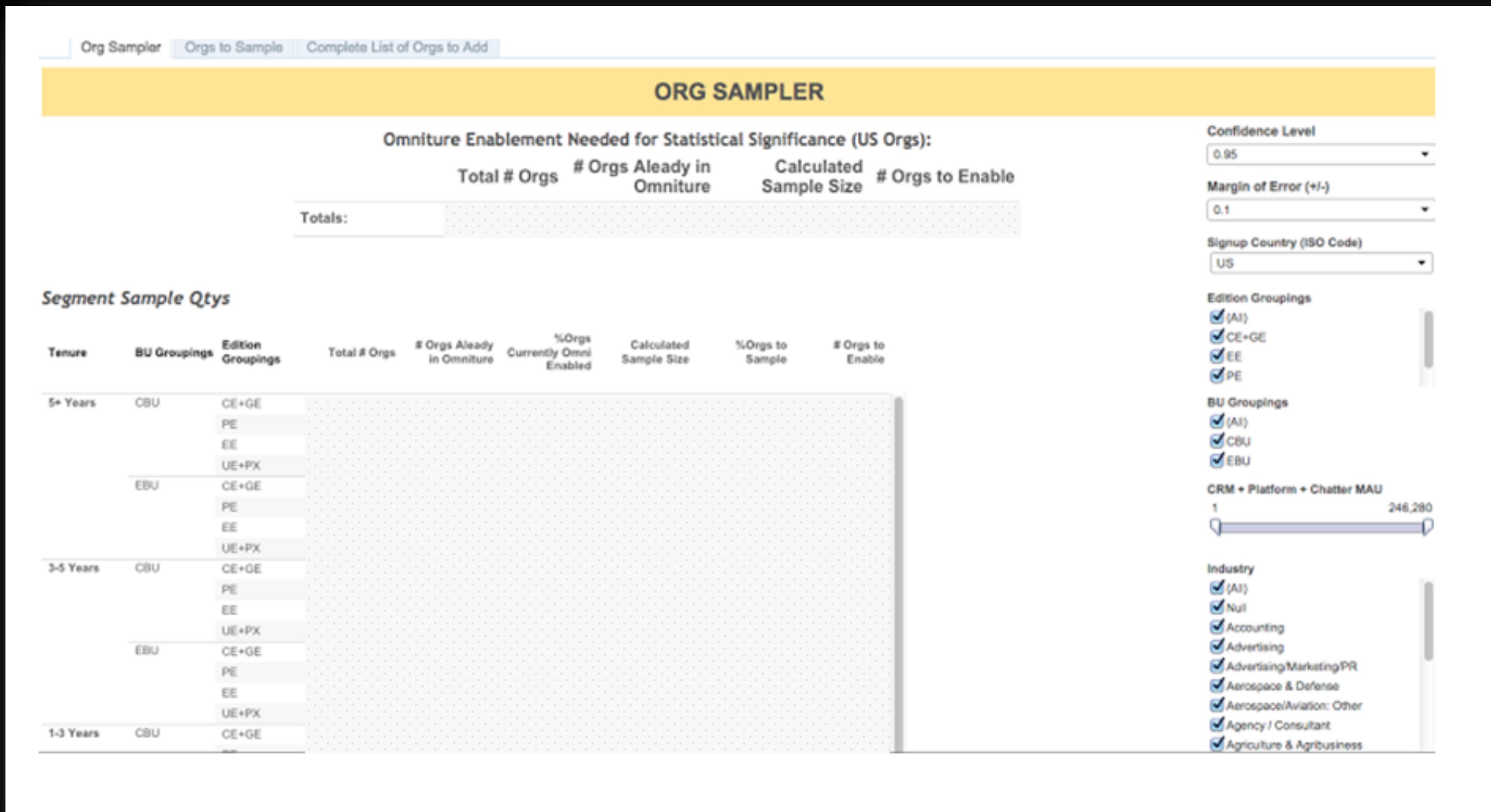
- Automated reporting and analysis capabilities



Ronny Kohavi’s key note “Online Controlled Experiments: Lessons from Running A/B/n Tests for 12 Years”

