AMENDED Exhibit 871

PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

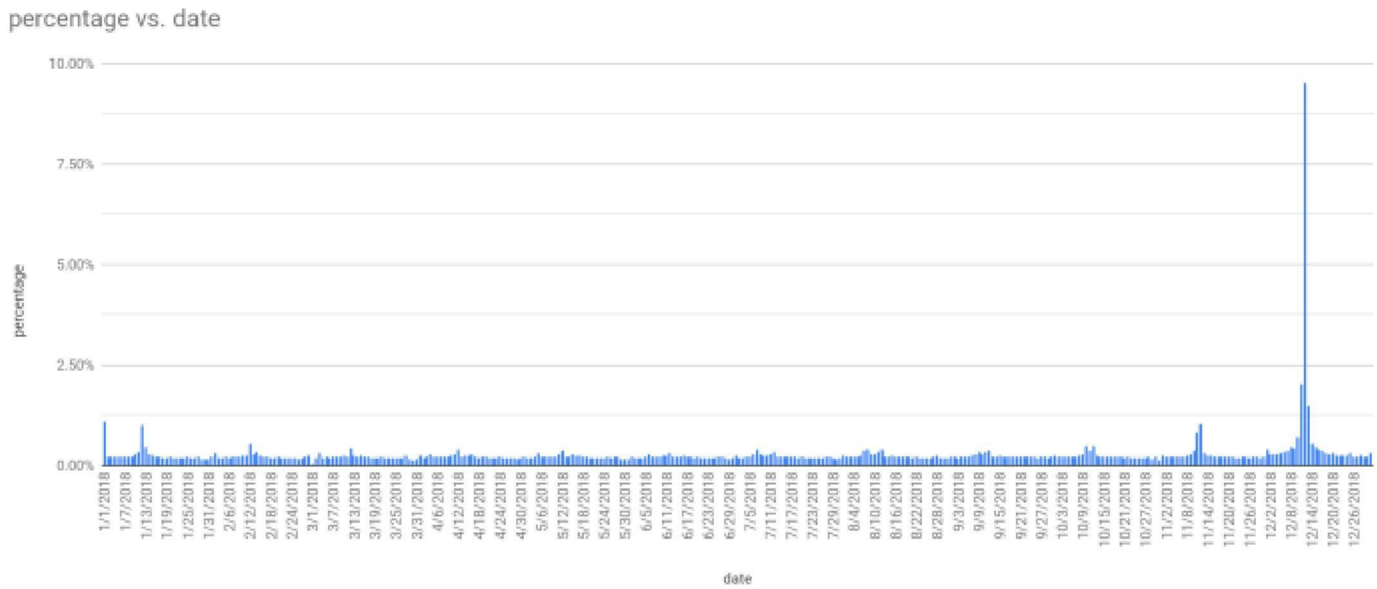
Age Inference

Last Update: 03/01/2019

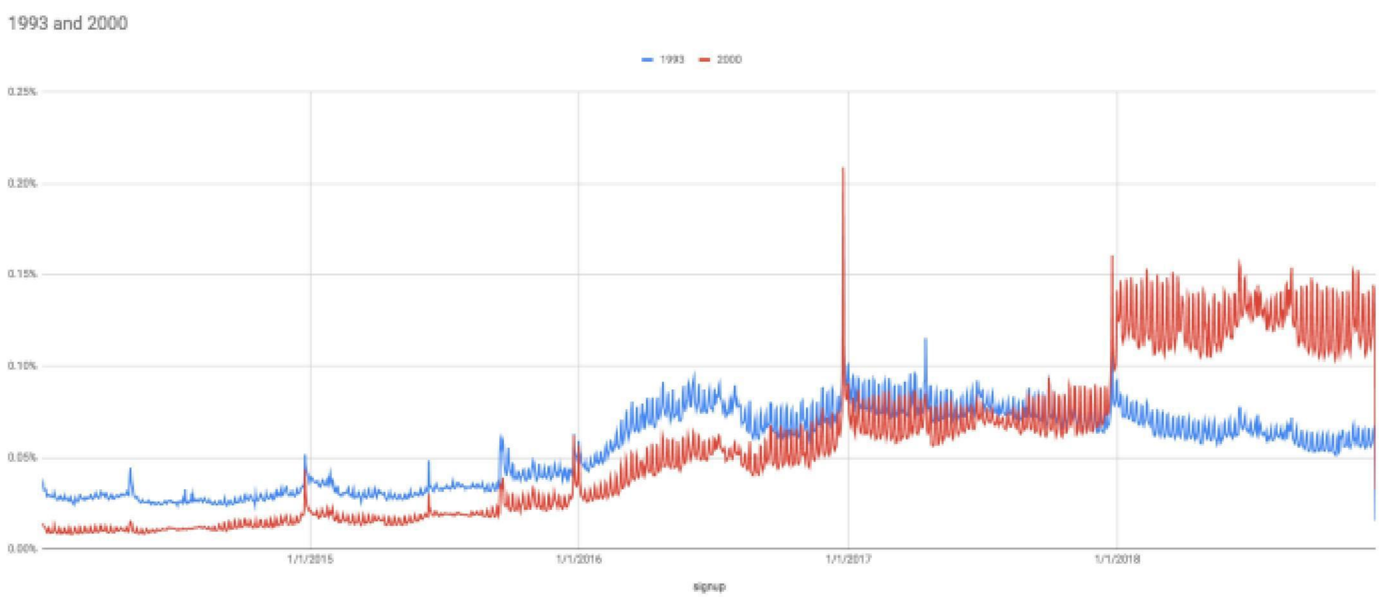
Why do we run age inference?

Deeply understanding our users benefits the broader business along many fronts – this is especially true for a user's age as this information is used to handle regulations and restrictions (e.g. no AlcBev ads for Snapchatters < 21).

Unfortunately, based on our existing new user sign-up flow, we've observed scenarios where we're highly confident the inputted age is *incorrect*. For example, we see nearly 10% of sign-ups from 12/12/2018 have 12/12 as the *MM/DD* for the age of birth – basic intuition tells us that every MM/DD should account for ~1/365 of total sign-ups.



Similarly, when we made a corresponding UX change that defaulted birth year from 1993 to 2000, we saw a large shift in age *YY*. These two highly biased data points indicate that there are certain users who are not inputting their actual DOB.



