



Digital Technologies and Value Creation

Dr. Philippe Blaettchen
Bayes Business School (formerly Cass)

Overview – subject to change

Overarching theme	Week	
Introduction	1	Introduction to analytics applications and coding basics
Gathering data	2	Scraping web data
Gathering data / descriptive analytics	3	Data pre-processing and descriptive analytics
Gathering data / descriptive analytics	4	APIs, social media engagement, and other marketing metrics
Descriptive analytics	5	Descriptives in people analytics
NO LECTURE	6	NO LECTURE
Predictive analytics	7	Retaining employees and customers
Predictive analytics	8	Valuing a (social media) customer base
Predictive analytics	9	Segmenting customers and positioning products
Prescriptive analytics	10	Optimizing products and organizations
Prescriptive analytics	11	A/B-testing in practice

Learning objectives of today

Goals: Understand Application Programming Interfaces (APIs) and the power of social media data

- How can we access APIs, especially to get social media data?
- How can we get important insights from social media data?
- How can we use social media data in marketing analytics?

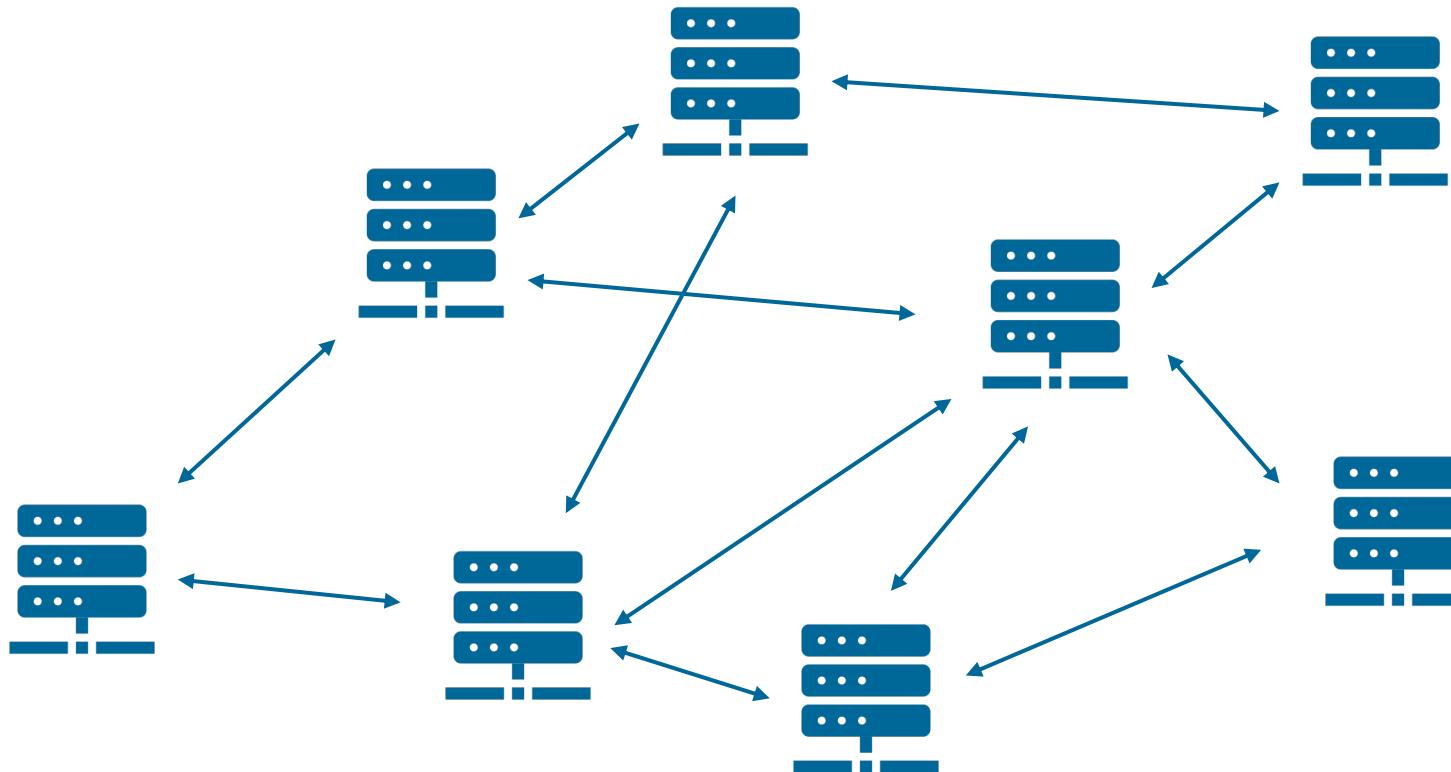
How will we do this?

