B9651 – Marketing Analytics

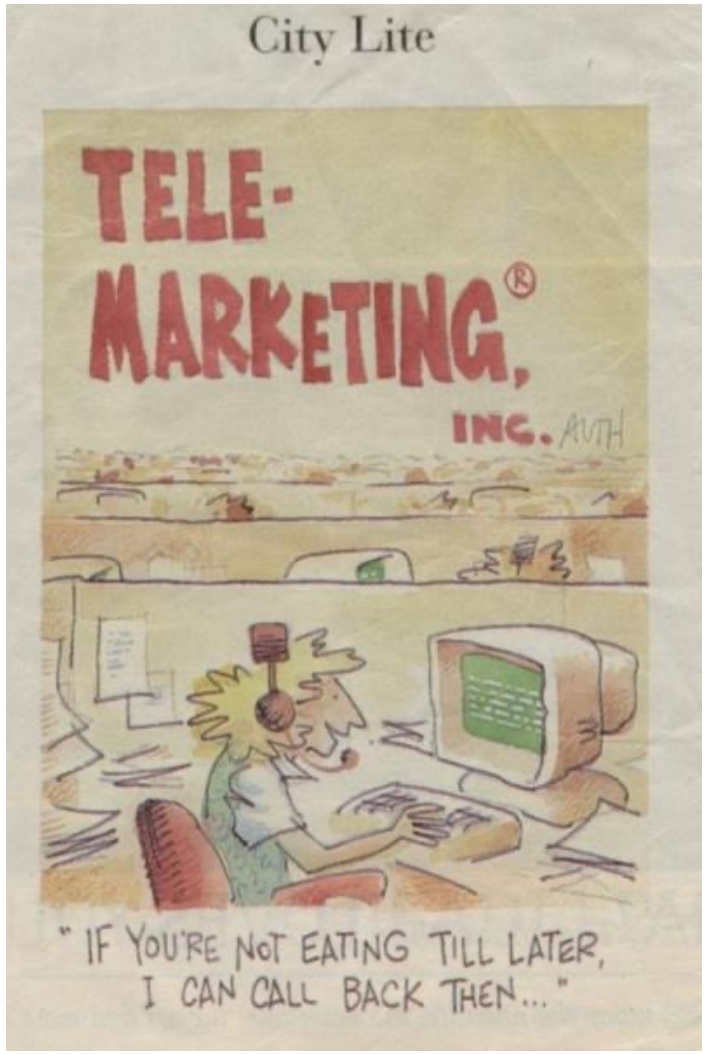
Session 1: Course Introduction + Marketing Datasets

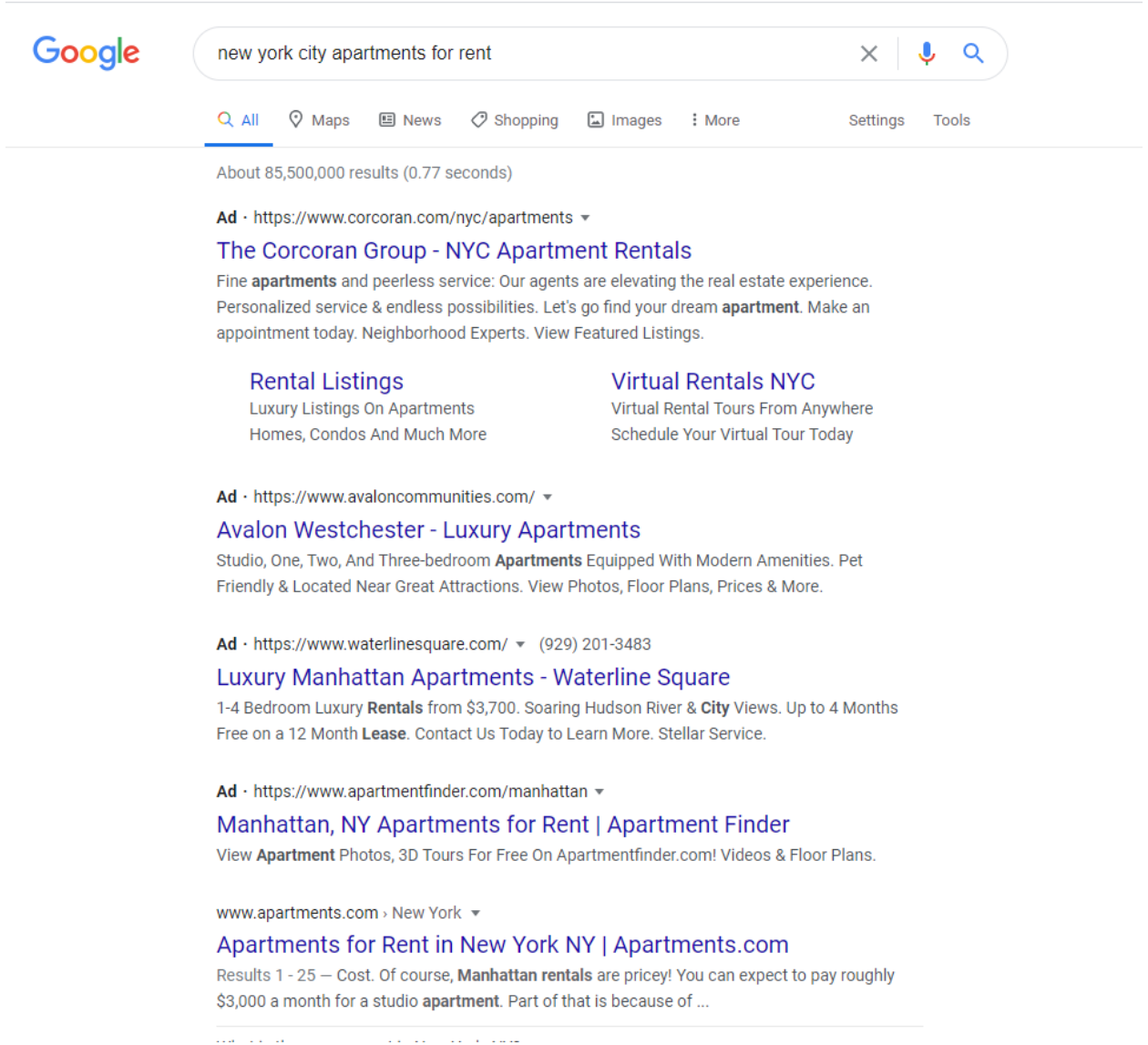
Professor Hortense Fong

A Primer on Marketing Problems: Online Banking

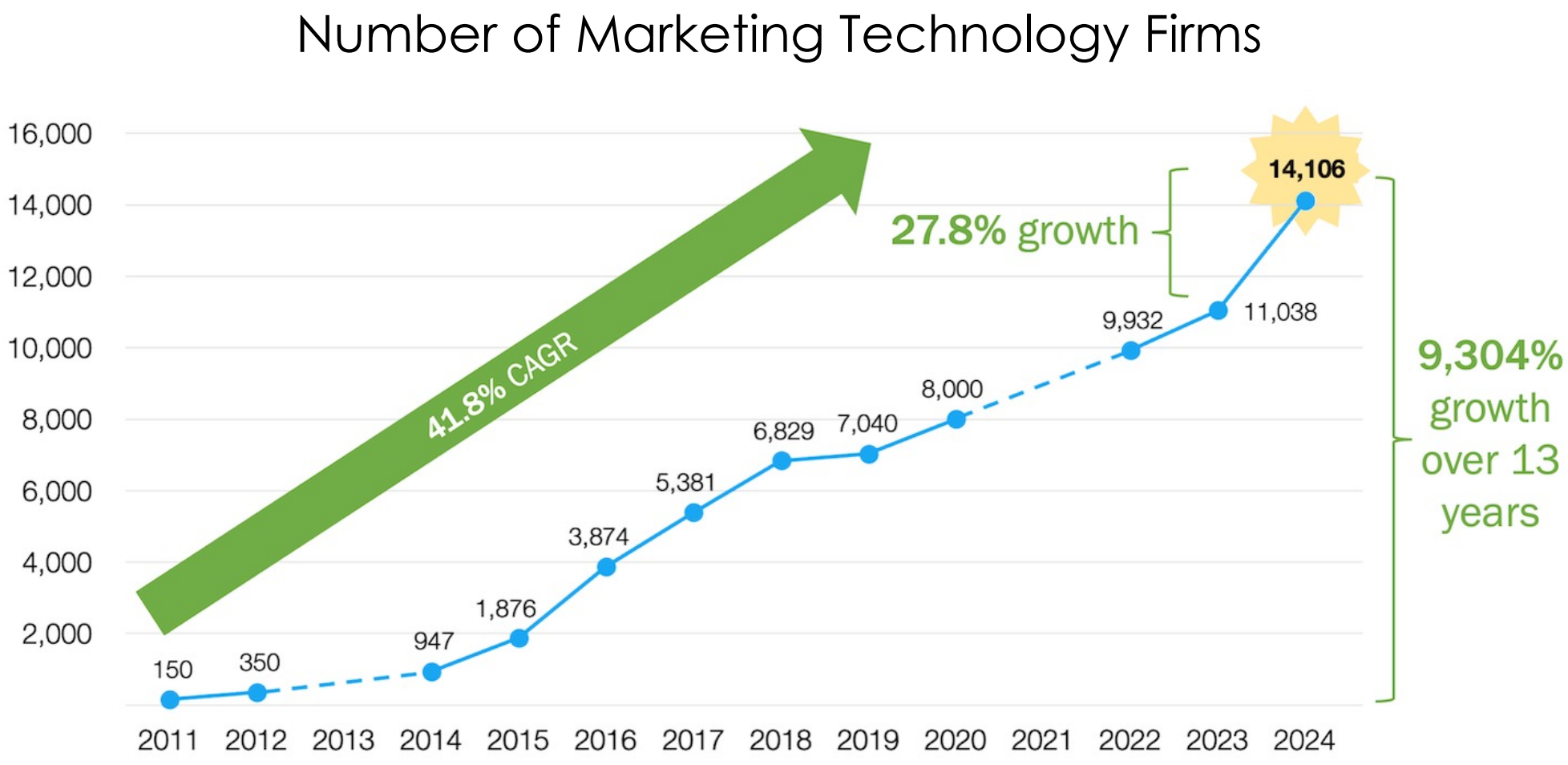
- 1990's: many firms started preparing for the "digital revolution"
  - Example: banks started investigating online banking
  - First analysis: compared profits generated by offline (traditional) vs online customers
    - Wanted to assess whether the new online channel increases profits
    - What happened? Offline customers were much more profitable than online!
  - What went wrong? In-depth analysis showed that online customers were much younger
    - Should compare younger offline customers with younger online customers and older offline customers with older online customers
    - → Found online customers more profitable!
- By the end of this class, you should be able to:
  - Ask relevant marketing questions
  - Know where and how to collect information/datasets
  - Extract meaningful insights!

What is Marketing?


City Lite
TELE-
MARKETING®
INC.
"IF YOU'RE NOT EATING TILL LATER,
I CAN CALL BACK THEN..."

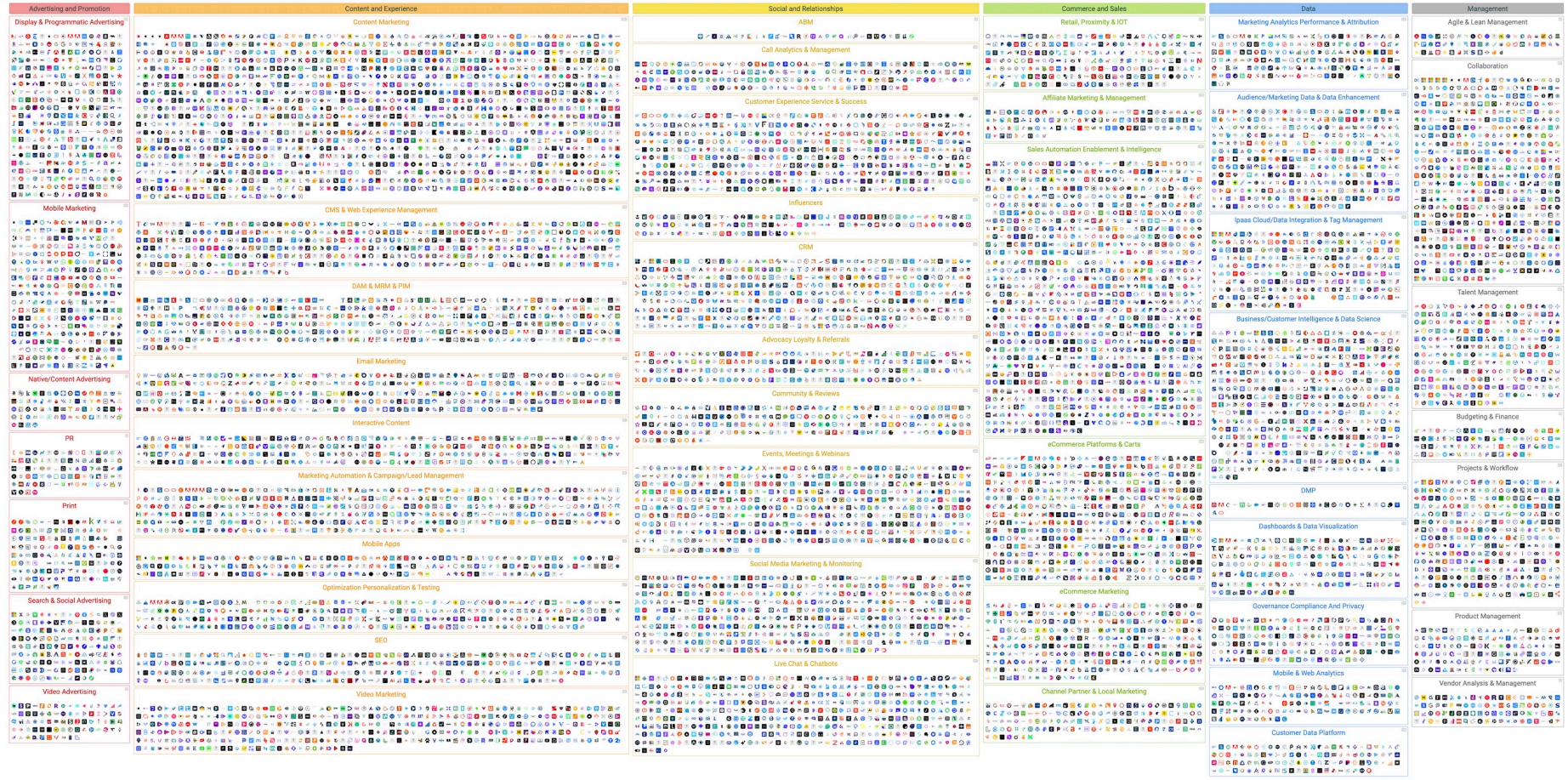

Google
new york city apartments for rent
All
Maps
News
Shopping
Images
More
Settings
Tools
About 85,500,000 results (0.77 seconds)
Ad · https://www.corcoran.com/nyc/apartments
The Corcoran Group - NYC Apartment Rentals
Fine apartments and peerless service: Our agents are elevating the real estate experience. Personalized service & endless possibilities. Let's go find your dream apartment. Make an appointment today. Neighborhood Experts. View Featured Listings.
Rental Listings
Luxury Listings On Apartments
Homes, Condos And Much More
Virtual Rentals NYC
Virtual Rental Tours From Anywhere
Schedule Your Virtual Tour Today
Ad · https://www.avaloncommunities.com/
Avalon Westchester - Luxury Apartments
Studio, One, Two, And Three-bedroom Apartments Equipped With Modern Amenities. Pet Friendly & Located Near Great Attractions. View Photos, Floor Plans, Prices & More.
Ad · https://www.waterlinesquare.com/ (929) 201-3483
Luxury Manhattan Apartments - Waterline Square
1-4 Bedroom Luxury Rentals from $3,700. Soaring Hudson River & City Views. Up to 4 Months Free on a 12 Month Lease. Contact Us Today to Learn More. Stellar Service.
Ad · https://www.apartmentfinder.com/manhattan
Manhattan, NY Apartments for Rent | Apartment Finder
View Apartment Photos, 3D Tours For Free On Apartmentfinder.com! Videos & Floor Plans.
www.apartments.com › New York
Apartments for Rent in New York NY | Apartments.com
Results 1 - 25 — Cost. Of course, Manhattan rentals are pricey! You can expect to pay roughly $3,000 a month for a studio apartment. Part of that is because of ...

