INTERNAL ARCHIVAL RESEARCH

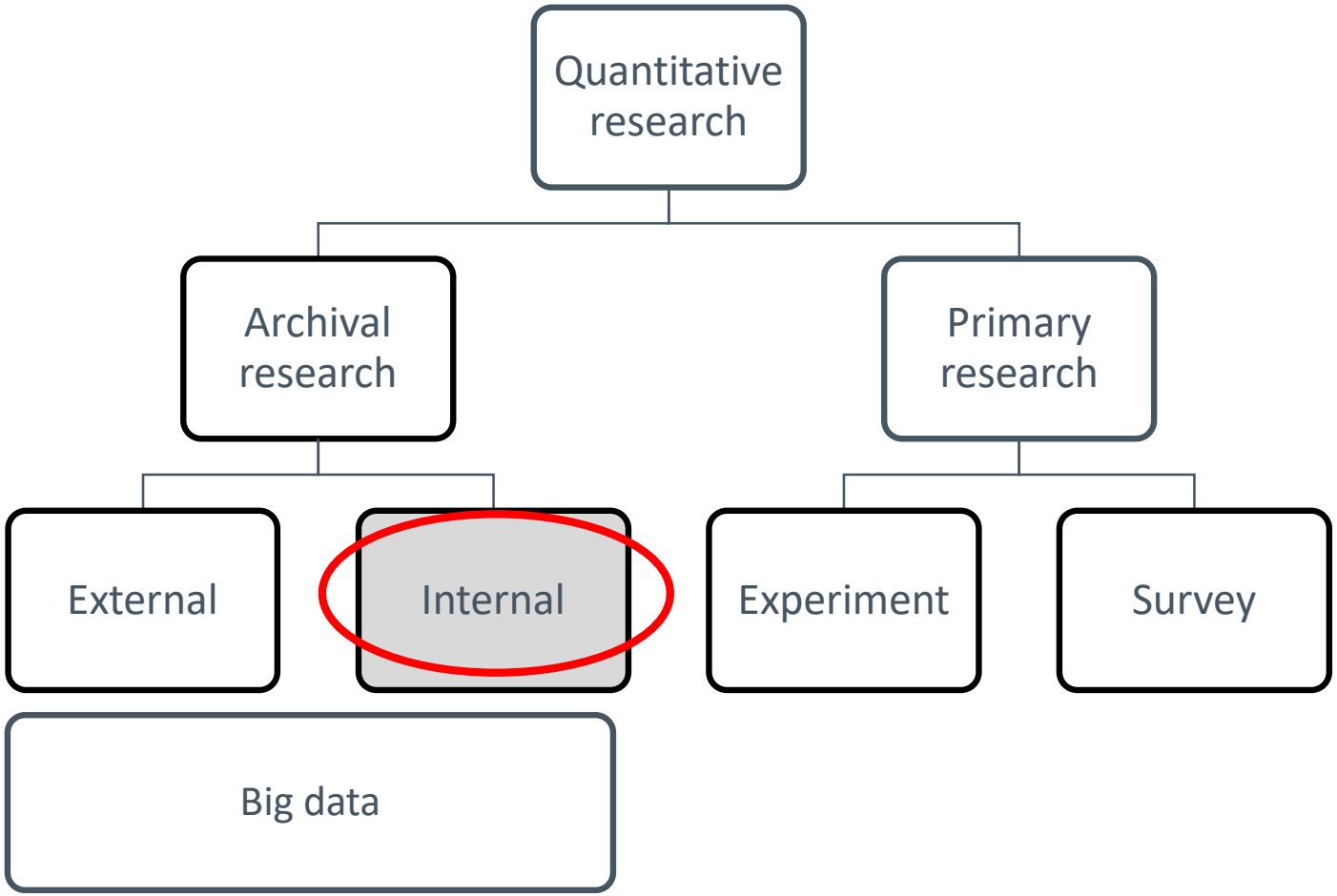
Business

Research

Techniques

Aurélie Lemmens

Where are we?


Quantitative research
Archival research
Primary research
External
Internal
Experiment
Survey
Big data

Agenda

1. Examples of business problems that require internal archival data

2. A more detailed example

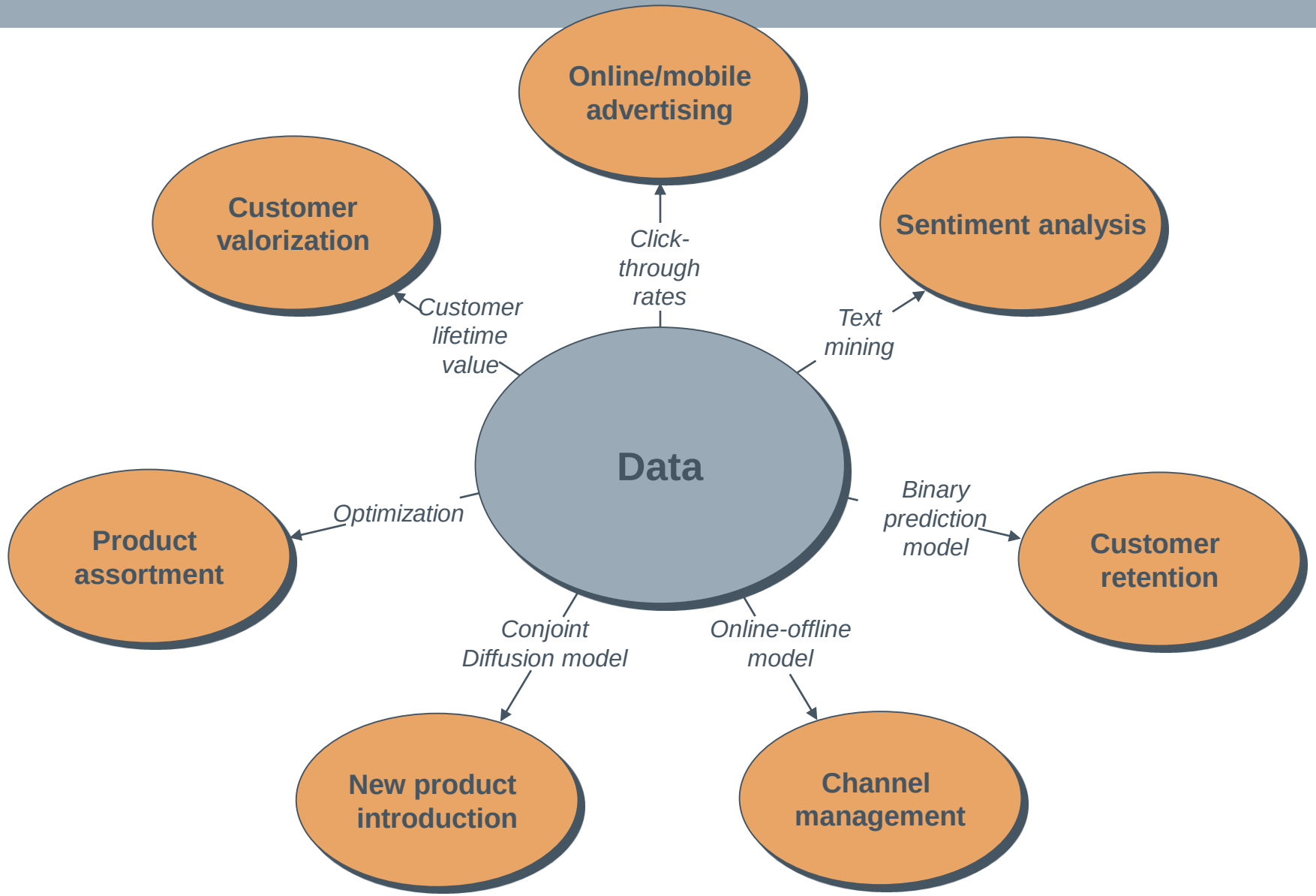
3. What is internal archival research?

