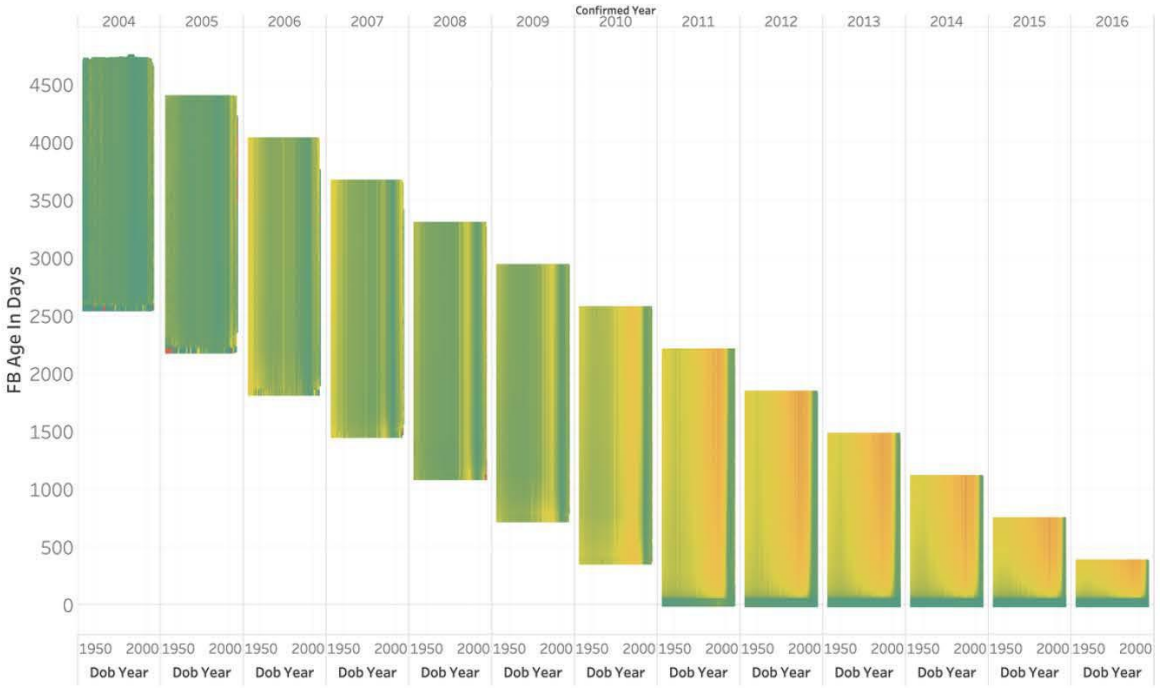
Long Term Retention: The Young Ones are the Best Ones and Other Learnings

By Santosh Kumar Panda and Robert Chen

(cover photo below)



TL;DR

We observe very stark differences in **long term retention** in the United States depending on **(1) Age at which you join Facebook** and **(2) Year in which you join Facebook**.

Establishing **long term retention as a standard metric** for Facebook to observe and analyze across our analytics and research teams, in addition to product teams more generally, is a major objective of this work. We see that year-of-birth/age and confirmation year are important dimensions through which to understand how different people are using Facebook over time.

PX-01325, Page 1 of 35

Tennessee v. Meta
Case No. 23-1364-IV
PX-01325

CONFIDENTIAL
CONFIDENTIAL

META3047MDL-047-00995641
METATNAG-019-01492550

Our three major findings and associated implications are below.

1. **Tweens (approximate age 10-12) are special. People who join Facebook as tweens have the highest long term retention out of all age groups.**
    a. Observations
        i. As an example, for people who joined Facebook in 2016, 12 year-olds (born in 2004) have long term retention (LTR) MAP_28 of 72% at an FB age of 365. 32 year-olds (born in 1984) have long term retention of 25% at the same FB age. This is a **3x difference** for a 12 y/o versus a 32 y/o! The difference is similar in magnitude across multiple confirmation years.
        ii. You can also see this very clearly in the Chart 2 below. There is a green band (indicating high retention) on the right in every confirmation year. The exceptions are in the very early days of Facebook (2004 and 2005) when high schoolers did not have access to Facebook.
        iii. Caveat: These findings rely on the Age Affinity model, and while it is the best we have for age, there are biases in the model which may influence the results.
    b. Implications
        i. **ACTION:** Given that tweens have long term retention multiple times (~3X) more than adults, understand why this is the case.
            1. Facebook as a whole should consider investing more heavily in bringing in larger volumes of tweens.
            2. For the Youth Team, **prioritize tweens over all other age groups**.
        ii. **ACTION:** Track down the former tweens who have incredibly high retention on Facebook and find out why they joined Facebook when they did and what keeps them on Facebook till this day. Can we replicate some of those conditions so that we have a greater number of highly retentive tweens in our pipeline?
2. **The earlier the confirmation year (the year in which a user joined Facebook), the higher the long term retention.**
    a. Observations
        i. For example, for somebody who was born in 1953, this user's long term retention is markedly higher the earlier the user joined Facebook
            1. Joined 2009: 76% at Day 1,800
            2. Joined 2011: 54% at Day 1,800
            3. Joined 2013: Projected ~47% at Day 1,800 (48% at Day 1,400)
            4. Joined 2015: Projected ~35% at Day 1,800 (42% at Day 700)
        ii. You can also see this clearly in the Chart 2 below. The early confirmation years are much greener (LTR closer to 100%) than the latter confirmation years despite the early confirmation years despite having users who are of much greater Facebook age.
    b. Implications
        i. Facebook's long term retention deteriorates with each passing year. This is unsustainable, and we need to find out why it deteriorates.
            1. Discovering why will lead to different product changes. Some hypotheses:
                a. Competition: Other social services are better and better

PX-01325, Page 2 of 35

b. Nature of underlying population: The people who join Facebook later on are more marginal. Perhaps they have worse connectivity? Or perhaps they have negative sentiment toward Facebook?

c. Changes in Facebook's product/the code and pixels: Our product has changed in a way that does not encourage strong long term retention.

d. Changes in the friend graph on Facebook: People who joined earlier on have a different graph than those who joined later on. The graph of those who joined earlier on lends itself to better long term retention .

e. Changes in accounts created: There could be more secondary accounts (soon abandoned) or more fake accounts.

2. **ACTION:** People who joined Facebook earlier on are fundamentally different from the people who join today. What makes them or their situation different? Can we replicate what we had before so that people who join Facebook today have higher retention?

3. **In a recent confirmation year, 2016, the higher the population penetration by year of birth, the lower the long term retention.**

a. Observations

i. **This is counterintuitive.** Conventional wisdom is that as population penetration increases, retention increases.

ii. Skip to the section "A Fascinating Relationship Between Long Term Retention and Population Penetration" towards the end to find analysis on this.

b. Implications

i. If the relationship is a causal one, then it suggests that we do not understand how social media products work. Even if it is only correlational, we should seek to understand the drivers.

1. Is it:

a. Users who are signing up for Facebook now are extremely marginal users?

i. Why are they marginal? Don't like Facebook? Primarily use Snapchat? Have better uses of their time?

b. Users who are signing up for Facebook now are already existing Facebook users but are signing up secondary accounts?

i. Why doesn't our SUMA model detect this?

c. Users who are signing up are not real people

i. Why doesn't our fake accounts team detect this?

ii. **ACTION:** This finding appears to challenge our understanding of how social media is supposed to work. It behooves us as Facebook to understand what is happening.

Context and Definitions

We initially looked into long term retention in order to calculate the lifetime value for teens. What we found was a dataset that was interesting in of itself.

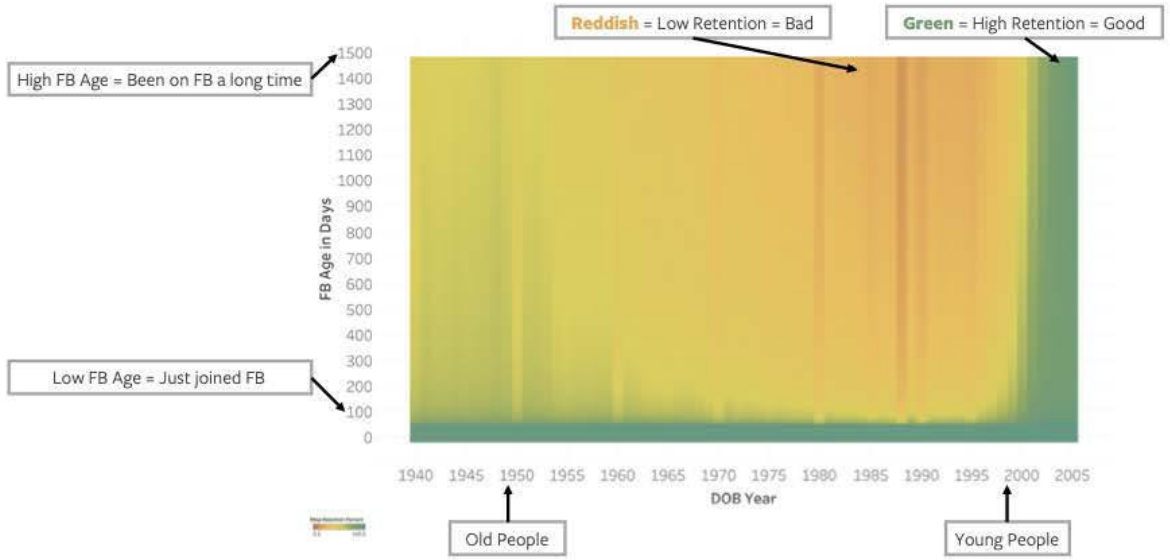
We define **long term retention** as the continued use of Facebook over very long periods of time, hundreds or even thousands of days. For the purposes of this note, we looked the percent of users who were MAP_28 after n number of days.