Marketing Evolution – More Analytical & Complex

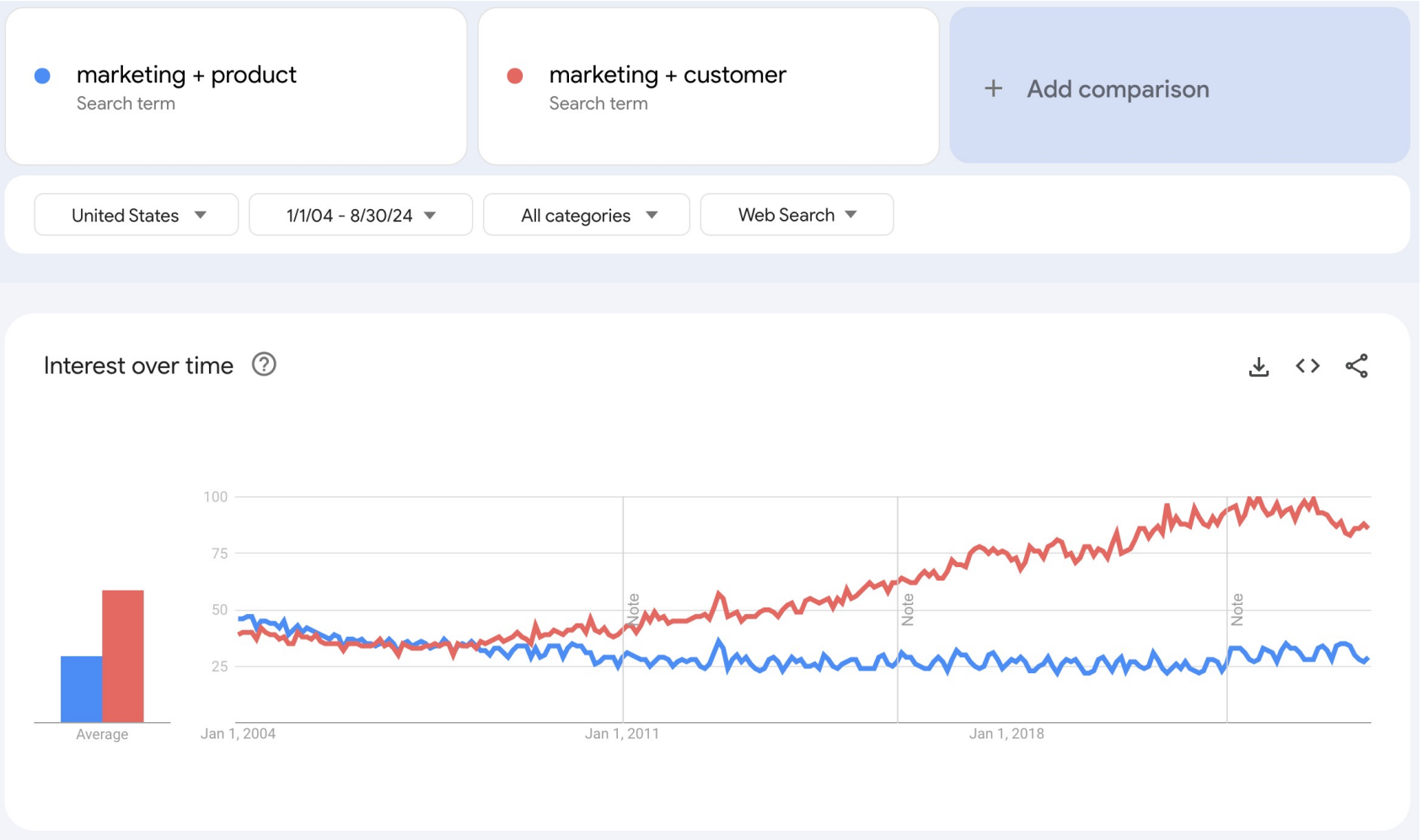


Source: https://chiefmartec.com/2024/05/2024-marketing-technology-landscape-supergraphic-14106-martech-products-27-8-growth-yoy/

Marketing Today



Google Trends on Marketing



This is (Mostly) the Past

Prior to the 90's

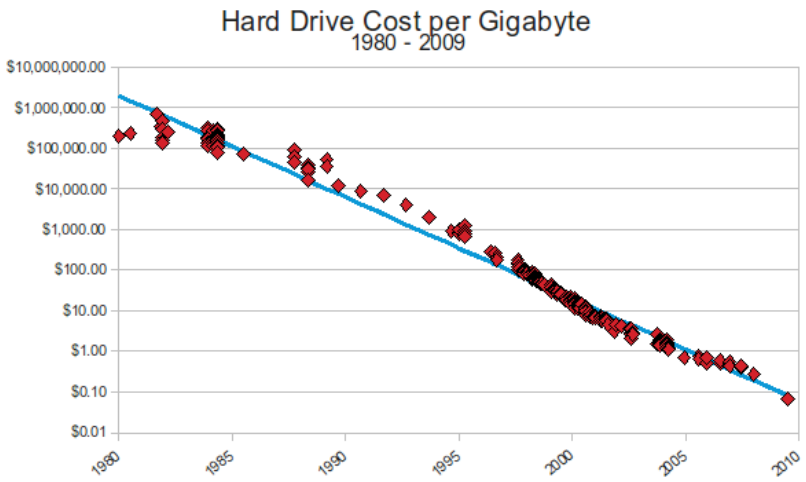
Products and brands were the center of the marketing activity (e.g., mass market, mass media, and impersonal transactions)

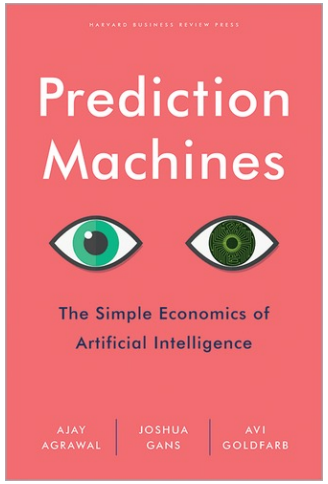
Build it and they will come!

(Customers exist to serve the marketplace)

A Few Technological Trends

- Internet – real time!
- Data – cheaper to store → more and more data
- Computing – faster and cheaper to process information → more computing





The 1990's

- Initial findings from financial services:
  - A high concentration of profits among customers

The profit distribution of a west coast US bank

Table 1.2 Percentage profit per household

$ Profit/ Household	% of Household	% of Balance	$ Profit
Over $600	0.19%	2.34%	$3 021 332
$550 to 599	1.45%	1.53%	$945 321
$350 to 549	1.78%	4.22%	$1 353 798
$200 to 349	2.80%	9.35%	$4 354 323
$150 to 199	3.88%	7.55%	$3 456 387
$100 to 149	6.03%	31.87%	$2 435 678
$0 to 99	13.88%	12.44%	$978 453
–$1 to –25	22.34%	14.32%	–$7 345 234
–$26 to –49	33.78%	9.90%	–$2 435 654
–$50 to –74	13.64%	5.50%	–$877 954
Under –$75	0.23%	0.98%	–$324 165
Total	100.00%	100.00%	$5 544 285

Not all customers are equal

The Result – Customer as a Unit of Analysis

- Aiming to proactively manage the ways individual customers, not only products, create profitability

Customer Centricity & Drivers of Revenue

- Before:
  - New product development
  - One-way street in terms of communication

- Now:
  - Acquisition (of new customers)
  - Retention (of current customers)
  - Expansion (of revenue from current customers)

The New Rules of Customer Centricity

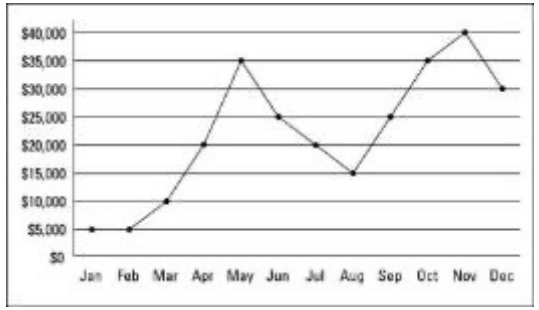
All Customers Are Created Equal

Customer Pyramid

DELTA

Trained service based on segment

Sales Statistics



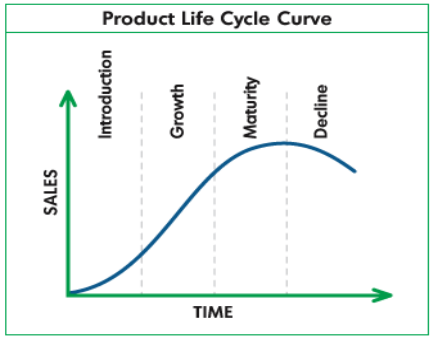
Customer (Predictive) Analytics

AMERICAN EXPRESS

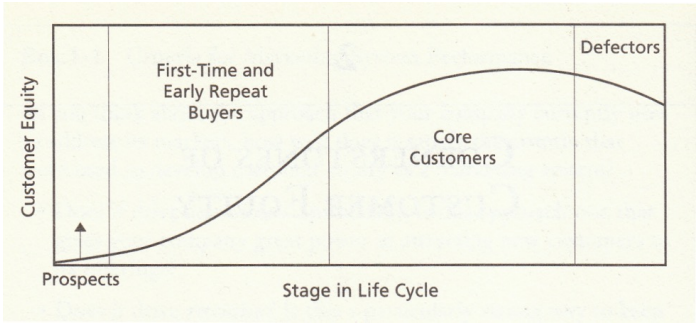
Promotions based on life events

The New Rules of Customer Centricity

Product Life Cycle

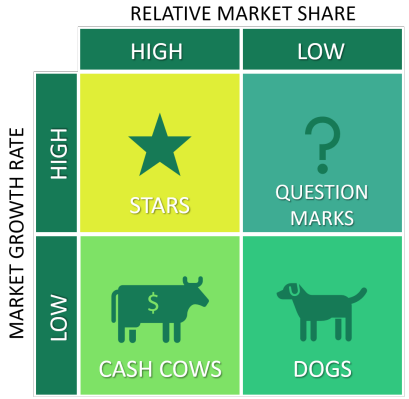


Customer Life Cycle

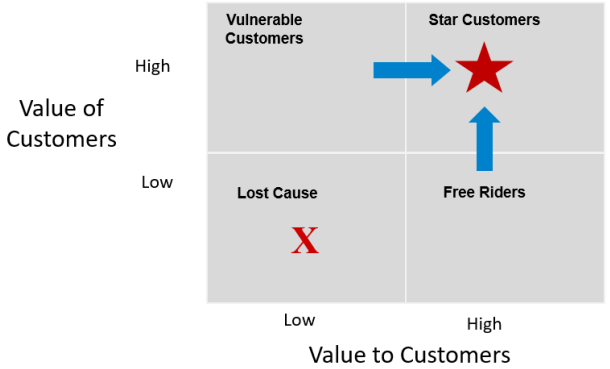


Boston Consulting Group Matrix

Product Portfolio



Customer Portfolio



Not All Customers are Equal

Costco employee stories

Current and former store employees chimed in to paint a picture of the types of returns customer service can expect to see on any given day. Here's a sample:

- Sony boombox, over a decade old
- Old mattress with clear signs of urine stains
- Container of bones and fat, all that was left of a ribeye steak
- 13-year-old fish that had been found in a freezer
- Playset because "the kids grew up"
- 4-year-old Whirlpool washing machine
- 10-year-old sneakers

Here's What Happens When You Return Too Many Items to Costco

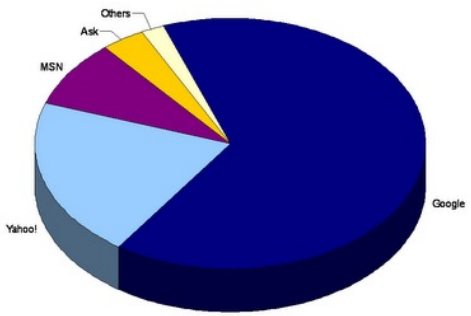
Published on May 4, 2023

- ✓ Too many returns to Costco and your account is likely to be flagged.
- ✓ Once flagged, you run the risk of losing membership privileges.

Source: https://www.fool.com/the-ascent/personal-finance/articles/heres-what-happens-when-you-return-too-many-items-to-costco

The New Rules of Customer Centricity

Market Share



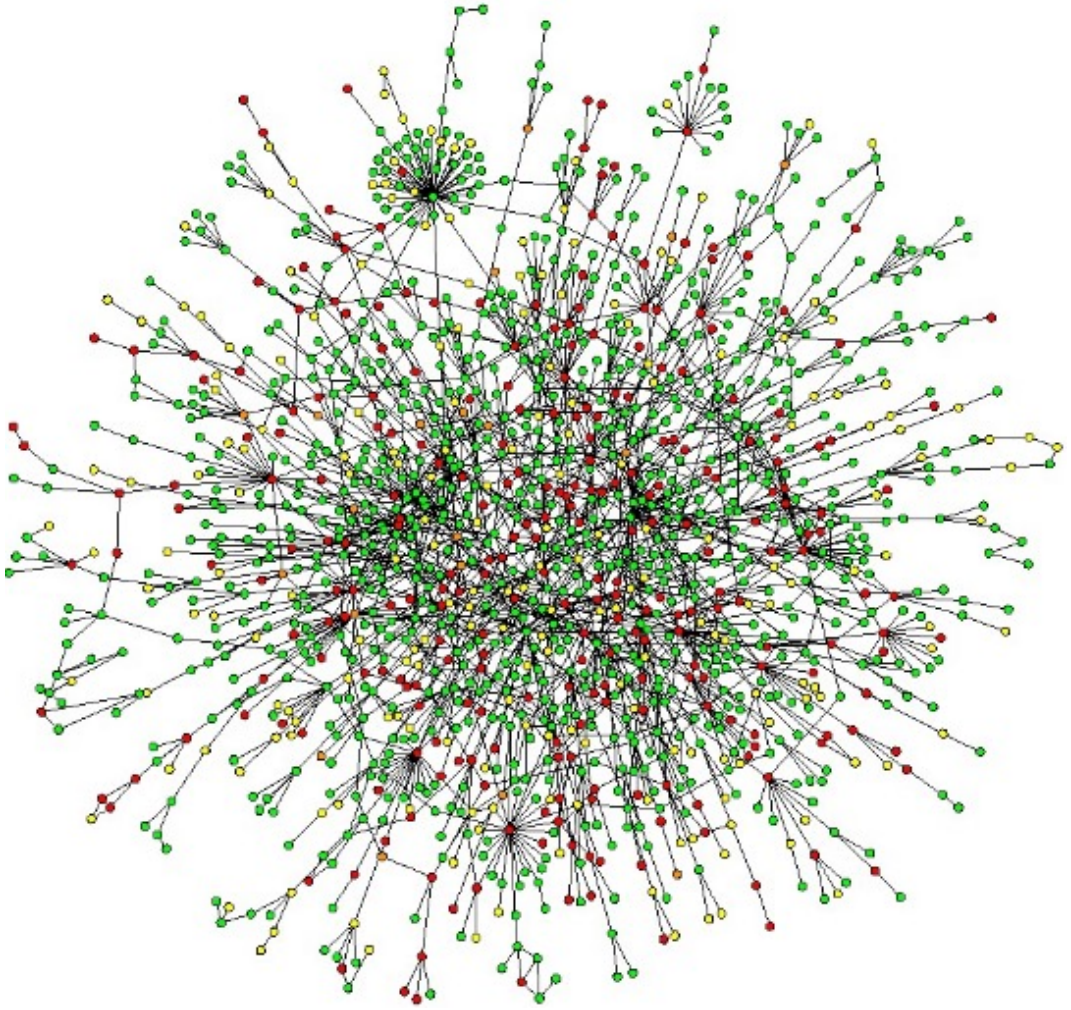
Share of Wallet

% of a customer's spending within a category that's captured by a given firm

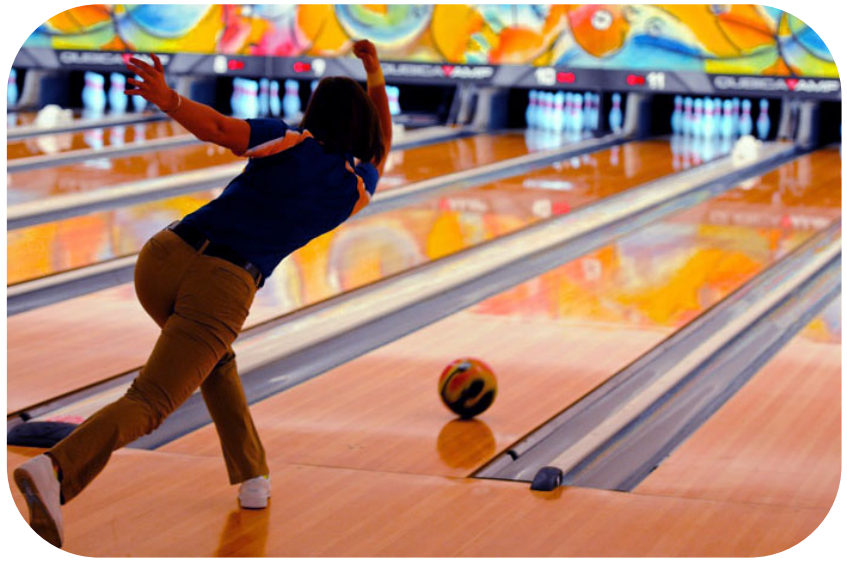
Segmentation

Customization

And Then Came The Social Networks…



Moving to a Non-linear World…



From bowling

To pinball

… to Complicate Things Further

Sotheby's – Pre-2020

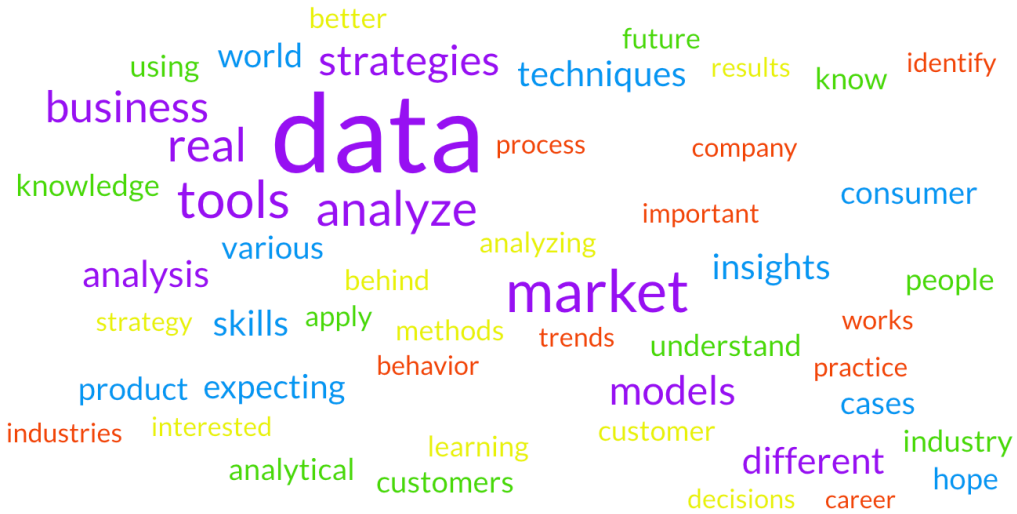
Sotheby's – 2020

Some practices will be completely changed in the foreseeable future…
creating new challenges and opportunities!