4. Internal archival data trends per industry

5. Reliability and validity

1. Examples of decision problems that require internal archival data

Data is at the core of any company decision

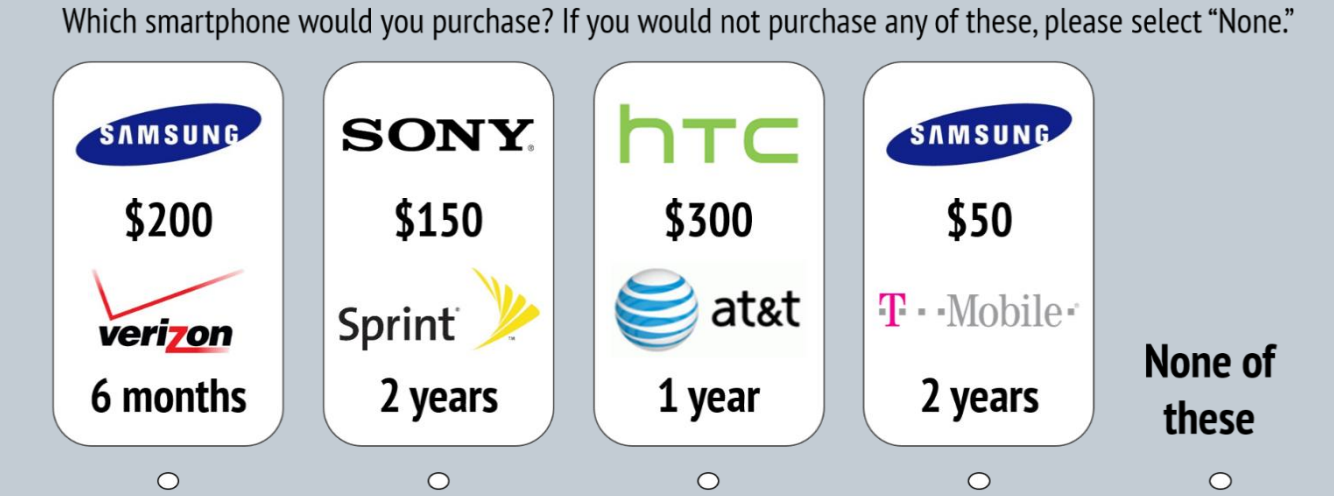


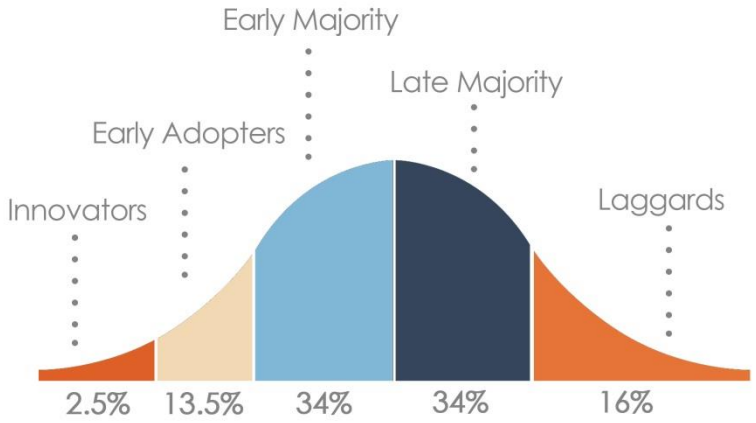
NEW PRODUCT INTRODUCTION

Conjoint Analysis

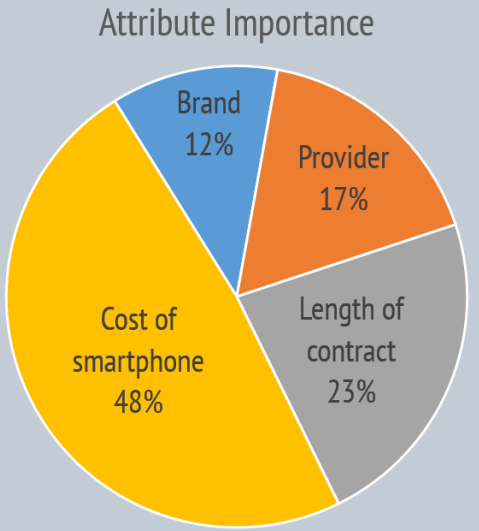
Diffusion Models

Sales Forecasting



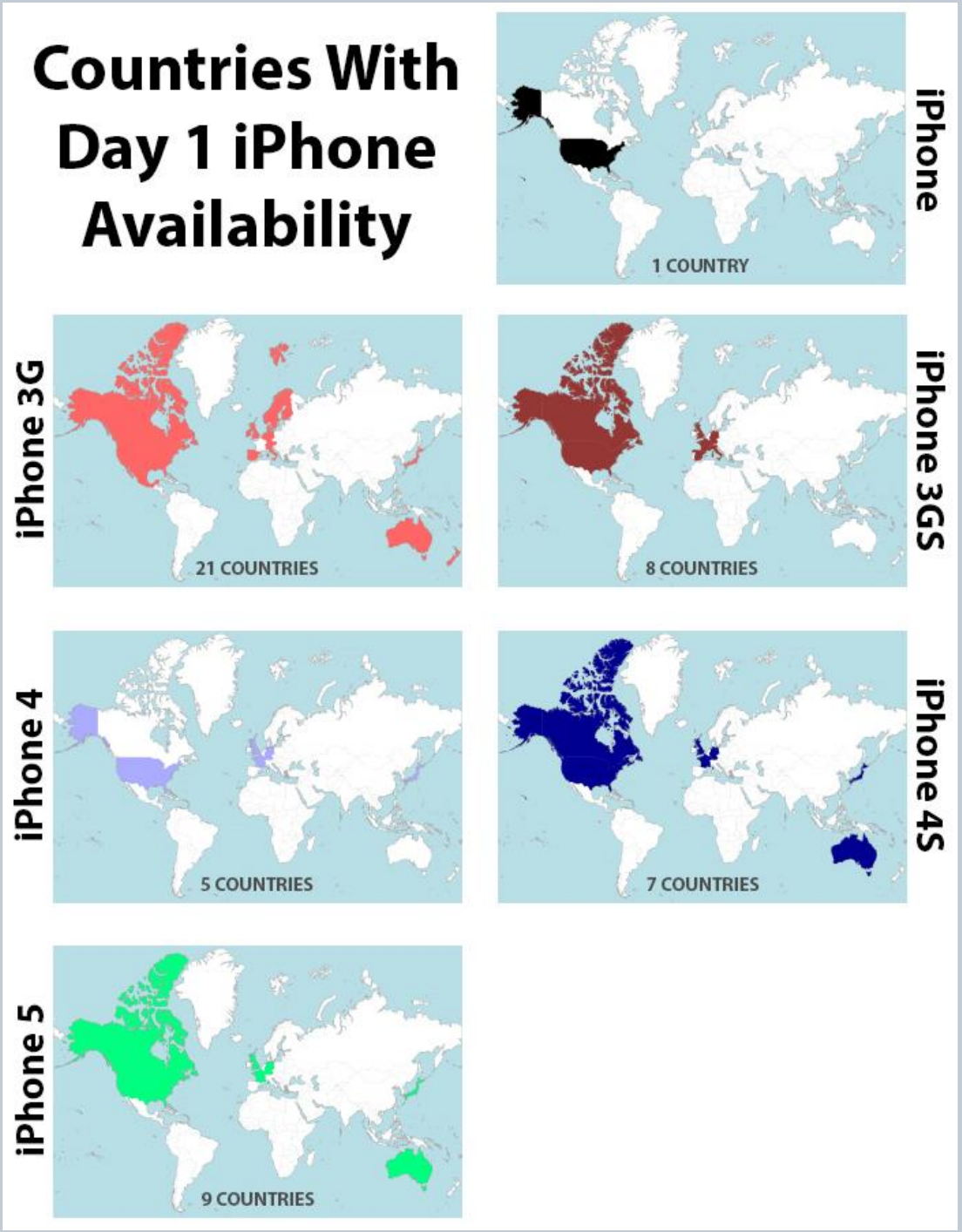


Rogers Diffusion Of Innovation Bell



INTERNATIONAL LAUNCHES

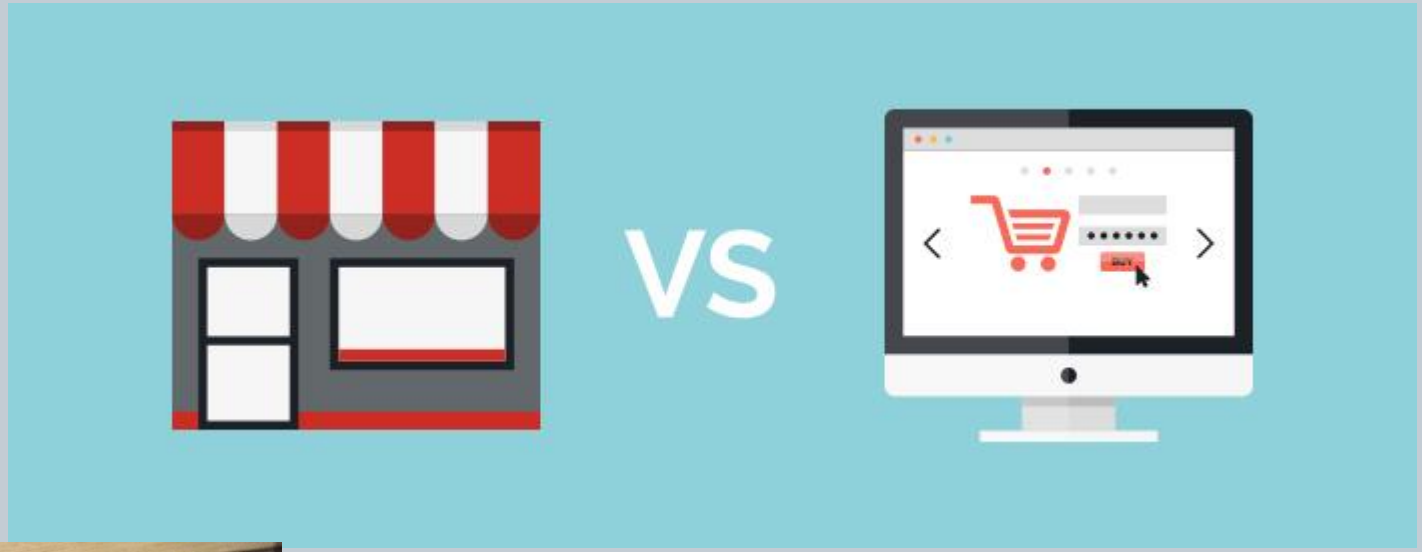
All at once (sprinkler) or waterfall strategies?

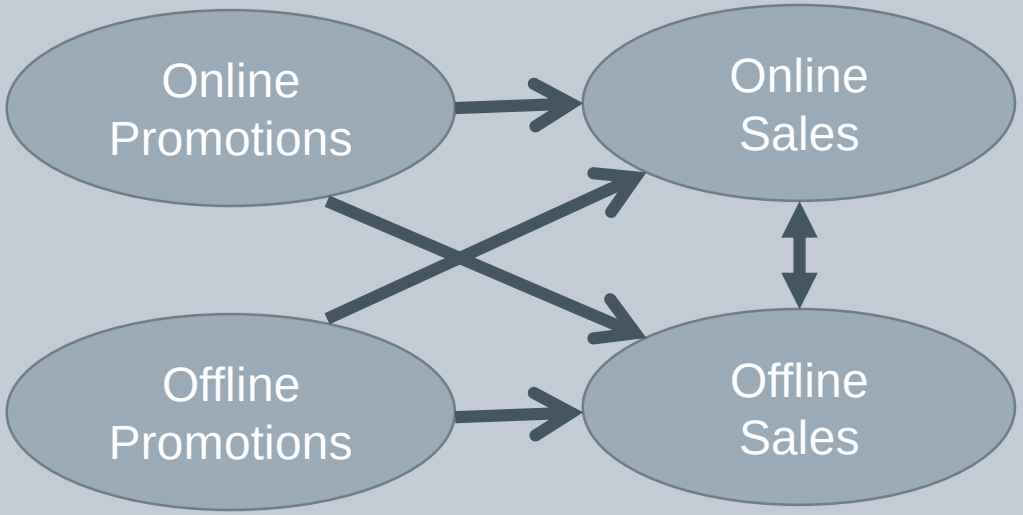


ASSORTMENT OPTIMIZATION

What if you could generate the optimal assortment for hundreds of planograms at the push of a button?

ONLINE VS OFFLINE





CUSTOMER LIFETIME VALUE

Which customers will be the most profitable?