Quick Primer on Charts

PX-01325, Page 3 of 35

These LTR color charts are similar in format to triangle charts. You can skim this section if familiar.

(Caption Image: Chart 1: Sample Chart for Confirmed Year 2013)



The color chart represents 3 dimensions of data.

- On the X Axis is the year the user was born, the date of birth (DOB) year
- On the Y Axis is the Facebook age—the number of days after the user joins Facebook
- On the "Z Axis" (the color of the data point) is the long term retention percent
  - 0%, the lower bound, is red; low retention is bad
  - 50%, the midpoint, is yellow
  - 100%, the upper bound, is green; high retention is good
- In some of the color charts, the data is blank for the earlier Facebook ages. This is because our current dataset does not go that far back in time. E.g., we didn't have dim_all_users in 2005.

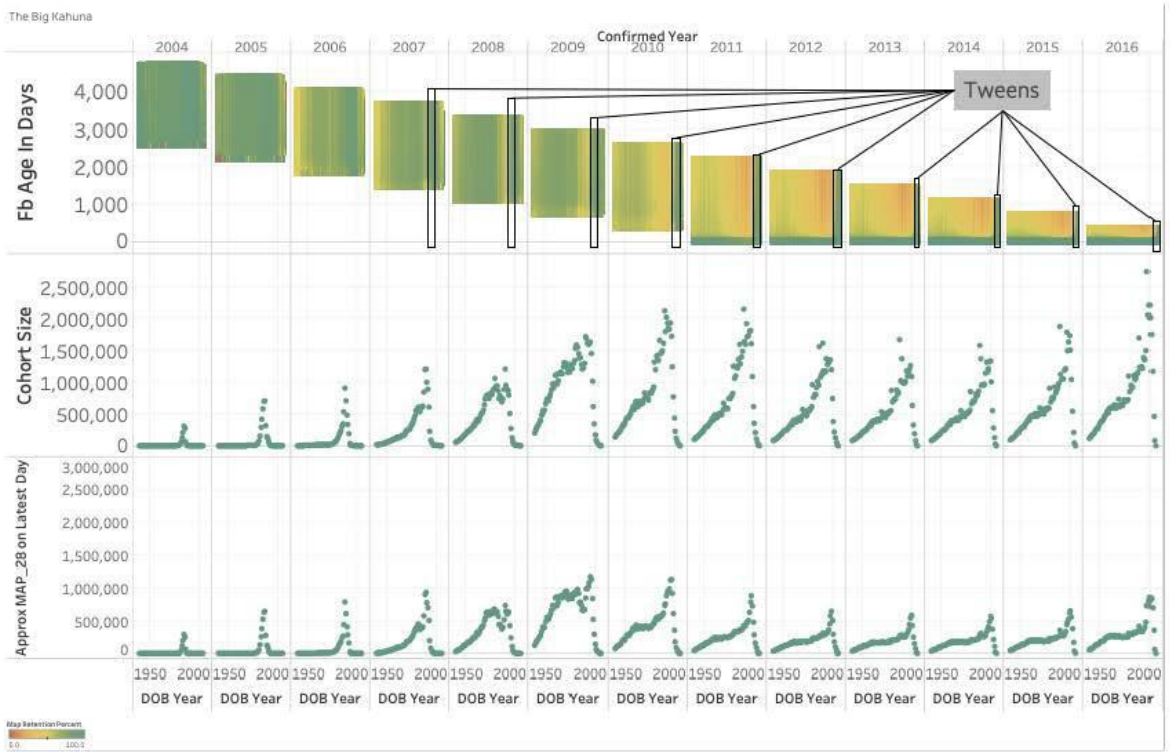
Long Term Retention in the US since 2004 by Year of Birth

We will start by looking at the big picture, birth years 1940 to 2006 and confirmed years 2004 to 2016.

All birth Years

(Caption Image: Chart 2: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for All Birth Years and All Confirmed Years)

PX-01325, Page 4 of 35



Note on reading Chart 2:

- The tops charts are what you saw in Chart 1 but instead of showing 1 confirmation year, we see all confirmation years
- The middle charts showing "Cohort Size" indicate the number of users on Day 0
  - The number of users that are MAP on Day X divided by the number of users on Day 0 = the retention %
- The bottoms charts show the "Approximate MAP_28 on Latest Day"
  - This shows the number of MAP_28 for each confirmation year x YOB on the latest DS
  - The # divide by cohort size gives the topmost color chart line for retention %

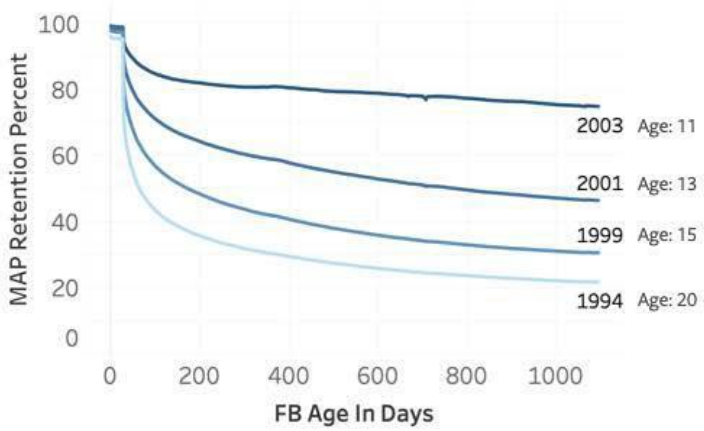
At a high level, we observe the following:

- The most retentive cohorts (the most green) are:
  - **Tweens (10-12 years old)**
  - Users from early confirmation years (2004 to 2009)
  - Further discussion in both the TL;DR and the 1981-2006 "zoom-in" section below
  - See Chart 3 below for an alternative view

PX-01325, Page 5 of 35

- The least retentive cohorts (the most red) are:
  - **Late teens/young adults**—we see this pattern emerging as early as 2007/2008!
    - **Implication**: Social product for those joining as young adults are unlikely to be retentive
  - Over time, the band of less retentive users expands in both directions
    - Younger and younger teens become less and less retentive. Hypothesis here is that people are joining social media earlier and earlier in life and habits are established early on
    - Older and older adults become less and less retentive. Hypothesis here is that most of the young and middle aged adults already joined a social media service earlier on so the people who are joining Facebook in recent years are more "marginal," either because they are already on an alternative service (e.g., Snapchat, Instagram) or because they naturally have less affinity toward social media
    - Another hypothesis, which we go into more detail at the "A Fascinating Relationship Between Long Term Retention and Population Penetration" section toward the bottom, is that retention goes down as we population penetration increases

Chart 3: MAP Retention for Select YOB's in Confirmation Year 2014. The people who join as tweens do much better than people who join as teens or young adults. Note that the ages indicated are the age at which they joined Facebook. Lines do not start at exactly 100% due to deactivated accounts.



