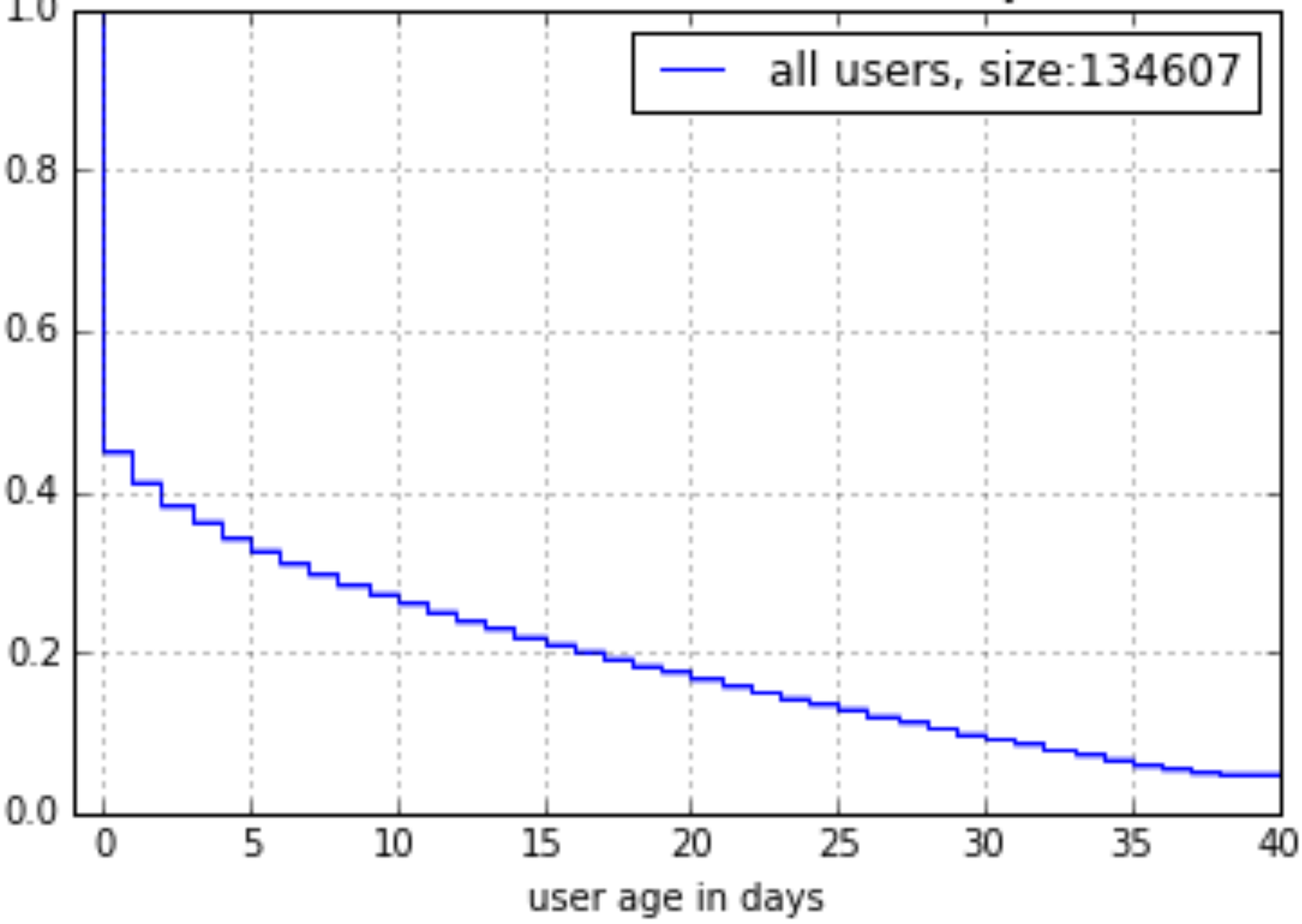
Customer churn in a news app

Lev Konstantinovskiy

Disclaimer: To comply with NDA the graphs in this presentation are based on toy data that somewhat resembles the original dataset.

What makes users stay?