How to attract them? ***Acquisition***

How to grow their revenues? ***Cross-selling***

How to retain them? ***Retention***

BOOST (ONLINE) SALES

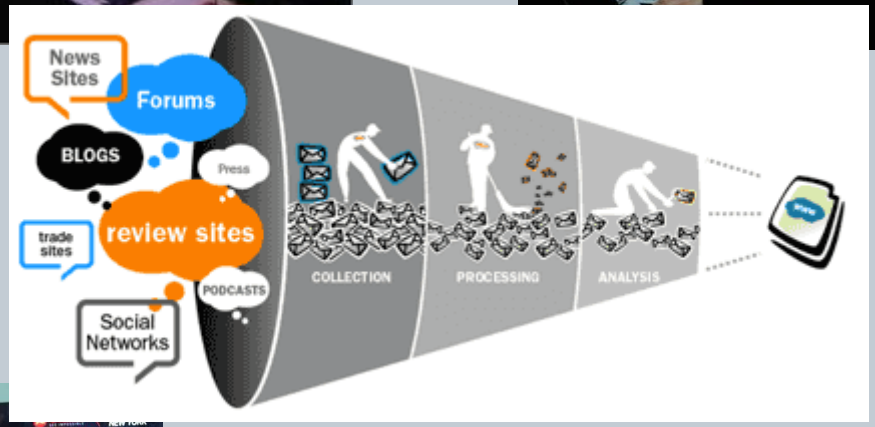
Designing the best online and/or mobile advertising campaigns (e.g. banner)

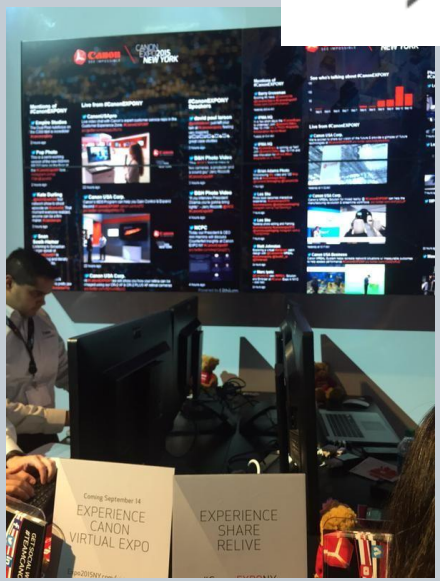
SENTIMENT ANALYSIS

Text mining

Online monitoring of e-WOM on the brand

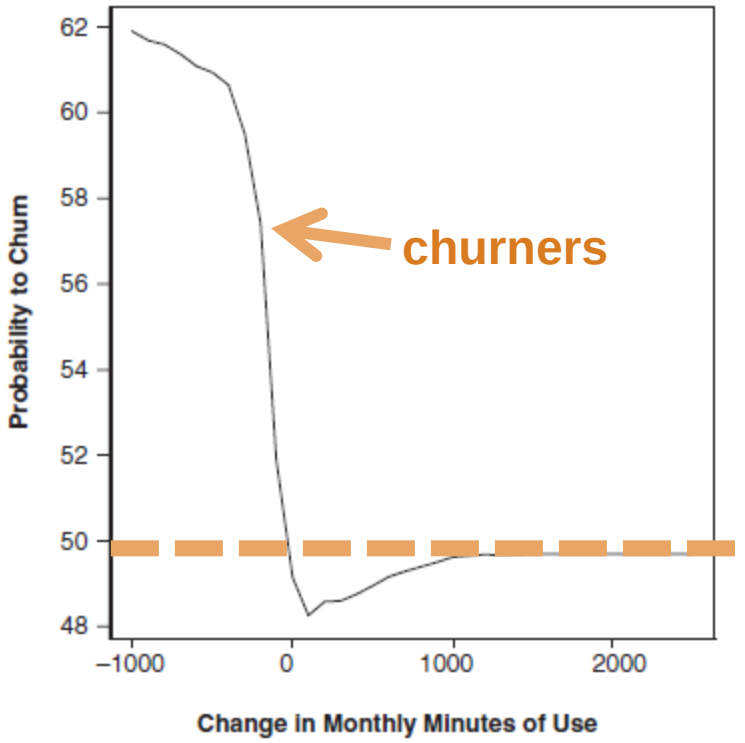
Social media listening rooms

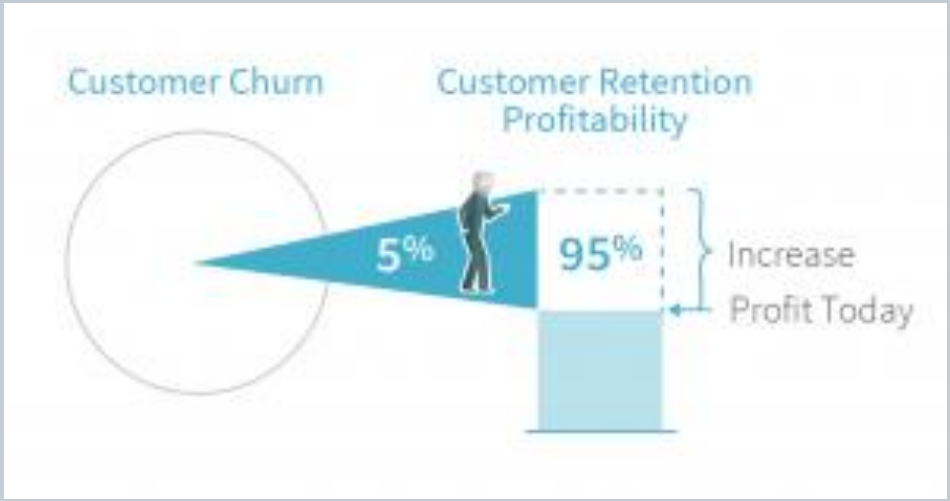




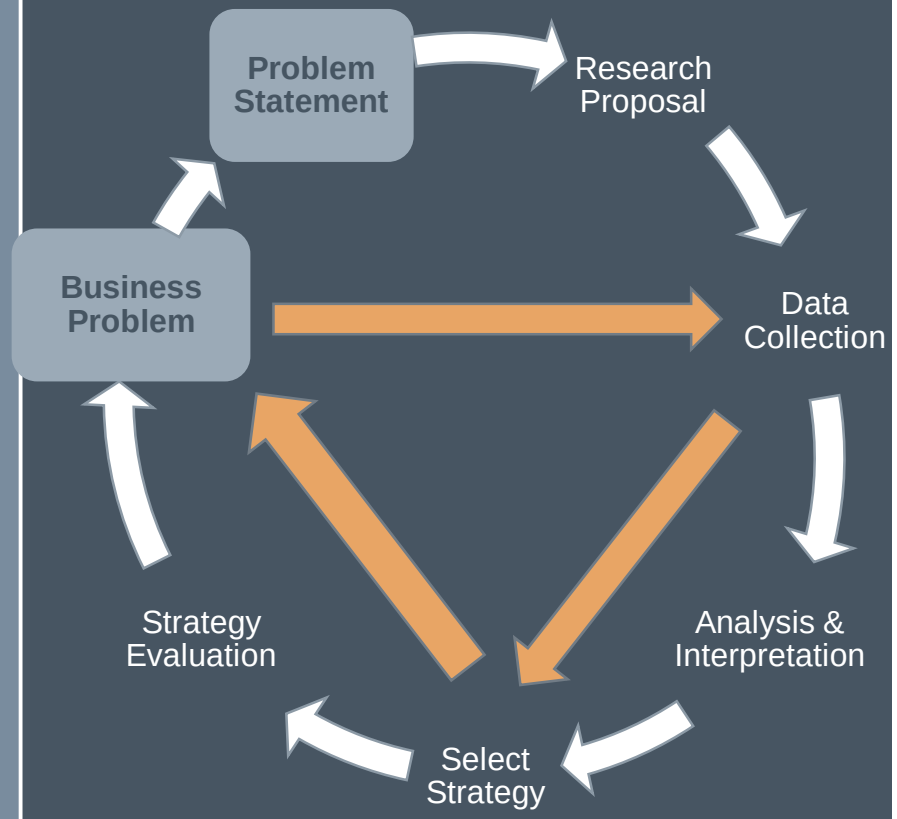
CUSTOMER RETENTION

Data can tell you which customers are most likely to churn





One example: Customer defection at T-Mobile

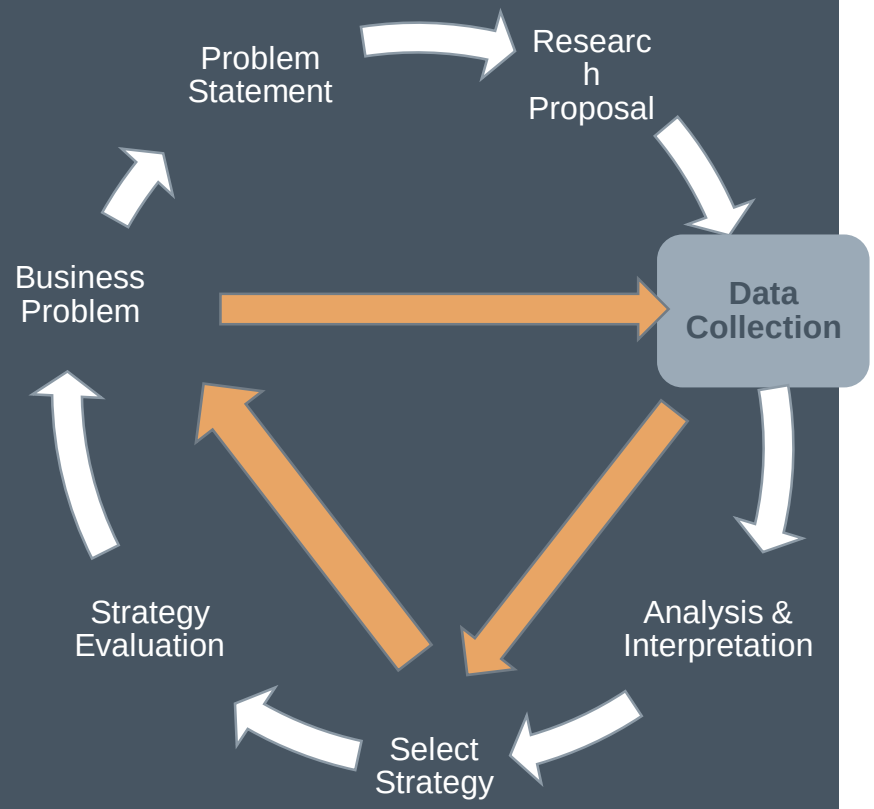


- **Business problem**: 2% monthly churn rate
- **Problem statement**: can we predict which customers are the next defectors?

One example: Customer defection at T-Mobile®

- **Research proposal:**
  - Gather customer information
    - Which data sources?
  - Estimate a churn prediction model
    - Which prediction model?