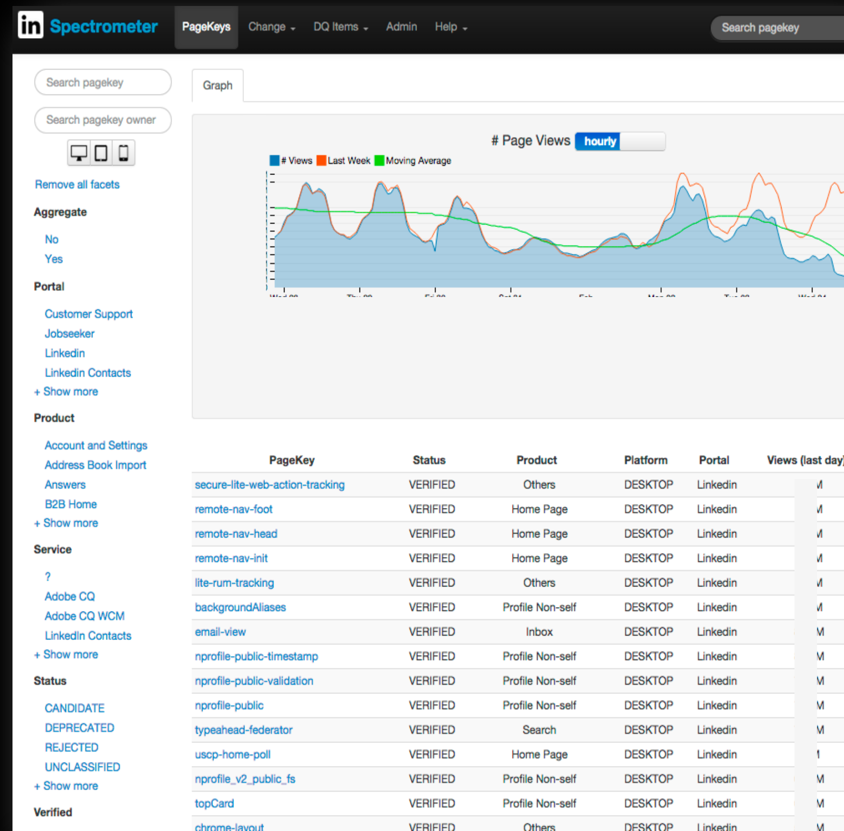
Ya Xu et al.’s talk “From Infrastructure to Culture: A/B Testing Challenges in Large Scale Social Networks”

Streams of Innovation around A/B Testing



3. Tracking and Validation

- Joint ownership between data scientist, engineer, QA and product manager
- Quarterly sign-off process from product owners
- Created a tool for ongoing monitoring



2. Holistic Evaluation Criteria

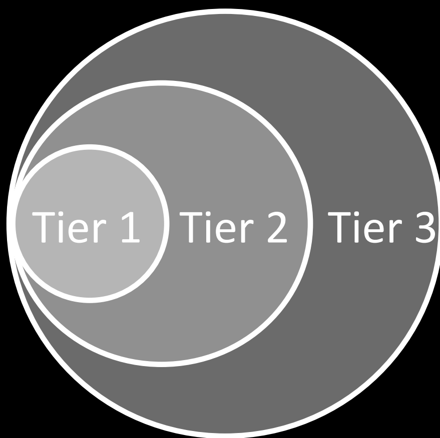
Opportunity for strong thought leadership from data scientists

Past: focus on product-specific metrics

- Relevance change ⇒ CTR
- Profile redesign ⇒ # Profile Views

Now: standardized, tiered metric system

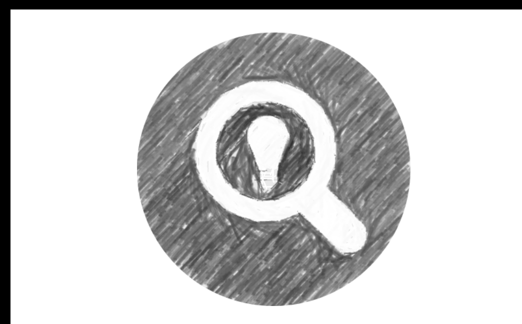
- Site-wide Tier 1 metrics
- Product-specific Tier 2 / Tier 3 metrics



1. Hypotheses & Questions

Components of a good analysis

DATA



INSIGHTS

RECOMMENDATIONS

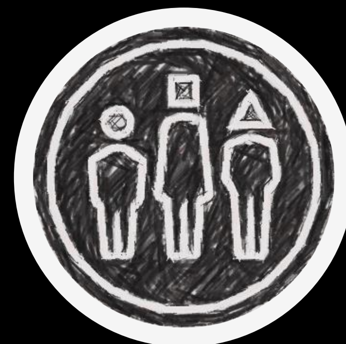
What we would like to see

ACTION!