Marketing According to You

What you think Marketing is?

What you want to learn?



So… What is Marketing?

Marketing is the art and science of identifying, delivering, and capturing value by understanding customers and their needs.

Marketing Analytics is about bridging the gap between quantitative tools and substantive marketing questions.

Why Marketing Analytics Hasn't Lived Up to Its Promise

"Why Marketing Analytics Hasn't Lived Up to Its Promise"

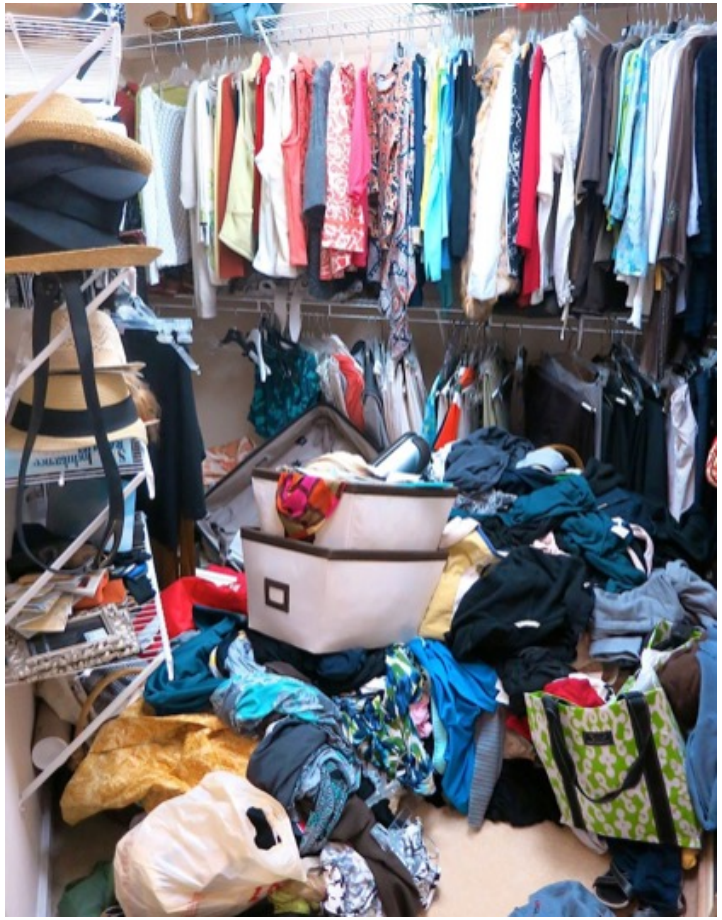
2018: 198% increase in planned allocation of marketing budgets to analytics in 3 years

But the effect of analytics on company-wide performance is modest (4.1 out of 7, on average).

Why are firms increasing marketing budgets when analytics does not contribute that much?

"Why Marketing Analytics Hasn't Lived Up to Its Promise"

- **Data: Too much data = too little *information***
  - Ubiquitous, but difficult to obtain insight (cluttered closet)
  - Not integrated across platforms (e.g., mobile & PC browsing)
  - Can't infer causality (e.g., correlation between search ad and purchase does NOT mean the ad caused the sale)



"Why Marketing Analytics Hasn't Lived Up to Its Promise"

Key solution for an organized closet…

Decide what to do FIRST, then decide which data you need

"Why Marketing Analytics Hasn't Lived Up to Its Promise"

- **Analyst Talent: Skill boundaries**
  - Clearly define the problem
  - Identify best tools
  - Spend time with customers IRL, *outside* data
  - Communicate insights, not jargon
  - Span skill boundaries (coding AND communication)
  - Takes time and practice → this class is the first step

Goal	Tool
Find causality (e.g., does ad increase sales)	Experiment
Prediction (e.g., expected future sales)	Supervised machine learning
Qualitative social media analysis (e.g., customer segments)	Unsupervised machine learning

Today's Agenda

Course Outline

Expectations

Marketing Datasets

Today's Agenda

Course Outline

Expectations

Marketing Datasets

Course Roadmap

STP Analytics (Identify Value)	Customer Analytics (Deliver Value)	4P Analytics (Capture Value)
Module 1	**Module 2**	**Module 3**
What datasets can we use? How can we segment and target our customers? How should we position our products/services?	How much are our customers worth? Are our customers leaving? How do our customers make choices?	How do we build a new product? How should we price our products? How do we distribute them? How do we quantify the impact of our promotions?

By the end of this class, you will become an informed…

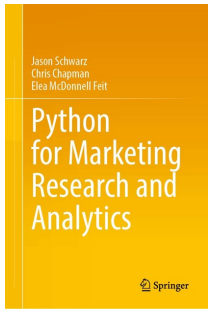
Provider of Data & Analysis	Provider and Consumer of Data & Analysis	Consumer of Data & Analysis
• Marketing research • Marketing analytics • Business analytics • Data scientists	• Tech • Start-ups • Consulting	• Product management • Brand management • Venture capital • Finance

Evaluation

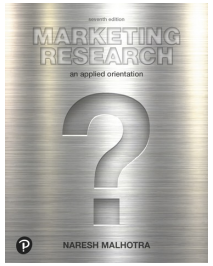
- Participation & Attendance (10%)
  - In-class discussions
  - Pre-class case surveys and/or concept checks
    - Graded on completion
- Assignments (50%)
  - Three individual (10% each)
  - Two group (groups of four formed randomly) (10% each)
  - **No late assignments will be accepted**
- Midterm & Final Exam (40%)
  - Midterm – closed book (15%), in-class Oct 22-23
  - Final – one-page cheat sheet (25%), Week of Dec 9

Textbooks

- *Python for Marketing Research and Analytics* textbook available free online @ clio.columbia.edu



- Optional textbook for reference—Marketing Research: An Applied Orientation by Naresh Malhotra



Today's Agenda

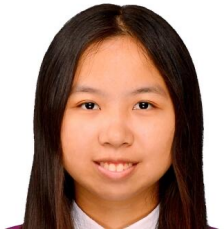
Course Outline

Expectations

Marketing Datasets

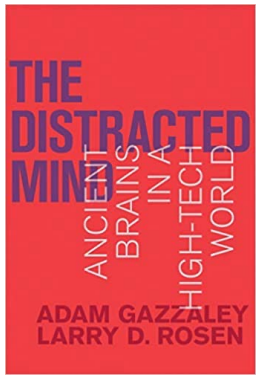
What should you expect from us?

- Emails will be returned usually within 24-48 hrs
- Fair evaluation of student work
- Professor Office Hours Wed 2-3:30PM by appointment, TA Office Hours Thurs 10-11AM by appointment
  - Google doc sign-up
- Slides posted before class
- TAs
  - Penny Chen (YChen26@gsb.columbia.edu)
  - Eli Sugerman (Esugerman25@gsb.columbia.edu)
  - Yihan Luo (yl5426@columbia.edu)



What do I expect from you?

- Brush up on Python/Excel if needed
- Attendance is required
- Come to class prepared
  - Do the pre-class surveys and readings
  - Prepare the assigned cases
- Need full attention in class
  - Cellphones are turned off
  - No laptops, tablets are okay for notes
  - Have a pen and paper for note taking
- Be engaged in class discussion with classmates
- Challenge me: make sure you understand!
- Submit assignments on time – no late assignments will be accepted



Today's Agenda

Course Outline

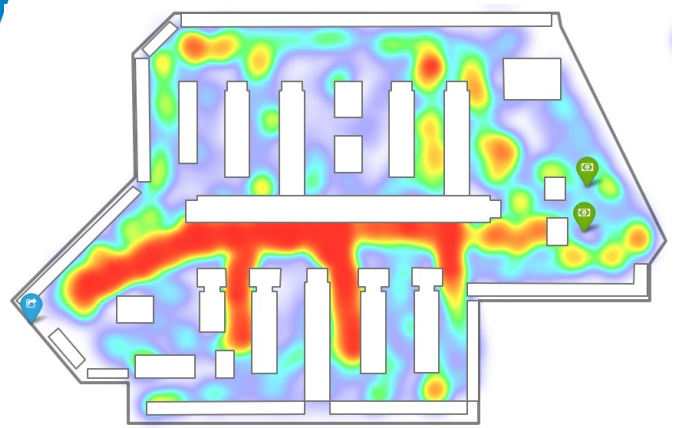
Expectations

Marketing Datasets

Course Roadmap

STP Analytics (Identify Value)	Customer Analytics (Deliver Value)	4P Analytics (Capture Value)
Module 1	**Module 2**	**Module 3**
What datasets can we use? How can we segment and target our customers? How should we position our products/services?	How much are our customers worth? Are our customers leaving? How do our customers make choices?	How do we build a new product? How should we price our products? How do we distribute them? How do we quantify the impact of our promotions?

The Starting Point: Data



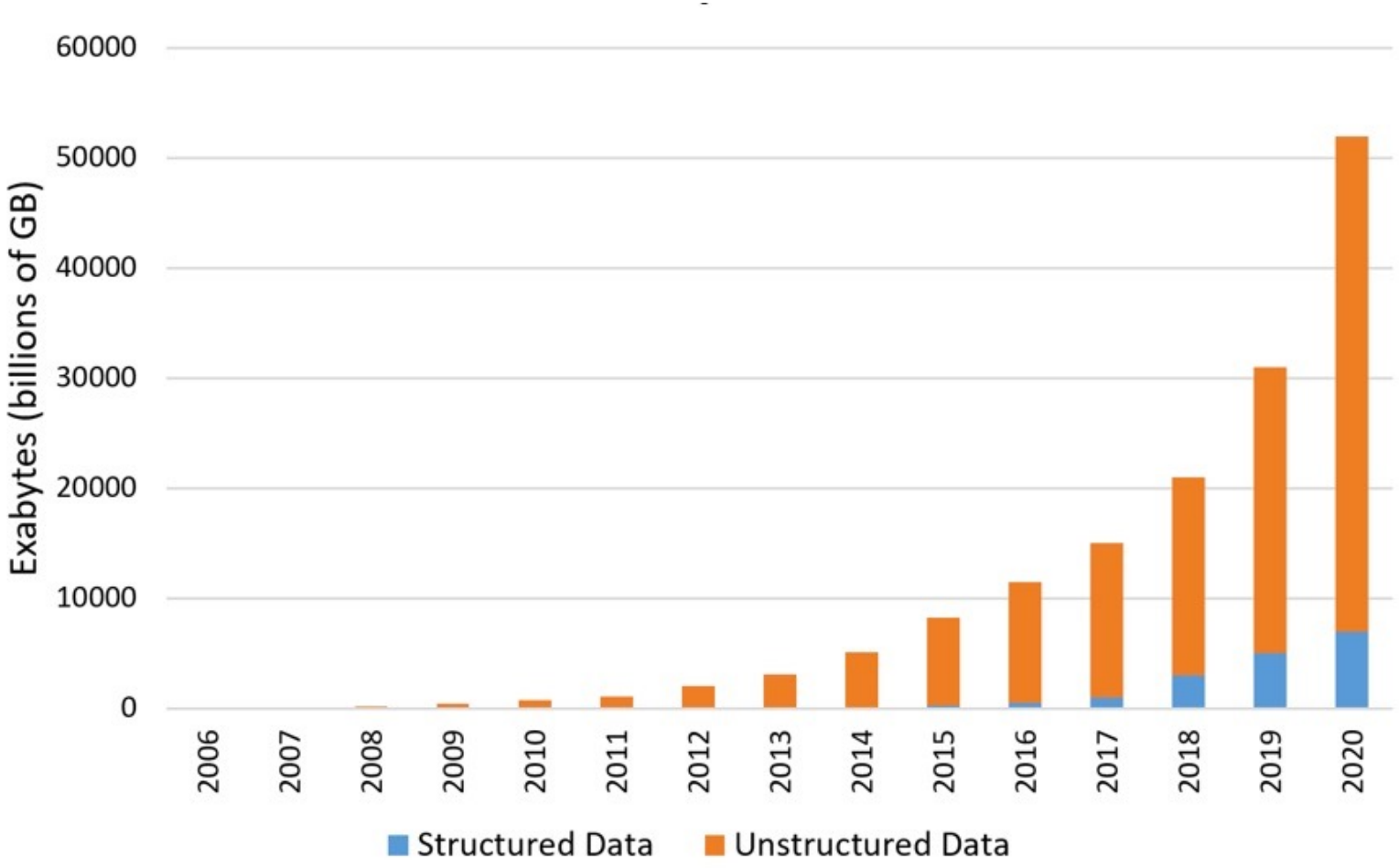
We Generate TONS of Data!

- ~44 zettabytes of data in the world beginning of 2020
  - One zettabyte = 10^{27}
- Predicted to be ~175 zettabytes of data by 2025

40x more data than observable stars in the universe!

Source: https://seedscientific.com/how-much-data-is-created-every-day/

Data Explosion!



Source: https://www.eetimes.com/author.asp?section_id=36&doc_id=1330462

Login to Poll Everywhere

Step 1: Go to **pollev.com/register** in your browser.

Step 2: Enter **hortensefong** in the dialog box and tap Search.

Why do we need data? Are there more words that begin with the letter K or more words that have K as their third letter?

0

Words that begin with the letter K

Words that have K as their third letter

But…Why do we need data?

A typical text contains twice as many words that have K as the third letter rather than the first

K first

1. Kitchen
2. Kangaroo
3. Kale

K third

1. Ask
2. Cake
3. Biking
4. Fake
5. Hiking
6. Ink

Availability Bias:
People assess the likelihood of something by the ease with which instances of that thing can be brought to mind.

Source: Tversky, Amos, and Daniel Kahneman. "Availability: A Heuristic for Judging Frequency and Probability." Cognitive Psychology 5, no. 2 (1973): 207–32.

Why do we need data/analysis to make decisions?

1. Humans have limited memory, and a limited ability to process and synthesize data
2. People are bad at learning from experience
    - **Overconfidence bias:** *subjective confidence is greater than objective accuracy*
    - **Confirmation bias:** *the tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs*
3. People are bad at judging probabilities

We (also) Need Structure

- Each dataset has pros and cons
  - What are they?
  - What type of question can I answer?
- Important to quickly know what is and is not possible with your data
- We need data and question taxonomies!

Data Taxonomy

	Primary Data *Data that is gathered by the researcher for the purpose of answering a specific question.*	**Secondary Data** *Data that was gathered for a purpose other than answering the specific question.*
Structured *Data that can be easily and meaningfully represented and manipulated in a traditional database (spreadsheet). Typically numeric or "choice" data.*	Surveys (ratings, choice) Experiments	Transaction logs Scanner panel data Ad tracking Product usage data
Unstructured *Data that cannot be meaningfully stored in a traditional data structure (spreadsheet) without further processing. Examples include text, images, video, and voice.*	Focus groups Interviews Surveys (free response) Observation Eye tracking Physiological/neural	Online reviews Social media Most digital content Call logs

Types of Marketing Research

Exploratory Research

(Ambiguous Problem)

"Our sales are declining and we do not know why."

Descriptive Research

(Aware of Problem)

"What kinds of people are buying our products?"

"Who buys our competitors' products?"

Causal Research

(Problem Clearly Defined)

"Will buyers purchase more of our product in a new package?"

Types of Marketing Research

Exploratory Research

(Ambiguous Problem)

"Our sales are declining and we do not know why."

Use to…

1. Develop initial hunches or insights
2. Run a pilot study

Tools: secondary data, focus groups, survey opinion leaders, observation, etc.

Types of Marketing Research

Descriptive Research

(Aware of Problem)

"What kinds of people are buying our products?"

"Who buys our competitors' products?"

Use to…

1. Generate data describing characteristics of relevant customers
2. More specific and systematic than exploratory

Tools (similar to exploratory): secondary data, focus groups, surveys, etc.

Types of Marketing Research

Causal Research

(Problem Clearly Defined)

"Will buyers purchase more of our product in a new package?"

Use to…

1. Identify **cause-and-effect** relationships: If I do X, then Y.

Tools: Usually requires an **experiment** (e.g., A/B testing), there are quantitative techniques to identify causal relationships without an experiment

Example: New Coke, 1985

Taste tests showed that consumers preferred the taste of New Coke over old Coke and Pepsi, but…

How could this have happened?