- We will start by highlighting the connections and differences between scraping and APIs
- We will then gather social media data through an API and use it for analysis
- Finally, we will apply our learnings to marketing to measure customer engagement



From web scraping to APIs

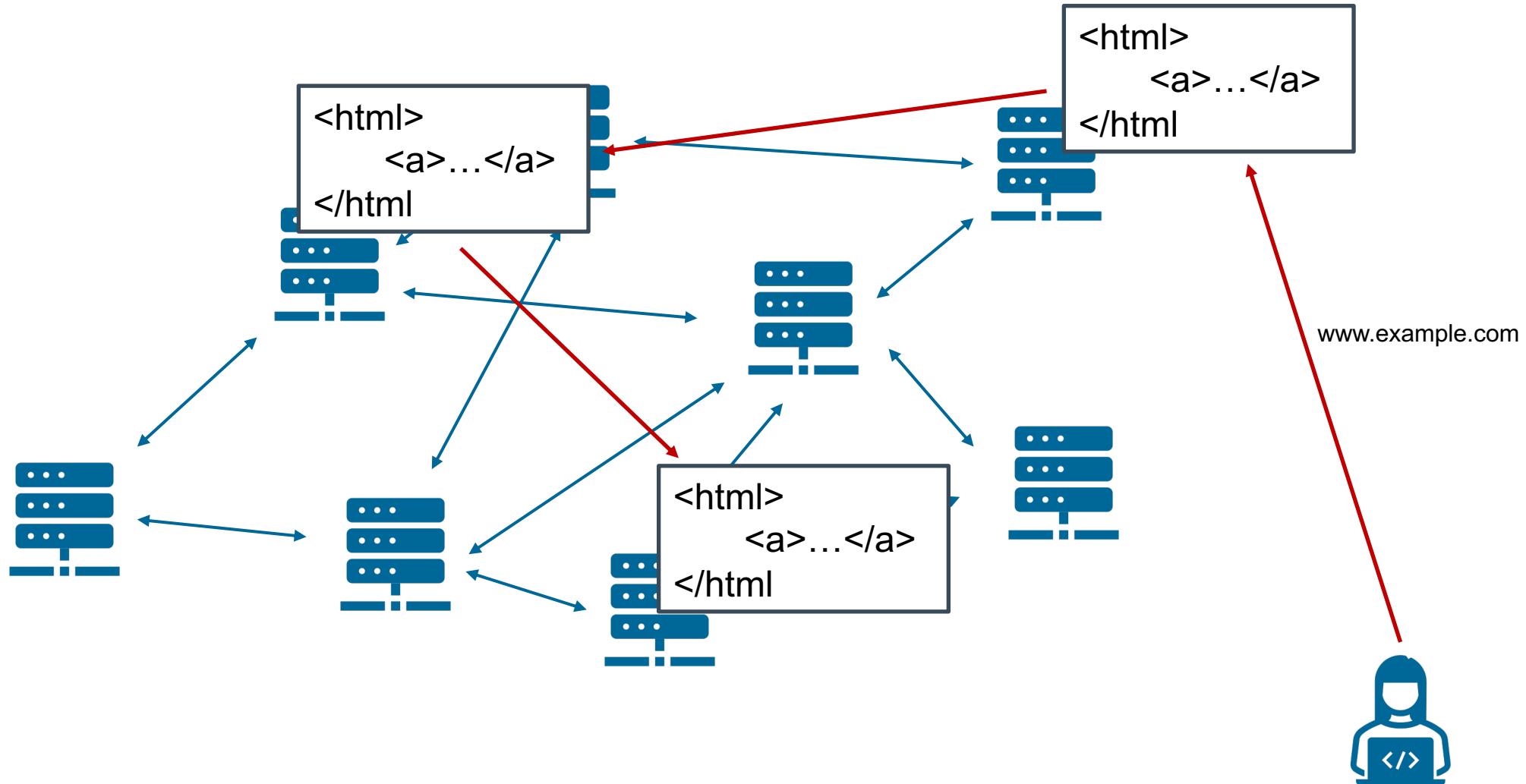
The internet



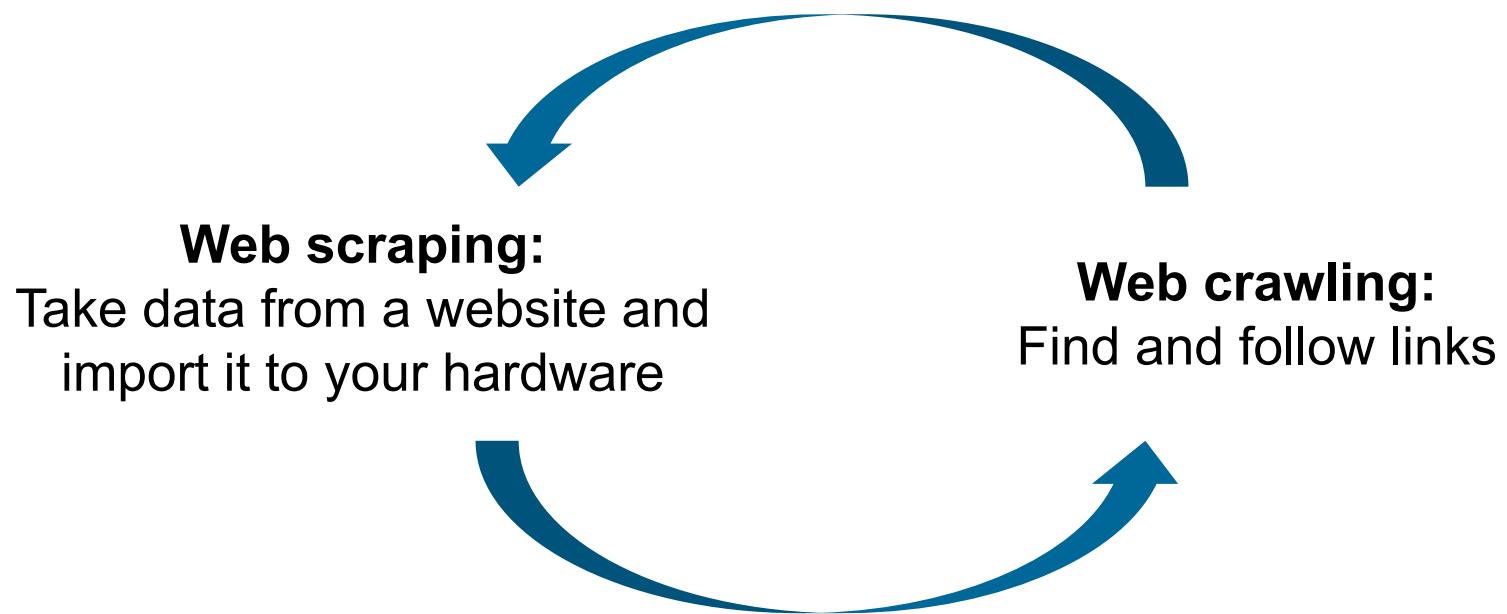
Computers need to speak a common language to talk to each other:

- Encodings: ASCII, UTF-8
- Sending data: XML, JSON

Building on top of the internet: the (world wide) web



Scraping and crawling



The key scraping (and crawling) technologies in Python

Static website
content

BeautifulSoup



Dynamic
website content
(JavaScript)



One-off / Speed
doesn't matter

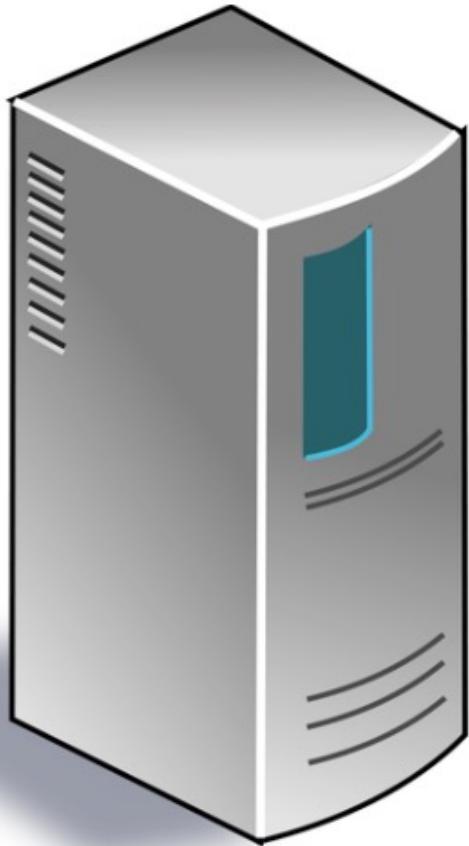
Big projects /
Speed matters

However, web sites are designed for human users!

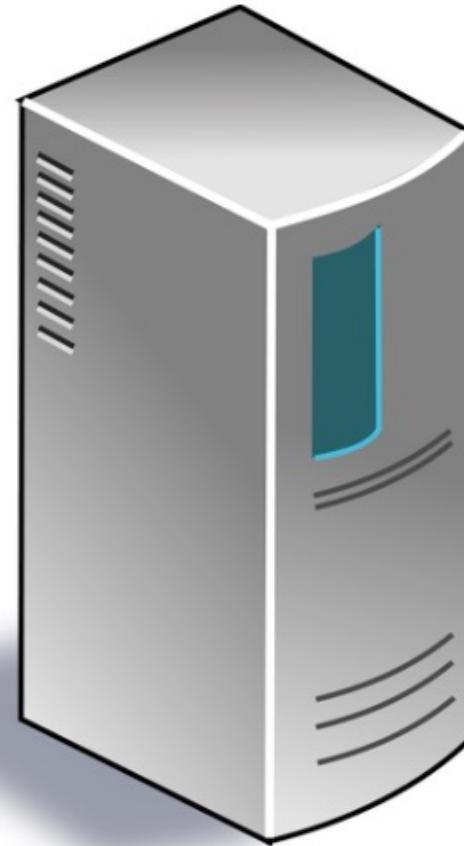
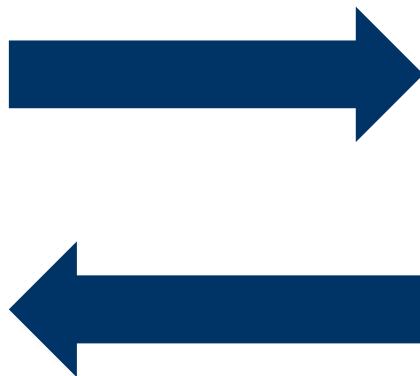
We are essentially appropriating a **user interface** to use with a program
instead of a human user

But: frequently, there are interfaces specifically made for programs →
Application Programming Interfaces

Application Programming Interface (API)