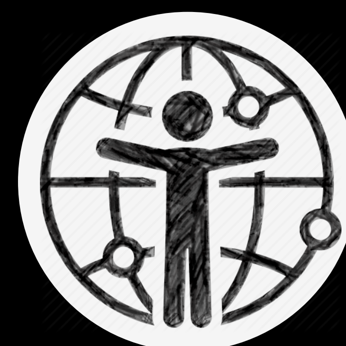
Effective Ways to Drive Actions

Outline expected benefits

Verify by testing with low-cost prototypes

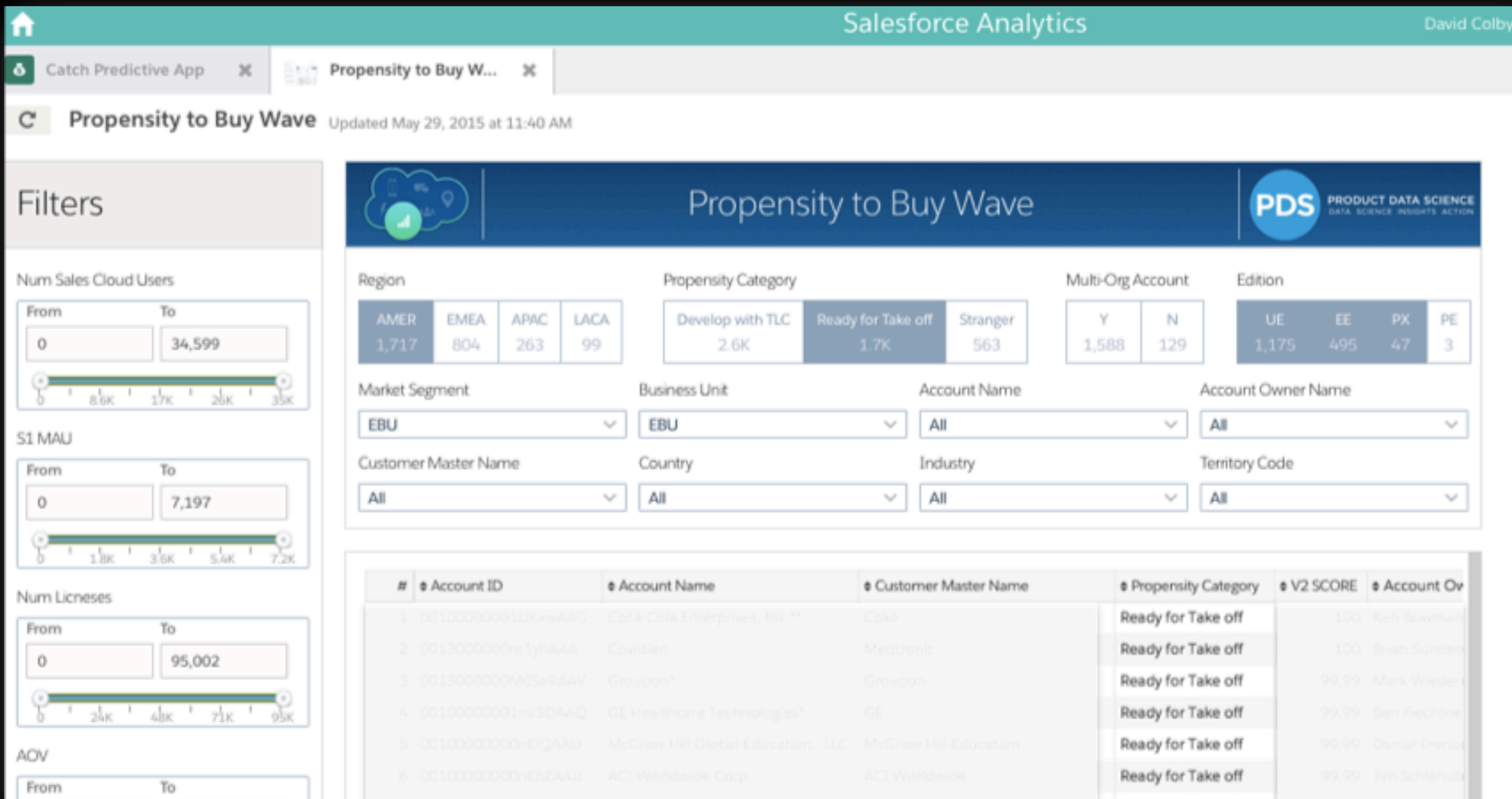


Consider the scale of target audience and technology



Secure a sponsor and build buy in

Data Science in Action!



OrgDNA

Abstract:

Develop prescriptive tool to help admins configure orgs. Use machine learning to identify the relationships between perms, prefs, user behavior and adoption metrics.

Value Creation:

- Increased product adoption
- Increased customer engagement

Accomplishments:

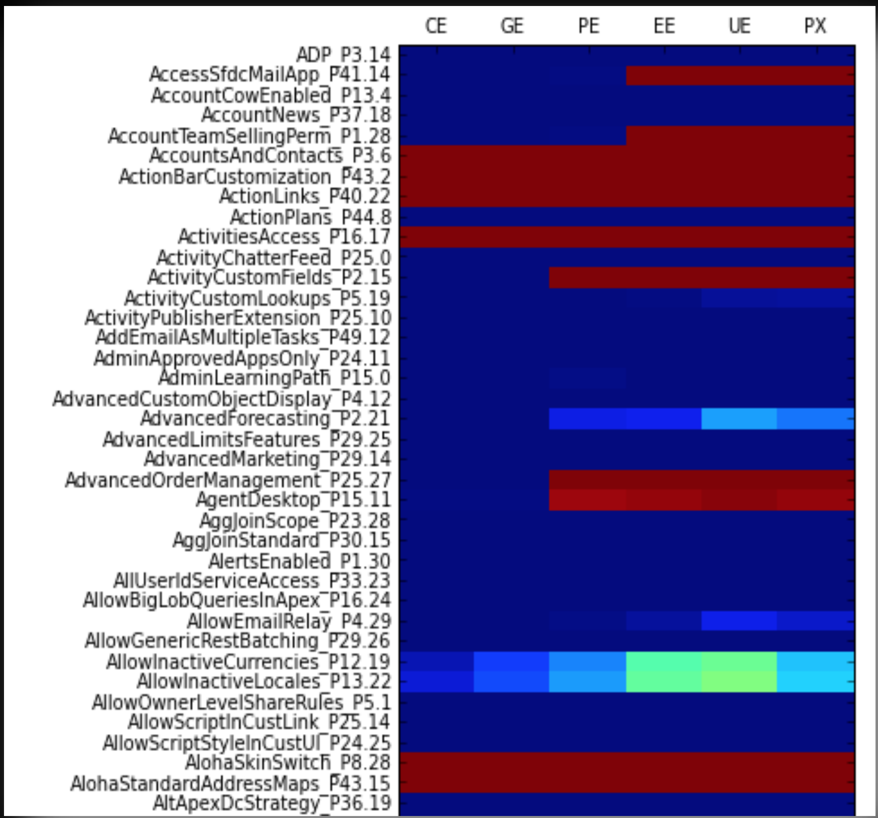
- Decoding of user permissions and preferences into flat binary table for analysis
- Joined in initial test target metrics (red accounts, tenure) and behavioral variables
- Data QA and sanity checks complete (explore distributions and initial correlations)

Sponsor: SVP of Mobile

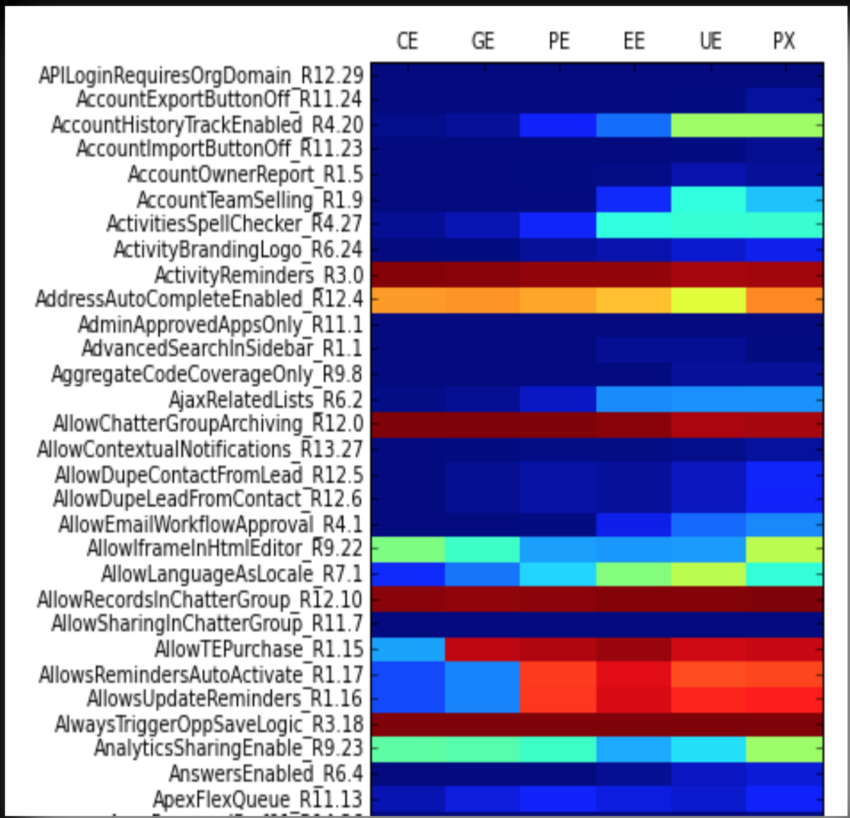
Benefiting Audience: Salesforce Administrators at Customer Locations