Explanations

- Drinking experience (short sip versus full can)
- None of the marketing research informed the tasters that New Coke was going to replace the old Coke
  - Coke had a strong brand identity and consumers identified with the old brand
  - Consumers wanted to retain 'The Real Thing'
- Target market: bringing new customers in vs. pleasing existing ones

In short: they mistook exploratory for conclusive!

Primary Data

Unstructured

Data Taxonomy

	Primary Data *Data that is gathered by the researcher for the purpose of answering a specific question.*	**Secondary Data** *Data that was gathered for a purpose other than answering the specific question.*
Structured *Data that can be easily and meaningfully represented and manipulated in a traditional database (spreadsheet). Typically numeric or "choice" data.*	Surveys (ratings, choice) Experiments	Transaction logs Scanner panel data Ad tracking Product usage data Geospatial tracking
Unstructured *Data that cannot be meaningfully stored in a traditional data structure (spreadsheet) without further processing. Examples include text, images, video, and voice.*	**Focus groups** **Interviews** Surveys (free response) **Observation** **Eye tracking** **Physiological/neural**	Online reviews Social media Most digital content Call logs

Insights from Primary Marketing Data

- **Demographic characteristics** (Who)
- **Attitudes / opinions** (Thoughts)
- **Awareness** (Consideration set)
- **Intentions** (Intended actions)
- **Motivation/Protocol** (Why)
- Can explain discrepancy between:
  - Intention vs behavior ("I intended to buy, but…")
  - Attitude vs intention ("I like this product, but I won't buy it")

How to Gather Primary Data?

Ask: *what do people say/think*

Qualitative methods → *interpret* what people say

- In-depth interviews
- Focus groups

Quantitative methods → *measure* what people think/say

- Surveys

Observe: *what do people do*

- Direct observation
- Field experiments

How to Gather Primary Data?

Ask: *what do people say/think*

Qualitative methods → *interpret* what people say

- In-depth interviews
- Focus groups

Quantitative methods → *measure* what people think/say

- Surveys

Observe: *what do people do*

- Direct observation
- Field experiments

Qualitative Research

- **Qualitative research** aims to ***interpret what people say*** about their experiences.

- Studies tend to be **exploratory.** Use when you want to:
  - Probe attitudes and behaviors, looking for new opportunities
  - Establish basis for quantitative research

- Methods enable participants to express themselves openly/without constraint

Qualitative Research: In-depth Interviews

- **In-depth interviews** attempt to obtain detailed insights into consumers' lived experiences.

- **Format** is flexible (conducted at place of business, home, point of consumption, etc.

- Can aid in:
  - ***New Product Development:*** Understand a market and identify gaps
  - ***Repositioning:*** Understand how a customer perceives the brand

And yet, Febreze didn't sell…

"The first inkling came when they visited a home outside Phoenix. They could smell her **nine cats** before coming inside. The house's interior, however, was **clean** and **organized**. She was somewhat of a **neat freak**, the woman explained… when they walked into the living room, where the cats lived, **the scent was so overpowering** that one of them gagged."

Q: What do you do about the cat smell?
A: It's usually not a problem

Q: How often do you notice the smell?
A: Oh, about once a month

Q: Do you smell it now?
A: No

Qualitative Research: In-depth Interviews

Potential concerns with interviews:

- ***Observer Interference:*** being observed changes what we say
- ***Spreading activation:*** the brain retrieves memories by moving through networks of ideas. The ideas we prompt first determine the path
- ***Representativeness:*** does the sample match our customer base?
- ***Leading questions:*** risk pushing participants in a particular direction

Qualitative Research: Focus Groups

Focus groups attempt to capture the dynamics of consumer attitudes, feelings, beliefs, experiences, and reactions.

Format typically includes 8-10 *carefully chosen* (and incentivized) individuals, 1-2 moderators, for 1 hour.

Can aid in: developing product concepts, products, ad copy, script for further studies, questionnaire design

Qualitative Research: Focus Groups

Potential concerns with focus groups:

- ***Observer Interference:*** being observed changes what we say
- ***Spreading activation:*** the brain retrieves memories by moving through networks of ideas. The ideas we prompt first determine the path.
- ***Representativeness:*** does the sample match our customer base?
- ***Leading questions:*** risk pushing participants in a particular direction
- Verbal and nonverbal ***power and dominance cues*** from participants

Qualitative Research: Focus Groups

Guidelines for conducting a focus group:

- When an idea comes up for discussion, stick to that idea until the group finishes with it
- One speaker at a time
- Everyone has to participate
- Encourage participants to disagree, if they do
- Have and follow a script
- Collect demographics at the end

Interviews vs. Focus Groups

	Focus Groups	Interviews
Social effects	Interpersonal idea stimulation	No peer pressure or power/dominance
Information content	More people per unit time (breadth)	More information per respondent (depth)
Logistics	Difficult to schedule, especially with targeted recruiting Small interviewer commitment Moderate analysis cost	Easy to schedule, especially with targeted recruiting Huge interviewer commitment High analysis cost?

Data Taxonomy

	Primary Data *Data that is gathered by the researcher for the purpose of answering a specific question.*	**Secondary Data** *Data that was gathered for a purpose other than answering the specific question.*
Structured *Data that can be easily and meaningfully represented and manipulated in a traditional database (spreadsheet). Typically numeric or "choice" data.*	Surveys (ratings, choice) Experiments	Transaction logs Scanner panel data Ad tracking Product usage data Geospatial tracking
Unstructured *Data that cannot be meaningfully stored in a traditional data structure (spreadsheet) without further processing. Examples include text, images, video, and voice.*	**Focus groups** **Interviews** Surveys (free response) **Observation** **Eye tracking** **Physiological/neural**	Online reviews Social media Most digital content Call logs

How to Gather Primary Data?

Ask: *what do people say/think*

Qualitative methods → *interpret* what people say

- In-depth interviews
- Focus groups

Quantitative methods → *measure* what people think/say

- Surveys

Observe: *what do people do*

- Direct observation
- Field experiments

Direct Observation

Observation attempts to watch how customers interact with products/services in the "real word."

"Hidden observer" – researcher disguised as a shopper

Can identify:

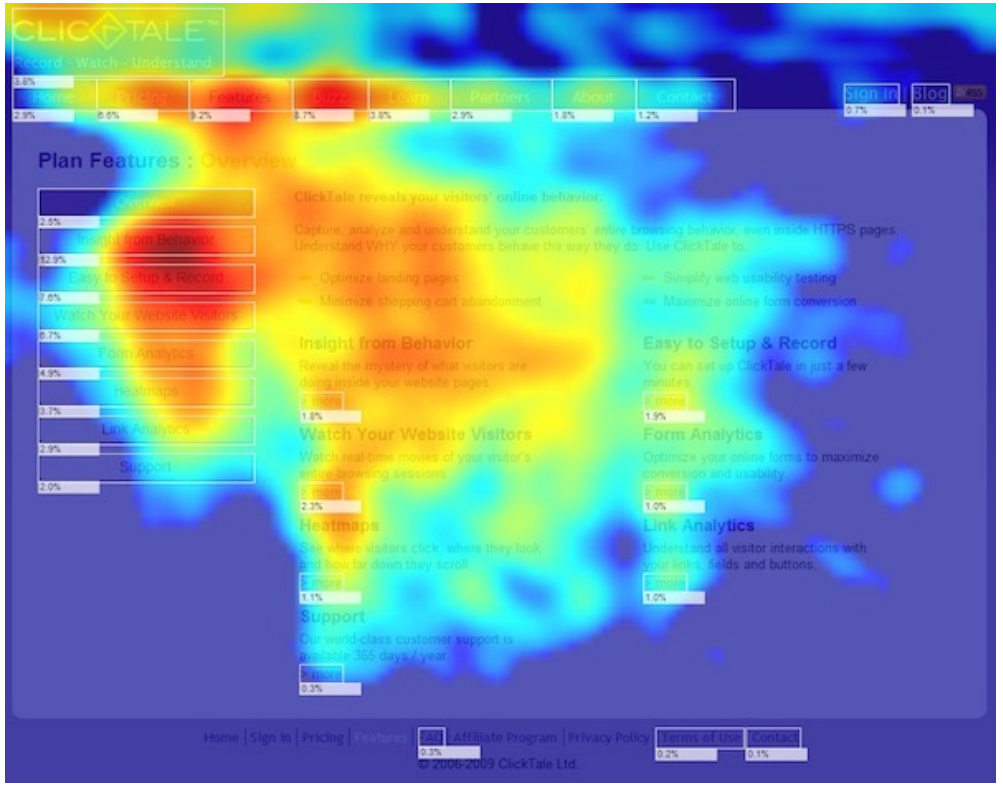
- How long are lines?
- What are people looking at?
- How easy is it for them to find products?

Direct Observation

Observation attempts to watch how customers interact with products/services in the "real word."

Observing product usage

- In-person experiences
- Online user testing
- Biometrics

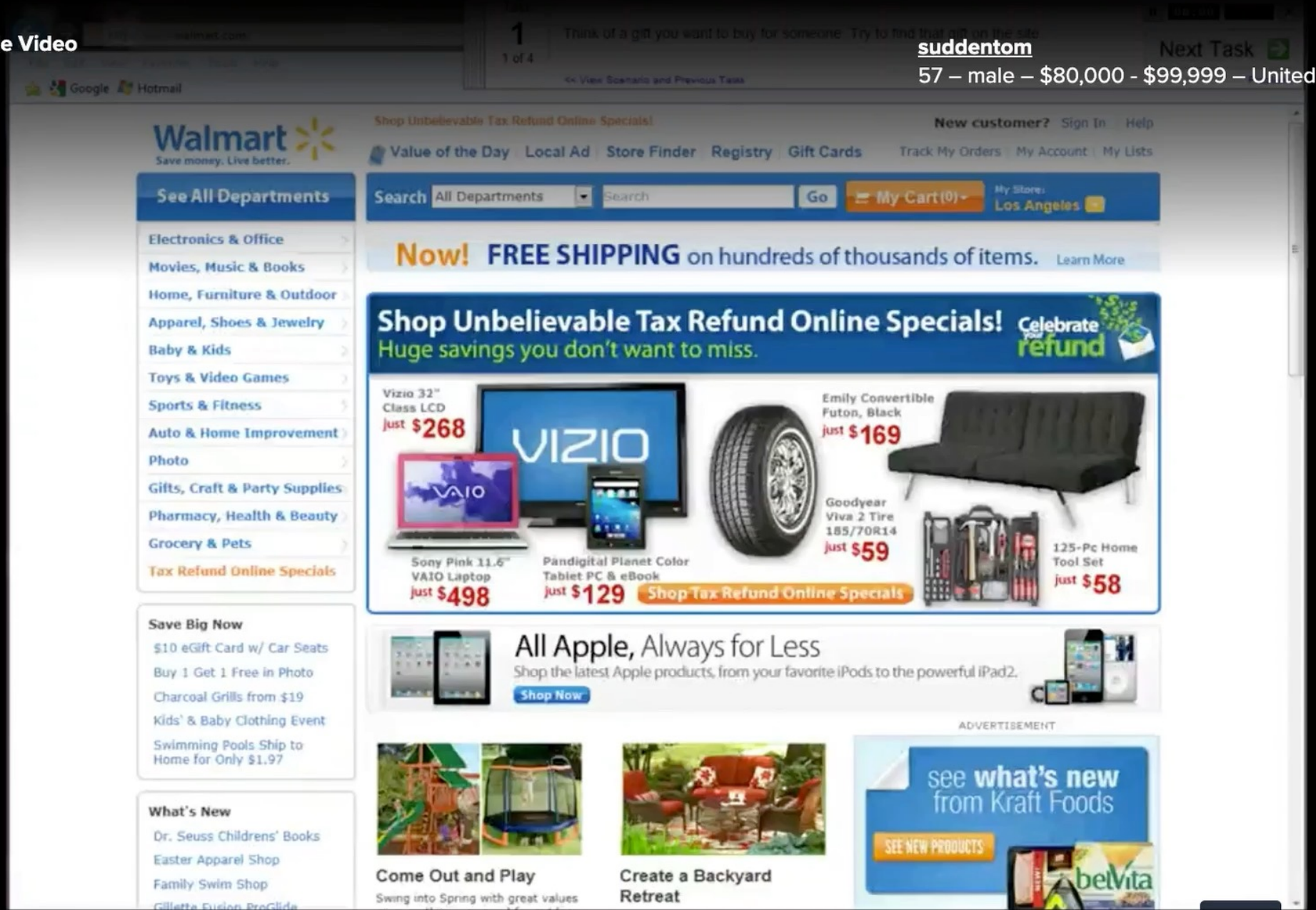


UserTesting

Walmart Sample Video

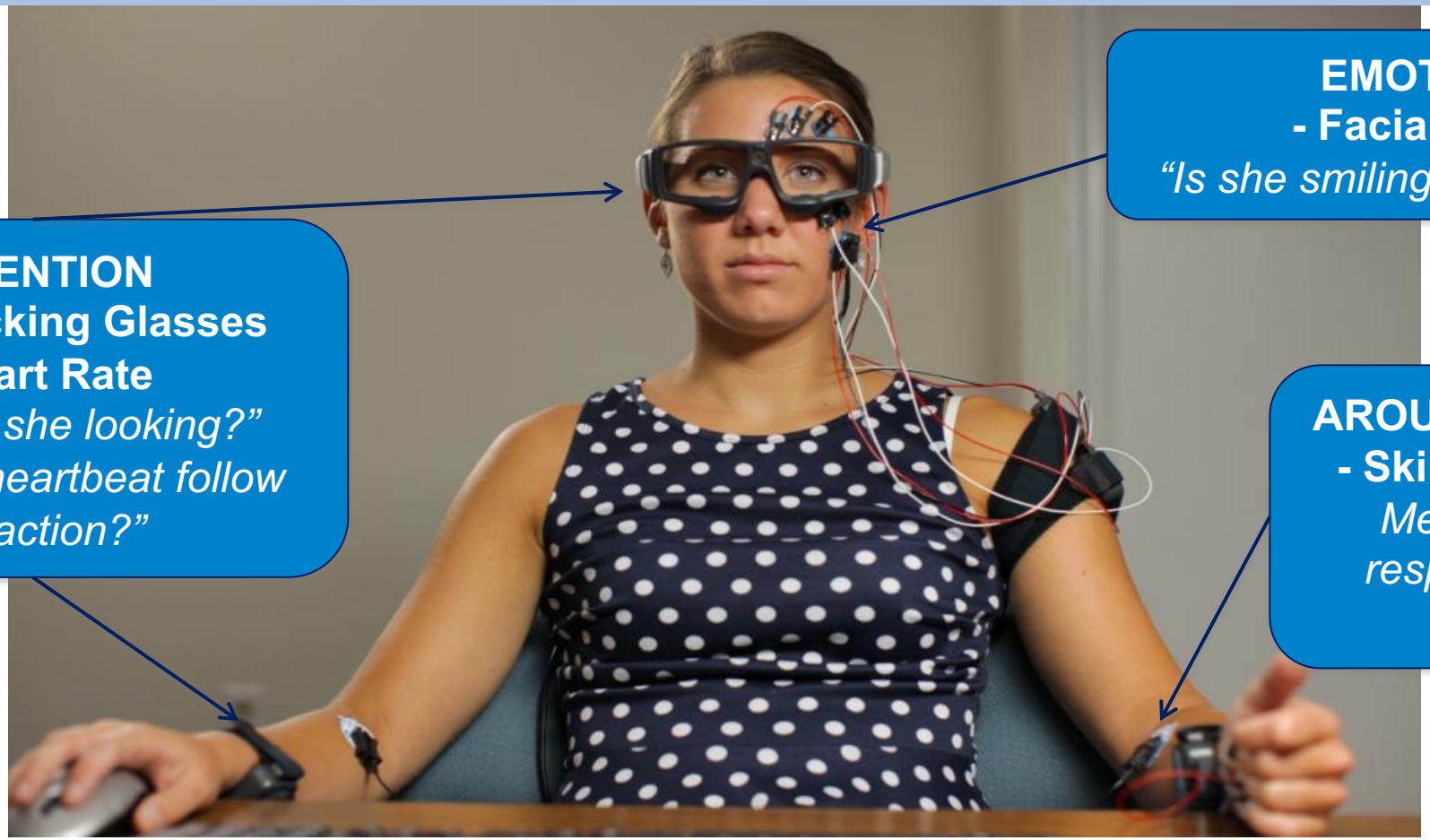
suddentom

57 – male – $80,000 - $99,999 – United States



Direct Observation – Biometrics Data

Obvious benefit of biometric data: don't have to rely on self report



EMOTION
- Facial EMG
"Is she smiling? Frowning?"

ATTENTION
- Eye-Tracking Glasses
- Heart Rate
"Where is she looking?"
"Does her heartbeat follow the action?"

AROUSAL/INTENSITY
- Skin Conductance
Measures stress response, anxiety, arousal

Primary Data for Exploratory Research

- Questioning
  - Qualitative methods:
    - In-depth interviews
    - Focus groups
  - Quantitative methods:
    - Surveys
- Observation:
  - Direct observation
  - Field Experiments

Pros and Cons

	Questioning	Observation
Versatility	+	-
Realness (Accuracy)	-	+
Respondent Convenience	-	+
Depth of Insights	+	-

Types of Marketing Research

Exploratory Research

(Ambiguous Problem)

"Our sales are declining and we do not know why."

Descriptive Research

(Aware of Problem)

"What kinds of people are buying our products?"

"Who buys our competitors' products?"