E.g., brokerage server



E.g., stock exchange

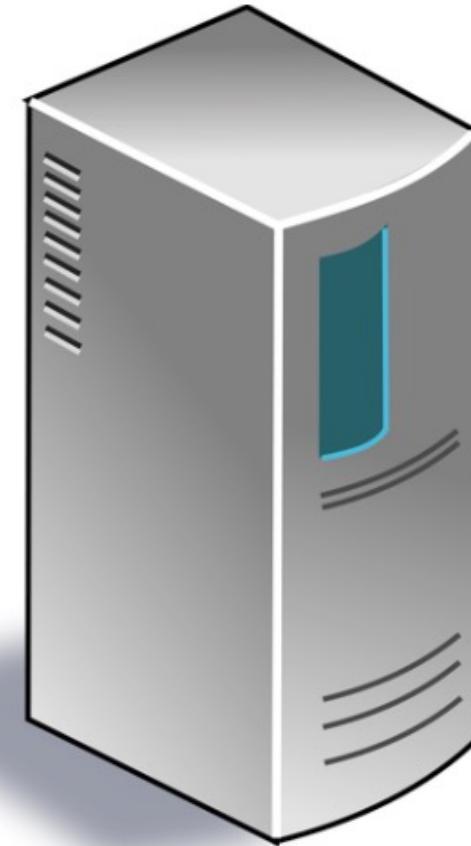
Using APIs to get data



Request for specific data, action



Response nicely formatted
(often, JSON, XML,...)



Using APIs versus Webscraping

API

Webscraping

Using APIs versus Webscraping

API	Webscraping
May not always exist	Usually, data and website come hand in hand
Made to be easily navigable	Have to replicate a human user
Designed for calling up data, including access controls	Legal risks and grey zones
Access controls can be limiting content and speed	Sometimes, implicit access controls
Access can be costly	There is a reason that the API access is costly 😊

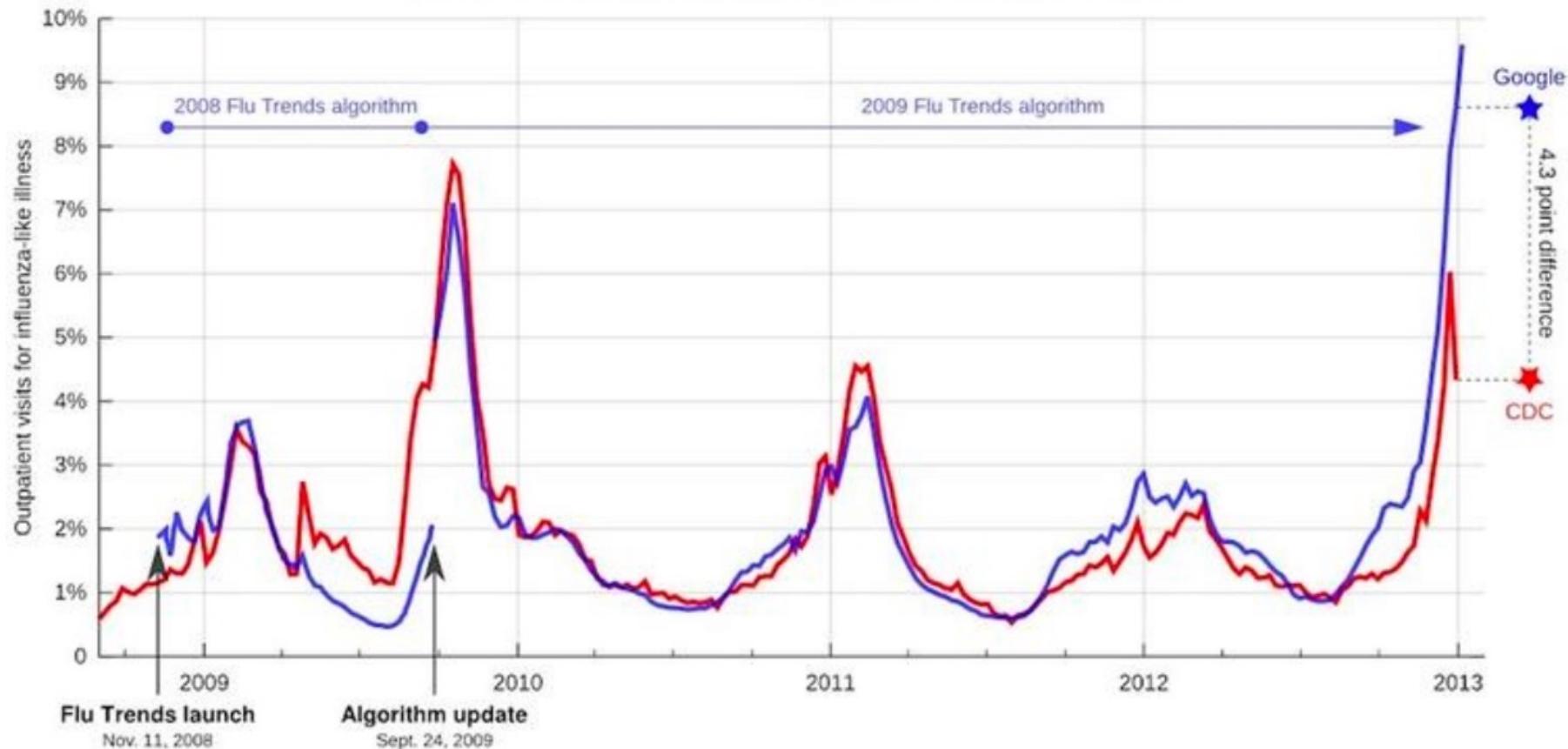
→ If the data is available through an API, it's generally the preferred option