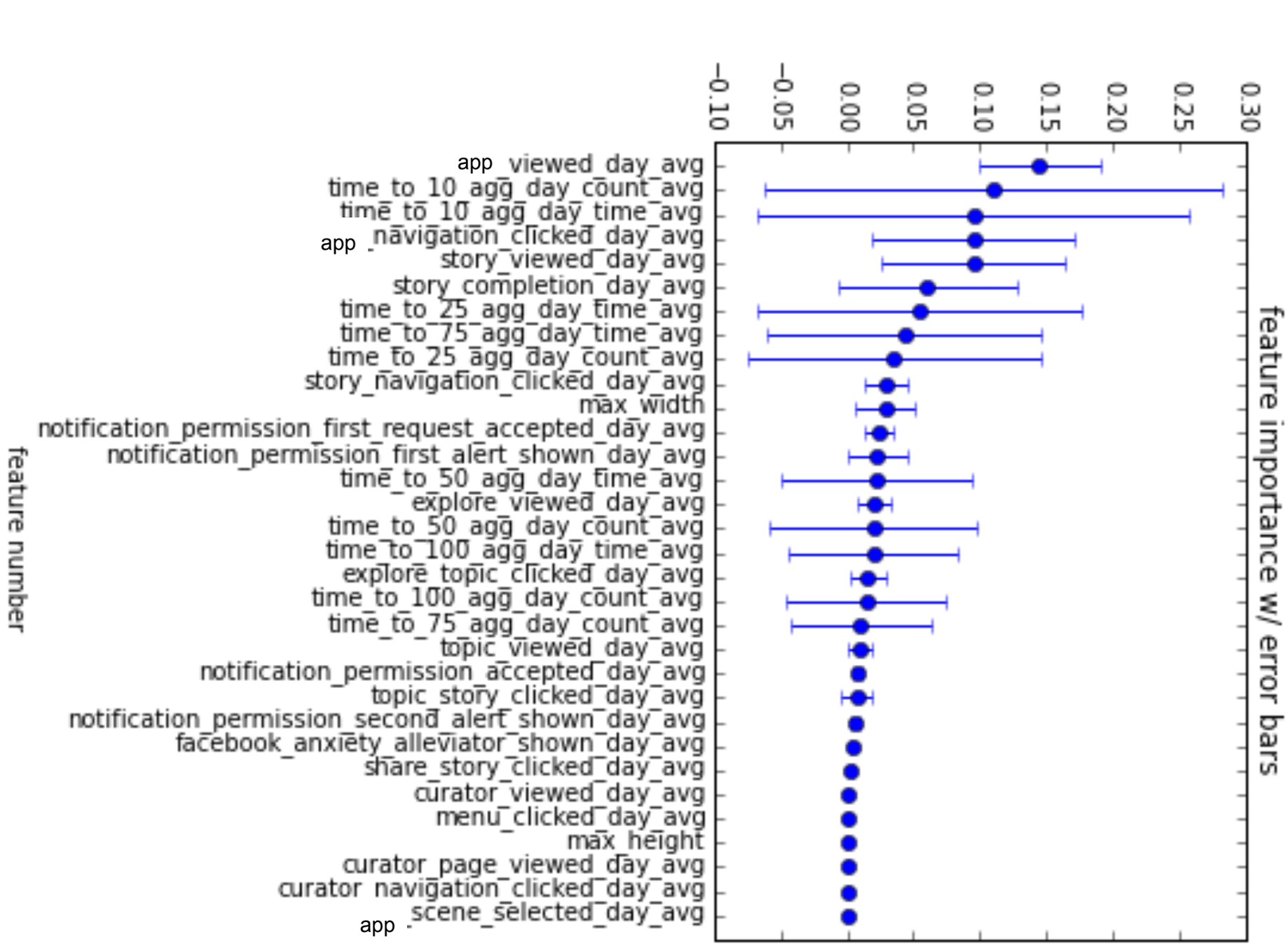


Survival over user's life:

most people visit only once

6 weeks

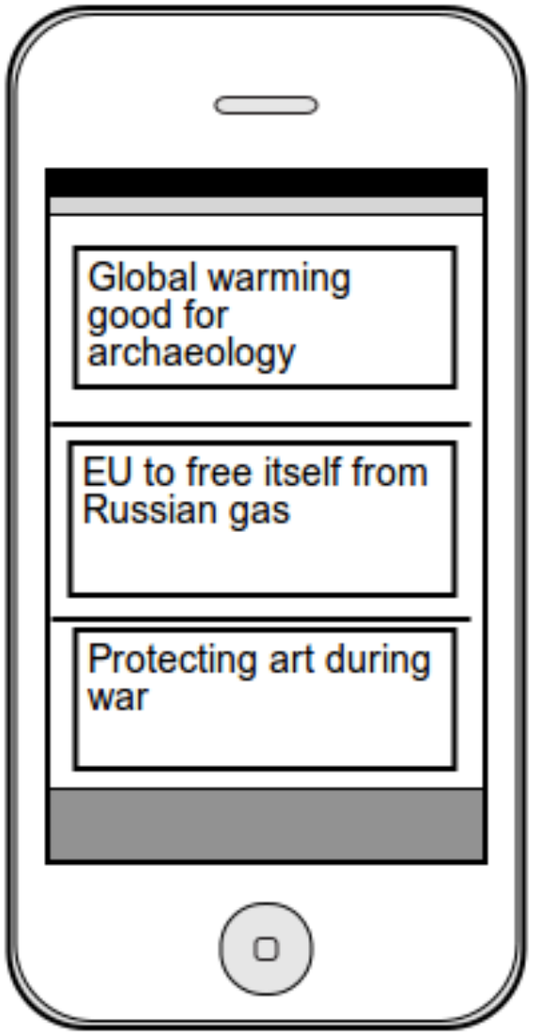
8m of all kinds of events, 300k user_id's, 4Gb