Just the Beginning…

Permission Frequency By Edition

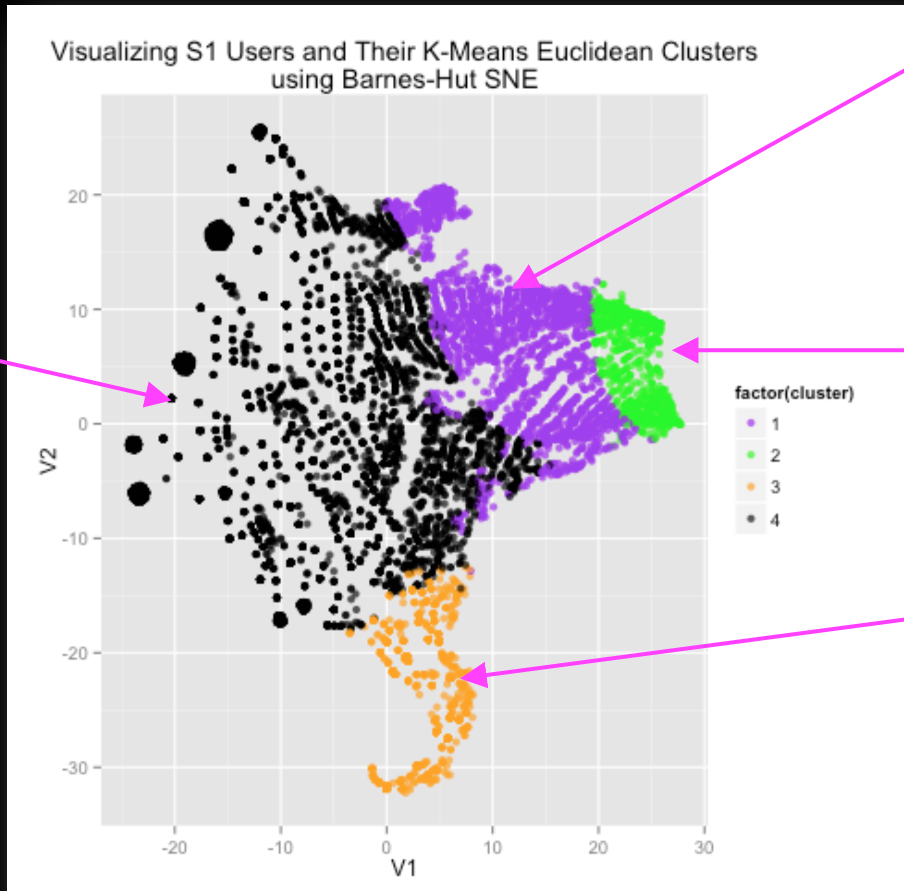


Preference Frequency By Edition



Just the Beginning…

NMI Score = 0.082



High Retention Group (cluster 1)

--Days_S1 between 8 and 16

--CHURN 5%

Super High Retention Group (cluster 2)

--Days_S1 > 16

--CHURN 1%

High Activity, Low Retention Group (cluster 3)

--Days_S1 < 4

--Activities Per Day > 6

--CHURN 46%

Lower Retention Group (cluster 4)

--Days_S1 < 8

--CHURN 34%

Cluster Averages

	cluster	RETAINED	ACT_PER_DAY	Days_S1	Days_S1_Feed	Open_Ntfn_Tray	count
1	1	0.95	3.90	11.69	7.05	2.30	1970
2	2	0.99	3.97	22.27	19.17	7.14	710
3	3	0.54	9.95	1.63	1.00	0.62	1207
4	4	0.66	4.07	3.42	1.55	0.50	6114

Data Science as a Partner

Data Science as a Partner

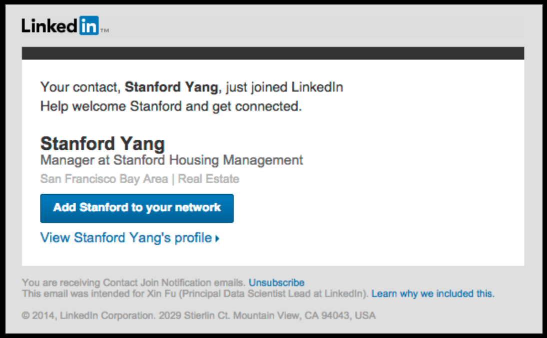
- Focus on Product and balance between
  - Product innovation and operation
  - Velocity and scale
  - Theoretical research and practical impact
- Leverage technology for reliability and sustainability
  - Reliability is key
  - Reduce "one-offs"
  - Speed matters

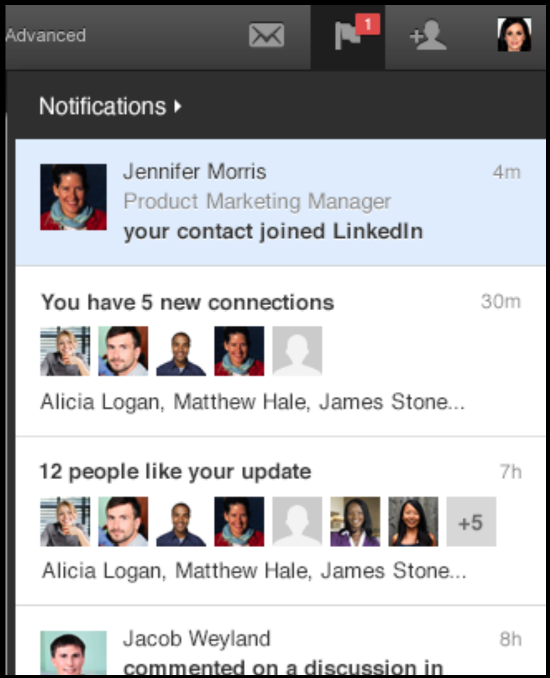
Case Study

Data-Driven Product In Action

In order to increase the connection density of new users, we will lookup the existing members whose address books contain these new users as contacts

The existing members will be notified that their contacts just joined LinkedIn and will be prompted to connect with them