Google Trends and Covid

Google Flu Trends

Google Flu Trends U.S. may have diverged again from the CDC data it predicts, but too early to be sure.



Sources: <http://www.google.org/flutrends/us>, CDC iLine data from <http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>, Cook et al. (2011) Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic. PLoS ONE 6(8): e23610. doi:10.1371/journal.pone.0023610.

Data as of Jan. 12, 2013. Keith Winstein (keithw@mit.edu)

Google Trends

- A continuously evolving dataset
- Trends data is an unbiased sample of Google search data. It's anonymized, categorized, and aggregated
- Index for **interest over time** is normalized and can only inform on **relative** changes

Let's access the Google Trends API through Python

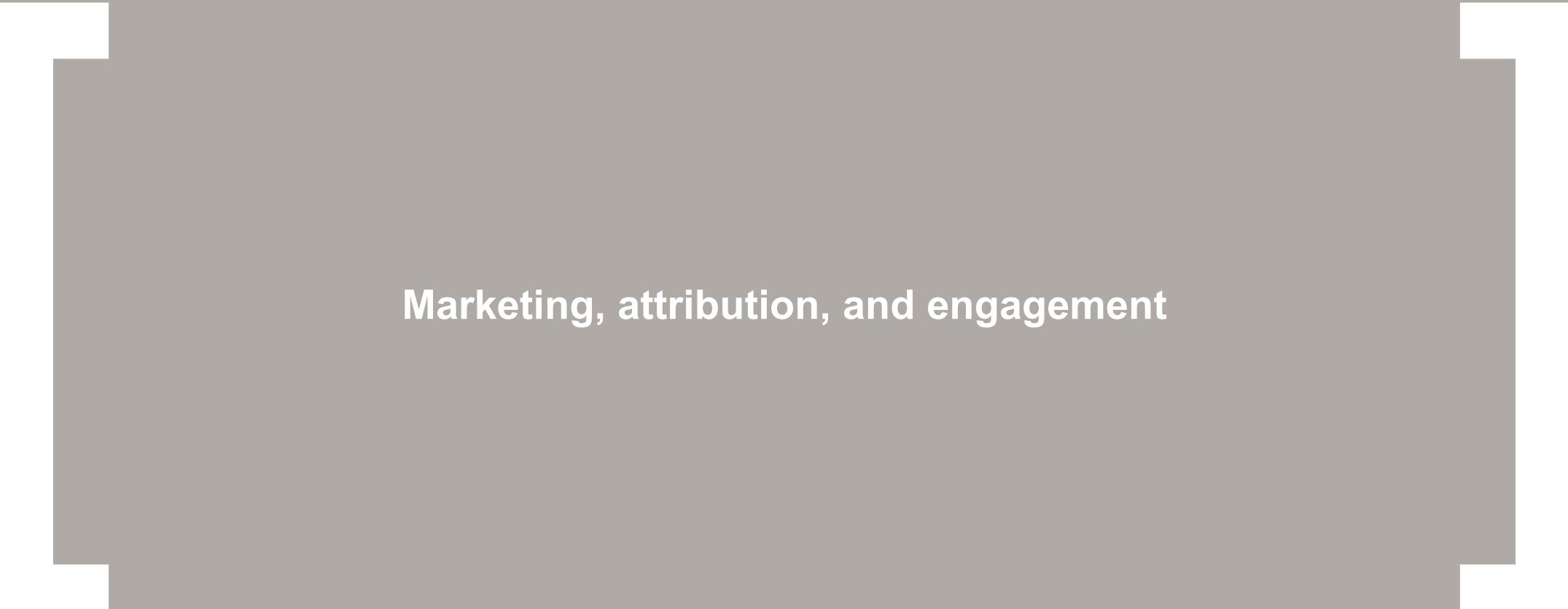
Google Covid Trends



Debrief

Follow up:

- Using Google search to detect new COVID outbreaks
- Hal Varian (Chief Economist at Google) predicting economic activities (carsales) in the US using Google search trends
- Exploring COVID symptoms with Google searches
- Predicting the short squeeze of the GameStop stock in early 2021



Marketing, attribution, and engagement