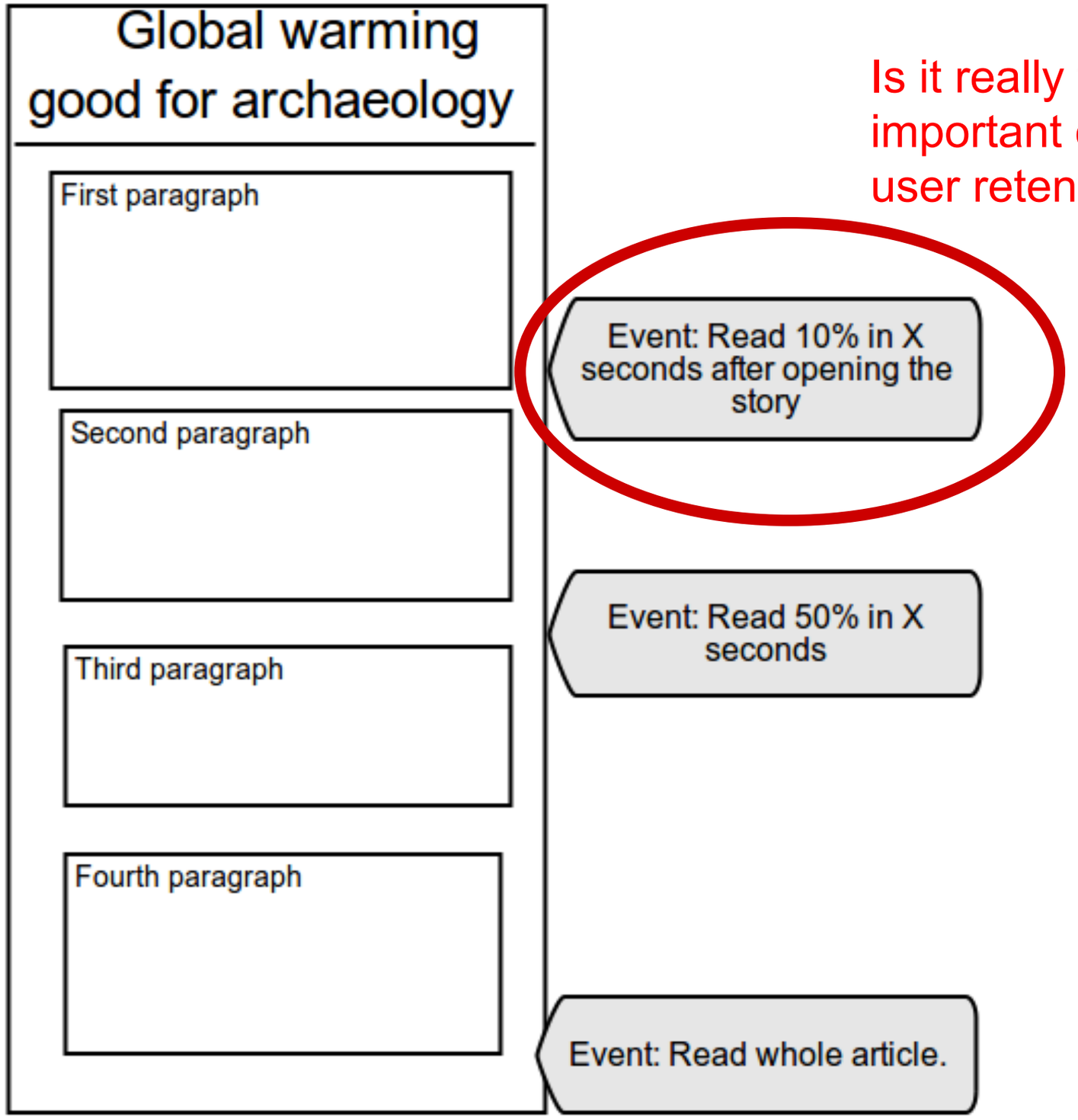
Users who stay vs users who leave: main difference is “reading to 10%”.