- MAP_28 long term retention tends to be stable after a few hundred days
- Long term retention can be very high despite small cohort sizes
  - This is the case with the early cohorts from 2004-2005 as well as the small cohorts for the very old and very young in almost all the confirmed years
- Facebook's largest influx of users (cohort size) occurred in 2009, 2016, 2010, 2015, 2011
  - However, the users from earlier confirmed years are much more retentive
  - Therefore, the biggest confirmations years by proportion of current MAP are 2009 (22%), 2010 (14%), and 2008 (11%)
  - For the large number of confirmed users in the more recent years, we wonder if a significant portion of those are spam/fake accounts

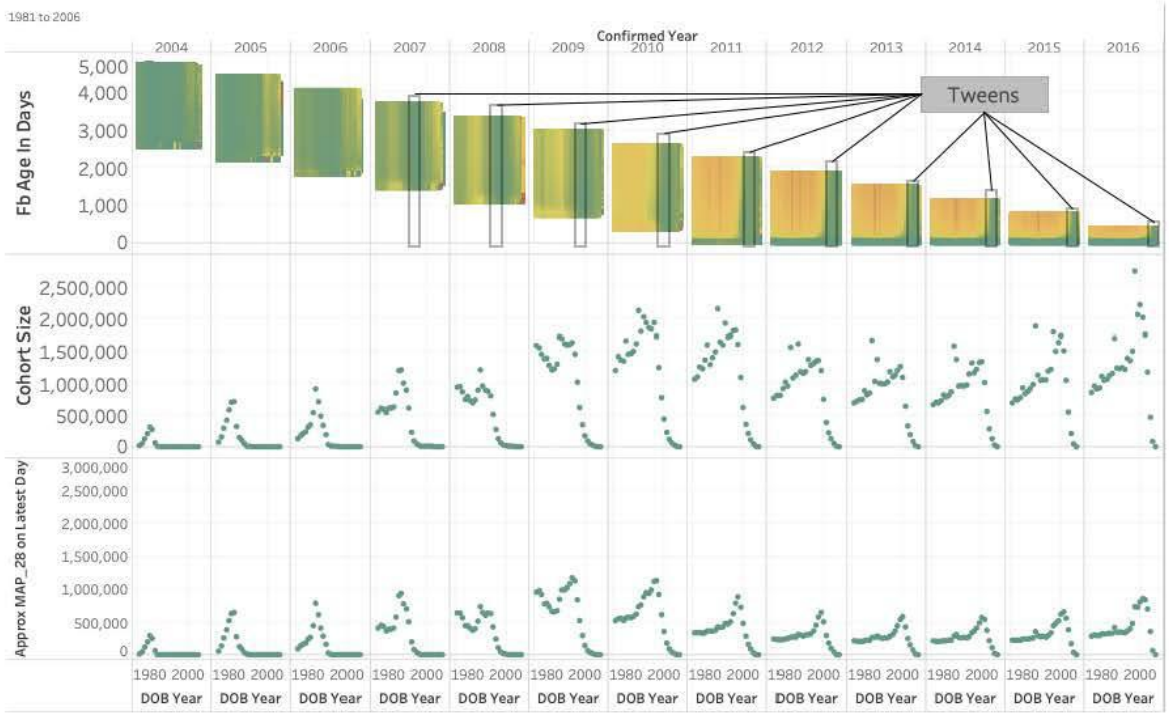
PX-01325, Page 6 of 35

- Difference in the shape of the curve between the Cohort Size (middle charts) versus Approx MAP_28 on Latest Day (bottom charts) shows you which cohorts are relatively more retentive
    - Conventional wisdom is that Facebook does poorly with older people—however, our long term retention data seems to say that older people who join Facebook, whether in the earlier years (e.g., 2009) or even more recently (e.g., 2016), tend to continue using Facebook

Next we zoom in on particular birth years. We start with 1981 to 2006. Then 1961 to 1980 and finally 1940 to 1960.

(Chart 4: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for Birth Years 1981 to 2006 and All Confirmed Years)

1981 to 2006



- About 51% of MAP is represented by birth years 1981 to 2006
- As early as confirmation year 2007 (Facebook was opened up to all users in September 2006), we see that people who joined FB as **tweens have the highest long term retention. This is true from confirmation year 2007 to 2016.**
    - E.g., people born in 1995 and confirmed in 2007, approximately 12 y/o at confirmation, have **92%** MAP retention on Day 3,652!
    - E.g., people born in 2005 and confirmed in 2016, approximately 11 y/o at confirmation, have **87%** MAP retention on Day 365
- Users joining as young teens used to have high LTR, but that has declined drastically in more recent confirmation years

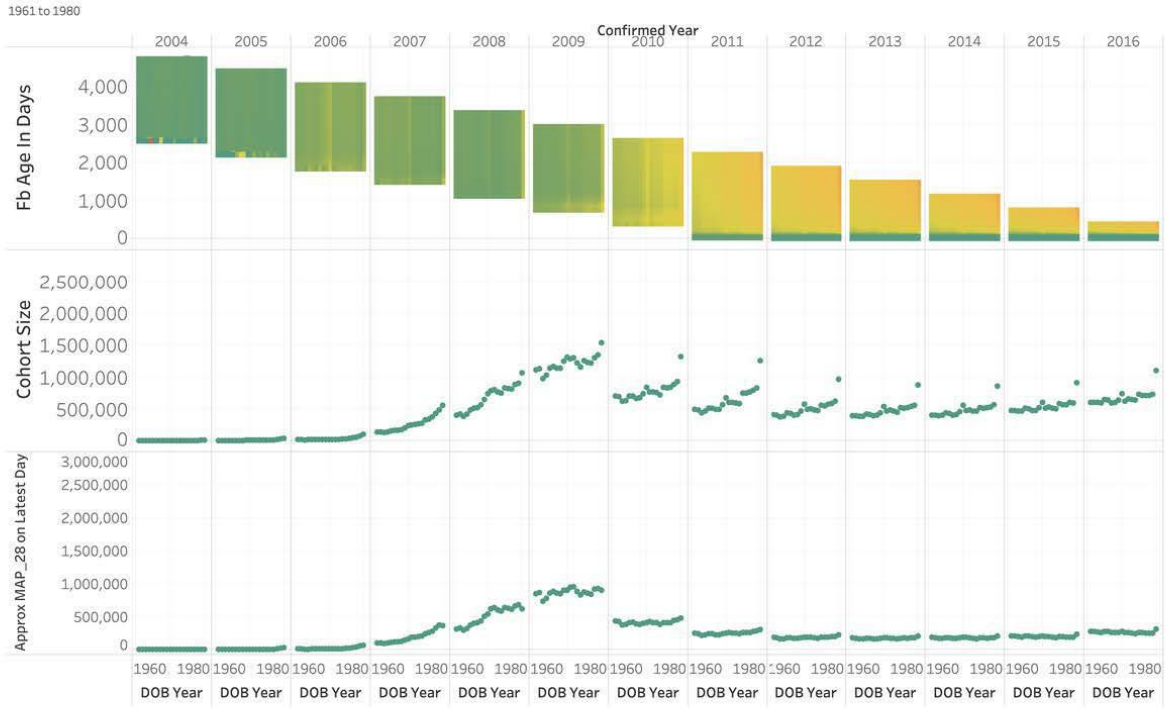
PX-01325, Page 7 of 35

- E.g., people born in 1996 and confirmed in 2010 (approx 14 y/o at confirmation), have 58% MAP retention at Day 2,000
- But for people born in 1997 and confirmed in 2011 (approx 14 y/o at confirmation) have dropped to 41% at Day 2,000. Even on the same date (Day 1,635), for those born in 1997 and confirmed in 2011, their MAP retention was 42%
- And even more recently, for people born in 2002 and confirmed in 2016 (approx 14 y/o at confirmation) have dropped to 40% as early as Day 365 and projected to drop to around 25% by Day 2,000

- In recent years, **people who join as young adults (19-29) have the worst long term retention** of all ages
  - However, this group is also fully penetrated on a MAP/population basis. So that leads us to ask:
    - Who is creating new accounts if it seems like almost all people in the age range are already FB MAP?
    - Are they creating secondary accounts?
    - How many of these are fake/spam accounts?

1961 to 1980

(Chart 5: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for Birth Years 1961 to 1980 and All Confirmed Years)



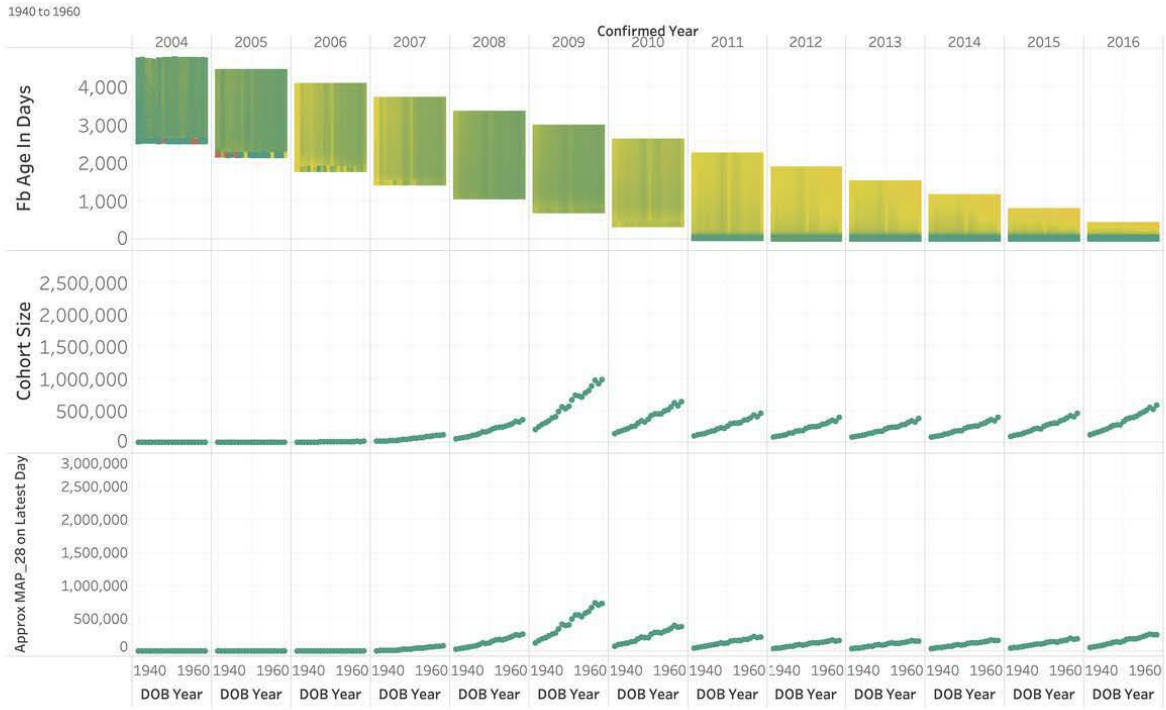
- About 32% of MAP is represented by birth years 1961 to 1980
- During the first few years of Facebook, this group had very high long term retention
- However, this group's long term retention has dropped considerably