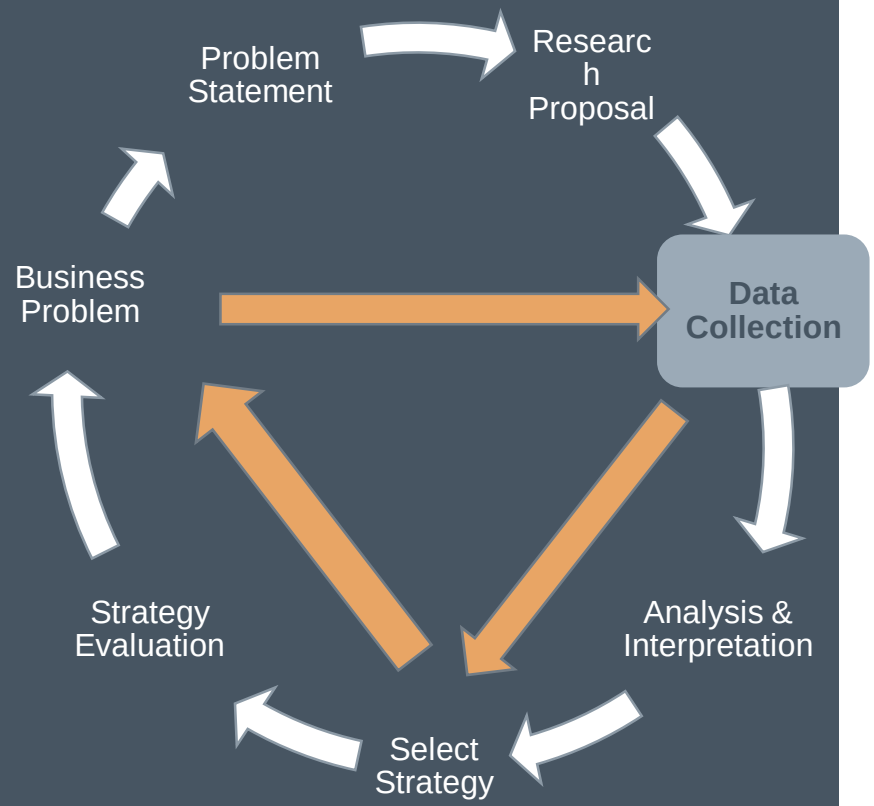
One example: Customer defection at T-Mobile®



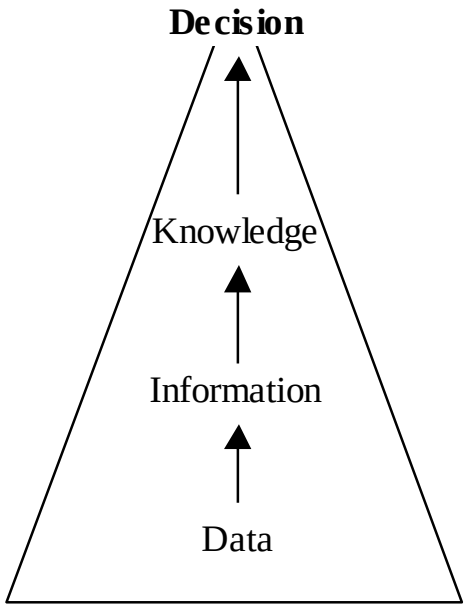
Data

1. Collected from numerous sources
   - Sales, customer characteristics, marketing mix, customer service, social networks, online, …
   - As many as possible!
2. Filtered and assembled into databases
   - Merging and matching on customer IDs or based on similar customer characteristics
   - Can be complex!
3. Turned into business knowledge
   - Model estimation and translation into managerial insights
   - Need for analytic and managerial profiles!

Another example: Nestlé Purina PetCare



› Purina wants to know on which website to advertise. They combine online sales with other website visits to inform their decisions.



Let's put banner ads on about.com

Dog owners who see ads online are likely to buy Purina ONE. We know the sites they visit: about.com, www.petsmart.com.

1. Purina buyers are 20% more likely to visit about.com
2. 36% of dog owners who see Purina ads would buy the brand.

016030102
(Buyer 1 bought Purina puppy chow on March 1)

3. What is internal archival research?

Internal archival data

- **Internal**: retrieved from inside the company.
- Advantages:
  - Can be accessed quickly and easily
  - Less expensive
- Disadvantages
  - Incomplete information
  - Timeliness of information
  - Amount of information
  - Inappropriate to a particular question or situation
  - Need for sophisticated equipment and techniques
- 4 types
  - Accounting/finance
  - Sales
  - HR (employees)
  - Marketing

Internal archival data

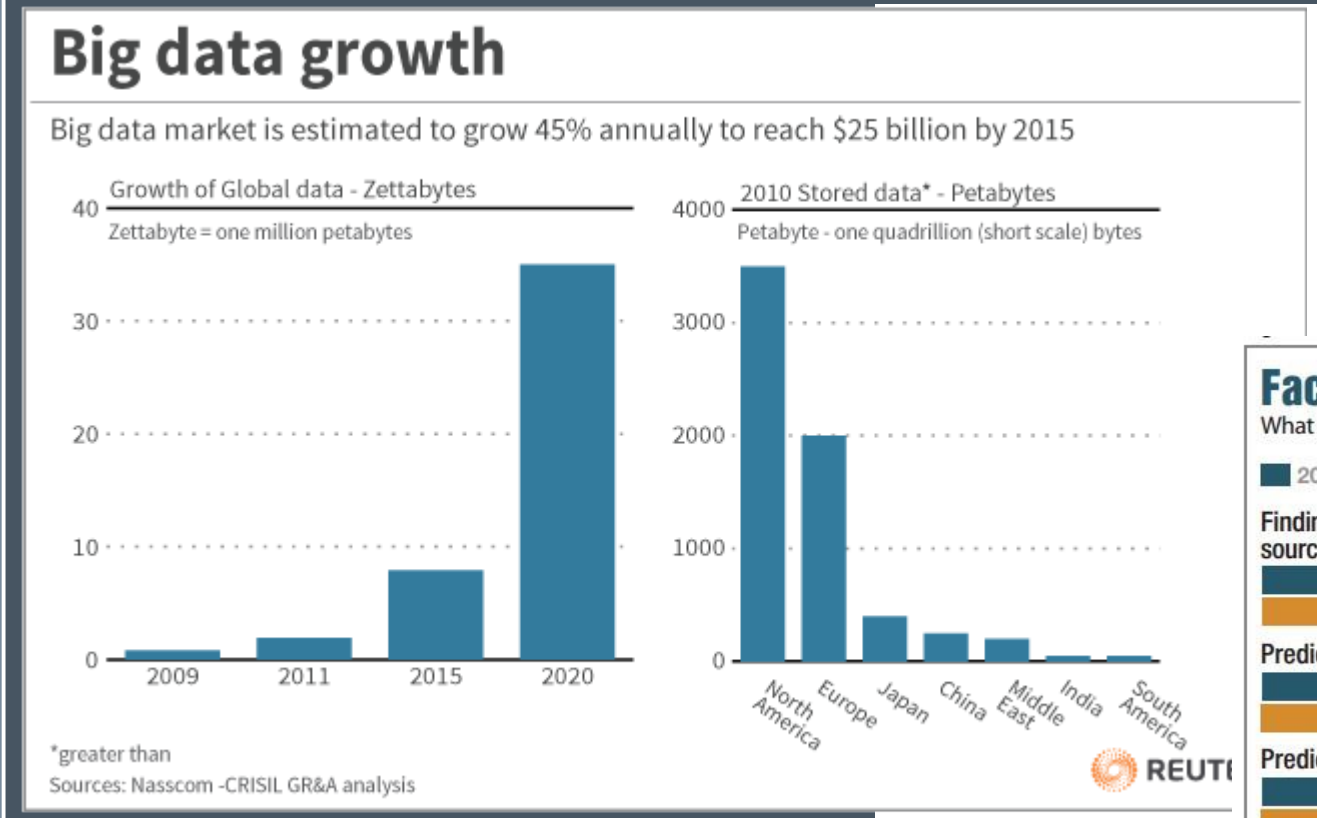
- Accounting/finance data:
  - Generates sales data, cash flows reports, production reports (costs), marketing expenses, profitability analyses
- Sales (force) data:
  - Sales information systems, using sales force automation software, allow sales reps to input results of sales calls to both prospects and current customers into the MIS
  - Can be decomposed into distribution channel, price point, geographic area, customer type and salesperson
  - Sales reps access the product and customer databases both for input and review of customer records while on the road from their laptops.
- HR data:
  - Contracts, salary, performance, …
- Marketing data
  - Customer information from all “touch points”

Customer characteristics and behavior

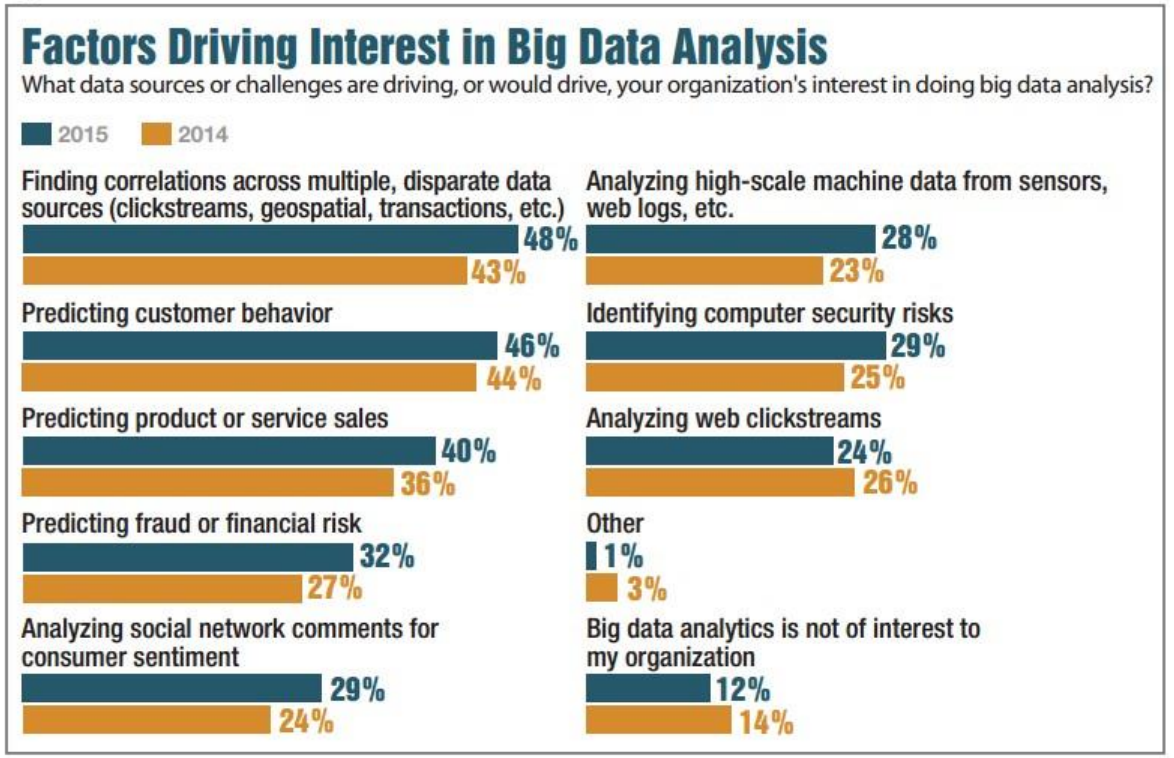
- Individual customer activity is the most important internal marketing data
- Collected via many touch points
  - Demographics
  - Payments
  - Scanner data (purchases)
  - Customer care calls
  - Complains
  - Website visits
  - Social media
  - …