(feature importance from a random forest with a binary target “active/not active in the last week”)

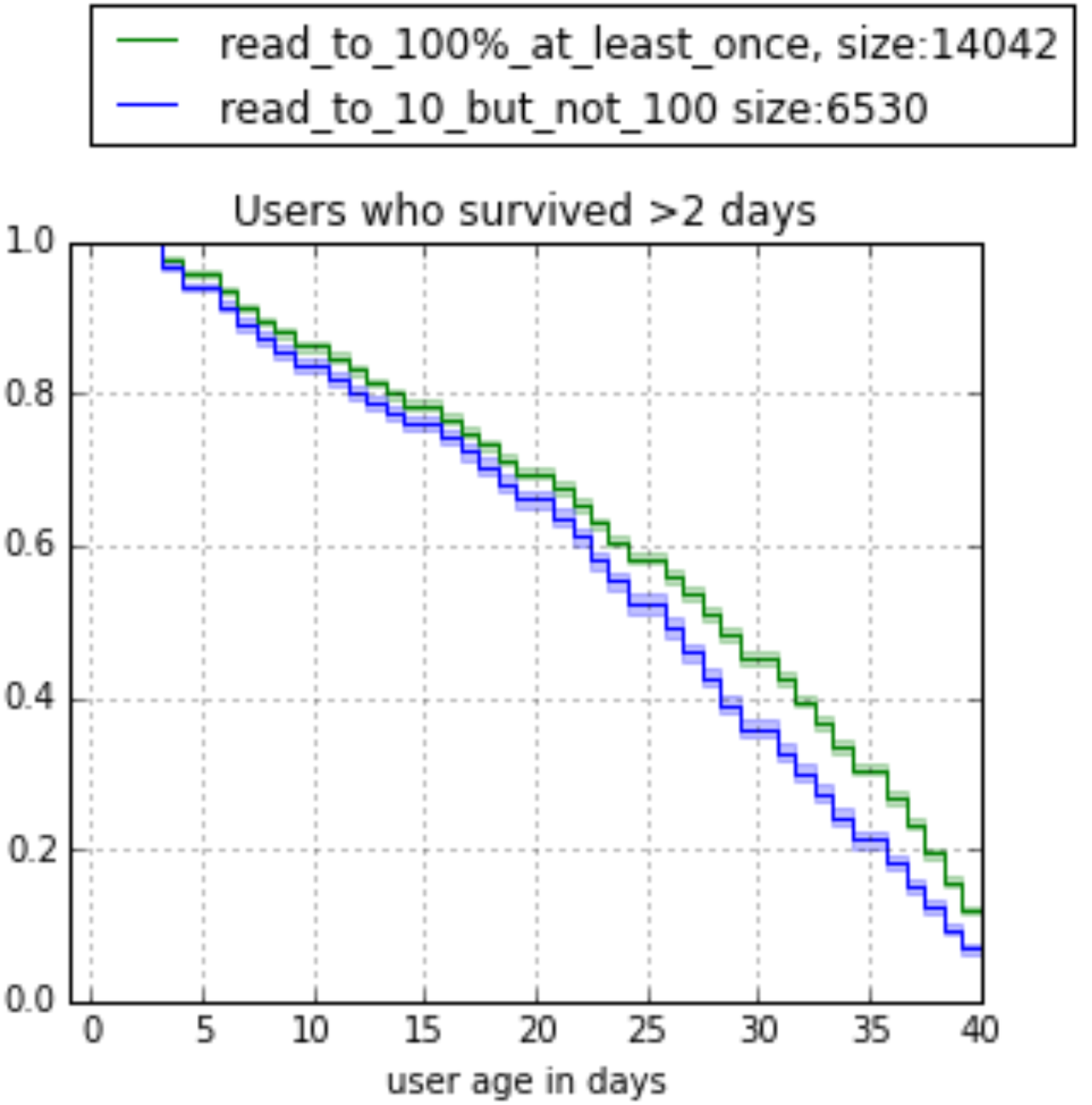
Homepage



A news story



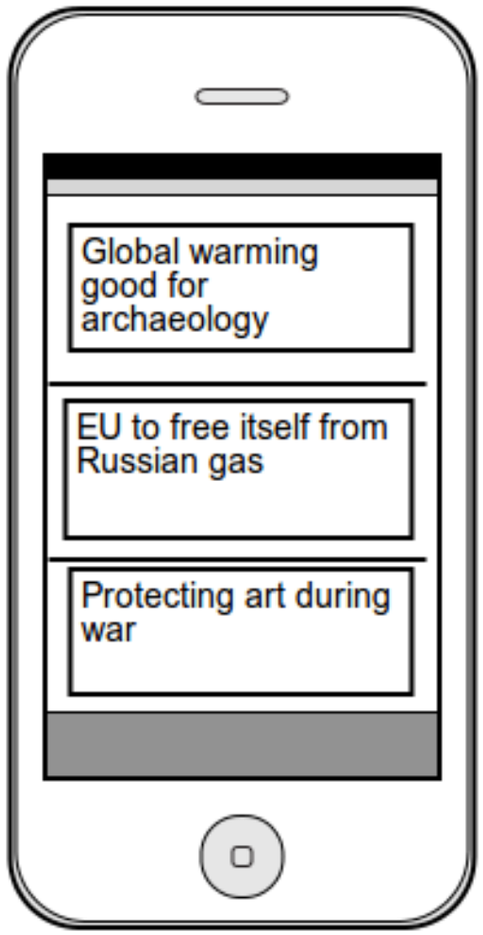
Is it really the most important event for user retention?