Page 1 Comments

L1 [redacted] Can you clarify how this logic works? Do you take the minimum of reported and inferred age for this logic or something different?
_Assigned to [redacted]
[redacted], *3/8/2019 09:13 AM*

D2 +[redacted] – that's correct. When advertisers select "Regulated Content" we use the MIN(supplied_age, inferred_age). This is stored in its own field in the User Profile ("age_group_regulated")
[redacted], *3/8/2019 05:09 PM*

How do we infer a user's age?

The guiding principle behind our approach is that of *homophily* – you are friends with people who are like you. Meaning, the best estimate of your age is that of your friends'. While these are the most influential inputs into our model, we also account for a user's historical engagement activity (e.g. a user who routinely watches CNN[1] is likely to be in an older cohort).

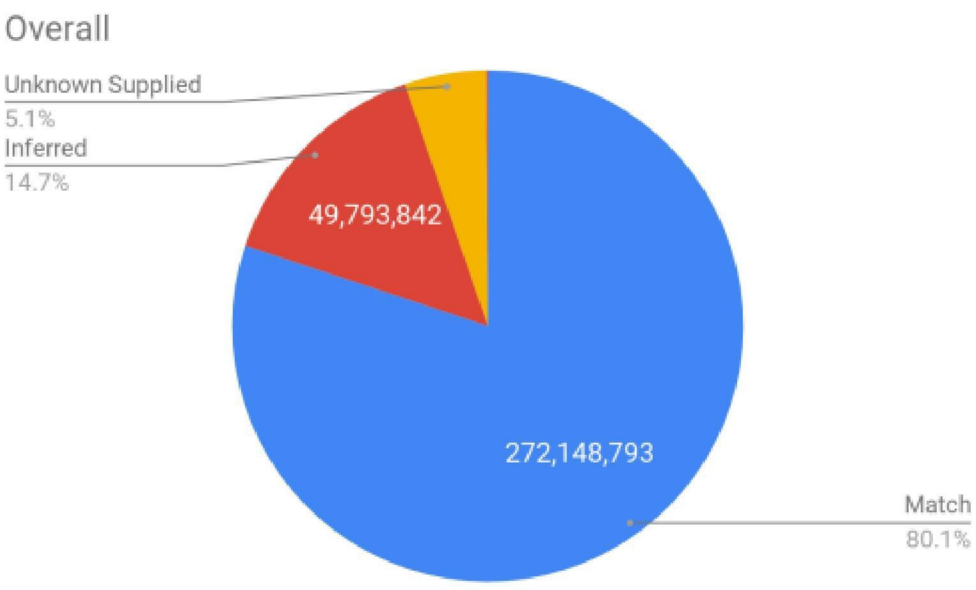
Age inference happens in five steps:

1. **Identify users for which we have have low confidence in their supplied age – this is based on a set of "Trusted vs. Untrusted" heuristics[2]**
2. **Train a model that classifies users into age groups**
   a. Use only "Trusted" users in model training
   b. Compute model performance/validation against a holdout and Experian data
3. **Train a model to generate user age point estimates**
   a. Use only "Trusted" users in model training
   b. Compute model performance/validation against a holdout and Experian data
4. **Finalize results leveraging a flowchart of heuristics**
   a. Fallback to supplied age whenever we're unable to generate a prediction (e.g. lack of available data for that user) or lack confidence in our prediction (e.g. model assumes equal likelihood that a user falls into multiple age buckets)
   b. Only overwrite supplied age with our inferred age when the model is very confident
5. **Validate output and upload age to the User Profile**

See https://snapchat.quip.com/73DlATnlZUWi for more details.

How often do we override a user's supplied age?

In aggregate, `[Inferred_Age == Supplied_Age || Confident_Inferrence == FALSE]` ~80% of the time (Blue - Match). For the remaining 20% of users, we override the supplied age – 15% (Red - Inferred) coming from when an explicit supplied age is provided and another 5% (Orange - Unknown Supplied) for when no supplied age exists (note: this bucket of users is largely comprised of highly tenured users who signed-up before we required DOB).



[1] Google slides with additional examples
[2] Trusted == FALSE (aka Untrusted) when [1] birth year is older than 1940, [2] sign-up month = birth month and signup day = birthday, and [3] selected date of birth is among the top 20 most selected date of births of the day of signup

Page 2 Comments

L3 +[redacted] Can you link the table where we can see non-bucketed user age at the user-level? I.e. these are the users we infer are 18, 19, etc. rather than 18-20.
[redacted], *3/8/2019 09:14 AM*

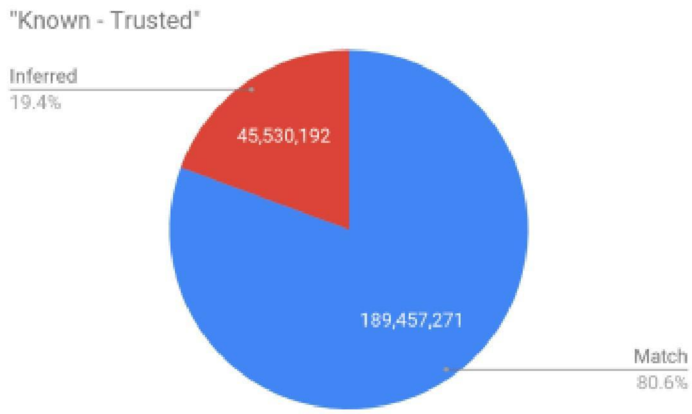
R4 +[redacted] In userprofile BQ table, demographics.age field is the point estimate
[redacted], *3/8/2019 05:18 PM*

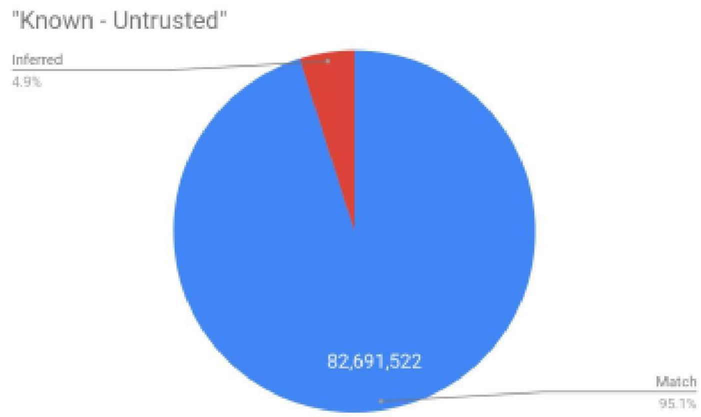
J5 when did snapchat start to require DOB?
[redacted], *8/17/2022 04:31 PM*

R6 @d[redacted] do you remember?