→ Information overload

Big Data era



Reuters graphic/Catherine Trautman 05/10/12



Capitalizing on Big Data:

Strategies outperforming companies are taking to deliver results

Leaders are **166%** more likely to **make most decisions based** on data

And they are **2.2x** more likely to have **formal career path** for analytics

75% of Leaders cite **growth as the key source of value** from analytics

Leaders **measure the impact** of analytics investments

Leaders have **predictive analytics** capabilities

Leaders have some form of **shared analytics resources**

Join the conversation on Twitter at #ibmanalytics and follow @IBMIBV

Source: *Analytics: A blueprint for value - Converting big data and analytics insights into results*
IBM Institute for Business Value. © IBM 2013.

ibm.co/9levers

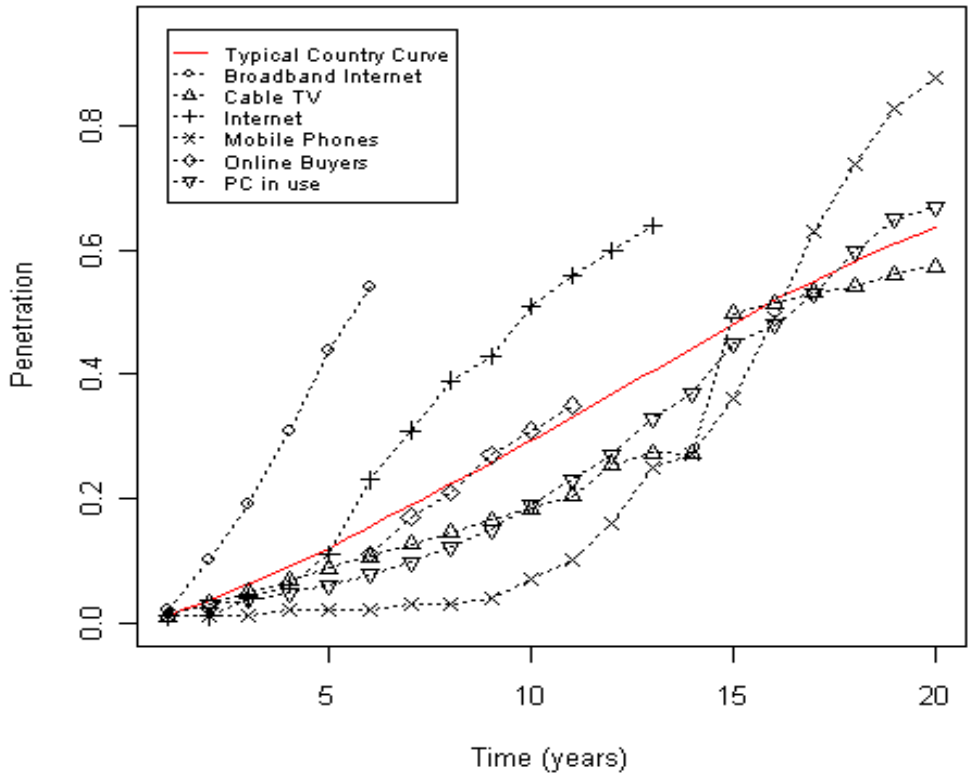
IBM

4. Internal archival data trends per industry

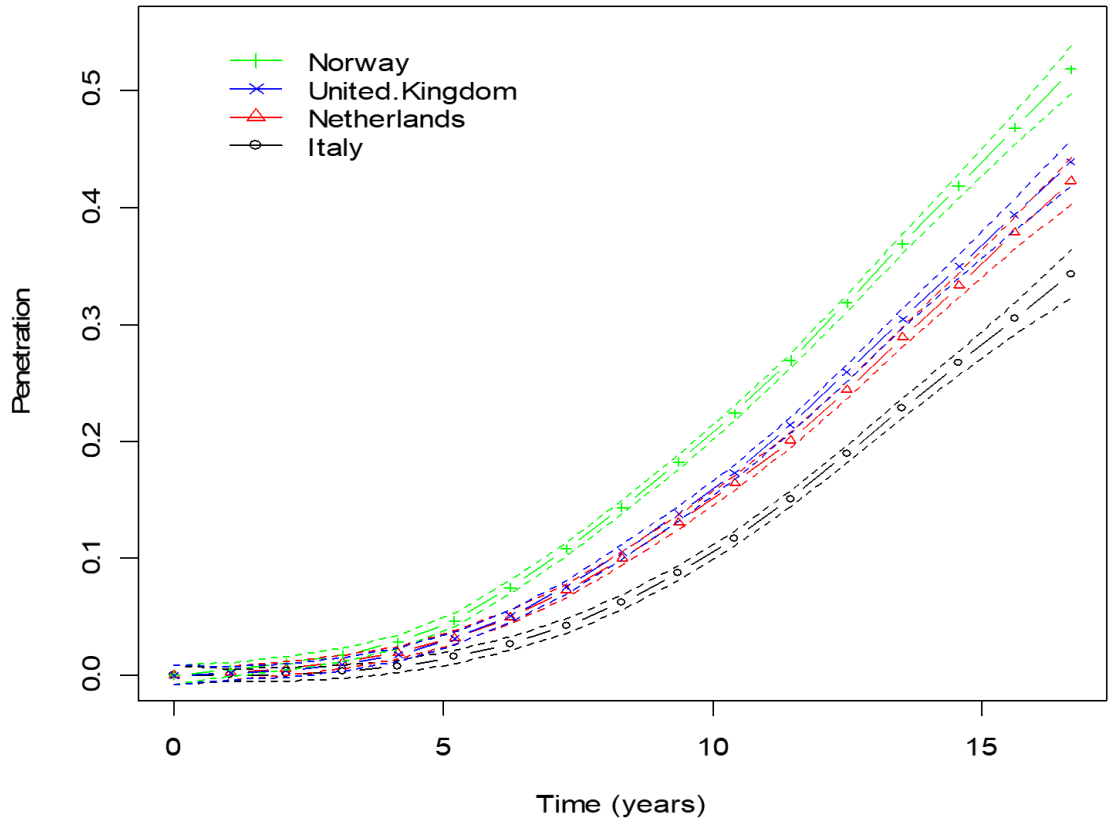
Data trends per industry: **Manufacturing**

› From predicting new product success from using historical data…

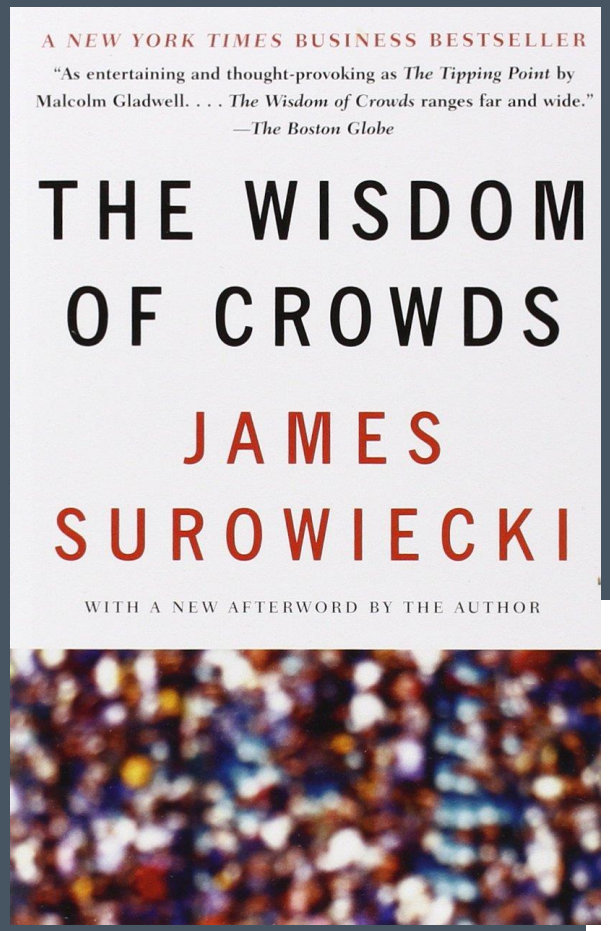
Sales prediction using similar products



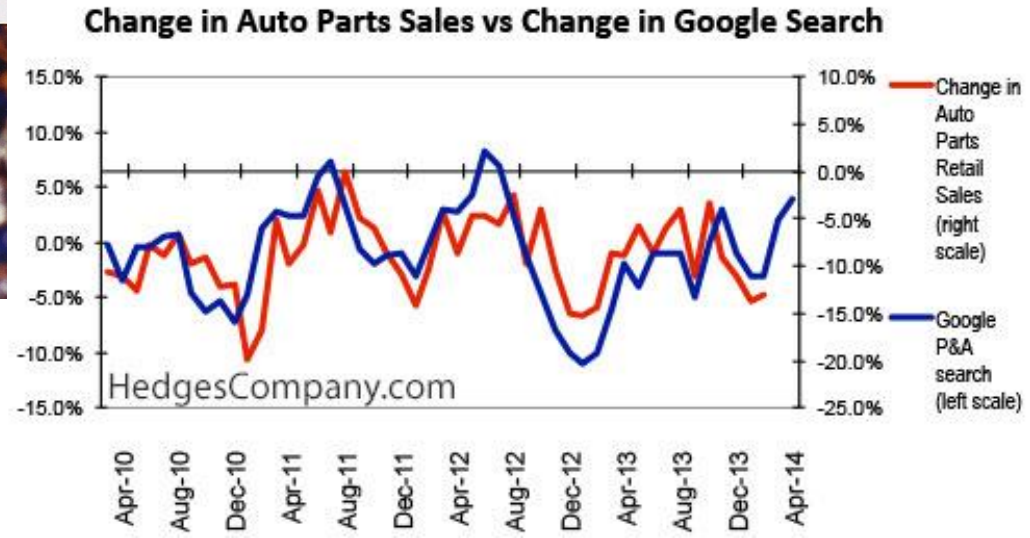
Sales prediction using similar countries