PX-01325, Page 8 of 35

- In contrast to the younger ages, the number of recent account confirmations is considerably lower

1940 to 1960

(Chart 6: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for Birth Years 1940 to 1960 and All Confirmed Years)



- About 17% of MAP is represented by birth years 1940 to 1960
- Separate from those who join as tweens (10-12 y/o), those who join in old age have the highest long term retention
  - We see this in the yellow-green coloring for those in this age group
- **People who join Facebook at older ages do better than people who join as young adults**
  - Conventional wisdom is that Facebook does poorly with older people—however, our long term retention data seems to say that older people who join Facebook, whether in the earlier years (e.g., 2009) or even more recently (e.g., 2016), tend to continue using Facebook

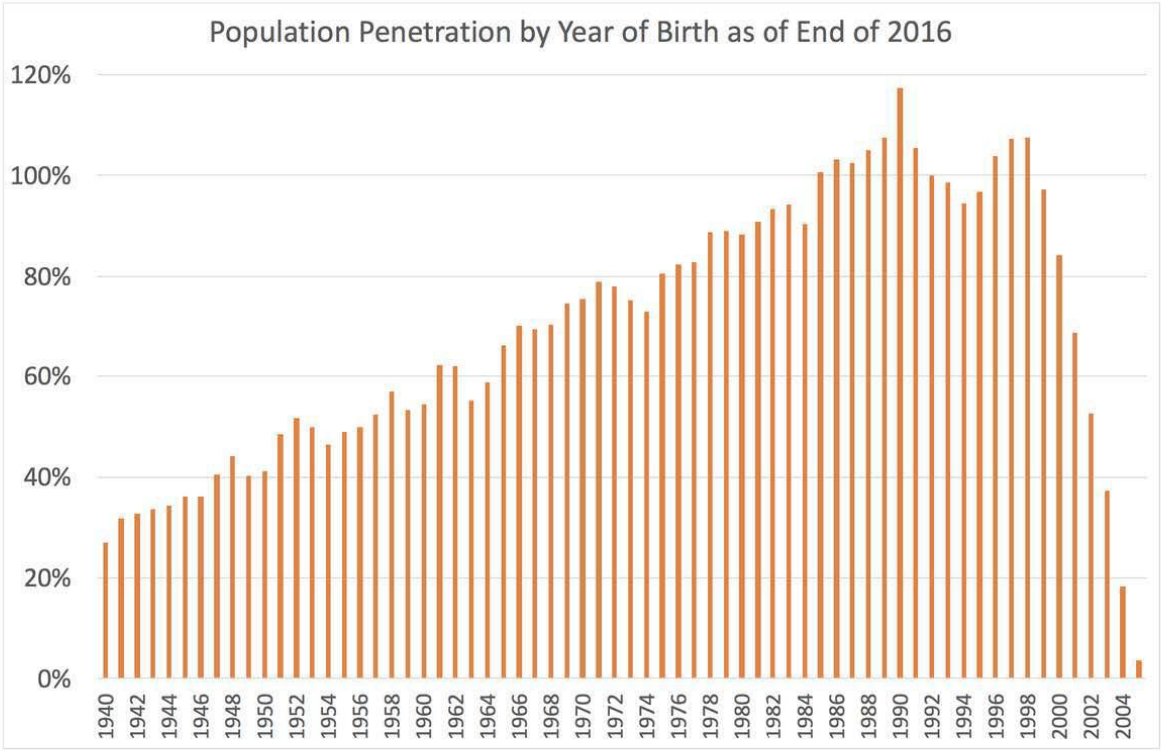
A Fascinating Relationship Between Long Term Retention and Population Penetration

The MAP_28 population penetration (MAP_28 / total population for that year of birth) is eerily an inverse of the long term retention chart in recent confirmation years. This leads to an unexpected result. The higher the MAP_28 population penetration, the lower the long term retention of new cohorts. We would expect the opposite effect. A simple linear regression (with population penetration as the independent

PX-01325, Page 9 of 35

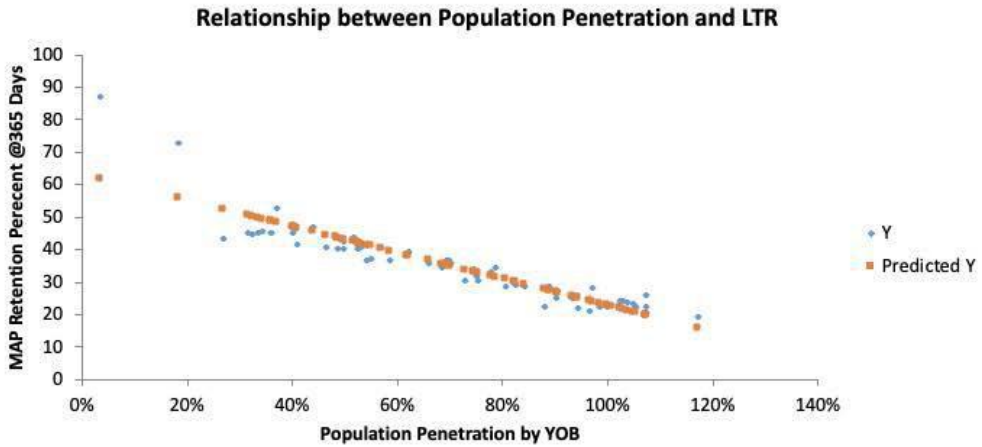
variable and LTR as the dependent variable) has an **r-squared = 0.84.** For confirmation year 2016, **the relationship is that the higher the population penetration, the lower the long term retention.** This runs counter to our expectation that as population penetration increases, the value of the social service goes up because more people you know are on the service. It seems like the countervailing forces such as differences in the underlying population (e.g., type of users who join later), fake accounts, secondary accounts, and/or other factors may be more important than the increased value from the greater % of people from your year of birth being on the service. Perhaps making the world more open and connected, beyond a certain point, is not that valuable.

(Chart 7: Population Penetration by Year of Birth as of End of 2016)



(Chart 8: Relationship between Population Penetration and Long Term Retention. r-squared = 0.84)

PX-01325, Page 10 of 35



Recommendations and Concluding Thoughts

Product Recommendations

Who precisely are we building for?

- Given the extremely high long term retention of tweens, what can we do to safely and appropriately onboard high volumes of tweens onto Facebook?
  - And can we push those growth while maintaining high retention?
- When we launch new products, do we intend for these new products for high-tenure or low-tenure people?
  - Because the LTR for high-tenure people continues to be high, we could have greater focus on building for low-tenure people
  - However, we would want to monitor our high-tenured people in case their retention drops
- Can we replicate the earlier days of Facebook and recreate the very high long term retention that Facebook used to have?

Research and Analytics Recommendations

Let's understand long term retention and these Confirmation Year x YOB cohorts much more deeply.

- Who are the users in the different confirmation year X YOB cohorts and what makes them the same or different?
  - By confirmation year:
    - If they are the same as the users today, is it the environment (e.g., competition) or product that has changed?
    - If early confirmation year users are different from those joining today, in what ways are they different and what can we build for them?
  - By age:

PX-01325, Page 11 of 35

    - What is it about people who join as tweens make them have such high long term retention?
    - Why do we do the poorest with those who join as young adults?
    - Why do we do better with people who join when they are old compared to young adults?
- Operationalize confirmation year into Deltoid
    - The differences in usage are starkly different by confirmation year so it makes sense to monitor metrics by confirmation year

Because of the difficulty in assessing long term retention, nobody has done it until now. With this dataset, we can understand Facebook and other family apps more deeply than before.