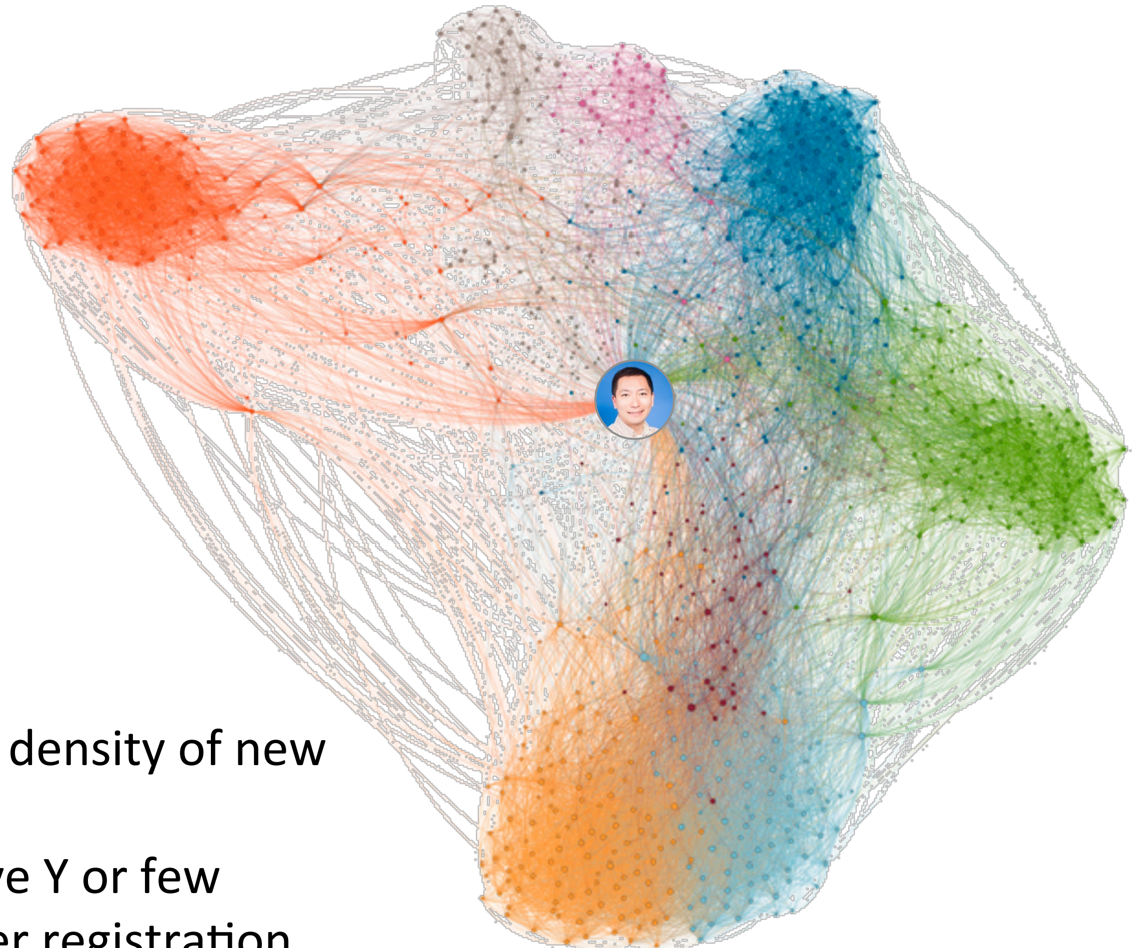




Anatomy of a Data-Driven Product

1. Problem Statement
2. Opportunity Analysis
   – Identify target audience
   – Verify feasibility
3. Holistic Evaluation Criteria
   – Including impact estimate
4. Low-cost Testing with Tracking
5. Analysis of Test Results
6. Hand off for Scalable Implementation