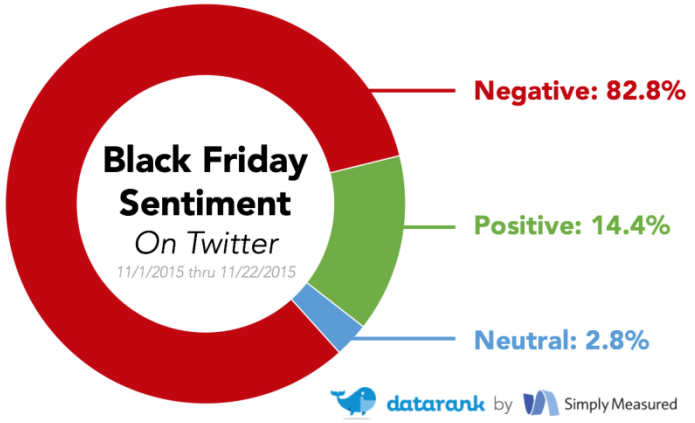


Data trends per industry: **Manufacturing**

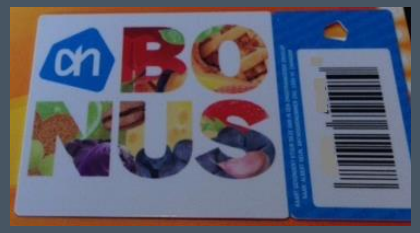


- …To using online activity across many users
  - Wisdom of the crowd to maximizing stock market performance
  - Early signal using google search and twitter feeds

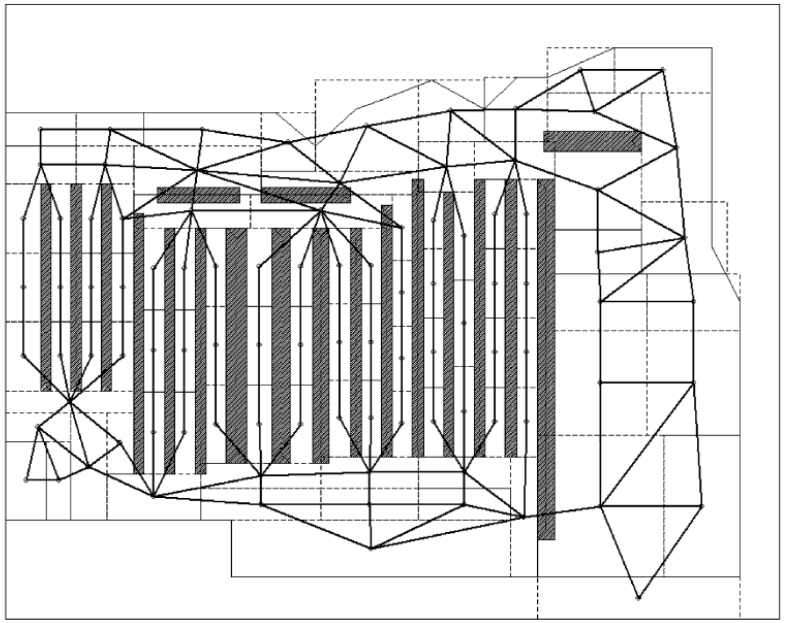




Data trends per industry: **Retailing**

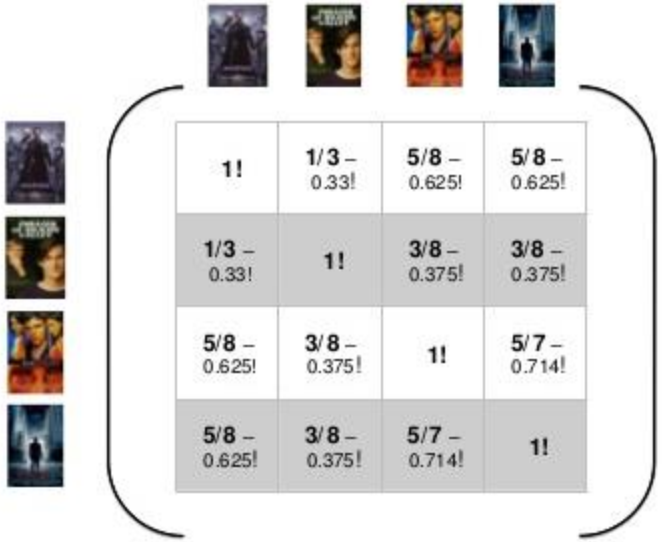


- From brick and mortars…
  - Aggregate sales data
  - With the emergence of loyalty programs:
    - Individual-level purchases
    - Non-purchase data
  - GPS on shopping carts, mobile scanners and apps, …
    - In-store tracing of the shopping paths



Data trends per industry: **Retailing**

› … To online retailers
  – Recommendation systems
  – Collaborative filtering
  – Cross-selling



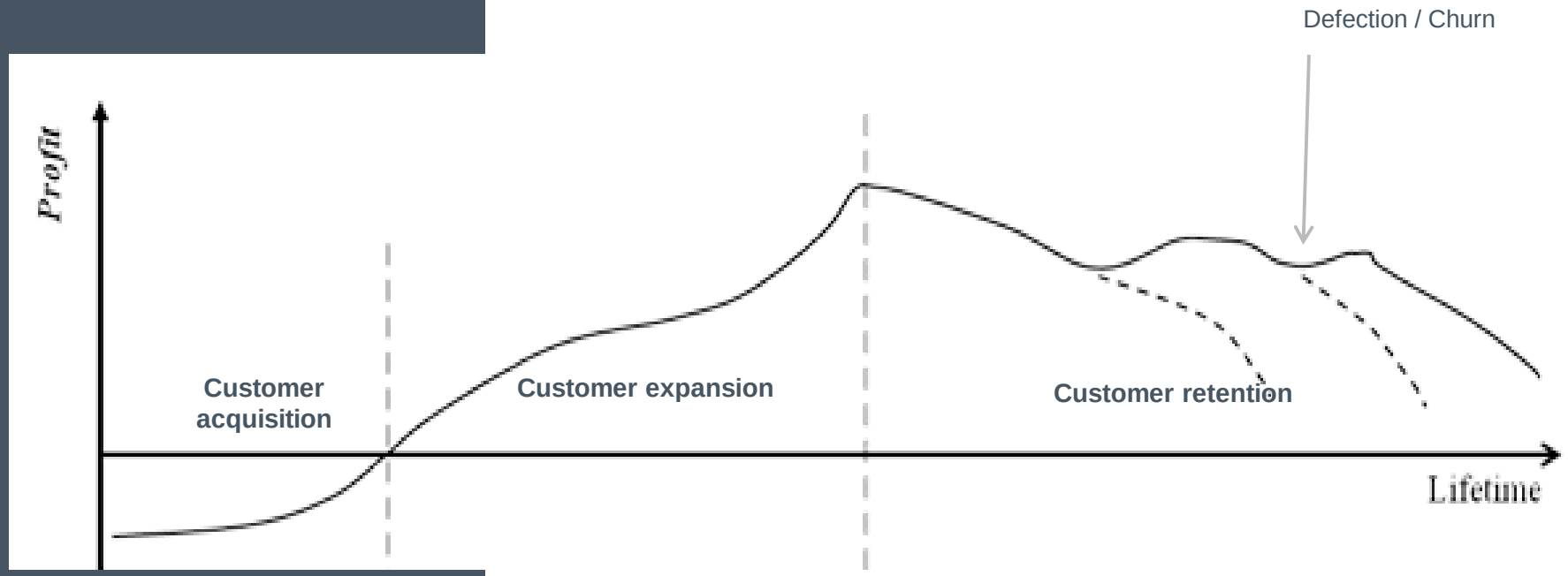
Tanimoto Coefficient!

$$T(a,b) = \frac{N_c}{N_a + N_b - N_c}$$

N_A – Number of Customers who bought Product A!
N_B – Number of Customer who bought Product B!
N_c – Number of Customer who bought both Product A and Product B!

Data trends per industry: **Services**

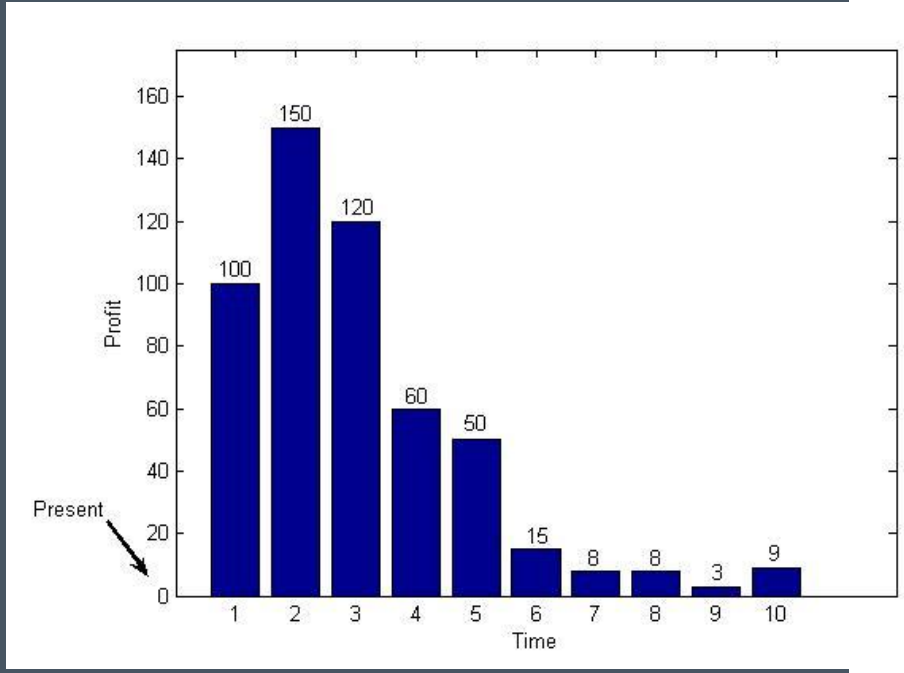
- Customer lifecycle



- RFM: Recency, Frequency, Monetary value

Data trends per industry: **Services**

Customer Lifetime Value (CLV)

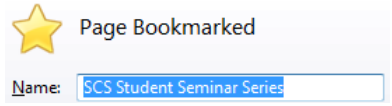


Customer Retention

Customer	Demographics	Consumption history	Marketing actions	Churn
1				???
2				???
…	…	…	…	…
…	…	…	…	…
9,998				???
9,999				???
10,000				???

