UVA/Hum User
Engagement Project

Presenters: Niall Little,
Will Fortin Ph.D.,
Dylan DiGioia MBA

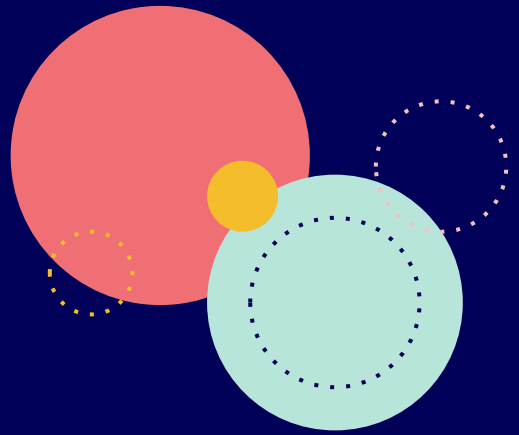
Engagement Time Machine –

Understand the past, predict the future

hum

Hum helps publishers understand their content and audience

Our goal is to move organizations to more efficient communication with users, i.e. reduce useless ads and emails



What is Hum?

Customer Data Platform (CDP)

A CDP:

- **Captures first-party data** from a publisher's separate platforms and systems, **normalizes** it, and **unifies** it into a **persistent, individual-level profile**.
- **Makes that unified data available to other platforms and systems** whenever and wherever needed to transform reader relationships and drive growth.
- **Activates the data across touchpoints and publisher systems** in support of a wide range of growth-focused initiatives, including:
  - Author and reviewer recruitment
  - Enhanced reader engagement and experience
  - Targeted advertising
  - Analytics and modeling
  - Content intelligence and digital product development

Hum Brings Your Audience and Content Data Together

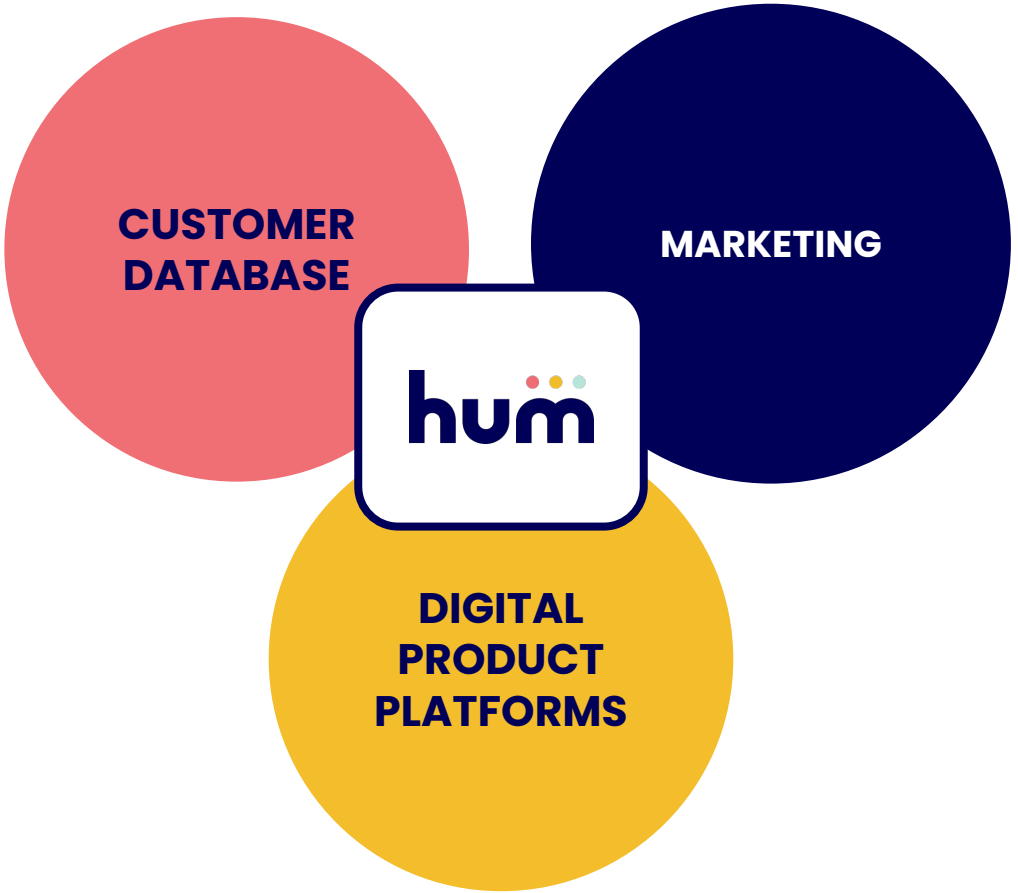
Streams data in realtime

Hum interprets “events”