What's Next

We are planning to look at long term retention outside the United States and also for Instagram. We also hope this work spurs others to look into long term retention themselves as the breadth of this data across time and products lends itself to further analysis.

Methodology

The long term retention dataset provides statistics and analyses on retention of users on facebook over time. The primary retention metric for the current dataset is overall FB xAP and it is limited to the users in US only. The dataset is populated all the way from Jan 2012 till March 2018. The dataset provides a wide variety of dimensions such as the Age, Year of Birth and Registration Date among other things. The dataset also tracks the Facebook Age as well as the age of the user though time which are aimed at providing generational insights into user retention.

Data Derivations

Almost all of the data including most of the dimensions and all of the xAP metrics are derived from dim_all_users. It is one of the most trusted and well monitored datasets which is critical for data quality. Age related attributes are derived from Dim Age Affinity (see below). Accounts that are deactivated are included in the cohort until the day of de-activation. Accounts which are marked as SUMA or Fake are excluded from any all calculations.

Age Calculations

The Teen/Non-Teen and the age of a user are determined using Age Affinity. Instead of using the Point Estimates - we estimate the distribution of the sample mean of the user being in a certain age/age-group using the bootstraps. For example to determine if the user is a Teen - we use the following 3 step formula

1. Calculate Age for a bootstrap n being calculated as $A(n)$ = @DATEDIFF(CURRENT_DS, AA_DOB(n))
2. $P(n) = \{A(n) \leq 18\}$ - then 1 else 0.
3. $P(\text{teen}) = (\sum P(n))/100$

The same is extended onto age specific calculations.

We also froze the Age Affinity partition to the 07/01/2018 partition to avoid inconsistencies in the dataset due to predictions changing over time. Since this analysis is primarily for long term retention - the impact of AA not having accurate predictions for new users is minimal.

PX-01325, Page 12 of 35

SUMA

One of the key assumptions in the dataset is that the rate of SUMA has been fairly constant over time. Based on the results of the high level analysis - the current SUMA model only flags about 10% of accounts as SUMA. For example in for YoB 2012 - there were 12M accounts created on FB for US users but only 4M people born in US.

Dataset Details

Primary Summary Table - ytg_cohort_user_metrics_summary_agg
Dash - https://our.intern.facebook.com/intern/unidash/dashboard/youth_growth/cohort_analysis_by_dob/
Sample Query - For people born on 1999 who were between the age of 10 and 30 - how is their retention across time?

```
SELECT
    fb_age_in_days,
    age_at_registration as dim,
    SUM(map_28) as metric_val,
    SUM(total_users) AS cohort_size,
    (SUM(1.0 * map_28) / SUM(total_users))*100 AS retention_percent
FROM
    ytg_cohort_user_metrics_summary_agg
WHERE
    ds between '2012-01-04' AND '2017-12-31'
    AND age_at_registration BETWEEN 10 AND 30
    AND fb_age_in_days IS NOT NULL
    AND dob_year = '1999'
    AND confirmed_year IS NULL
    AND age_in_years IS NULL
    AND age_in_years_months IS NULL
GROUP BY
    1,
    2
ORDER BY
    1, 2
```

Thanks

Thank you to those who shared feedback on earlier versions of this work: Danny Ferrante, Mike Winters, Samarth Mohanty, Shrivats Iyer, Steve DeLucia, Tom Cunningham.

Thanks to our respective managers Ali Imam and Keval Patel who gave us the time to pursue a novel type of metric and analysis.

Appendix

CONFIDENTIAL
CONFIDENTIAL

META3047MDL-047-00995653
METATNAG-019-01492562

Previous presentation on the topic: Youth Leads Review

[Not part of the note]
People who might be interested in this:

- fake accounts team
- SUMA team
- core growth
- teens team members
- Nadine (interest in helping out older people)
- segmentation peeps (earl and alan)
- growth marketing peeps
- Messenger Kids

[removed from note to make shorter]

- The example above is the 2013 Confirm Year color chart
  - As evidenced by the dark green band on the right, young teens/tweens have the highest long term retention
  - The yellow-red blob in the middle shows that older teens and younger adults tend to have the lowest long term retention
  - The yellow-green band on the left shows that much older adults tend to have reasonably good long term retention (in between the tweens and young adults)
  - We also see columns that stand out of place for 1940, 1950, 1960, 1970, 1980, 1990, and 2000. This could be a reflection of human psychology when creating "fake" accounts per Spencer Burns

Contact Support

Conversation History

Quipbot#08 left the thread *Aug 9 at 6:56 am*

Quipbot#08 locked edits on the document *Aug 9 at 6:51 am*

Quipbot#08 *Aug 9 at 6:51 am*

=====IMPORTANT=====

PX-01325, Page 14 of 35

This Quip file has been copied to Google Workspace (https://docs.google.com/document...) as part of the Quip2Google migration efforts. Quip is moving to **Read-Only**. Please work in the new Google file moving forward.

Learn more

https://fb.workplace.com/groups...

Questions?

Visit the Quip2Google Migration Feedback Group (https://fb.workplace.com/groups...).

Note

New edits in this Quip file will NOT be automatically synced to Google Workspace.

Quipbot#08 added Quipbot#08 to the thread *Aug 9 at 6:56 am*

Quipbot#08 added you to a document *Aug 9 at 6:51 am*

Robert Chen (PM) edited *Dec 5, 2018 at 4:14 pm*
(Chart 5 6: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for Birth Years 1940 to 1960 and All Confirmed Years)
(Chart 6 7: Population Penetration by Year of Birth as of End of 2016)
(Chart 7 8: Relationship between Population Penetration and Long Term Retention. r-squared = 0.84)

Robert Chen (PM) edited *Dec 5, 2018 at 7:35 am*

Establishing **long term retention as a standard metric** for Facebook to observe and analyze across our analytics and research teams, in addition to product teams more generally, is a major objective of this work. Additionally, we We see that year-of-birth/age and confirmation year and year-of-birth/age are important dimensions through which to understand how different people are using Facebook over time.

Our three major findings and associated implications are below.

1. **Tweens (approximate age 10-12) are special. People who join Facebook as tweens have the highest long term retention on an absolute basis out of all age groups.**
    a. Observations
        i. As an example, for people who joined Facebook in 2016, 12 year-olds (born in 2004) have long term retention (LTR) MAP_28
    b. (LTR)
        i. of 72% at an FB age of 365. 32 year-olds (born in 1984) have long term retention of 25% at the same FB age…
        ii. model, and while it is the best we have for age, there are biases in the model which may influence the results.
    c. Implications

PX-01325, Page 15 of 35

i. **ACTION:** Given that tweens have long term retention multiple times (~3X) more than adults, understand why this is the case.

1. Facebook as a whole should consider investing more heavily in bringing in larger volumes of tweens.

2. For the Youth Team, **prioritize tweens over all other age groups**.

ii. **ACTION:** Track down the former tweens who have incredibly high retention on Facebook and find out why they joined Facebook when they did and what keeps them on Facebook till this day. Can we replicate some of those conditions so that we have a greater number of highly retentive tweens in our pipeline?

2. **The earlier the confirmation year (the year in which a user joined Facebook), the higher the long term retention.**

a. Observations

i. For example, for somebody who was born in 1953, this user's long term retention is markedly higher the earlier the user joined Facebook…

ii. (LTR closer to 100%) than the latter confirmation years despite the early confirmation years despite having users who are of much greater Facebook age.

b. Implications

i. Facebook's long term retention deteriorates with each passing year. This is unsustainable, and we need to find out why it deteriorates.

1. Discovering why will lead to different product changes. Some hypotheses:

a. Competition: Other social services are better and better

b. Nature of underlying population: The people who join Facebook later on are more marginal. Perhaps they have worse connectivity? Or perhaps they have negative sentiment toward Facebook?

c. Changes in Facebook's product/the code and pixels: Our product has changed in a way that does not encourage strong long term retention.

d. Changes in the friend graph on Facebook: People who joined earlier on have a different graph than those who joined later on. The graph of those who joined earlier on lends itself to better long term retention .

e. Changes in accounts created: There could be more secondary accounts (soon abandoned) or more fake accounts.

2. **ACTION:** People who joined Facebook earlier on are fundamentally different from the people who join today. What makes them or their situation different? Can we replicate what we had before so that people who join Facebook today have higher retention?

3. **In a recent confirmation year, 2016, the higher the population penetration by year of birth, the lower the long term retention.**

a. Observations

i. **This is counterintuitive.** Conventional wisdom is that as population penetration increases, retention increases.

b. Perhaps making the world more open and connected, beyond a certain point, is not that valuable

i. Skip to the section "A Fascinating Relationship Between Long Term Retention and Population Penetration" towards the end to find analysis on this.

c. Implications

i. If the relationship is a causal one, then it suggests that we do not understand how social media products work. Even if it is only correlational, we should seek to understand the drivers.