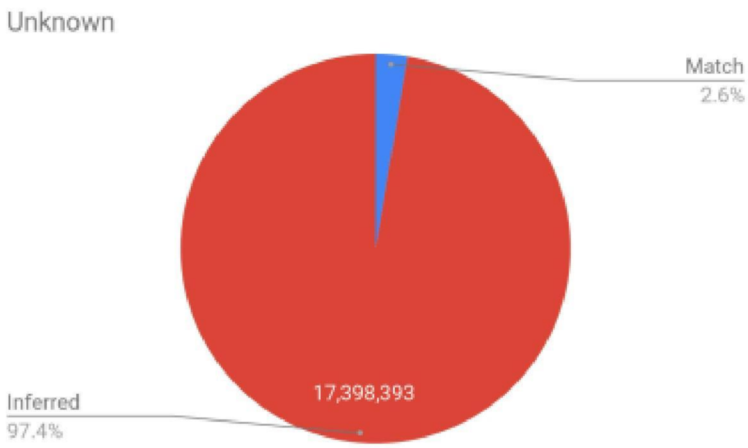
IIRC this was 2016 or earlier
[redacted] *8/17/2022 07:28 PM*

When breaking down the composition for “Known - Trusted” vs. “Known - Untrusted”, we see similar Blue - Match rates, but for “Known - Trusted”, we override the `Supplied_Age` with the `Inferred_Age` a larger percentage (19.4% vs. 4.9%) of the time. The intuition here is that, for “Known - Trusted” users, we have a better understanding of their friends, engagement behavior, etc. and thus are better equipped to make more confident predictions (i.e. `Confident_Inferrence == TRUE` more often which empowers us to leverage inferred).





When looking specifically at users who have NEVER provided an explicit `Supplied_Age` (i.e. `Supplied_Age = UNKNOWN`), we accept our `Inferred_Age` ~98% of the time. While this seems somewhat paradoxical, these Unknown users are highly engaged with the app which provides valuable predictive power for our models. Unfortunately, for the remaining ~2% of the time, we’re unable to ascribe any DOB (i.e. user provided no input and we lack sufficient confidence in our predictions).



Page 3 Comments

P7 So you took the overall population above and broke it into these two groups?
Peter Sellis, 3/1/2019 07:27 PM

R8 yup. There are 3 groups. these 2 where user gave us some dob and we flag it as "trusted" or "untrusted" and the one below where no dob is provided
[redacted] *3/1/2019 07:55 PM*

D9 [redacted] there's a footnote on the previous page for the exact heuristic.
[redacted]

P10 Gotcha - I would make this paragraph more of a transition paragraph to let the reader know this. I didn't get that exactly in my first reading. (The data are great btw.)
Peter Sellis, 3/2/2019 06:10 PM

J11 what do we do with this 2% users?
[redacted] *8/17/2022 07:00 PM*

R12 they just remain "unknown", they become unreachable by certain types of ads
[redacted] *8/17/2022 07:29 PM*