Updates profiles: person, content, topic, organization

Enable insights and exploration

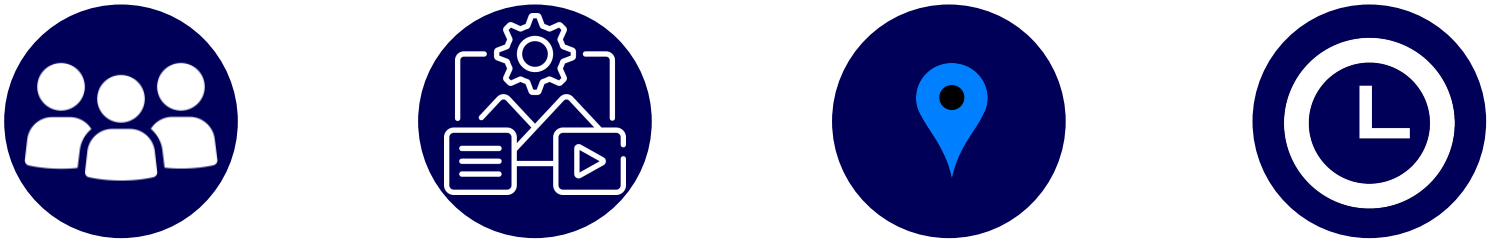
Supercharges other systems



Hum Listens for Events

And interprets those events according to infinitely configurable rules that are core + client-specific

PERSON does **THING** in **PLACE** at **TIME**



What are we doing *here*?

Use historical data to **build models for recommendations** to our clients (publishers and associations)

- Recognize and encourage high value events
  - What 5-10 things happen before a high value event?
  - Build a tool to automatically identify and encourage users along a journey toward the high value path
- User retention
  - Who is in danger of dropping out?
  - Build a tool to identify and provide value again to that user

What are our tools?

- Client provided data
  - Usually limited, some metadata
- Hum CDP and events
  - First-party profiles
  - User “events”
  - Unemployed Fractional Attribution model
  - Keywords / Topics
  - User engagement
- Let’s look at the CDP
- Snowflake
- AWS SageMaker
- Sigma Computing

What are likely to be our biggest hurdles?

- Folks who use these tools are in marketing
  - Statistics and model nuance aren't their specialties
  - Need an intuitive product
- Standard model VS creating one for individual clients
  - Toggles so clients can tweak model or try different preset ML models?
- Defining "high value" and where we can achieve them
  - Remember: "Everybody wants advice" - John Challice

User Engagement

What is it all about?

- User engagement models aim to understand and balance what content individuals interact with and how it affects their journey with published content

- e.g. A student reads a few papers from a journal, sees a social media post relevant to their field, gets an email about a few meetings adjacent to their field. Later, the student considers joining an academic society.
    - What value was brought to the student due to these touchpoints?
    - Can we reduce the amount of junk the student sees and reduce the efforts of the publisher?

Why this matters

- Know when to reach out to people primed to join
- Gauge how audience responds to content
- Determine conversion rates for customer goals
- Hone return on investment
- Reframe engagement score as a metric of who is getting value out of a product

Engagement score basics

- Simple representation, usually a number, of how much an individual interacts with content
- Most often based on digital behavior, chosen weights, and number of users

$$Interaction = W_1 n_1 + W_2 n_2 + W_3 n_3 + \ldots + W_i n_i$$

or $$Interaction = \sum_i W_i n_i$$

$$Engagement = Interactions \frac{10000}{Users^{0.8}}$$

*standard Facebook engagement scoring

Updated equation

Doesn’t have all solutions yet

$$Score = \sum_i W_i \left(\sum_j n_j \left(1 - r_i \right)^{\left(t - c_j \right)} \right) \left(\frac{100}{U^{0.8}} \right)$$

- W are vector weights
- n is number of user actions
- r is rate of decay
- t is current time, c is click time
- U is number of users

hum engagement metrics

Vector weights and decay

- Decay could be function of weight, e.g.