PX-01325, Page 16 of 35

1. Is it:
    a. Users who are signing up for Facebook now are extremely marginal users?
        i. Why are they marginal? Don't like Facebook? Primarily use Snapchat? Have better uses of their time?
    b. Users who are signing up for Facebook now are already existing Facebook users but are signing up secondary accounts?
        i. Why doesn't our SUMA model detect this?
    c. Users who are signing up are not real people
        i. Why doesn't our fake accounts team detect this?

ii. **ACTION:** This finding appears to challenge our understanding of how social media is supposed to work. It behooves us as Facebook to understand what is happening.

Implications

1. There seems to be something about tweens (age 10-12) that makes them special.
    a. Why is their long term retention multiple times larger than adults?
        i. Perhaps Facebook as a whole should invest more heavily in bringing in larger volumes of tweens
    b. **ACTION:** Track down the former tweens who have incredibly high retention on Facebook and find out why they joined Facebook when they did and what keeps them on Facebook till this day. Can we replicate some of those conditions so that we have a greater number of highly retentive tweens in our pipeline?
2. Facebook's long term retention deteriorates with each passing year.
    a. Why? Some hypotheses:
        i. Competition: Other social services are better and better
        ii. Nature of underlying population: The people who join Facebook later on are more marginal. Perhaps they have worse connectivity? Or perhaps they have negative sentiment toward Facebook?
        iii. Changes in Facebook's product/the code and pixels: Our product has changed in a way that does not encourage strong long term retention
        iv. Changes in the friend graph on Facebook: People who joined earlier on have a different graph than those who joined later on. The graph of those who joined earlier on lends itself to better long term retention
    b. Should we invest more in keeping the people that we have or acquiring new people?
    c. **ACTION:** People who joined Facebook earlier on are fundamentally different from the people who join today. What makes them or their situation different? Can we replicate what we had before so that people who join Facebook today have higher retention?
3. There is a strong negative relationship between population penetration by year of birth and long term retention
    a. Why? Is it:
        i. Users who are signing up for Facebook now are extremely marginal users?
            1. Why are they marginal? Don't like Facebook? Primarily use Snapchat? Have better uses of their time?
        ii. Users who are signing up for Facebook now are already existing Facebook users but are signing up secondary accounts?
            1. Why doesn't our SUMA model detect this?

PX-01325, Page 17 of 35

iii. Users who are signing up are not real people

1. Why doesn't our fake accounts team detect this?

b. **ACTION:** This finding appears to challenge our understanding of how social media is supposed to work. It behooves us as Facebook to understand what is happening.

Santosh and I We initially looked into long term retention in order to calculate the lifetime value for teens. What we found was a dataset that was interesting in of itself.

We define **long term retention** as the continued use of Facebook over very long periods of time, hundreds or even thousands of days. For the purposes of this note, we looked the percent of users who were MAP_28 after n number of days. It is possible to use other xAP metrics, such as DAP, or even other metrics, such as original broadcast sharing posts (OBPS), if properly defined for the retention context.

- On the X Axis is the year the user was born, the date of birth (DOB) year
- On the Y Axis is the Facebook age—the number of days after the user joins Facebook
- On the "Z Axis" (the color of the data point) is the long term retention percent
  - 0%, the lower bound, is red; low retention is bad
  - 50%, the midpoint, is yellow
  - 100%, the upper bound, is green; high retention is good
  - **Green** = **"good"** and red = "bad"

(Caption Image: Chart 2: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for All Birth Years and All Confirmed Years)

Note on reading Chart 2:

- The tops charts are what you saw in Chart 1 but instead of showing 1 confirmation year, we see all confirmation years
  - The # divide by cohort size gives the topmost color chart line for retention %
  - Users from early confirmation years (2004 to 2009/2010) 2009)
  - See Further discussion in both the TL;DR and the 1981-2006 "zoom-in" section above for further discussion below
  - See Chart 3 below for an alternative view
- The least retentive cohorts (the most red) are:
  - **Late teens/young adults**—we see this pattern emerging as early as 2007/2008!
    - **Implication**: Social product for those joining as young adults are unlikely to be retentive
    - Another hypothesis, which we go into more detail at the "A Fascinating Relationship Between Long Term Retention and Population Penetration" section toward the bottom, is that retention goes down as we population penetration increases

Chart 3: MAP Retention for Select YOB's in Confirmation Year 2014. The people who join as tweens do much better than people who join as teens or young adults. Note that the ages indicated are the age at which they joined Facebook. Lines do not start at exactly 100% due to deactivated accounts.

- MAP_28 long term retention tends to be stable after a few hundred days
- Long term retention can be very high despite small cohort sizes

PX-01325, Page 18 of 35

    - This is the case with the early cohorts from 2004-2005 as well as the small cohorts for the very old and very young in almost all the confirmed years
- Facebook's largest influx of users (cohort size) occurred in 2009, 2016, 2010, 2015, 2011
    - However, the users from earlier confirmed years are much more retentive
    - Therefore, the biggest confirmations years by proportion of current MAP are 2009 (22%), 2010 (14%), and 2008 (11%)
    - For the large number of confirmed users in the more recent years, we wonder if a significant portion of those are spam/fake accounts
- Difference in the shape of the curve between the Cohort Size (middle charts) versus Approx MAP_28 on Latest Day (bottom charts) shows you which cohorts are relatively more retentive

(Chart 4: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for Birth Years 1981 to 2006 and All Confirmed Years)
(Chart 5: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for Birth Years 1961 to 1980 and All Confirmed Years)
(Chart 5: Long Term Retention, Cohort Size, and Approximate MAP_28 on Latest Day for Birth Years 1940 to 1960 and All Confirmed Years)

- About 17% of MAP is represented by birth years 1940 to 1960
- Separate from those who join as tweens (10-12 y/o), those who join in old age have the highest long term retention
    - We see this in the yellow-green coloring for those in this age group
- **People who join Facebook at older ages do better than people who join as young adults**
    - Conventional wisdom is that Facebook does poorly with older people—however, our long term retention data seems to say that older people who join Facebook, whether in the earlier years (e.g., 2009) or even more recently (e.g., 2016), tend to continue using Facebook

The MAP_28 population penetration (MAP_28 / total population for that year of birth) is eerily an inverse of the long term retention chart in recent confirmation years. This leads to an unexpected result. The highest higher the MAP_28 population penetration, the lower the long term retention of new cohorts. We would expect the opposite effect. A simple linear… population penetration as the independent variable and LTR as the dependent variable) has an **r-squared = 0.84**. For confirmation year 2016, **the relationships relationship is that the higher the population penetration, the lower the long term retention.** This runs counter to our expectation that as population penetration… like the countervailing forces such as differences in the underlying population (e.g., type of users who join later), fake accounts, and/or secondary accounts, and/or other factors may be more important than the increased value from the greater % of people from your year of birth being on the service. Perhaps making the world more open and connected, beyond a certain point, is not that valuable.

(Chart 6: Population Penetration by Year of Birth as of End of 2016)

(Chart 7: Relationship between Population Penetration and Long Term Retention. r-squared = 0.84)

Because of the difficulty in assessing long term retention, nobody has done it until now. With this dataset, we can understand Facebook and other family apps more deeply than before.
Santosh and I We are planning to look at long term retention outside the United States and also for Instagram. We also hope this work spurs others to look into long term retention themselves as the breadth of this data across time and products lends itself to further analysis.
Add thank Thank you section to those who shared feedback on earlier versions of this work: Danny Ferrante, Mike Winters, Samarth Mohanty, Shrivats Iyer, Steve DeLucia, Tom Cunningham.

Thanks to our respective managers Ali Imam and Keval Patel who gave us the time to pursue a novel type of metric and analysis.

PX-01325, Page 19 of 35

[Not part of the note]
People who might be interested in this:

- fake accounts team
- SUMA team
- core growth
- teens team members
- Nadine (interest in helping out older people)
- segmentation peeps (earl and alan)
- growth marketing peeps
- Messenger Kids

[removed from note to make shorter]

- The example above is the 2013 Confirm Year color chart
  - As evidenced by the dark green band on the right, young teens/tweens have the highest long term retention
  - The yellow-red blob in the middle shows that older teens and younger adults tend to have the lowest long term retention
  - The yellow-green band on the left shows that much older adults tend to have reasonably good long term retention (in between the tweens and young adults)
  - We also see columns that stand out of place for 1940, 1950, 1960, 1970, 1980, 1990, and 2000. This could be a reflection of human psychology when creating "fake" accounts per Spencer Burns

---

Robert Chen (PM) *Dec 5, 2018 at 2:16 am*

yes

Robert Chen (PM) *Dec 5, 2018 at 2:15 am*

yes, will do

Robert Chen (PM) *Dec 5, 2018 at 2:15 am*

probably?