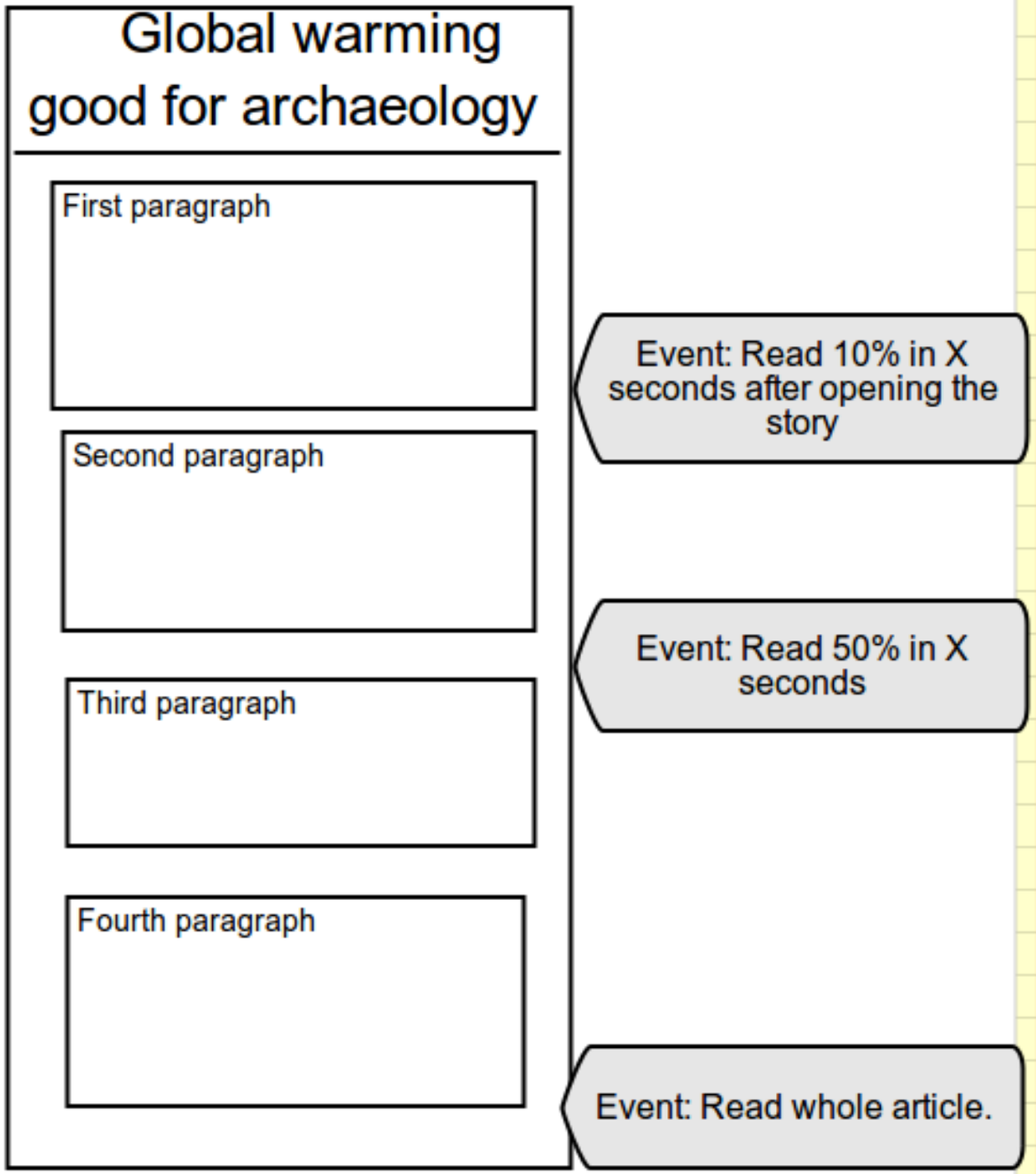
Survival over time:

There must be more difference between 10% and 100%!
We have all these dashboards that monitor how many got to 100% on a story.
But this graph shows it is not a good metric for engagement.
How come?

Homepage



A news story



These are "touch" events. Only triggered once.