Causal Research

(Problem Clearly Defined)

"Will buyers purchase more of our product in a new package?"

Data Taxonomy

	Primary Data *Data that is gathered by the researcher for the purpose of answering a specific question.*	**Secondary Data** *Data that was gathered for a purpose other than answering the specific question.*
Structured *Data that can be easily and meaningfully represented and manipulated in a traditional database (spreadsheet). Typically numeric or "choice" data.*	**Surveys (ratings, choice)** **Experiments**	Transaction logs Scanner panel data Ad tracking Product usage data Geospatial tracking
Unstructured *Data that cannot be meaningfully stored in a traditional data structure (spreadsheet) without further processing. Examples include text, images, video, and voice.*	Focus groups Interviews **Surveys (free response)** Observation Eye tracking Physiological/neural	Online reviews Social media Most digital content Call logs

Primary Data

Structured

How to Gather Primary Data?

Ask: *what do people say/think*

Qualitative methods → *interpret* what people say

- In-depth interviews
- Focus groups

Quantitative methods → *measure* what people think/say

- Surveys

Observe: what do people do

- Direct observation
- Field experiments

Quantitative Research

Goal: ***observe*** and ***measure*** how people think and behave (results in numerical output)

Studies tend to be **descriptive** or **causal/confirmatory**:

- Identify the target base of customers (who are they)
- Identify what and why customers do what they do

Methods narrow in on a particular attitude or behavior:

- Observation
- Surveys
- Experiments

Quantitative Research: Survey

Surveys directly measure what consumers think, feel, or intend to do.

We need to think about a few things when designing surveys:

- How do we ask the question? (to avoid unsavory effects)
- Should the questions be open-ended or close-ended?
- What question type should we use? (measurement scales)
- Are the questions reliable? Are they valid?
- Is the sample representative of our customers?

Quantitative Research: Survey

How do we ask the questions?

Quantitative Research: Survey

Surveys directly measure what consumers think, feel, or intend to do.

When designing surveys, the way in which you ask questions is ***crucial***.

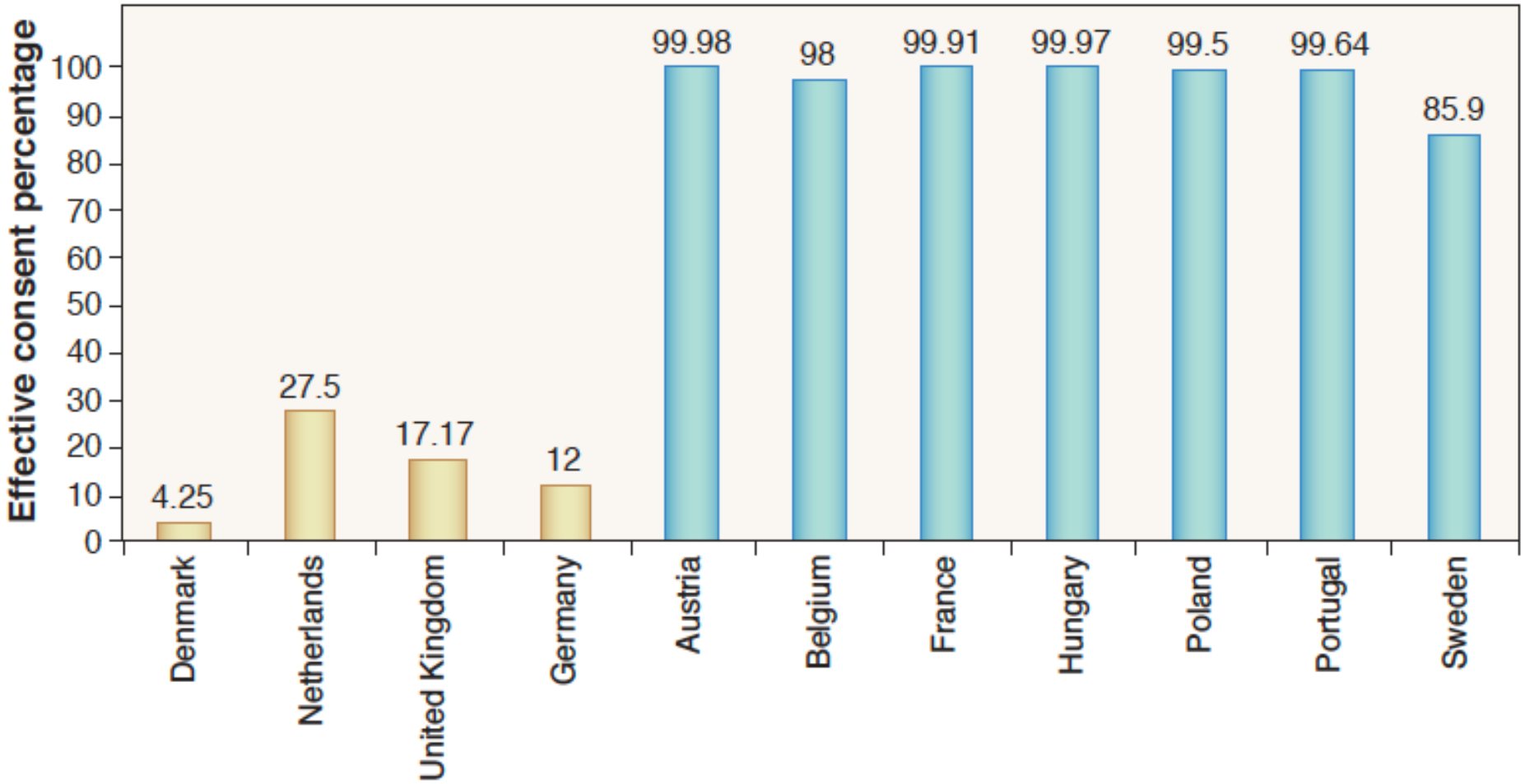
- **Priming effects:** preceding questions may influence response to focal questions
- **Framing effects:** the way you frame the question may change the answer
- **Leading questions:** when the question suggests a desired answer
- **Demand effects:** respondents may figure out your hypothesis and try to "help" you out

Priming Effects

Preceding questions may influence responses to the focal question:
How interested are you in buying this product?

Questions preceding the buying interest question	**"Very Much Interested" in Buying**
1. No question asked	2.8%
2. Asked only about advantages	16.7%
3. Asked only about disadvantages	0.0%
4. Asked about both advantages and disadvantages	5.7%

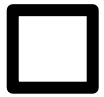
What Explains this Difference in Organ Donations?



Source: Johnson & Goldstein (Science, 2003)

Framing Effects

This is an example of a **default effect** – the tendency of people to stick with the default option.



Check this box if you want to opt in

Uncheck this box if you want to opt out

Framing Effects

A company designed a new automated car. Although the company worked hard on the security issues, it is **expected to injure 600 people**. To reduce this number, they designed two algorithms.

Assume the exact scientific estimate of the consequences of the algorithm are as follows:

Algorithm A: 200 people will be spared

Algorithm B: There is a 1/3 probability that 600 people will not be injured, and a 2/3 probability that no people will be spared.

Which algorithm do you prefer?

0

A

B

Framing Effects

A company designed a new automated car. Although the company worked hard on the security issues, it is **expected to injure 600 people**. To reduce this number, they designed two algorithms.

Assume the exact scientific estimate of the consequences of the algorithm are as follows:

72% **Algorithm A:** 200 people will be spared

28% **Algorithm B:** There is a 1/3 probability that 600 people will not be injured, and a 2/3 probability that no people will be spared.

Framing Effects

A company designed a new automated car. Although the company worked hard on the security issues, it is **expected to injure 600 people**. To reduce this number, they designed two algorithms.

Assume the exact scientific estimate of the consequences of the algorithm are as follows:

22% **Algorithm A:** 400 people will be injured

78% **Algorithm B:** There is a 1/3 probability that nobody will be injured, and a 2/3 probability that 600 people will be injured

Framing Effects

Positive Framing		Negative Framing	
72%	Algorithm A	22%	Algorithm A
28%	Algorithm B	78%	Algorithm B

Negative versus positive framing can make a huge difference, even if the options are identical!

Leading Questions

Leading: when the question suggests a desired answer

"Are you more likely to purchase now?" (Leading) vs.
"How likely are you to purchase now?" (Neutral)

"How good was your experience?" (Leading) vs.
"How do you feel about your experience?" (Neutral)

Could potentially cue participants to answer consistent with what you want to hear and increase demand!

Quantitative Research: Survey

Open-ended vs. Close-ended

Quantitative Research: Survey

Open-ended questions
(Unstructured)

Advantages:
- Similar to qualitative methods (depth)
- Captures things in the consumer's own language

Disadvantages:
- Depends on consumer's ability to articulate
- Difficult to analyze

Close-ended questions
(Structured)

Advantages:
- Easy to use, analyze, quantify
- Less threatening for respondent
- Less interviewer bias
- May jog respondent's memory

Disadvantages:
- Difficult to design
- Limited scope
- Usually requires pre-testing

Quantitative Research: Survey

What question type should we use?
(measurement scales)

Quantitative Research: Survey

Nominal (Categorical)	Ordinal	Interval (Ratings, Likert)	Ratio (Real numbers)

Quantitative Research: Survey

Nominal (Categorical)	Ordinal	Interval (Ratings, Likert)	Ratio (Real numbers)
When craving fast food, which of the following are you most likely to order? 1. Burger and fries 2. Pizza 3. Fried chicken **Labels have no order or meaning**			

Quantitative Research: Survey

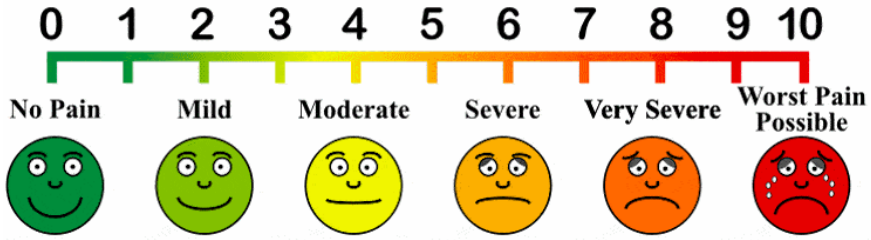
Nominal (Categorical)	Ordinal	Interval (Ratings, Likert)	Ratio (Real numbers)
When craving fast food, which of the following are you most likely to order? 1. Burger and fries 2. Pizza 3. Fried chicken **Labels have no order or meaning**	*What is the highest degree or level of education you have completed?* 1. Some high school 2. High school 3. Bachelor's Degree 4. Master's Degree 5. Ph.D. or higher **Options have an order, but no other numeric meaning**		

Quantitative Research: Survey

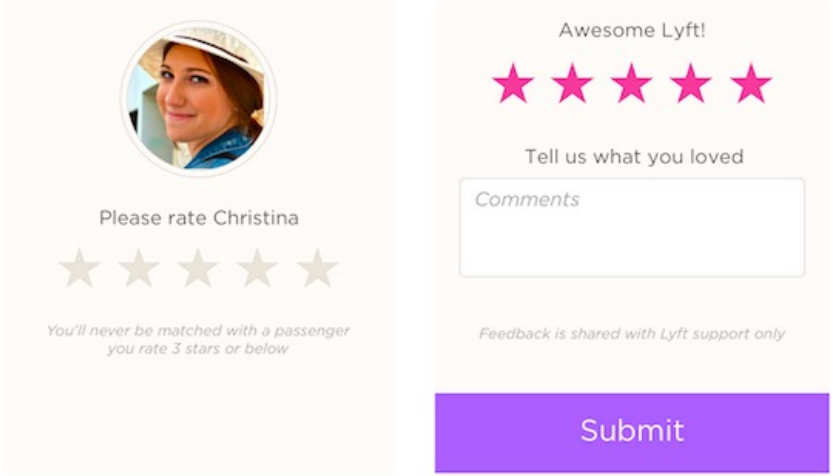
Nominal (Categorical)	Ordinal	Interval (Ratings, Likert)	Ratio (Real numbers)
When craving fast food, which of the following are you most likely to order?	*What is the highest degree or level of education you have completed?*	*Rate the extent to which you agree with the following: I love going to B9651.*	
1. Burger and fries 2. Pizza 3. Fried chicken	1. Some high school 2. High school 3. Bachelor's Degree 4. Master's Degree 5. Ph.D. or higher	1. Strongly agree 2. Agree 3. Neutral 4. Disagree 5. Strongly Disagree	
Labels have no order or meaning	**Options have an order, but no other numeric meaning**	**Options have an order, and the *intervals* between options are equal**	

Interval Scales: Considerations

- Individual-level perceptions:



- Are scale differences truly equal?



Quantitative Research: Survey

Nominal (Categorical)	Ordinal	Interval (Ratings, Likert)	Ratio (Real numbers)
When craving fast food, which of the following are you most likely to order?	*What is the highest degree or level of education you have completed?*	*Rate the extent to which you agree with the following: I love going to MKTG-212.*	*What is your annual pre-tax income?*
1. Burger and fries 2. Pizza 3. Fried chicken	1. Some high school 2. High school 3. Bachelor's Degree 4. Master's Degree 5. Ph.D. or higher	1. Strongly agree 2. Agree 3. Neutral 4. Disagree 5. Strongly Disagree	_____
Labels have no order or meaning	**Options have an order, but no other numeric meaning**	**Options have an order, and the *intervals* between options are equal**	**Response is a true number, with a meaningful zero value**

Quantitative Research: Survey

The scale determines which analyses are possible!

Central tendency: what is the most representative response?

- **Nominal scale:** can only use ***mode*** (e.g., most people crave pizza)
- **Ordinal scale:** ***median*** and mode (e.g., the median education level of this class is a Bachelor's Degree)
- **Interval and ratio scale:** ***mean (average)***, mode, median (e.g., the average pre-tax income for members of this class is…)

Common mistake: using mean for nominal scale

Quantitative Research: Survey

Are the questions reliable? Are they valid?

Quantitative Research: Survey

Reliability: Do people understand the question and how to respond?

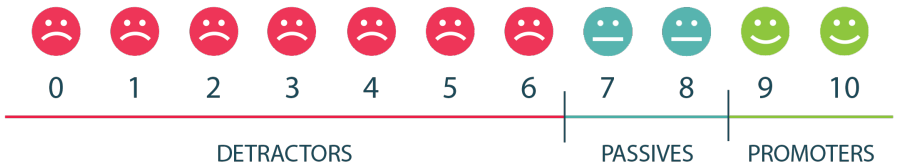
- ***Test-retest reliability:*** is the measure consistent *over time?*
  - If you were to re-measure brand love next year, would you get the same thing (assuming nothing changes in the environment)?
- ***Internal consistency:*** is the measure consistent *across related items?*
  - If you were to use 3 questions to measure brand love, do the answers move in the same direction?
- ***Inter-rater reliability:*** is it consistent across different researchers/judges?
  - If you were to ask two experts in branding whether the question measures brand love, will they agree?

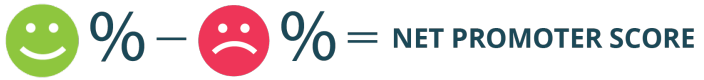
Always a good idea to pretest!

Quantitative Research: Survey

Validity: Is the question measuring something meaningful?

E.g., what is a good measure of satisfaction?





How likely are you to recommend us to a colleague or friend?

Quantitative Research: Survey

Validity: Is the question measuring something meaningful?

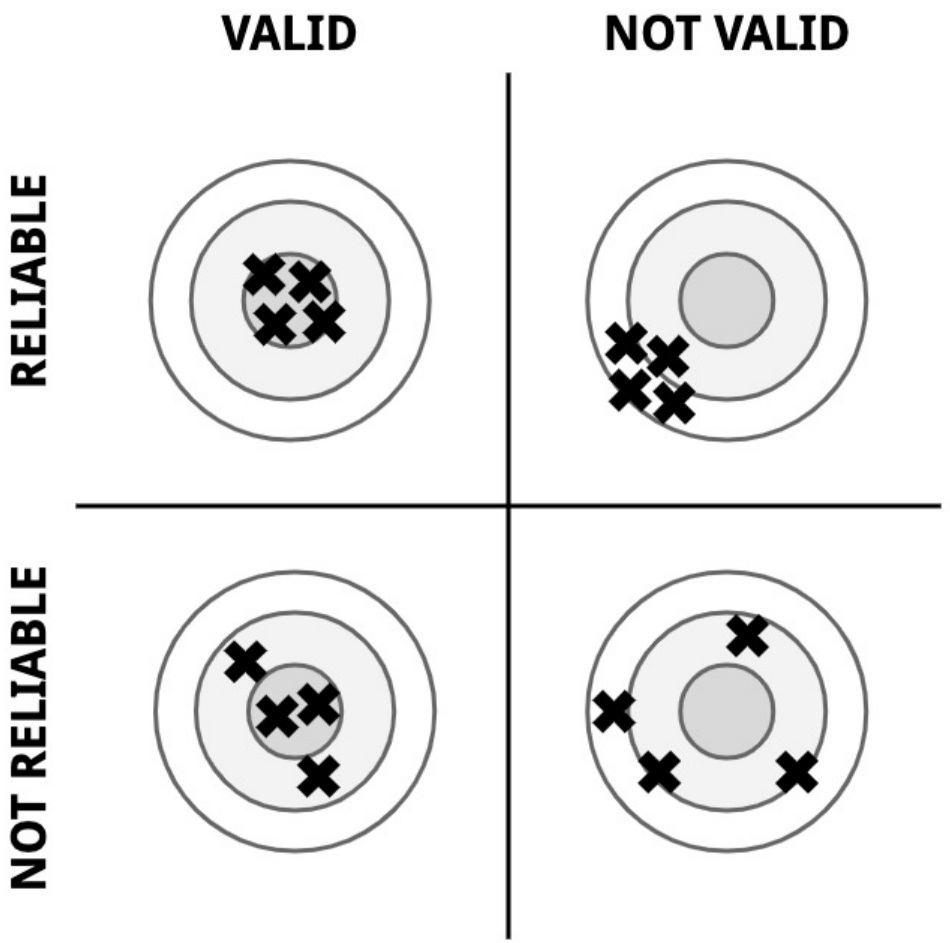
A few things to think about:

- What is the metric ***capturing***?
    - E.g., are you *really* measuring satisfaction if you ask, "Do you like our brand?"
- How does it ***compare*** with other metrics?
    - E.g., is there a better way to measure satisfaction?
- How does the metric link with **managerial outcomes**?
    - E.g., does this measure of satisfaction have implications for how you do business?