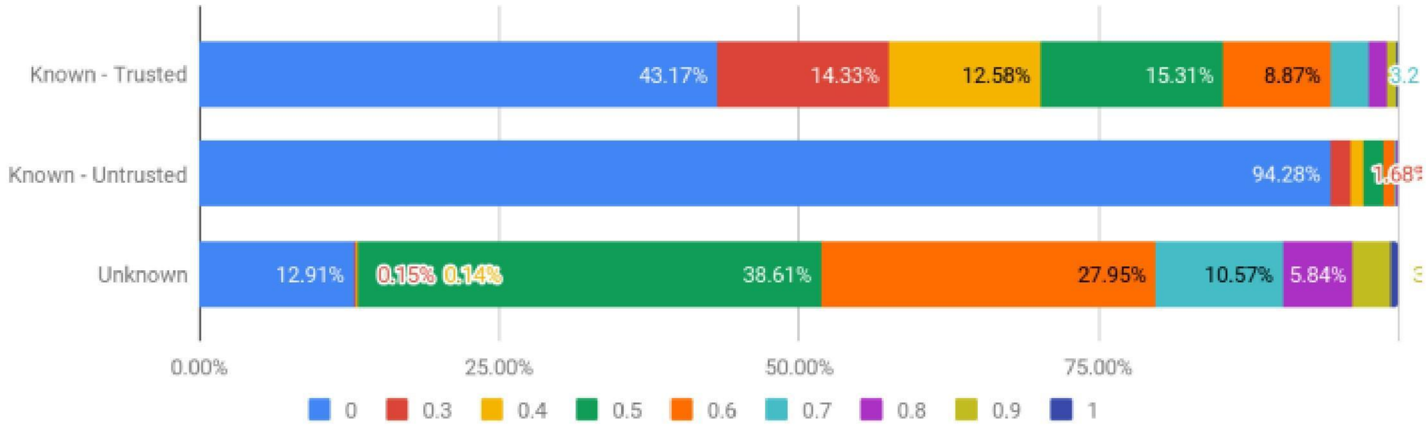
The following table shows the Age group composition before/after age inference for MAU as of March 5, 2021 across all countries:

Age group	Supplied	Final Age (most confident between supplied and inferred)
13-17	117 801 240	125 594 699
18-20	117 562 040	117 821 581
21-24	101 314 761	97 860 491
25-34	121 459 587	118 805 238
35+	95 902 786	94 112 403
unknown	656 035	502 037

Why do we continue to use supplied age in scenarios where we don't trust the DOB?

The TL;DR is that, for "Untrusted" users, we often don't have sufficient information to make reasonable (re: confident) predictions. In these situations, we have no other choice than to use the supplied age.

For starters, when we look at the predicted probability that a user "belongs" to a certain age bucket, we see that for "Known - Untrusted" users, we see that ~95% of the time, the age bucket with the greatest probability falls between [0, 0.3) – meaning, our models aren't providing much additional insight compared to picking completely randomly (i.e. 5 age buckets maps to a probability = 0.2 for each bucket). For "Known - Trusted" we see the outcome of our predictions have more mass (re: confidence) for specific age buckets[3]. Expanding on the point from above, for Unknown, we see we're even more confident (than "Known - Trusted") which speaks to their high degree of engagement.



The "more evenly generated probabilities" scenario (i.e. our model is not that predictive) usually arises because important features are missing (more common from "Known - Untrusted" users). When comparing overall platform engagement information, "Untrusted" users have significantly less session time and active hours and fewer number of friends (see appendix for specific engagement differences).

[3] For full confidence, our model would assign all the probability weight into a single bucket (the darkest blue bucket on the right) – we see this happens very infrequently

That said, note how `SUPPLIED_AGE == UNKNOWN` have very strong engagement characteristics. The theory here is that these are the OG users who signed-up (and are still very active) before we required DOB.