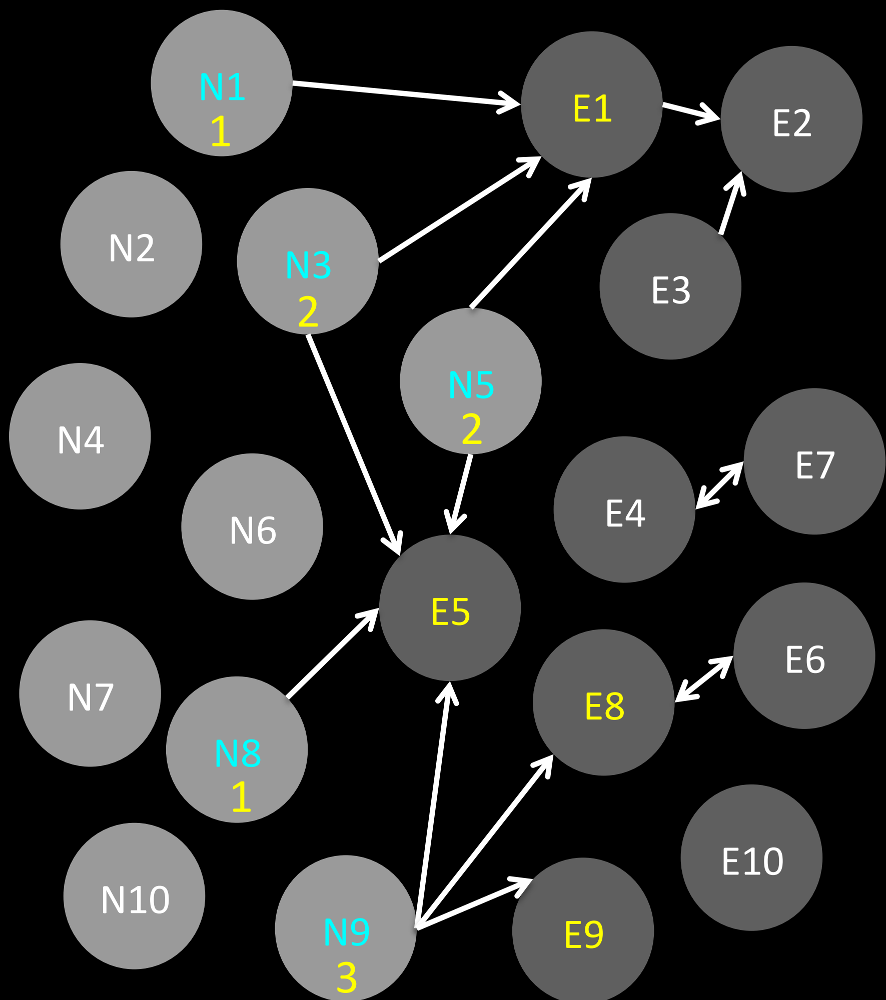
Step 1. Problem Statement



Goal: Increase the connection density of new LinkedIn members. Why?

- X percent of new users have Y or few connections at 30 days after registration
- High connection density => Better engagement

Step 2. Opportunity Analysis



Target Audience: 5 (50%)
Of the 10 new members, 5 are eligible

Feasibility Analysis:
Potential Upside: X new connections for each new member
For the 5 eligible New Members, median number of potential 'welcomer' is 2
X = Median (1,2,2,1,3)

Notification System Load: Of the 10 existing members, 4 are potential 'welcomers'

Member Overload: The volume of notifications to receive by potential welcomers has a distribution with Median=2 and Max=3

A→B : A appears in B's address book   N: New Member   E: Existing Member

Step 3. Holistic Evaluation Criteria

- ***Primary Metrics (New Members)***
  - connection density
  - retention rate
- ***Secondary Metrics (Existing Members)***
  - Connection invitations sent
  - Connection density
  - Engagement
- ***Impact Estimate***
  - Need to make assumptions about conversion rate

Step 4. Low-Cost Testing

- ***Email Test***
  - Daily offline Hadoop job to generate the list of members eligible to receive the email (with eng. review)
  - Randomize for the recipients
- ***Partner Involvement***
  - Work with Engineering and QA to set up tracking
  - Work with Ops, SRE and Customer Service teams to set up monitoring