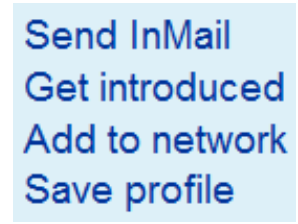
Data trends per industry: **Online**

- Click-through
- Browser action

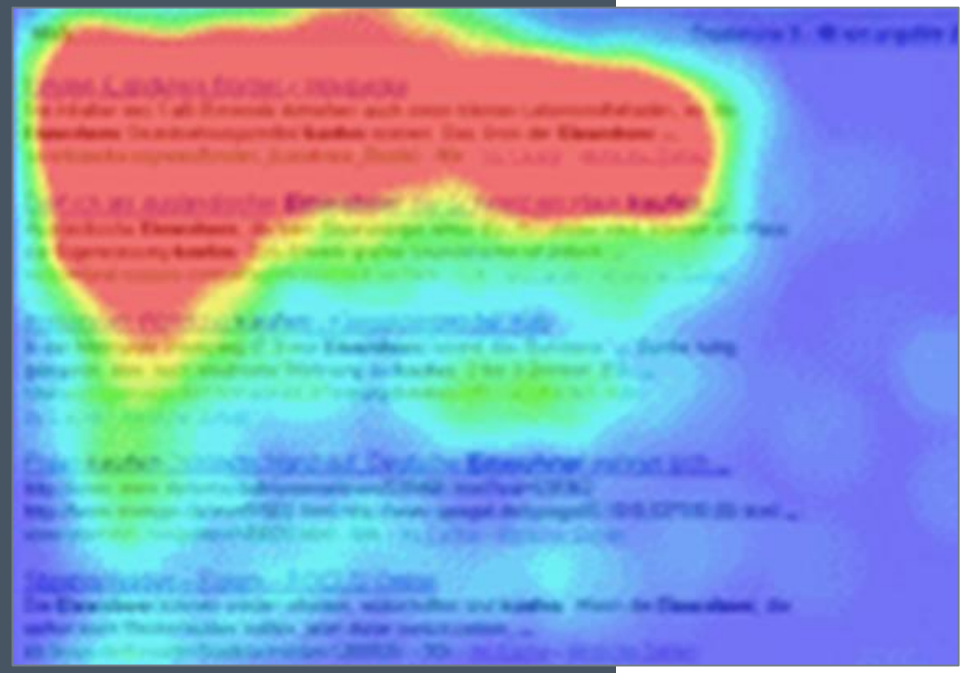


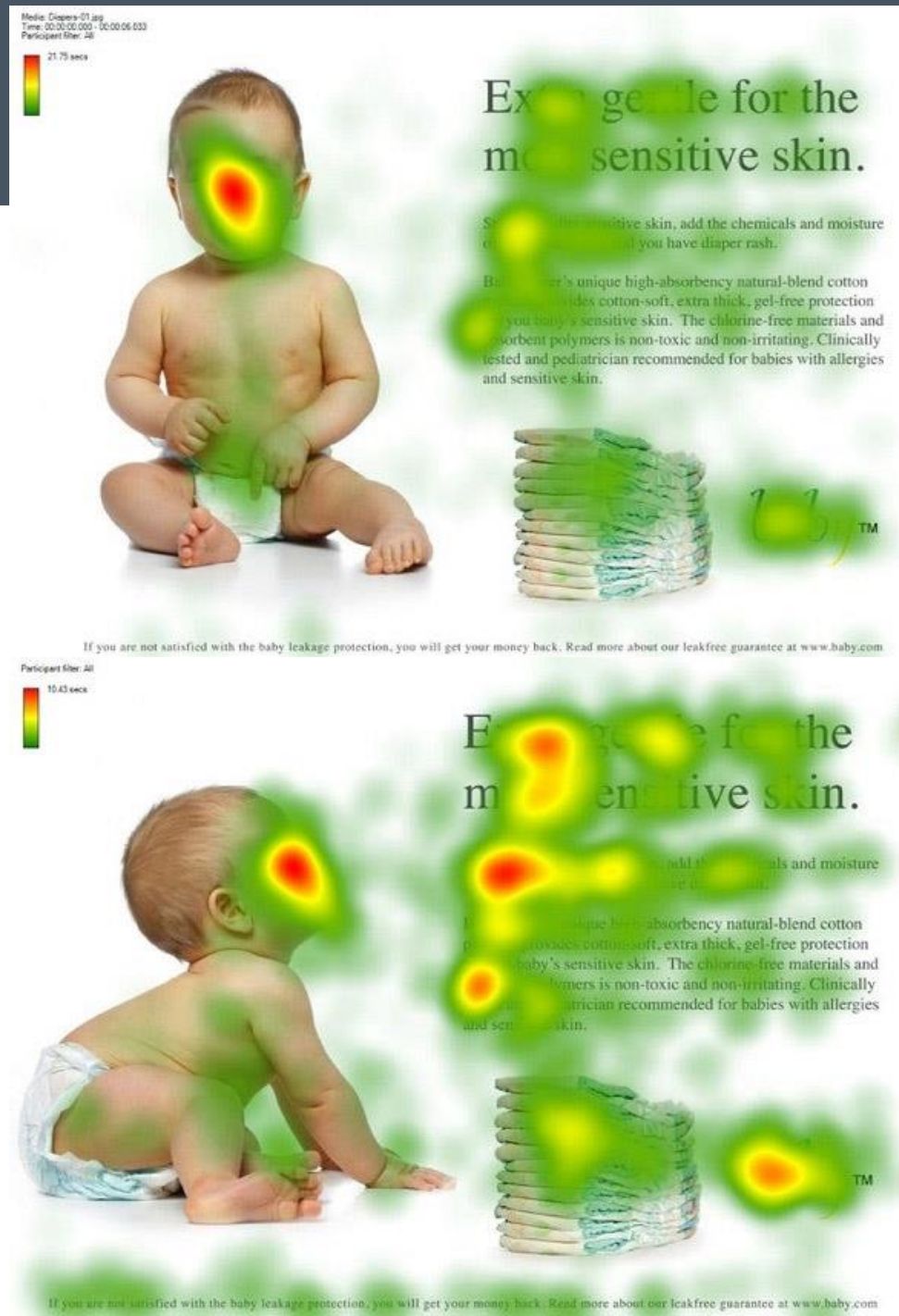
- Dwelling time
- Explicit judgment

- Reviews
- Other page elements

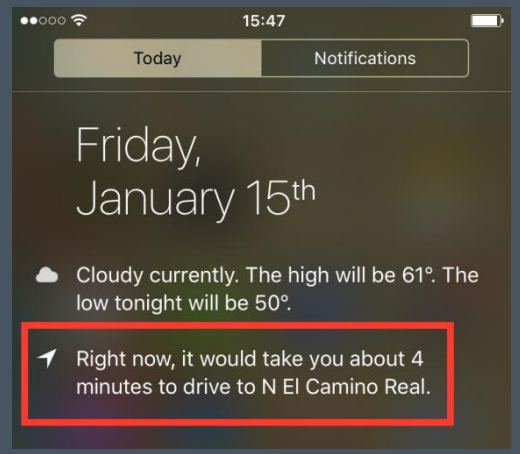


Data trends per industry: **Online**



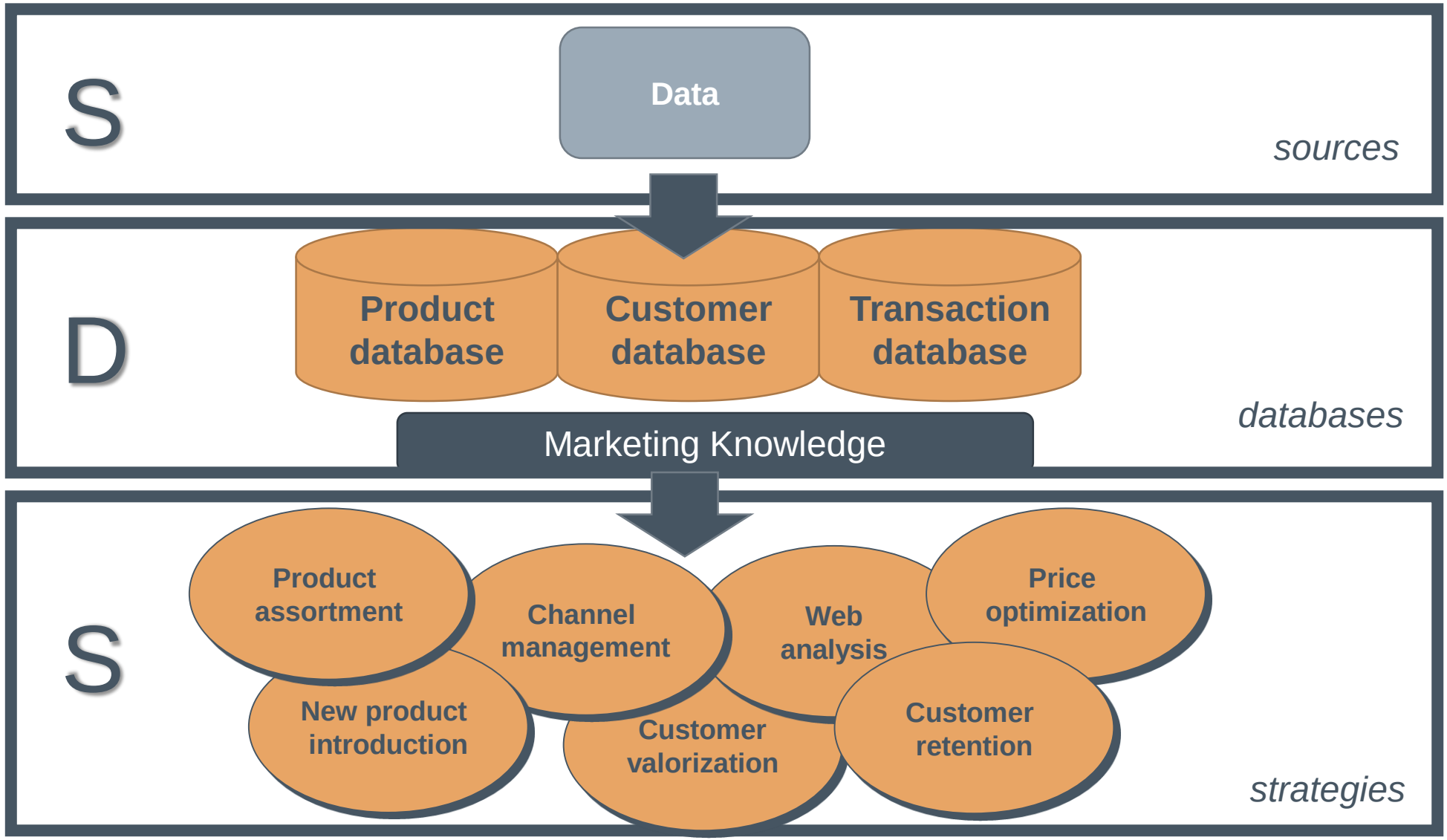


Data trends per industry: **Lifestyle and entertainment**

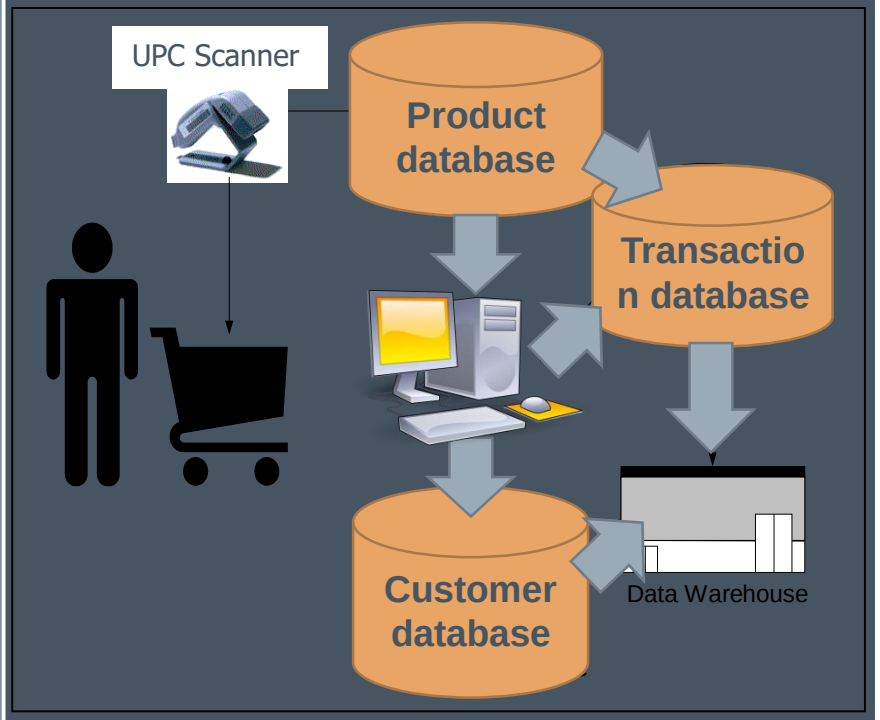


- Geo-localization
- https://disneyworld.disney.go.com/plan/my-disney-experience/bands-cards/

From Sources to Databases to Strategy (SDS Model)



Data Warehouses



- Stored into customer databases, transaction databases, and product databases.
  - Product databases: product features, prices, and inventories
    - attributes
  - Customer databases: customer characteristics and behavior
  - Transaction databases: …
- Data warehouses:
  - Store entire organization's historic data
  - Designed specifically to support analyses necessary for decision making
  - The data in the warehouse are separated into specific sub-parts, called data marts, and indexed for easy use.

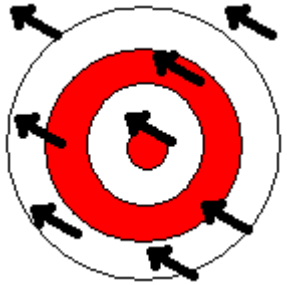
5. Reliability and validity

Data quality

Criteria for Judging Quantitative Research
validity
reliability
generalizability

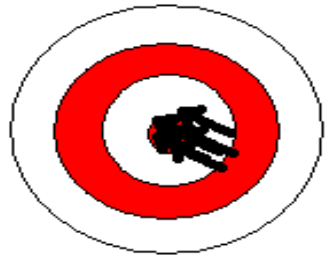
Data quality

Measurement validity ***Measurement reliability***



Not valid
Not reliable

Not valid
Reliable



Valid
Reliable

Data quality: Validity

Criteria for Judging Quantitative Research
Validity
Reliability
Generalizability

- Degree to which the tool measures what it claims to measure
- Think about:
  - Actual purchase data
  - Sentiment and text mining

Data quality: Reliability

Criteria for Judging Quantitative Research
Validity
Reliability
Generalizability

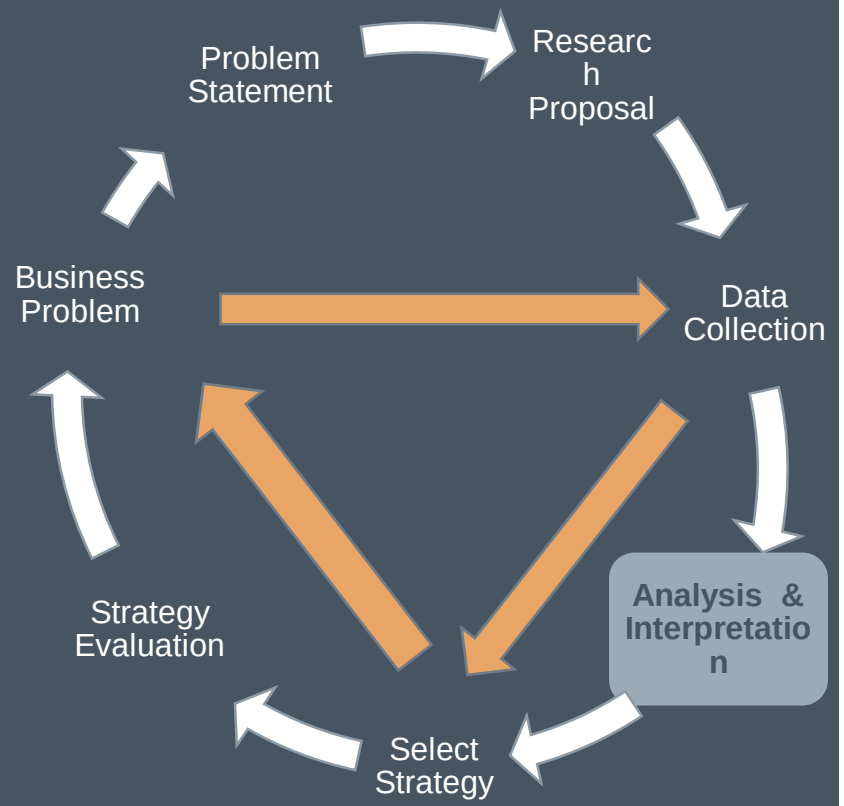
- Consistency of findings, the extent to which similar observations can be made by other researchers
- Think about
  - Persistent coding mistakes
  - Measurement error of the measurement tool (GPS tracker)