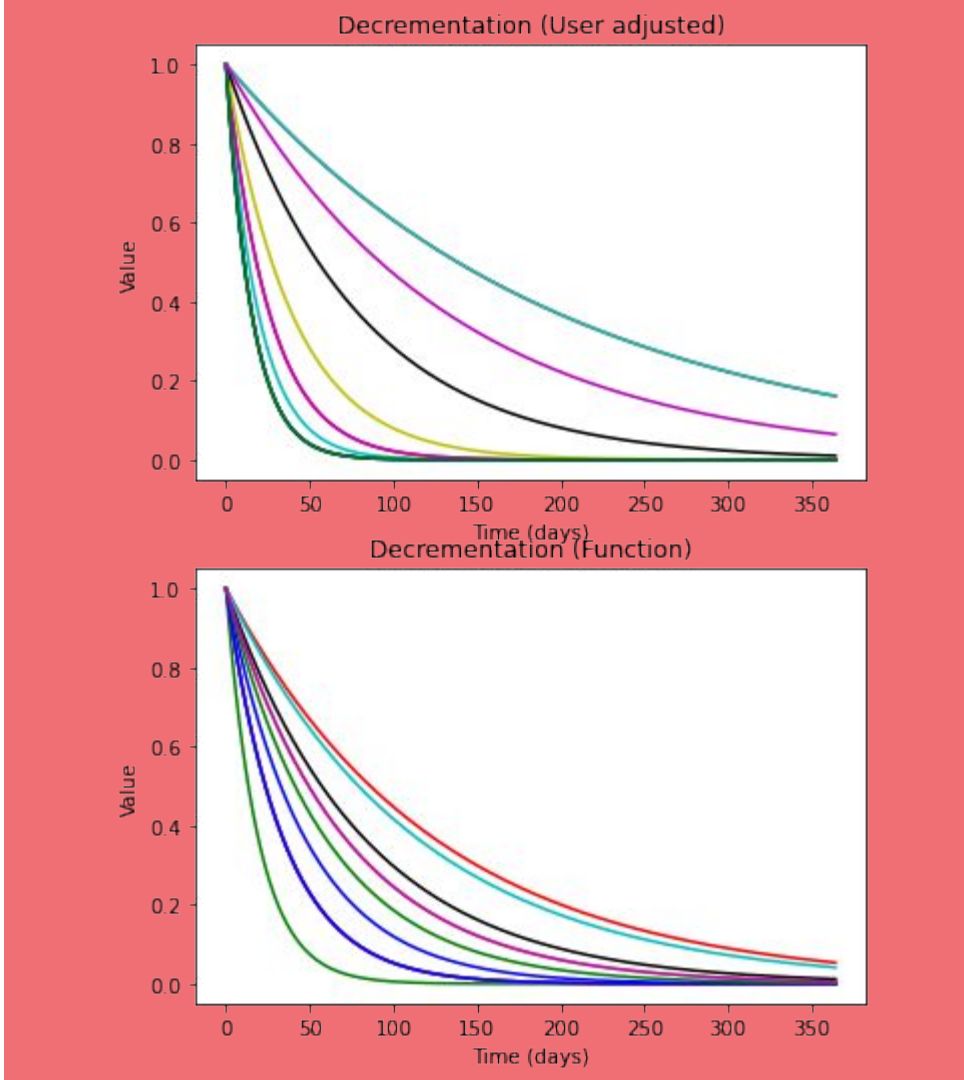
$$decay = \frac{1}{60} W^{-0.8}$$

Vector	Weight (W)	Decay (r)
Like	3	.25
Page View	1	.25
Page Scroll	2	.15
Start Read	2	.20
Page Skim	2	.15
Half Read	5	.10
Deep Read	6	.05
Open	2	.25
Click	4	.25
Earned Badge	10	.02
Took Test	9	.02
Login to Test System	5	.03

Decrementation over time

Top - User defined.
Pros: Highly customizable
Cons: Endless debates, potential for errors

Bottom - Function
Pros: Easy to implement, decay tied to weight of vector
Cons: Less variability



Project Flow

Model Goals:
(1) minimize loss of interest due to stale content and flow of too much junk
(2) maximize return on investment (this isn't just dollars)
(3) utilize time series for understanding human motivation

- Build known models to understand the data and give time for students to explore and start their creative process
  - Basic engagement scores, our proprietary algorithm, other factors
- Discuss which aspects students suspect are most influential
  - And why
- Students develop a new model based on their findings
  - We have a few ideas if anyone gets stuck

Background reading

Some starting inputs, we'll surely add to these topics

- CDPs
  - link - oldish 2020 so think creatively
  - Browse https://www.hum.works/
- First Party data
  - link
- User engagement scoring
  - link
- Fractional Attribution
  - link - simple but solid
- Snowflake / Snowpark
  - link
  - link2

- Plenty of rabbit holes to go down!
  - Keep a list of any questions
  - We can use Slack to post good reads as we find them

What makes this project a success?

Build tools, deploy them

- This is a research project, we'll have to stick and move
- Minimum for success
  - Build 2 models, undeployed, but ready for adaptation to various clients
- Stretch goal
  - A model for publishers that helps in author or reviewer/editor recommendations
- Together, we'll help academics and publishers streamline effort
  - We'll remove some junk from the internet that nobody wants to see
  - *Make human knowledge more interconnected and accessible*