Robert Chen (PM) *Dec 5, 2018 at 2:15 am*

PX-01325, Page 20 of 35

will mark the charts/add charts

Robert Chen (PM) added Alex Ilyin to the thread *Dec 4, 2018 at 11:53 pm*

Robert Chen (PM) added you to a document *Dec 4, 2018 at 11:48 pm*

Robert Chen (PM) *Dec 4, 2018 at 9:11 pm*

Mark where people are tweens. Mark where people are teens.

Robert Chen (PM) edited *Dec 4, 2018 at 9:52 pm*
Note on Chart 2:

- The middle charts showing "Cohort Size" indicate the number of users on Day 0
  - The number of users that are MAP on Day X divided by the number of users on Day 0 = the retention %
- The bottoms charts show the "Approximate MAP_28 on Latest Day"
  - This shows the number of MAP_28 for each confirmation year x YOB on the latest DS

Samarth Mohanty *Dec 4, 2018 at 7:16 pm*

this is also not clear from the charts

Samarth Mohanty *Dec 4, 2018 at 7:15 pm*

if the idea is just to make the point tween are retentive and late teen are not ... can we just plot the expected retention based on if you are a tween / late teen to highlight this point

---

Samarth Mohanty *Dec 4, 2018 at 7:12 pm*

could we mark this in the picture ,,, this is not superclear (at least for me) from the chart above

---

Samarth Mohanty *Dec 4, 2018 at 6:59 pm*

Can this be proved / disproved by 1) data usage or bandwidth correlations by join date

---

PX-01325, Page 21 of 35

Robert Chen (PM) *Dec 4, 2018 at 6:56 pm*

could you elaborate a bit more on what you think should change regarding what's written here?

Samarth Mohanty *Dec 4, 2018 at 6:55 pm*

this is includes 2 things in it , 1) the mere existence of other options , and 2) some of those options solve a niche use cases

---

Samarth Mohanty edited *Dec 4, 2018 at 7:16 pm*

Robert Chen (PM) edited *Dec 4, 2018 at 6:08 am*

  - As evidenced by the dark green band on the right, young teens/tweens have the highest long term retention
  - The green yellow-red blob in the middle shows that older teens and younger adults tend to have the lowest long term retention

- The least retentive cohorts (the most green) red) are:
- Operationalize confirmation year into Deltoid
  - The differences in usage are starkly different by confirmation year so it makes sense to monitor metrics by confirmation year

---

Samarth Mohanty opened your conversation with Santosh Kumar Panda, Hao Wu, Tom Cunningham, Robert Chen (PM), Sharmistha Singh and Alex Ilyin *Dec 4, 2018 at 2:02 am*

Robert Chen (PM) added Samarth Mohanty to the thread *Dec 4, 2018 at 1:58 am*

Robert Chen (PM) added you to a document *Dec 4, 2018 at 1:53 am*

Tom Cunningham opened your conversation with Santosh Kumar Panda, Hao Wu, Samarth Mohanty, Robert Chen (PM), Sharmistha Singh and Alex Ilyin *Dec 4, 2018 at 12:51 am*

Robert Chen (PM) added Tom Cunningham to the thread *Dec 4, 2018 at 12:56 am*

Robert Chen (PM) added you to a document *Dec 4, 2018 at 12:49 am*

Robert Chen (PM) *Dec 4, 2018 at 12:33 am*

PX-01325, Page 22 of 35

Hmm do you think there would be a way to assess this now?

Robert Chen (PM) *Dec 4, 2018 at 12:32 am*

How would you normalize against cohort size?

Robert Chen (PM) *Dec 4, 2018 at 12:32 am*

In general, across all YOB cohorts it declines

Robert Chen (PM) *Dec 4, 2018 at 12:32 am*

nah this is for people born in 1940 to 2000s

Santosh Kumar Panda *Dec 4, 2018 at 12:23 am*

I will double check this once again 2mrw

Santosh Kumar Panda *Dec 4, 2018 at 12:16 am*

MSI is the most notable example. The pivot away from viral videos I thought was a big contributor.

Santosh Kumar Panda *Dec 4, 2018 at 12:14 am*

Also have you tried to normalize this against the cohort size?

Santosh Kumar Panda *Dec 4, 2018 at 12:14 am*

Isn't this for Teens only?

Santosh Kumar Panda *Dec 4, 2018 at 12:07 am*

Is this Teens only?

Santosh Kumar Panda edited *Dec 4, 2018 at 12:22 am*

PX-01325, Page 23 of 35

Robert Chen (PM) edited *Dec 4, 2018 at 2:04 am*

(cover photo below)

1. **Tweens (approximate age 10-12) have the highest long term retention on an absolute basis.**
    a. As an example, for people who joined Facebook… is similar in magnitude across multiple confirmation years.
    b. You can also see this very clearly in the Chart 2 below. There is a red green band (indicating high retention) on the right in every confirmation year. The exceptions are in the very early days of Facebook (2004 and 2005)…
        i. 1,800 (42% at Day 700)
    c. You can also see this clearly in the Chart 2 below. The early confirmation years are much redder greener (LTR closer to 100%) than the latter confirmation years despite the early confirmation years despite having users who are of much greater Facebook age.
2. **In a recent confirmation year, 2016, the higher the population penetration by year of birth, the lower the long term retention.**
    a. This is counterintuitive. Conventional wisdom is that as population penetration increases, retention increases
    b. Perhaps making the world more open and connected, beyond a certain point, is not that valuable

1. There seems to be something about tweens (age 10-12) that makes them special.
    a. Why is their long term retention multiple times larger than…
    b. makes them or their situation different? Can we replicate what we had before so that people who join Facebook today have higher retention?
2. There is a strong negative relationship between population penetration by year of birth and long term retention
    a. Why? Is it:
        i. Users who are signing up for Facebook now are extremely marginal users?
            1. Why are they marginal? Don't like Facebook? Primarily use Snapchat? Have better uses of their time?
        ii. Users who are signing up for Facebook now are already existing Facebook users but are signing up secondary accounts?
            1. Why doesn't our SUMA model detect this?
        iii. Users who are signing up are not real people
            1. Why doesn't our fake accounts team detect this?
    b. **ACTION:** This finding appears to challenge our understanding of how social media is supposed to work. It behooves us as Facebook to understand what is happening.

- 0%, the lower bound, is green red
- 50%, the midpoint, is yellow
- 100%, the upper bound, is red green
- **Red Green** = **"good"** and green red = "bad"

- The example below is the 2013 Confirm Year color chart
    - As evidenced by the red green band on the right, young teens/tweens have the highest long term retention
    - The green blob in the middle shows that older teens and younger adults tend to have the lowest long term retention

PX-01325, Page 24 of 35

    - The yellow yellow-green band on the left shows that much older adults tend to have reasonably good long term retention (in between the tweens and young adults)

- The most retentive cohorts (the most red) green) are:
    - For example, if you start "red "green," you tend to remain red green
    - We see this in the yellow-green coloring for those in this age group
- **People who join Facebook at older ages do better than people who join as young adults**
    - Conventional wisdom is that Facebook does poorly with older people—however, our long term retention data seems to say that older people who join Facebook, whether in the earlier years (e.g., 2009) or even more recently (e.g., 2016), tend to continue using Facebook

The MAP_28 population penetration (MAP_28 / total population for that year of birth) is eerily an inverse of the long term retention chart in recent…. A simple linear regression (with population penetration as the independent variable and LTR as the dependent variable) has an r-squared = 0.84. For confirmation year 2016, **the relationships is that the higher the population penetration, the lower the long term retention.** This runs counter to our expectation that as population penetration increases, the value of the social service goes up because more people you know are on the service. It seems like the countervailing forces such as differences in the underlying population (e.g., type of users who join later), fake accounts, and/or secondary accounts may be more important than the increased value from the greater % of people from your year of birth being on the service. Perhaps making the world more open and connected, beyond a certain point, is not that valuable.

Recommendations and Concluding Thoughts

Establishing long term retention as a metric for Facebook to observe and analyze is a major objective of this work. Additionally, we see that confirmation year and year-of-birth/age are important dimensions through which to understand how different people are using Facebook over time.

Product Recommendations

Who precisely are we building for?

- Given the extremely high long term retention of tweens, what can we do to safely and appropriately onboard high volumes of tweens onto Facebook?
    - And can we push those growth while maintaining high retention?
- When we launch new products, do we intend for these new products for high-tenure or low-tenure people?
    - Because the LTR for high-tenure people continues to be high, we could have greater focus on building for low-tenure people
    - However, we would want to monitor our high-tenured people in case their retention drops
- Can we replicate the earlier days of Facebook and recreate the very high long term retention that Facebook used to have?

Research and Analytics Recommendations

Let's understand long term retention and these Confirmation Year x YOB cohorts much more deeply.