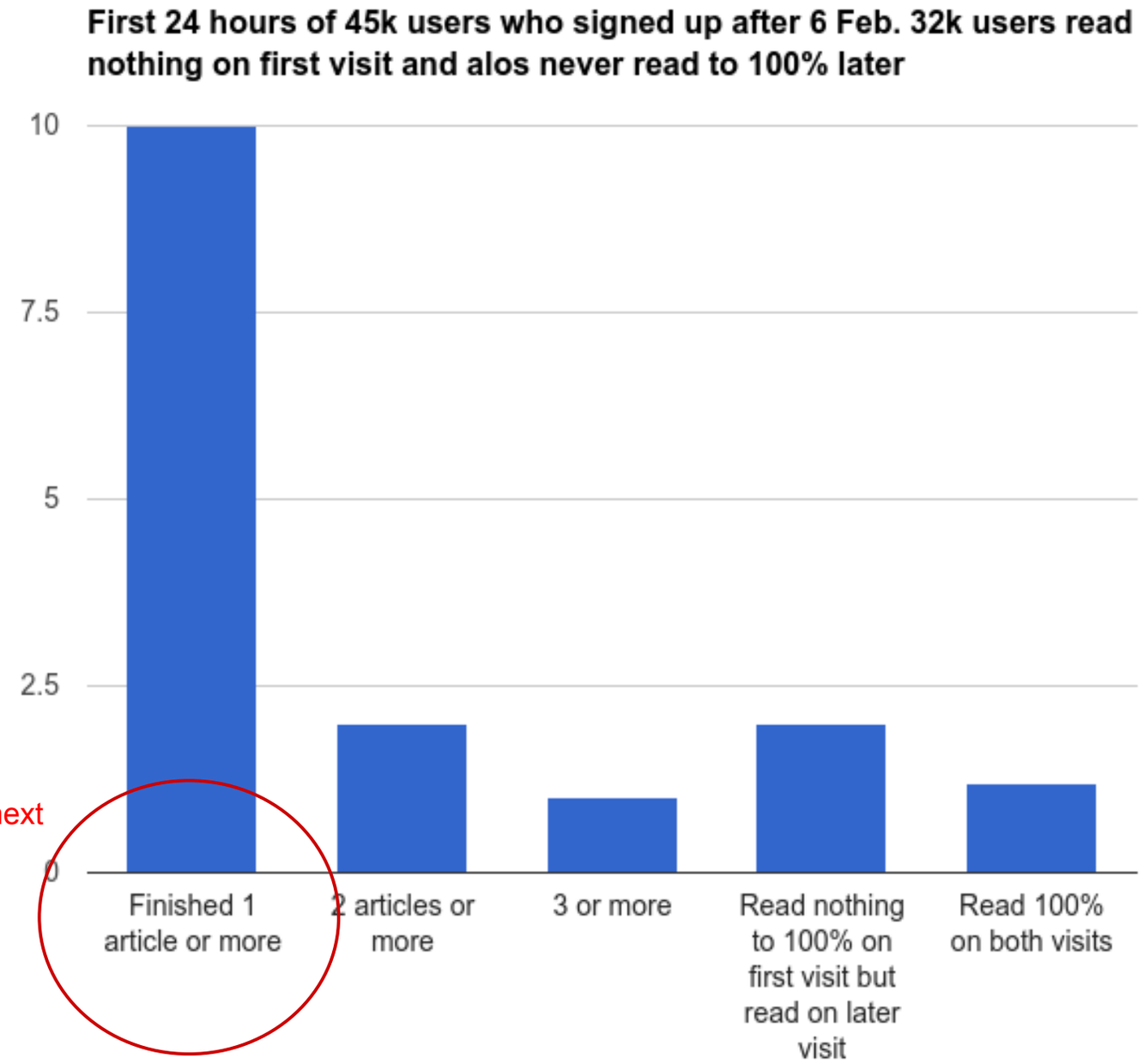
No events are coming from the user at all after they scrolled to the bottom.

Sometimes in 5 seconds, sometimes 5 minutes.

Code change: Events need to keep firing for as long as the user is on the page

Which stories make users return after first visit?

Getting through the biggest leak in the funnel.



10k users analysed on next slide

12% retention of those who read to the end

5% for those who didn’t.

Retention = “came back and read to 100”

Short and timely stories make users come back.

Story title	Date published, length in paragraphs	Predicted retention rate	Probability that prediction is true	Users read to 100% on first visit	% came back and finished another article at some point later (24 or more hours later)
Global warming good for archaelogy	Fri 5 Feb 7	17%	90%	200	31%
EU to free itself from Russian gas	Tue 10 Feb 6	17%	90%	297	31%
Protecting art during war	Fri 6 Feb 5	17%	90%	476	31%
Story A	Thur 29 Jan 7	17%	80%	115	31%
Story B	Tue 10 Feb 6	17%	80%	199	29%
Story C	Mon 16 Feb 9	17%	80%	125	32%
Story D	Thur 12 Feb 8	17%	80%	184	27%
Story E	Tue 3 Feb 6	17%	80%	143	30%

Eight stories get 40% lift over average 12%

Model assumption: read only one story on first visit. It is quite valid - in reality if a user reads one story, then they read 1.4 stories on average.

Next steps… Which stories are best at getting from 0 to 100?