Quantitative Research: Survey

Reliability: Do people understand the question and how to respond?

Validity: Is the question measuring something meaningful?



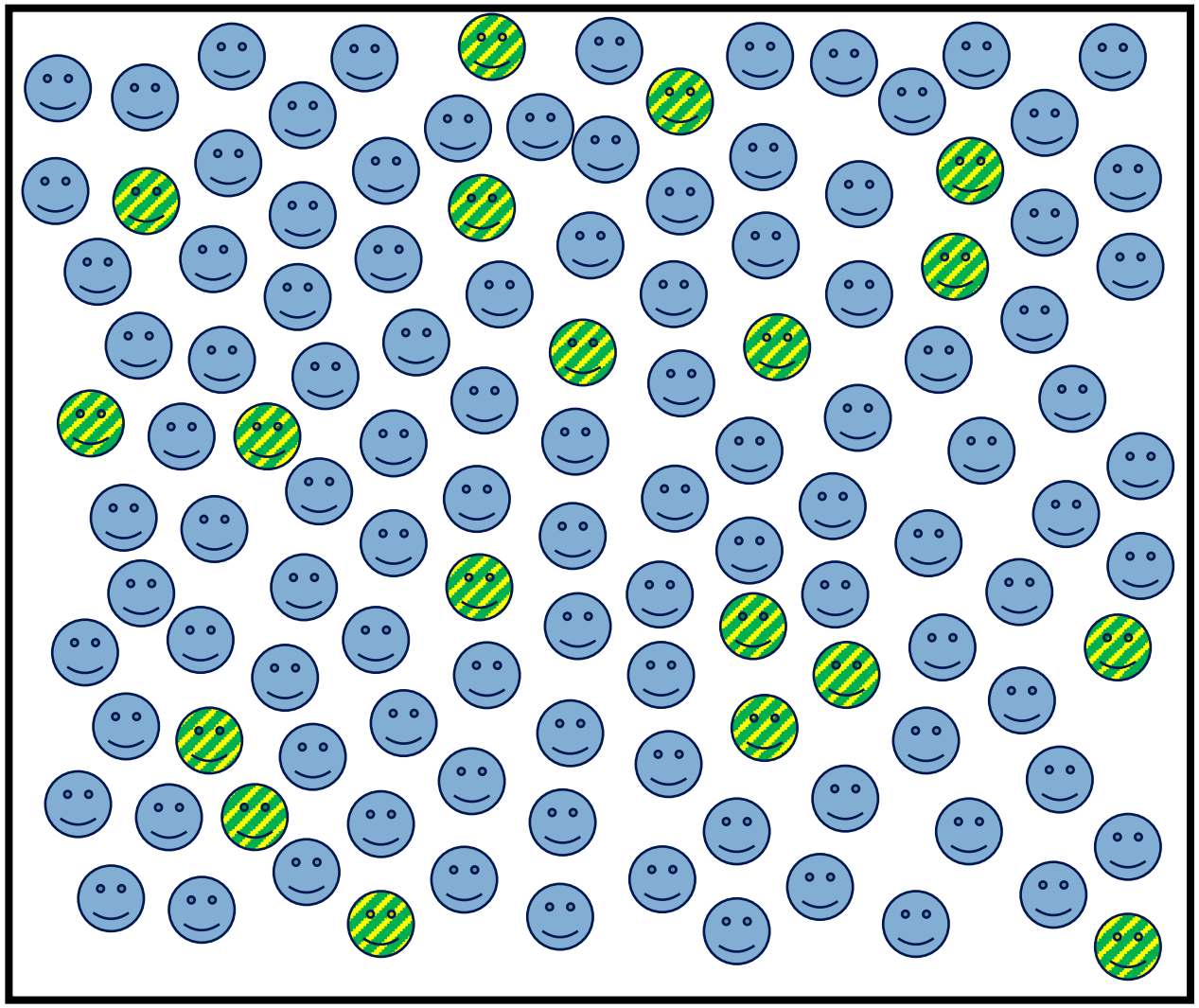
Quantitative Research: Survey

Is the sample representative of our customers?

Sampling and Representativeness

"This is interesting, 70% of the respondents to our survey said they don't respond to surveys."

Quantitative Research: Surveys



For survey results to be meaningful, the set of people who respond must **represent** your customers.

Idea: Randomly sample the population

- Every person in the population has an equal chance of being sent the survey

Types of Marketing Research

Exploratory Research

(Ambiguous Problem)

"Our sales are declining and we do not know why."

Descriptive Research

(Aware of Problem)

"What kinds of people are buying our products?"

"Who buys our competitors' products?"

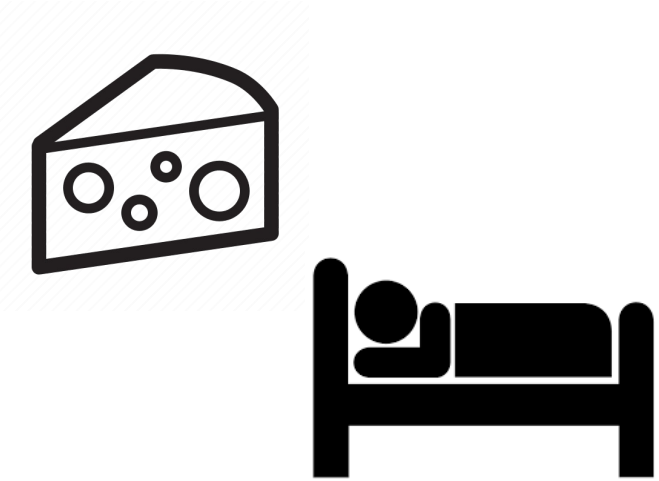
Causal Research

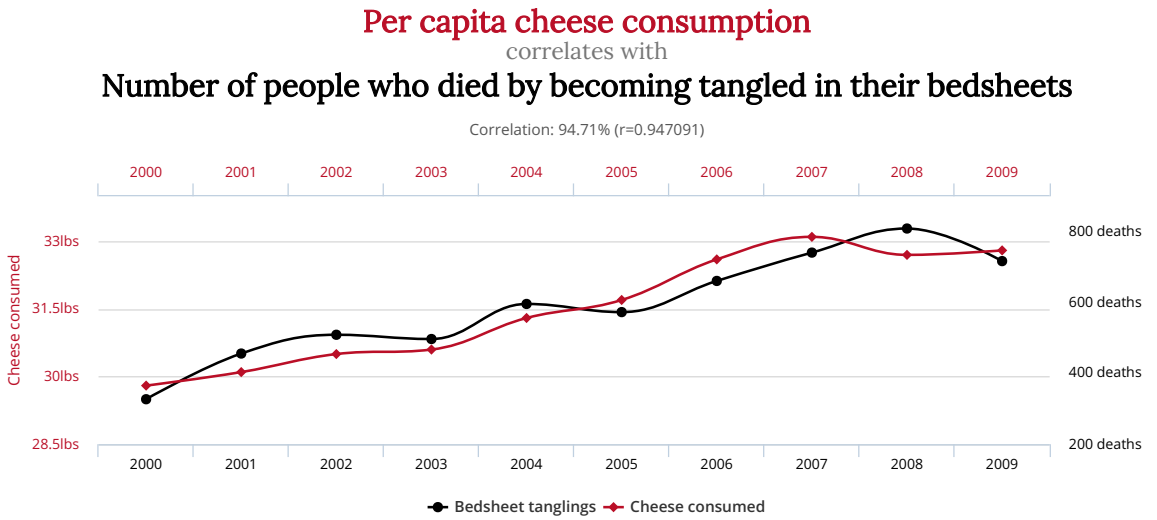
(Problem Clearly Defined)

"Will buyers purchase more of our product in a new package?"

Shortcomings of Surveys?

- Surveys are great at uncovering and describing a problem but:
  - Correlation (relation between X and Y) is not Causation (X causes Y)





  - It guides us but doesn't allow us to take a stand: e.g., will my promotion work?
- Experimentation aims to understand the counterfactual universes

Why can't surveys address those questions?

No causation without (quasi-)experimentation.

We can guarantee that an observed effect is a causal effect by running an experiment where …

No experimentation without manipulation.

… subjects are assigned to different conditions in which causal variables are systematically and differentially manipulated.

And randomization is necessary

However, more advanced statistical tools can be used when an experiment is not feasible.

Types of Marketing Research

Exploratory Research

(Ambiguous Problem)

"Our sales are declining and we do not know why."

Descriptive Research

(Aware of Problem)

"What kinds of people are buying our products?"

"Who buys our competitors' products?"

Causal Research

(Problem Clearly Defined)

"Will buyers purchase more of our product in a new package?"

How to Gather Primary Data?

Ask: *what do people say/think*

Qualitative methods → *interpret* what people say

- In-depth interviews
- Focus groups

Quantitative methods → *measure* what people think/say

- Surveys

Observe: *what do people do*

- Direct observation
- Field experiments

Causal Research

Experiments are everywhere, in practice:

- ***Field experiments***: large-scale studies in "real life"

  Facebook (secretly) manipulated users' mood by filling their news feeds with more negative vs. positive content from friends, then measured what type of content people posted (emotional contagion)

- ***A/B (split) testing:*** compare two or more versions of a variable (e.g., web page, advertisement, etc.)

  Facebook for Business has an a/b testing feature

Source: https://www.pnas.org/content/111/24/8788.full

Data Taxonomy

	Primary Data *Data that is gathered by the researcher for the purpose of answering a specific question.*	**Secondary Data** *Data that was gathered for a purpose other than answering the specific question.*
Structured *Data that can be easily and meaningfully represented and manipulated in a traditional database (spreadsheet). Typically numeric or "choice" data.*	Surveys (ratings, choice) Experiments	Transaction logs Scanner panel data Ad tracking Product usage data Geospatial tracking
Unstructured *Data that cannot be meaningfully stored in a traditional data structure (spreadsheet) without further processing. Examples include text, images, video, and voice.*	Focus groups Interviews Surveys (free response) Observation Eye tracking Physiological/neural	Online reviews Social media Most digital content Call logs

Secondary Data

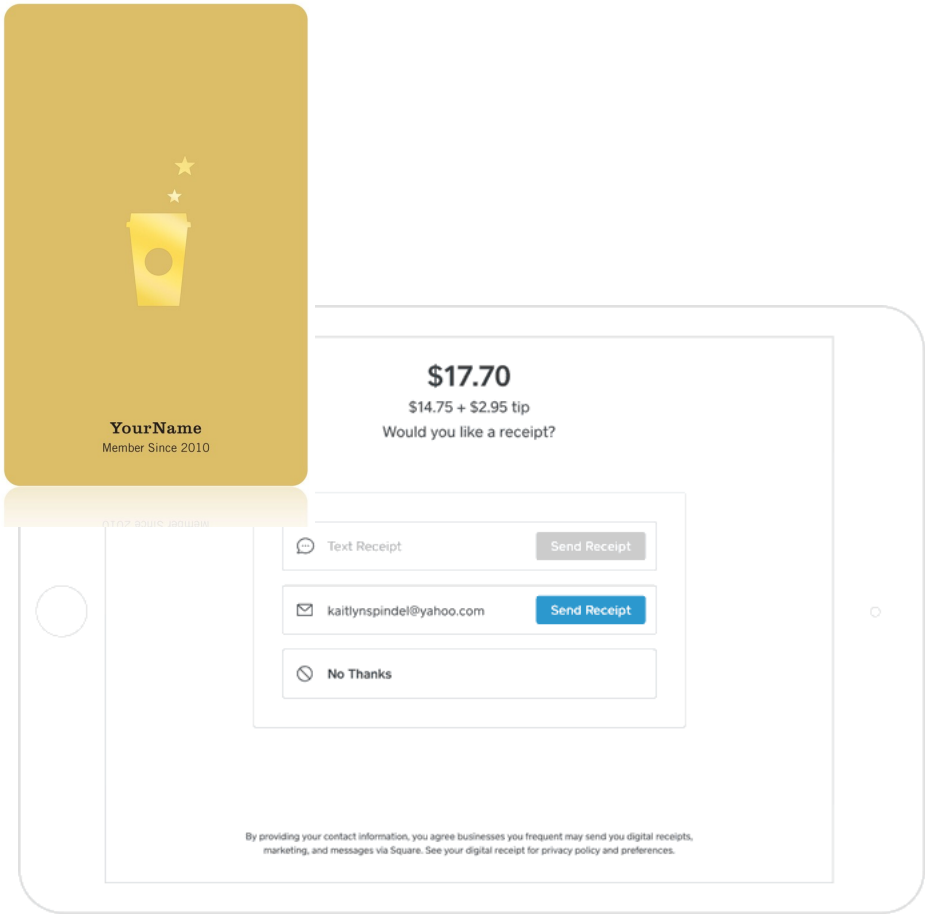
A few interesting ways to collect/utilize secondary data:

- Scanner & Point-of-Sale
- Geodemographics
- Media
- Web
- Mobile

Secondary Data

Structured

Secondary Data: Point-of-sale & Scanner Data



Secondary Data: Point-of-sale & Scanner Data

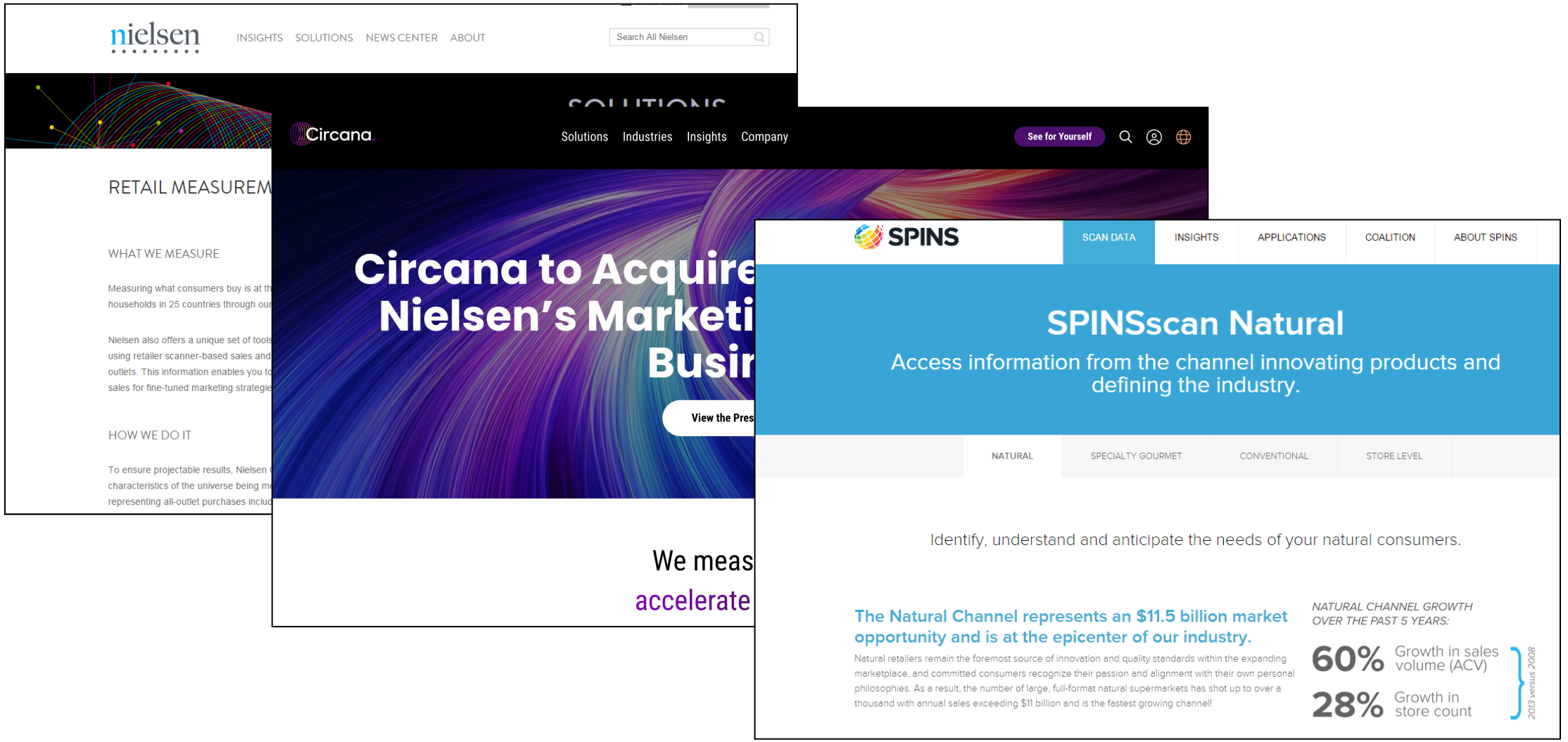
- The value of POS data:

  Geography x Product x Time x Variables (e.g., price, promo)

- Retail market: where is the data?
  - CPG manufacturers
  - major warehouse and distribution centers
  - supermarkets
  - households

Each layer has its own data, and its own questions...
which may require another layer's data!