  - Missing values

Data quality: Generalizability (external validity)

Criteria for Judging Quantitative Research
Validity
Reliability
Generalizability

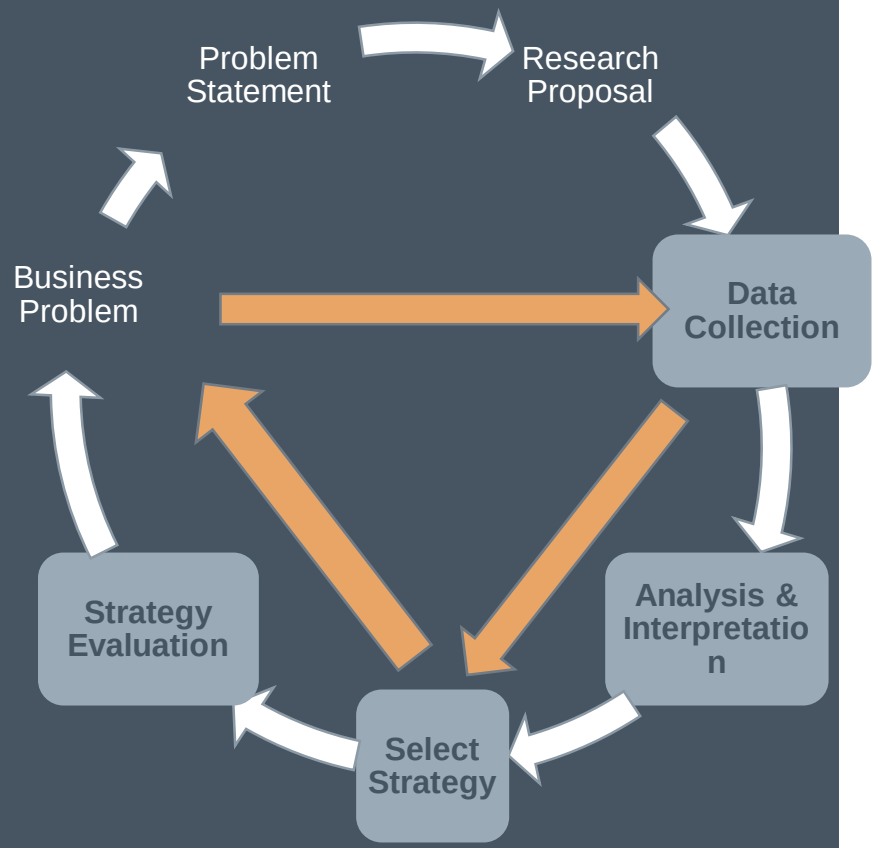
- Degree to which findings can be generalized to other people, contexts, places, times, settings
- How accurate will
  - Sales forecasts of other products be?
  - CLV Forecasts of new customers be?
- Timeliness of the data
- Is the firm policy or the market (structurally) different?
- Drawing causal inferences using field experiments

From data to knowledge



- Data is the necessary ingredient for a learning organization
- The CTO is responsible of collecting/maintaining the data for the CMO
- Marketing insight occurs between information and knowledge
  - Knowledge is more than information but resides in the employees
  - Employees create knowledge, computers are learning enablers

One example: Customer defection at T-Mobile



- Data: customer and transaction databases, customer satisfaction survey
- Analysis: logit model. Customers with a sudden decrease in consumption are most likely defectors.
- Strategy selection: call customers with a sudden change in consumption pattern
- Evaluation: implementation (field experiment). Churn rate decrease? Update strategy and continue…

Knowledge Management Metrics

- Business research is not cheap:
  - Need to weigh the cost of gaining additional information against the value of potential opportunities or the risk of possible errors from decisions made with incomplete information.
  - Storage cost of all those terabytes of data coming from the Web.
  - Proactive vs. reactive strategies?
- Two metrics are currently in widespread use:
  - **ROI.** Companies want to know:
    - Why they should save all those data.
    - How will they be used, and will the benefits in additional revenues or lowered costs return an acceptable rate on the storage space investment?
  - **Total Cost of Ownership (TCO).** Includes:
    - Cost of hardware, software, and labor for data storage.
    - Cost savings by reducing Web server downtime and reduced labor requirements.