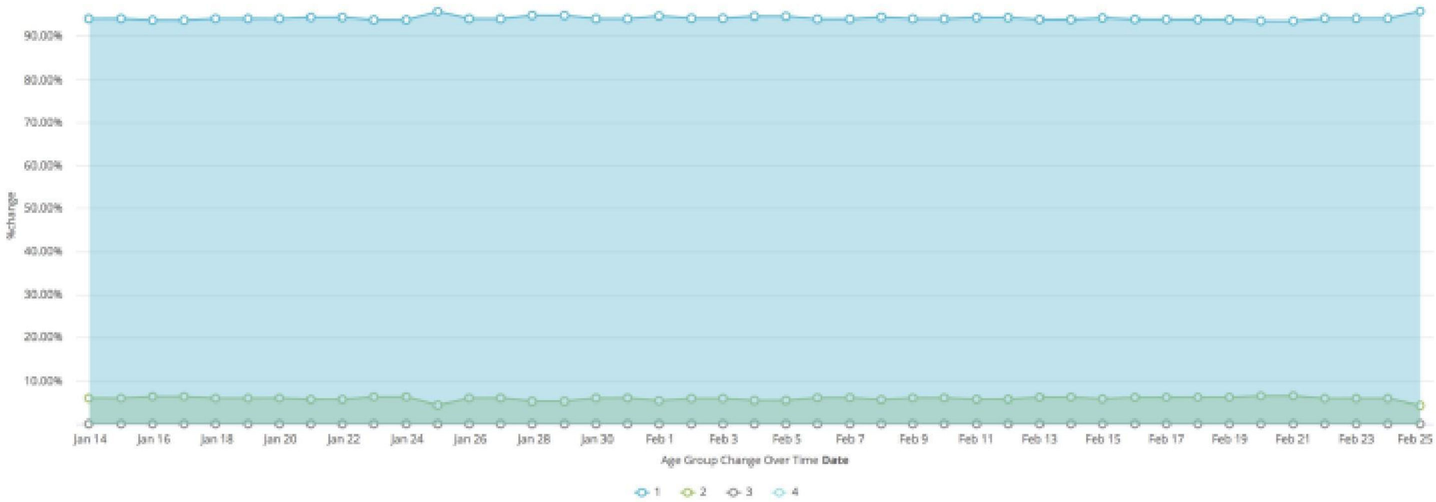
Supplied Age	Session Time[4]	Monthly Active Hours	Story Time Spent	Friend Count	MAU
Known -Untrusted	1398	7	0.00[5]	2	86,955,172
Known -Trusted	8437	61	1757.57	51	234,987,463
Unknown	10903	113	3576.57	95	17,871,714

Note: all numbers above represent medians

[4] Over last 28-days
[5] While 0.00s seems off, consider that this represents median story time spent and that the median user is spending ~7 hours per month

How often do we make changes?

- Our age inference model is trained and runs across **all users** (people who have a User Profile) 2x per week[J13]
- For a given user, age inference can only happen once per 30-days – meaning, a user's DOB can only ever be "set" (i.e. use existing supplied or override with inferred) once per 30-days
- The below graph highlights that, on any given day, **~6% of our total users change age**[D15] **(with respect to their previously `Inferred_Age` from 30-days ago)**



% users with changes

[J18] [J19]

How do we ensure the results are correct? How do we quantify improvement?

For starters, we first check (at a macro-level) the output of our model with matched data from 3rd party vendors such as Experian. While Experian data is strictly > 17yr, it still provides a good sense for how much better our model performed versus no intervention strategy (i.e. always taking supplied age).[J21]

Additionally, once our model is trained and the output ready, we perform the few final checks:[J22]

- Userprofile validator checks
  - Inferred age is outside inferred age group
  - Number of unknown age groups is not too high
- Feature updates to models are documented and analyzed for impact before release
  - E.g. https://snapchat.quip.com/oE3GAdbrftLT

Page 6 Comments

J13 Redacted--AC

[redacted] 5/22/2023 10:47 AM

L14 [redacted] Can you provide more detail on the 6%? This is only users that previously had an inferred age (so excluding new users getting their first assignment)? And it doesn't include the 0.27% that will change age every day due to birthdays?

[redacted], 3/8/2019 09:17 AM

D15 +[redacted] – good question. The 6% here captures the % of users on a given day for whom we [1] ran the inferred age algorithm and [2] the most recent inferred age != the previously inferred age (from the previous inference cycle).

Stated another way:
– On a given day, we run age inference on ~25% of L28 active users
** We only run age inference for a given user once every 30-days
– 6% of that 25% will have age_inference(t) != age_inference(t-1)