Providers of POS & Scanner Data



Why would firms pay for this data?

The biggest factor: **Completeness**

- Linking aggregate sales to marketing instruments
- Individual-level purchase and marketing mix data
- Obtaining a richer set of performance measures beyond market share

What can you do with this data?

- **Pricing:** What is the optimal price?
- **Promotions:** Are they working? What's the impact?
- **Display:** Which type of displays (e.g., end of aisle) work better?
- **Basket composition:** Which categories are substitutes / complements?

Limitations of "Old-school" POS Data

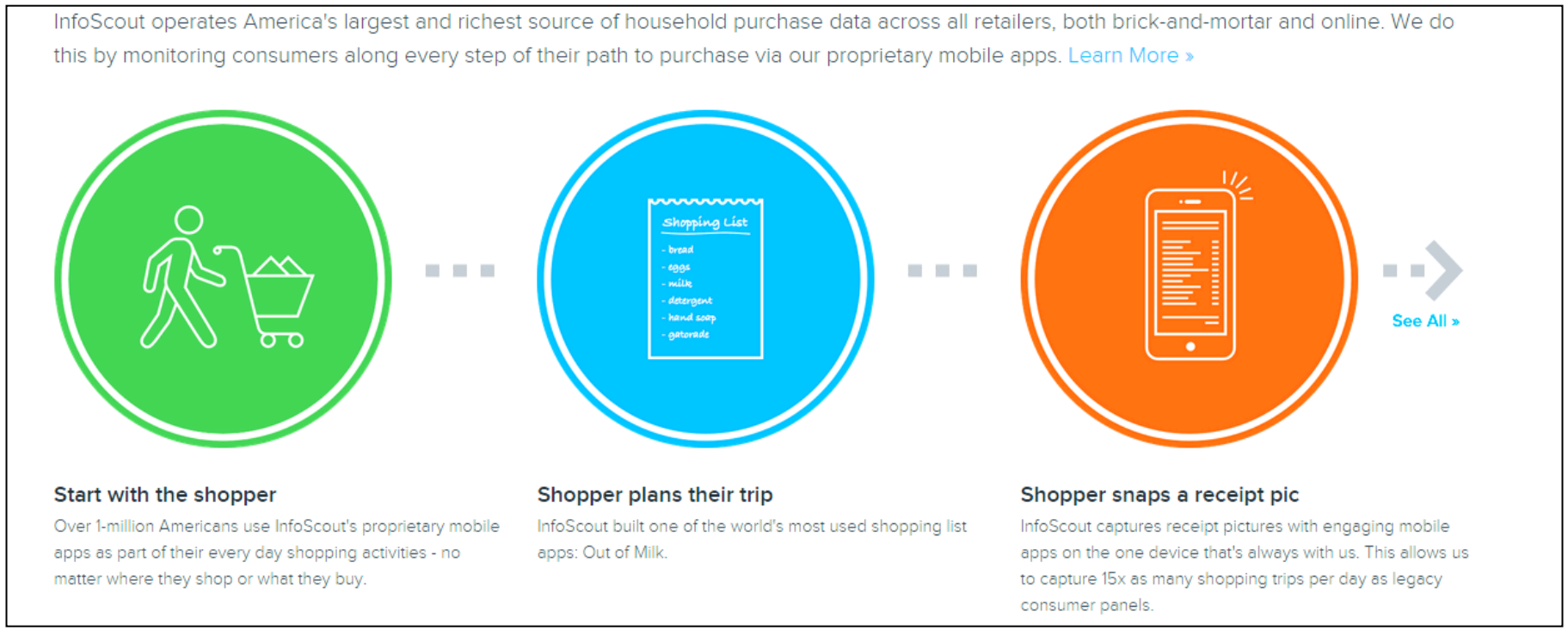
- Not totally comprehensive: some stores manage (and do not share) their POS data (e.g., Trader Joe's so Nielsen does not have access to this data)

- Not ***causal*** – if we change the price, what will happen?

- Not detailed:
  - Who are these customers? (Demographics, psychographics)
  - How many are new/existing customers?
  - What does the path-to-purchase look like?

POS Data Solutions

Solution 1: self- report scanner panel data (e.g., iNFOSCOUT)



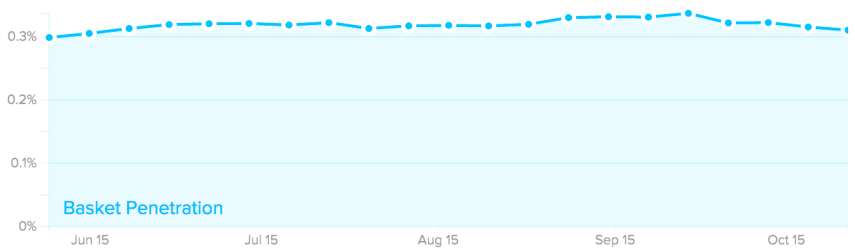
POS Data Solutions

Duracell Consumer Insights

Who buys Duracell?

Duracell Consumer Receipt Data

The sample data below is captured directly from our users' consumer shopping receipts.



Brand $ per Basket	% of Basket $ (median)	Total Basket $ (median)
$7.49	16.8%	$44.61

Users: 65+ years with $80—125k income

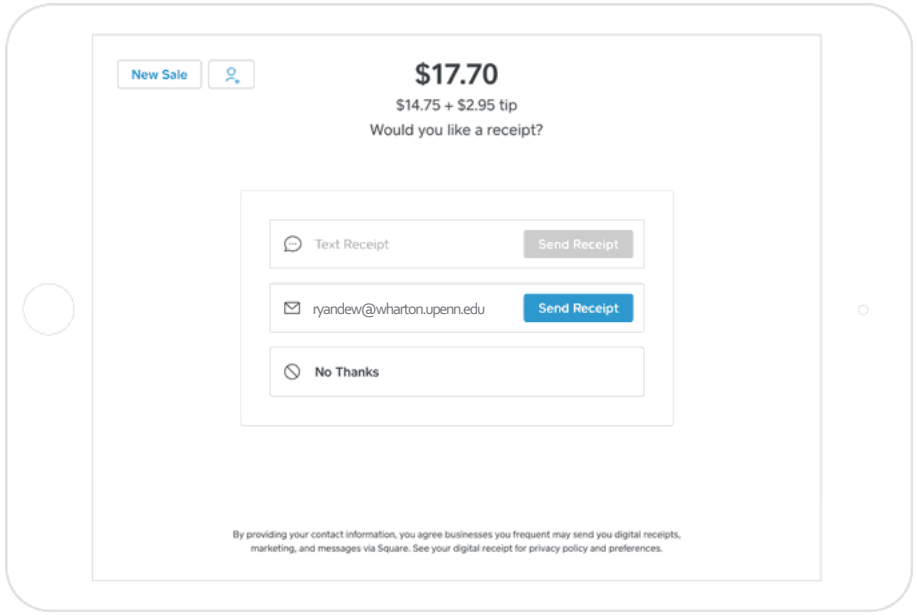
Duracell Consumer Demographics

demographic	index
Female	100
Male	102
<24	69
25-34	77
35-44	100
45-54	113
55-64	131
65+	181
African American	67
Asian	111
Caucasian	104
Hispanic	77

demographic	index
Has Kids	98
No Kids	104
- $20k	71
$20k-40k	81
$40k-60k	88
$60k-80k	98
$80k-100k	110
$100k-125k	115
$125k +	141
No College	89
College	99
Adv. Degree	119

POS Data Solutions

Solution 2: modern POS technology



Loyalty programs, email receipts

Match offline with online

Smart shelves, store beacons, apps

Track the whole in-store experience

Secondary Data: Geodemographics

Beware of demographic profiling! Similar demographics can conceal different lifestyles and interests.

Male
Caucasian
25 years old
Some college
<$80,000/year

Manhattan, NY

Manhattan, KS

To make information more insightful and actionable: combine demographics with geography (and other data, e.g., purchasing)

Secondary Data: Geodemographics

- Geography is **actionable**:
  - Market scouting (open a store!)
  - Advertising planning and targeting

- **Market sizing:** understanding potential contagion

- Correlate with other outcomes:
  - Where are my best performing stores?
  - Who are the customers there?

Secondary Data

Unstructured

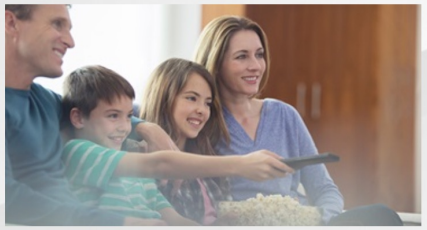
Secondary Data: Media Data

Menu KANTAR MEDIA

Global Country Login Contact Us

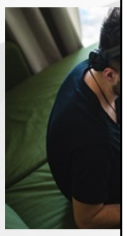
Audienc

We measure what media people actually cons
content owners a true understanding of the m
media strategies.



Television and Video

Changing TV, video and digital consumption habits make it increasingly important to understand audiences. We measure TV



Radio Anc

We have me
radio audien
different pla

RENTRAK PRECISELY MEASURING MOVIES & TV EVERYWHERE

HOME | PRESS | CONTACT US | CLIENT LOGIN Search

MOVIES & TV EVERYWHERE | OUR SERVICES | WHO ARE YOU? | ABOUT RENTRAK | RENTRAK CAREERS | INVESTOR RELATIONS

TV Essentials

Providing television measurement from more than 35 million screens and approximately 16 million households across the country, Rentra
the trusted source for networks, agencies and
advertisers, delivering the most precise and
reliable TV ratings all day, every day across th
largest media landscape.

TV ESSENTIALS OVERVIEW

Rentrak's TV Essentials® helps television buyers and sellers make smarter transactions by giving them a deeper understanding of the true value of their television viewing audience. Providing unparalleled TV measurement, this unique service gives agencies, advertisers and television networks access to massive amounts of exclusive viewership information for a level of granularity and stability absent from traditional television measurement services.

Advanced Demographics
Rentrak's Advanced Demographics allow users to more accurately pinpoint audiences they want to reach. By combining Rentrak's massive viewing intelligence with third-party consumer behavior information from sources like IHS, Experian, Simmons, and Epsilon, Rentrak users can dig deeper into TV

Exact Commercial Ratings
A game-changer for the nationa
television industry, the Exact Co
Ratings® data metric allows ad a
and advertisers to determine ho
specific national TV commercial
performs within an ad pod, adva
the national TV industry standar
measuring an average rating for
commercials within a telecast (c
providing individual commercia
When used along with the TV E
service, Exact Commercial Ratir
network advertisers exactly how
viewers were exposed to their s
commercials in a campaign, allo
them to maximize the results of
television spend.

EXACT COMMERCIAL RATINGS® OV

nielsen INSIGHTS SOLUTIONS NEWS CENTER ABOUT Search All Nielsen

SOLUTIONS

TELEVISION

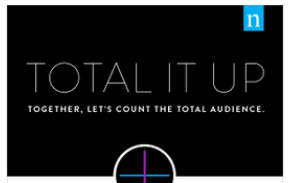
Television and the way we watch it have come a long way since Nielsen began measuring TV audiences in 1950. Today, the ability to watch our favorite shows at any time and on multiple devices amplifies the need for exceptionally adept and flexible audience measurement capabilities.

And there's a big reason why Nielsen is synonymous with television audience measurement. We invented it. Since day one, we've offered the media industry the expertise it needs to make the best marketing decisions possible. Today, our expansive and representative television measurement services capture video viewing across different screens: television and computers.

WHAT WE MEASURE

Consumers are changing with the times, and the same goes for us. As technology continues to evolve and media companies try new ways to attract viewers, understanding what consumers are watching—and what they're watching on—is more important than ever. Today, viewing video is a personal and online experience. This media fragmentation is both a challenge and an opportunity.

Our capabilities provide relevant metrics that are necessary to inform successful marketing and programming



RELATED SOLUTIONS

Advertising Effectiveness »

Changing Landscape: Streaming Media Planning

Secondary Data: Web Data

Psychological Language on Twitter Predicts County-Level Heart Disease Mortality

Johannes C. Eichstaedt[1], Hansen Andrew Schwartz[1,2], Margaret L. Kern[1,3], Gregory Park[1], Darwin R. Labarthe[4], Raina M. Merchant[5], Sneha Jha[2], Megha Agrawal[2], Lukasz A. Dziurzynski[1], Maarten Sap[1], Christopher Weeg[1], Emily E. Larson[1], Lyle H. Ungar[1,2], and Martin E. P. Seligman[1]

[1]Department of Psychology, University of Pennsylvania; [2]Department of Computer and Information Science, University of Pennsylvania; [3]Graduate School of Education, University of Melbourne; [4]School of Medicine, Northwestern University; and [5]Department of Emergency Medicine, University of Pennsylvania

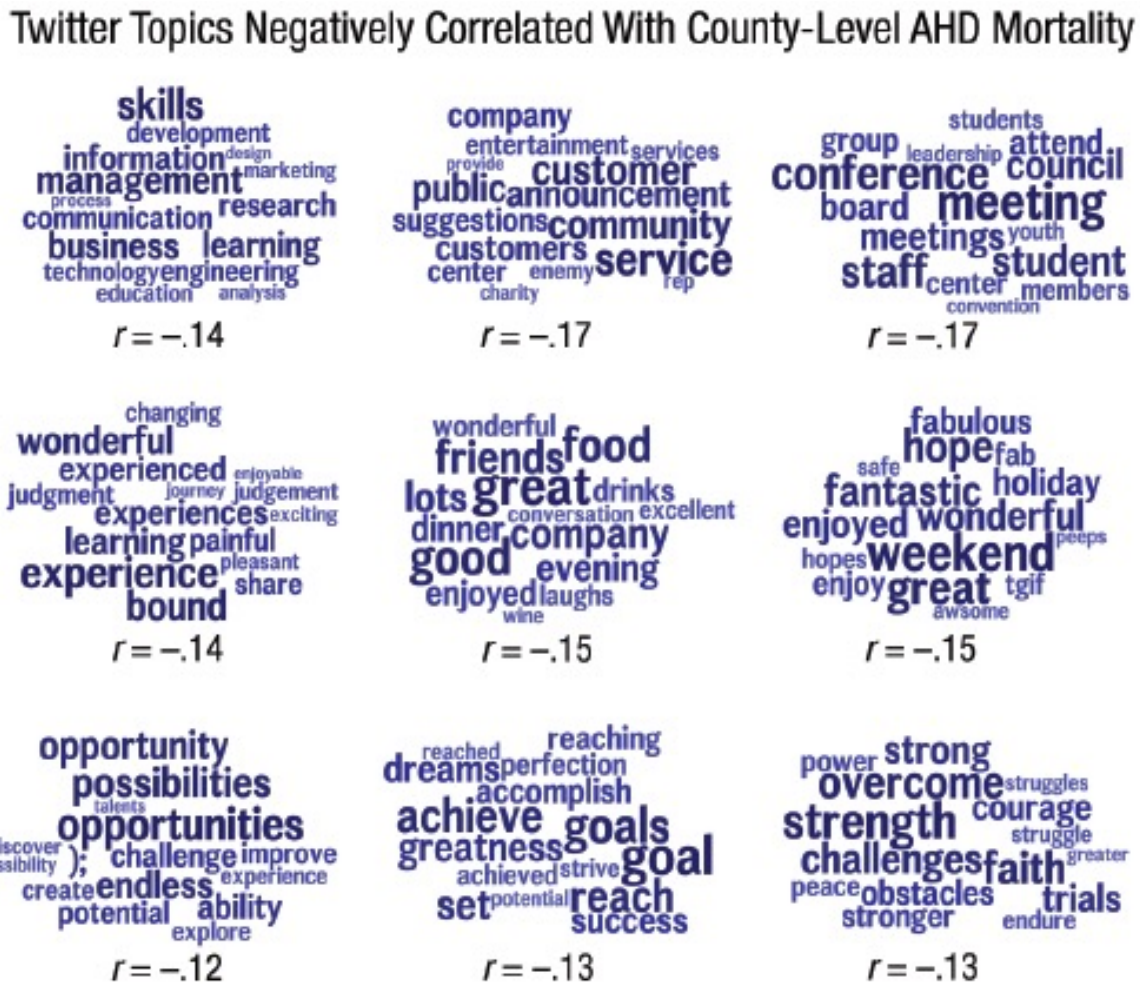
Secondary Data: Web Behavior Has Meaning

Hostility and chronic stress are known risk factors for heart disease, but ***costly to assess***.