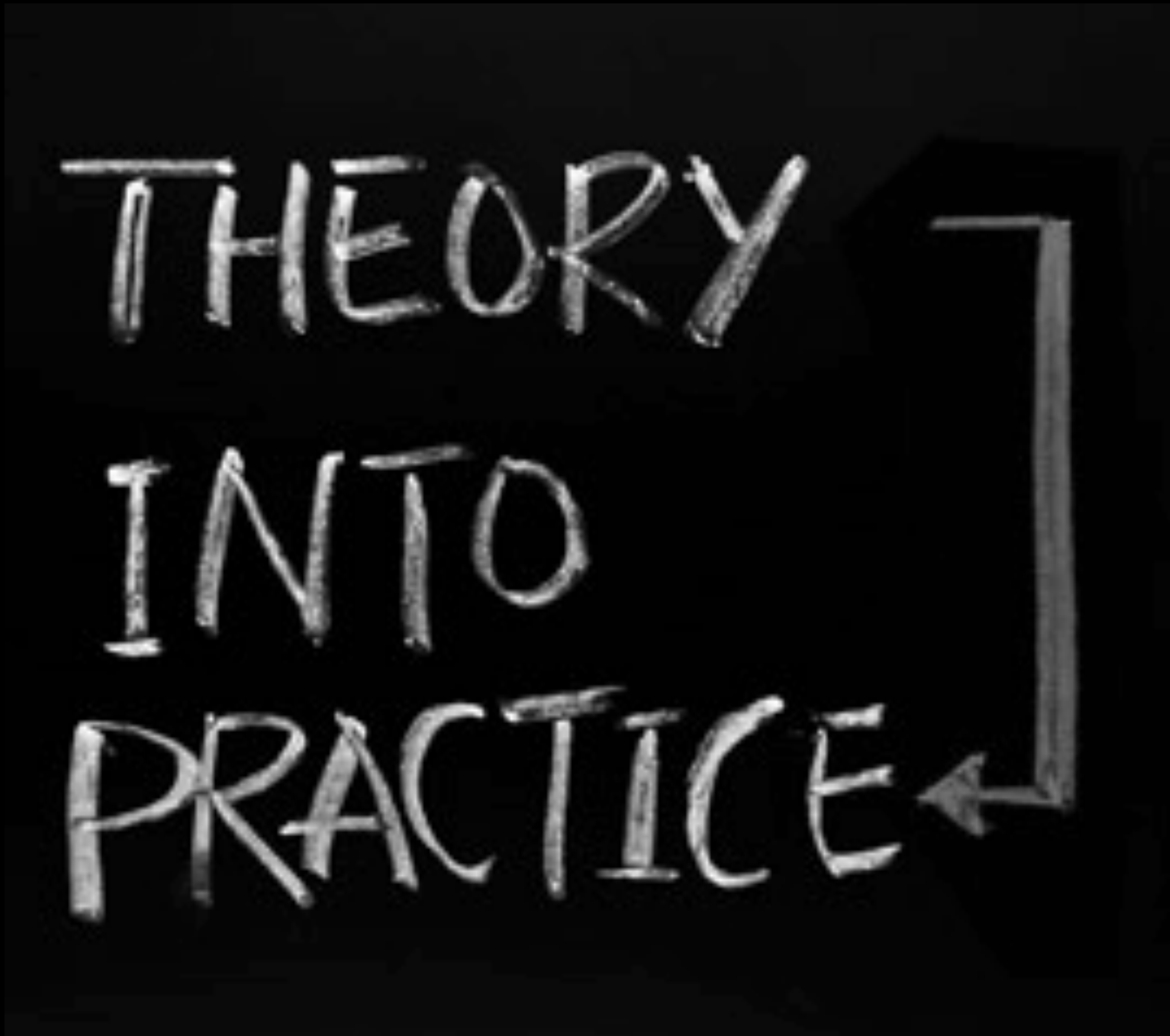
Step 5. Analysis of Test Results

- Email open rate, invite rate, invite acceptance rate
- Estimate impact on existing members
- Estimate impact on new members

Step 6. Scalable Implementation

Handed off to Engineering to scale up the impact:

- Expand target audience, e.g. to dormant members
- Offline email to online and mobile notifications
- Introduce relevance rules


THEORY
INTO
PRACTICE

Design Your Data-Driven Product

Please refer to the hand-out

- 1. Pick an idea for a data-driven product either for your institution, or for your favorite Web product
- 2. Outline your plan to evaluate the idea
  – Problem statement
  – Opportunity analysis (target audience, feasibility)
  – Holistic Evaluation Criteria and potential impact
  – Low-cost testing plan
- 3. Identify who you can partner with

Data Science's Role in a Product Org

Data Science ▼	Team Structure	Importance of Prioritization	Velocity	Sustainability	Proactive Approach
As an Owner	Single	◔	●	◔	●
As a Service	Center of Excellence	●	◔	◑	◔
As a Partner	Hub & Spoke	◕	◑	◕	◕

○ ◔ ◑ ◕ ●

Low — High

Data Science as Product Partner

PARTNERSHIP

- Start w/ credibility projects
- Context and ownership
- Scalable path to innovation

TECHNOLOGY

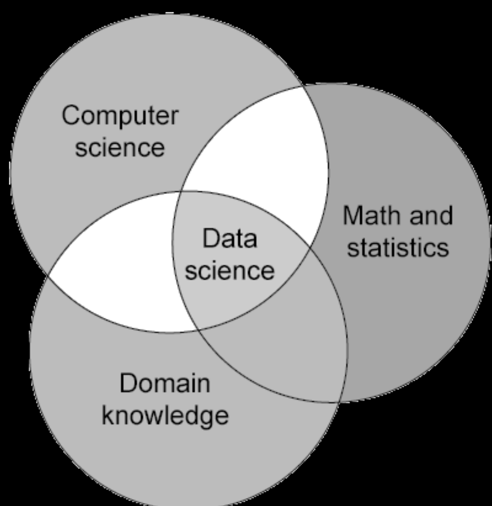
- Leverage technology to improve quality and speed
- Reliability is key
- Automate, Automate!

TALENTS

?

Data Science as Product Partner

Talents



- We put great emphasis on understanding and solving real business problems
  - Data sense: what is ***possible***
  - Product sense: what is ***valuable***
- We value people who think holistically and in scale
  - Cross-product impact
  - Viral effect

Tutorial Takeaway

Data Science as a Partner

PARTNERSHIP

- Start w/ credibility projects
- Context and ownership
- Scalable path to innovation

TECHNOLOGY

- Leverage technology to improve quality and speed
- Reliability is key
- Automate, Automate!

TALENTS

- ***Passion*** for product innovation
- ***Proactive*** in partnership
- ***Impact*** oriented

THANK YOU

Data-Driven Product Innovation

Xin Fu

LinkedIn

Hernán Asorey

Salesforce

KDD2015

21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining

10 - 13 August 2015, Hilton, Sydney