[redacted], 3/8/2019 05:14 PM

J16 Redacted--AC

[redacted] 5/22/2023 10:58 AM

J17 Redacted--AC

[redacted] 5/22/2023 11:00 AM

J18 Redacted--AC

[redacted] 5/22/2023 11:02 AM

J19 Redacted--AC

[redacted] 5/22/2023 10:55 AM

J20 Redacted--AC

[redacted] 5/22/2023 10:44 AM

J21 Redacted--AC

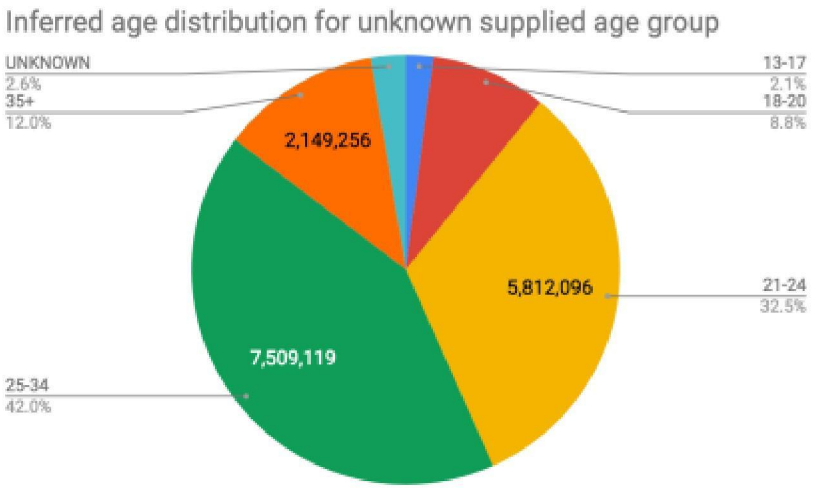
[redacted] 5/22/2023 10:45 AM

J22 Redacted--AC

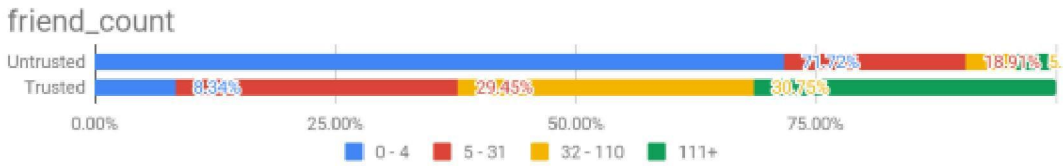
[redacted] 5/22/2023 10:48 AM

Appendix

- Age inference dashboard: https://lk-dev.sc-corp.net/dashboards/293
- Age inference documentation https://snapchat.quip.com/73DlATnlZUWi
- Supportive analysis for Untrusted vs trusted analysis: Google Sheet
- Date of birth analysis: Google Sheet
- Additional graph capturing how we leverage `Inferred_Age` by age bucket for `Supplied_Age == UNKNOWN`



- Additional model input/feature breakdown showcasing lack of information from “Untrusted” users
    - “Trusted” users have far more friends



    - Similarly, engagement with content (in this case Publisher Stories) is lower for “Untrusted” users

- When looking at user female-friend mean ages, we see the friends' supplied ages are much more in line with that of the users for "Trusted" users – stated another way, there is more mass along the diagonal (in bold).

Friend_age_mean_f:

Untrusted	13-17	18-20	21-24	25-34	35+	UNKNOWN
13-17	**7.43%**	16.51%	7.57%	5.50%	1.34%	61.63%
18-20	4.89%	**13.25%**	9.82%	8.26%	1.90%	61.87%
21-24	2.96%	9.65%	**18.36%**	15.34%	2.17%	51.52%
25-34	2.22%	6.15%	10.30%	**27.21%**	5.14%	48.98%
35+	2.33%	5.18%	6.86%	20.27%	**12.05%**	53.31%
UNKNOWN	0.52%	5.79%	24.36%	48.41%	13.16%	7.75%
Grand Total	**2.40%**	**8.40%**	**17.05%**	**30.25%**	**8.22%**	**33.68%**

Trusted	13-17	18-20	21-24	25-34	35+	UNKNOWN
13-17	**11.99%**	36.60%	11.47%	3.37%	0.72%	35.84%
18-20	4.57%	**36.04%**	21.75%	5.34%	0.93%	31.38%
21-24	1.95%	12.69%	**35.82%**	19.80%	1.43%	28.31%
25-34	1.36%	5.60%	10.62%	**44.79%**	7.96%	29.67%
35+	2.04%	4.32%	4.99%	24.09%	**29.18%**	35.38%
UNKNOWN	3.43%	8.05%	7.63%	9.01%	3.83%	68.06%
Grand Total	**4.15%**	**18.70%**	**16.83%**	**20.49%**	**7.55%**	**32.28%**

- We've seen that the model struggles to make an informed prediction when it lacks sufficient signal – the same holds true for the final heuristic decision flow. When we breakdown the outcomes from the final heuristic, we also observe that we rely more heavily on supplied age for "Untrusted" users.