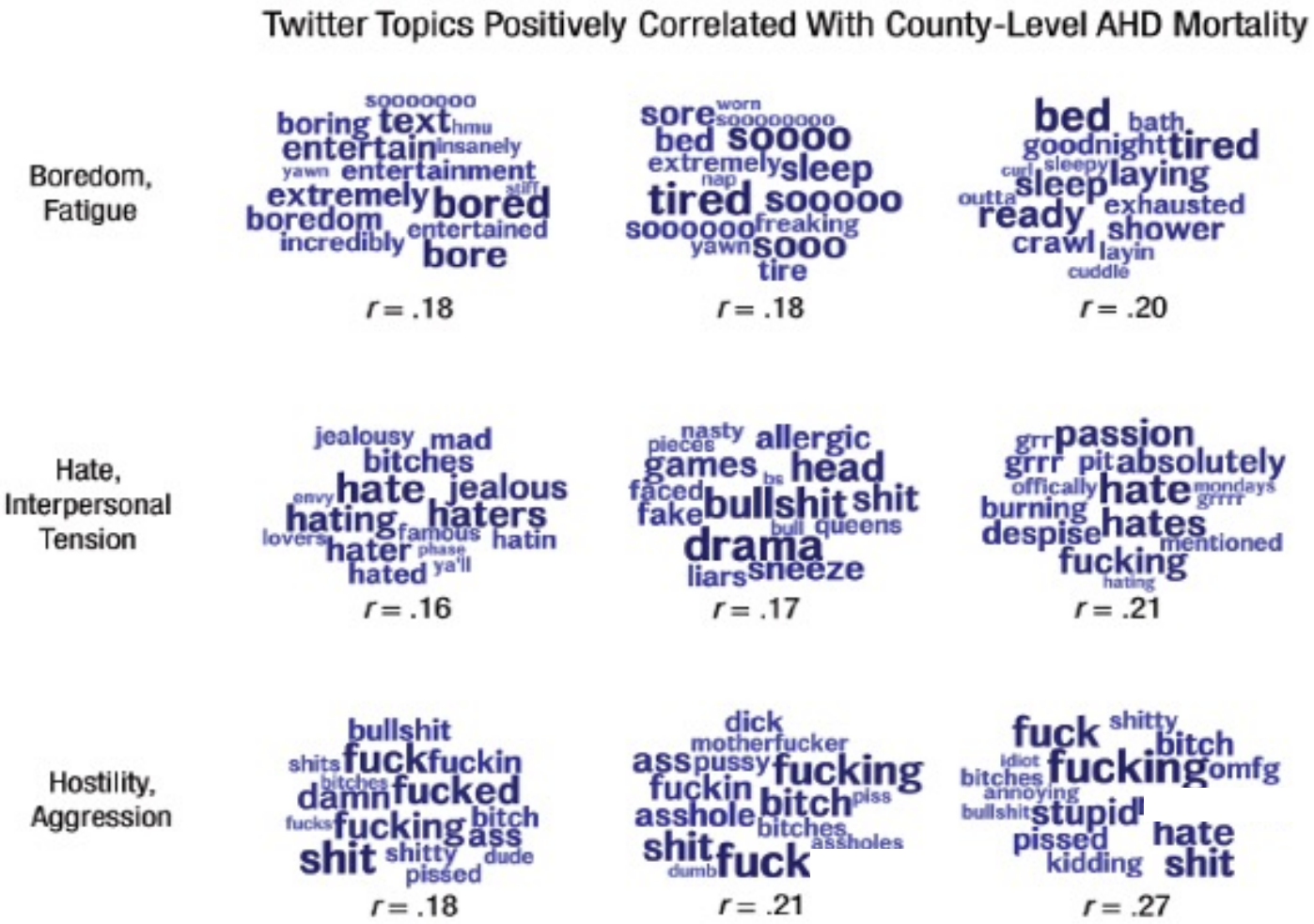
Can we capture ***community psychological characteristics*** through social media?

Specifically, ***can language patterns on Twitter help us predict mortality from heart disease***?

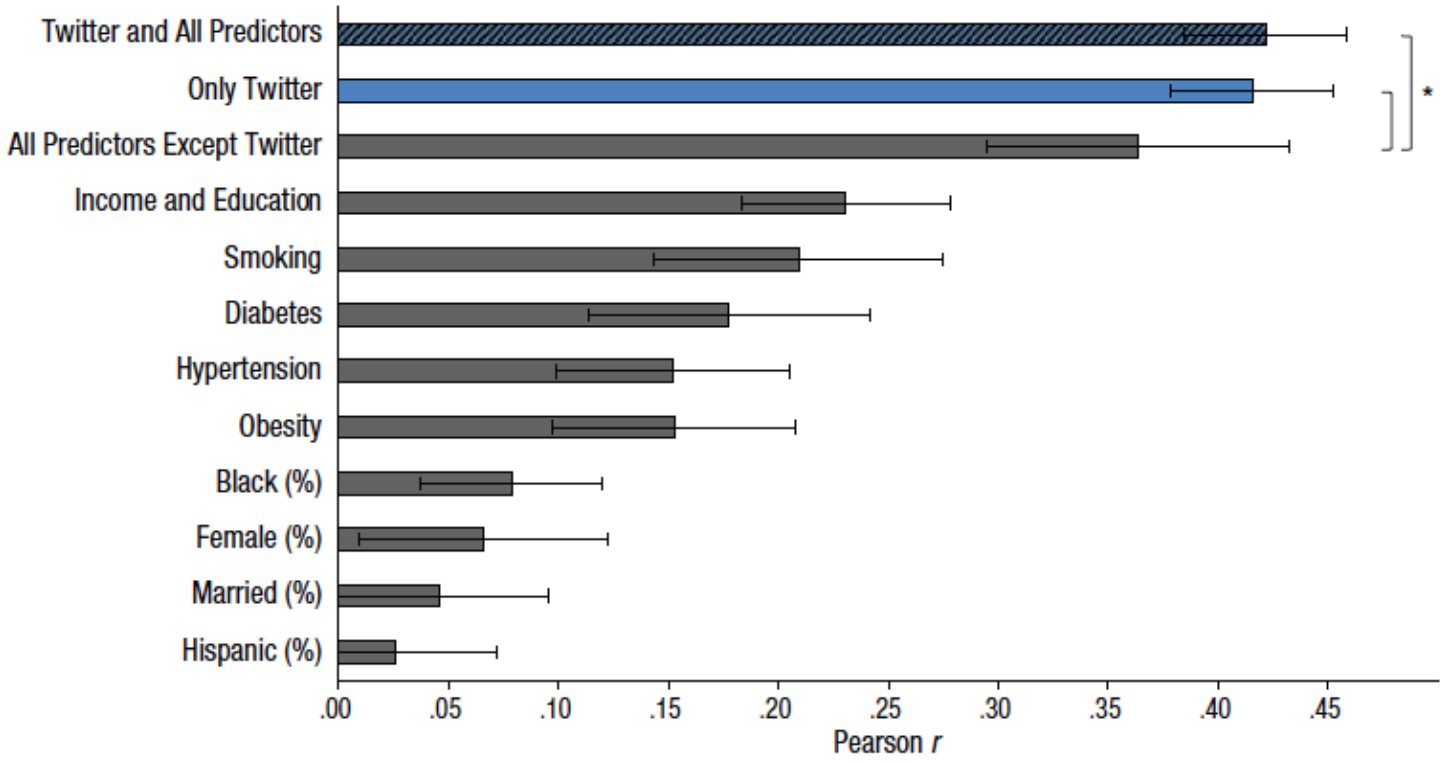
Secondary Data: Web Behavior Has Meaning



Secondary Data: Web Behavior Has Meaning



Secondary Data: Web Behavior Has Meaning



Takeaway: social media chatter and sentiment can be powerful predictors of real-life outcomes

Secondary Data: Web Behavior Has Meaning

Social Media Data Uses

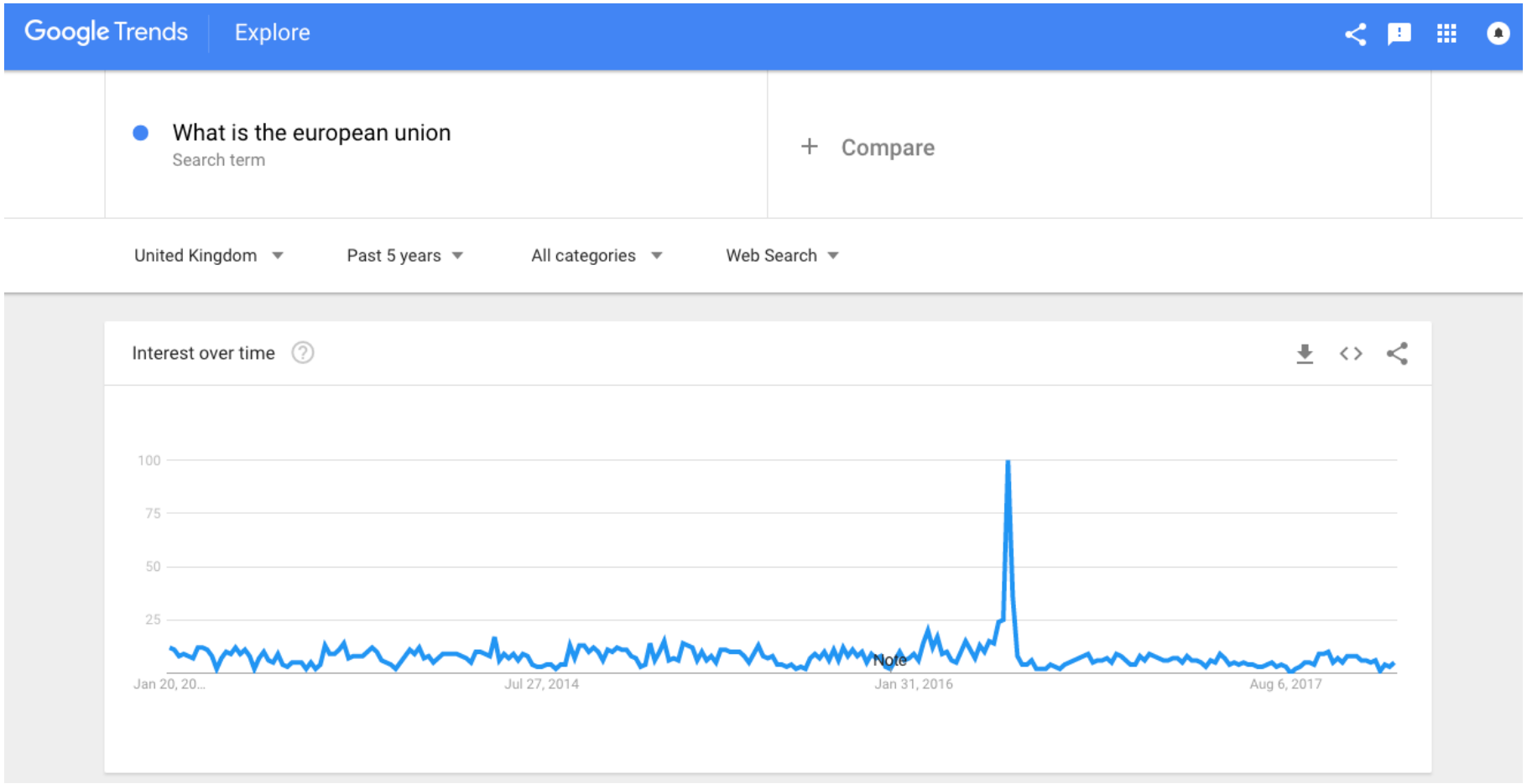
- **Audience engagement** for a campaign
- **Brand mentions** vs. competitors: tracking "buzz"
- **Sentiment analysis:** do people like or hate us?
- **Segmentation:** understanding consumers by their social media
- **Brand perception:** when do people tag us?

Secondary Data: Web Data

Web Search vs. Social Media:

What can search data reveal that social media data might not?

Secondary Data: Web Data



Secondary Data: Web Data

Web Search and Traffic Data Uses

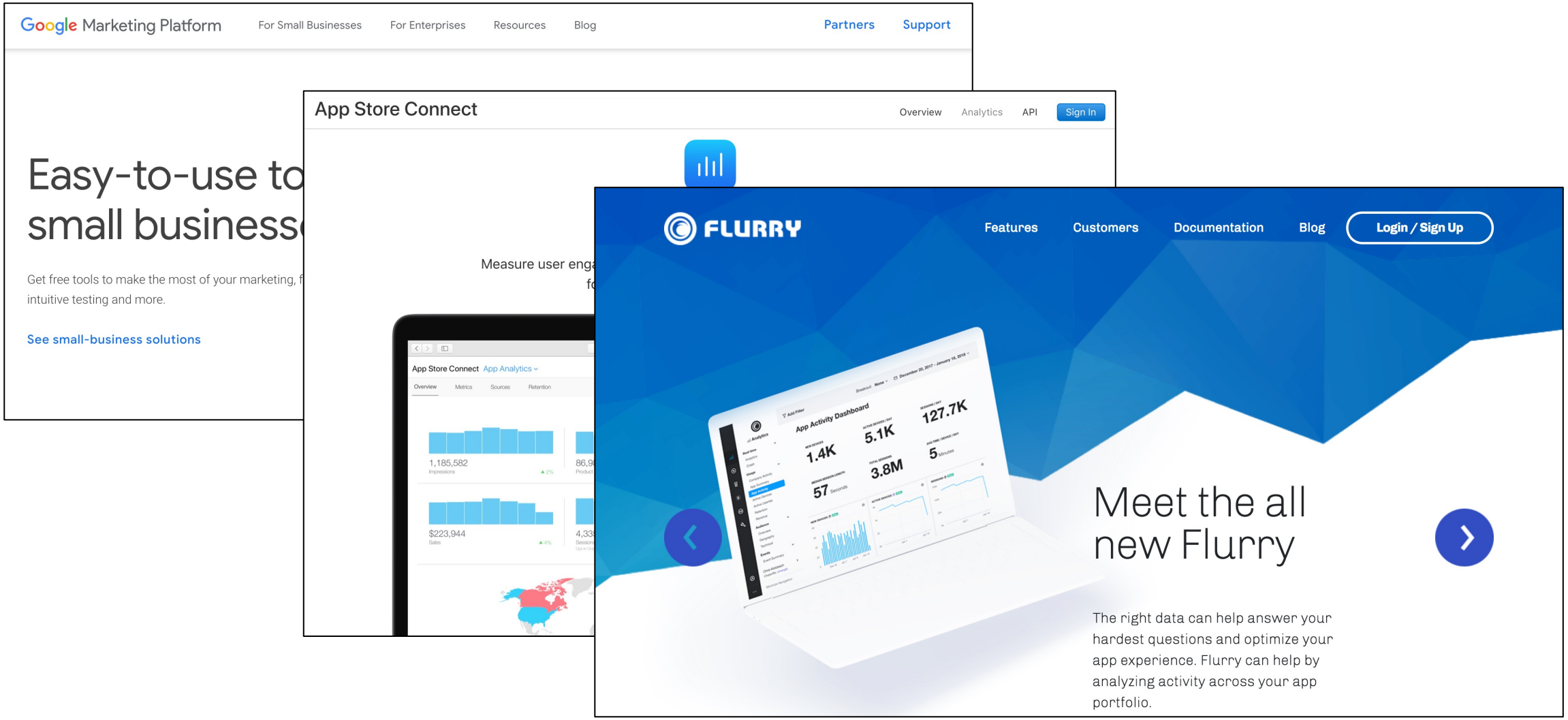
Exploratory research:

- The "ebbs and flows" of traffic
- Where are people coming from?

Measurement through the funnel:

- Who is noticing us?
- How do people become aware?
- Where do they go next?
- What else are they considering?

Secondary Data: Mobile Data



Types of Marketing Research

Exploratory Research

(Ambiguous Problem)

"Our sales are declining and we do not know why."

Descriptive Research

(Aware of Problem)

"What kinds of people are buying our products?"

"Who buys our competitors' products?"

Causal Research

(Problem Clearly Defined)

"Will buyers purchase more of our product in a new package?"

Data Takeaway

	Primary Data *Data that is gathered by the researcher for the purpose of answering a specific question.*	**Secondary Data** *Data that was gathered for a purpose other than answering the specific question.*
Structured *Data that can be easily and meaningfully represented and manipulated in a traditional database (spreadsheet). Typically numeric or "choice" data.*	Surveys (ratings, choice) Experiments	Transaction logs Scanner panel data Ad tracking Product usage data
Unstructured *Data that cannot be meaningfully stored in a traditional data structure (spreadsheet) without further processing. Examples include text, images, video, and voice.*	Focus groups Interviews Surveys (free response) Observation Eye tracking Physiological/neural	Online reviews Social media Most digital content Call logs

Marketing Research Takeaway

Exploratory Research

(Ambiguous Problem)

"Our sales are declining and we do not know why."

Descriptive Research

(Aware of Problem)

"What kinds of people are buying our products?"

"Who buys our competitors' products?"

Causal Research

(Problem Clearly Defined)

"Will buyers purchase more of our product in a new package?"

Marketing Datasets: Takeaways

This was a whirlwind tour of marketing datasets!

- The amount of data available today is unprecedented and has created many opportunities
- We only scratched the surface of marketing dataset available, and barely discussed analysis

In the remainder of this course:

- We will keep encountering some of those datasets but will focus on the analytical tools
- We will pay close attention to every type of marketing research question (exploratory, descriptive, causal)

Wrap-Up

- Next class: Market Segmentation + Targeting
  - Cluster Analysis
  - Read Python for Marketing Research and Analytics (chapter 10)
    - Textbook can be found online

- (Post-class) Slides uploaded to Canvas under Week 1
- Recording in Echo360
- Groups will be formed over the weekend and we'll post Group Assignment 1