A small can and a big stage



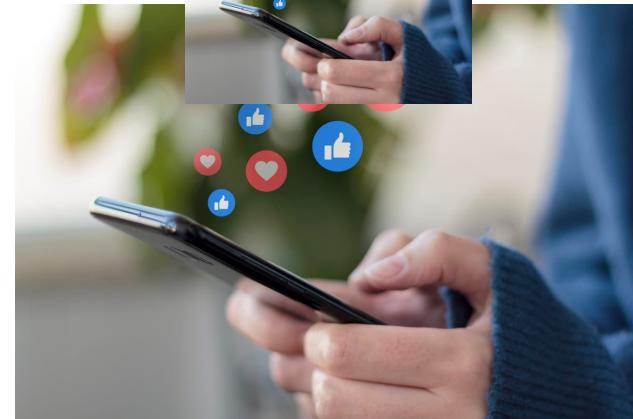
The problem of delayed effects



But we might know this (at least on the long term)



A more reactive measure



BAYES
BUSINESS SCHOOL
CITY UNIVERSITY OF LONDON

Let's ask Google trends

- Question 1: Does Red Bull create engagement with its Formula 1 expenditures?
- Question 2. Does the Pilot matter? I.e., is engagement related to success (or failure)?

Let's see whether customer engagement is driven by Formula 1 marketing



The Twitter API

A powerful tool, but with restrictions

- We can get pretty much any data from Twitter with relatively simple queries
 - users, tweets, followers, etc.
- However, access is controlled
 - authentication is needed → we need to agree to a bunch of conditions, especially regarding non-commercial use
 - we need to sign in to access the API (using tokens)
- There are many limitations on the process
 - E.g., if searching by hashtag, we can only find tweets in the last 9 days
 - We have a certain maximum number of requests per 15-minute window
 - Each request can give back only a limited number of items

Let's try it out





Until next week!