PX-01325, Page 25 of 35

- Who are the users in the different confirmation year X YOB cohorts and what makes them the same or different?
  - By confirmation year:
    - If they are the same as the users today, is it the environment (e.g., competition) or product that has changed?
    - If early confirmation year users are different from those joining today, in what ways are they different and what can we build for them?
  - By age:
    - What is it about people who join as tweens make them have such high long term retention?
    - Why do we do the poorest with those who join as young adults?
    - Why do we do better with people who join when they are old compared to young adults?

Because of the difficulty in assessing long term retention, nobody has done it until now. With this dataset, we can understand Facebook more deeply than before.

Santosh could you write up a short methodology section?
The long term retention dataset provides statistics and analyses on retention of users on facebook over time. The primary retention metric for the current dataset is overall FB xAP and it is limited to the users in US only. The dataset is populated all the way from Jan 2012 till March 2018. The dataset provides a wide variety of dimensions such as the Age, Year of Birth and Registration Date among other things. The dataset also tracks the Facebook Age as well as the age of the user though time which are aimed at providing generational insights into user retention.

Data Derivations

Almost all of the data including most of the dimensions and all of the xAP metrics are derived from dim_all_users. It is one of the most trusted and well monitored datasets which is critical for data quality. Age related attributes are derived from Dim Age Affinity (see below). Accounts that are deactivated are included in the cohort until the day of de-activation. Accounts which are marked as SUMA or Fake are excluded from any all calculations.

Age Calculations

The Teen/Non-Teen and the age of a user are determined using Age Affinity. Instead of using the Point Estimates - we estimate the distribution of the sample mean of the user being in a certain age/age-group using the bootstraps. For example to determine if the user is a Teen - we use the following 3 step formula

1. Calculate Age for a bootstrap n being calculated as A(n) = @DATEDIFF(CURRENT_DS, AA_DOB(n))
2. P(n) = {A(n) ≤ 18} - then 1 else 0.
3. P(teen) = (∑P(n))/100

The same is extended onto age specific calculations.

We also froze the Age Affinity partition to the 07/01/2018 partition to avoid inconsistencies in the dataset due to predictions changing over time. Since this analysis is primarily for long term retention - the impact of AA not having accurate predictions for new users is minimal.

SUMA

CONFIDENTIAL
CONFIDENTIAL
META3047MDL-047-00995666
METATNAG-019-01492575

One of the key assumptions in the dataset is that the rate of SUMA has been fairly constant over time. Based on the results of the high level analysis - the current SUMA model only flags about 10% of accounts as SUMA. For example in for YoB 2012 - there were 12M accounts created on FB for US users but only 4M people born in US.

Dataset Details

Primary Summary Table - ytg_cohort_user_metrics_summary_agg
Dash - https://our.intern.facebook.com/intern/unidash/dashboard/youth_growth/cohort_analysis_by_dob/
Sample Query - For people born on 1999 who were between the age of 10 and 30 - how is their retention across time?

```
SELECT
fb_age_in_days,
age_at_registration as dim,
SUM(map_28) as metric_val,
SUM(total_users) AS cohort_size,
(SUM(1.0 * map_28) / SUM(total_users))*100 AS retention_percent
FROM
ytg_cohort_user_metrics_summary_agg
WHERE
ds between '2012-01-04' AND '2017-12-31'
AND age_at_registration BETWEEN 10 AND 30
AND fb_age_in_days IS NOT NULL
AND dob_year = '1999'
AND confirmed_year IS NULL
AND age_in_years IS NULL
AND age_in_years_months IS NULL
GROUP BY
1,
2
ORDER BY
1, 2
```

Thanks

Add thank you section.
Previous presentation on the topic: Youth Leads Review

Robert Chen (PM) added Sharmistha Singh to the thread *Dec 3, 2018 at 7:31 pm*

Robert Chen (PM) added you to a document *Dec 3, 2018 at 7:24 pm*

PX-01325, Page 27 of 35

Robert Chen (PM) edited *Nov 30, 2018 at 10:56 pm*

Potential explanation for differences in A Fascinating Relationship Between Long Term Retention? and Population Penetration

The MAP_28 population penetration (MAP_28 / total population for that year of birth) is eerily an inverse of the long term retention chart in recent confirmation years. This leads to an unexpected result. The highest the MAP_28 population penetration, the lower the long term retention of new cohorts. We would expect the opposite effect. A simple linear regression (with population penetration as the independent variable and LTR as the dependent variable) has an r-squared = 0.84.

Santosh Kumar Panda opened your conversation with Hao Wu, Tom Cunningham, Samarth Mohanty, Robert Chen (PM), Sharmistha Singh and Alex Ilyin *Nov 30, 2018 at 8:04 pm*

Hao Wu opened your conversation with Santosh Kumar Panda, Tom Cunningham, Samarth Mohanty, Robert Chen (PM), Sharmistha Singh and Alex Ilyin *Nov 30, 2018 at 7:51 pm*

Robert Chen (PM) added Hao Wu to the thread *Nov 30, 2018 at 7:46 pm*

Robert Chen (PM) added you to a document *Nov 30, 2018 at 7:41 pm*

Robert Chen (PM) edited *Nov 30, 2018 at 8:09 pm*

1. Tweens (approximate age 10-12) have the highest long term retention on an absolute basis.
    a. As an example, for people who joined Facebook…
2. Facebook), the higher the long term retention.
    a. For example, for somebody who was born in 1953, this user's long term retention is markedly higher the earlier the user joined Facebook
        i. Joined 2009: 76% at Day 1,800
        ii. Joined 2011: 54% at Day…
    b. (LTR closer to 100%) than the latter confirmation years despite the early confirmation years despite having users who are of much greater Facebook age.

- Users joining as young teens used to have high LTR, but that's that has declined drastically in more recent confirmation years
- In recent years, **people who join as young adults (19-29) have the worst long term retention** of all ages
    - However, this group is also fully penetrated on a MAP/population basis. So that leads us to ask:
        - Who is creating new accounts if it seems like almost all people in the age range are already FB MAP?
        - Are they creating secondary accounts?

PX-01325, Page 28 of 35

- How many of these are fake/spam accounts?

- During the first few years of Facebook, this group had very high long term retention
- However, this group's long term retention has dropped considerably
- In contrast to the younger ages, the number of recent account confirmations is considerably lower
- Separate from the tweens (10-12 y/o), those born earlier have the highest long term retention

Potential explanation for differences in Long Term Retention?

Robert Chen (PM) edited *Nov 30, 2018 at 7:34 am*

1. There seems to be something about tweens (age 10-12) that makes them special.
    a. Why is their long term retention multiple times larger than adults?
    b. Perhaps Facebook as a whole should invest more heavily in bringing in larger volumes of tweens
    c. **ACTION:** Track down the former tweens who have incredibly high retention on Facebook and find out why they joined Facebook when they did and what keeps them on Facebook till this day. Can we replicate some of those conditions so that we have a greater number of highly retentive tweens in our pipeline?
2. Facebook's long term retention deteriorates with each passing year.
    a. Why? Some hypotheses:
        i. Competition: Other social services are better and…
        ii. on lends itself to better long term retention
    b. Should we invest more in keeping the people that we have or acquiring new people?
    c. **ACTION:** People who joined Facebook earlier on are fundamentally different from the people who join today. What makes them or their situation different? Can we replicate what we had before so that people who join Facebook today have higher retention?

Robert Chen (PM) edited *Nov 30, 2018 at 1:38 am*

1. Tweens (approximate age 10-12) have the highest long term retention on an absolute basis.
    a. As an example, for people who joined Facebook…
    b. year. The exceptions are in the very early days of Facebook (2004 and 2005) when high schoolers did not have access to Facebook.
    c. Caveat: These findings rely on the Age Affinity model, and while it is the best we have for age, there are biases in the model which may influence the results.
2. The earlier the confirmation year (the year in which a user joined Facebook), the higher the long term retention.
    a. For example,…

CONFIDENTIAL
CONFIDENTIAL

b. (LTR closer to 100%) than the latter confirmation years despite the early confirmation years despite having users who are of much greater Facebook age.

1. There is seems to be something about tweens (age 10-12) that makes them special.
    a. Why is their long term retention multiple times larger than adults?
    b. Perhaps…
        i. on lends itself to better long term retention
    c. Should we invest more in keeping the people that we have or acquiring new people?

What's Next

Santosh and I are planning to look at long term retention outside the United States and also for Instagram. We also hope this work spurs others to look into long term retention themselves as the breadth of this data across time and products lends itself to further analysis.

All birth Years

Then Next we zoom in on particular birth years. We start with 1981 to 2006. Then 1961 to 1980 and finally 1940 to 1960.

1981 to 2006

- About 51% of MAP is represented by birth years 1981 to 2006
- As early as confirmation year 2007 (Facebook was opened up to all users in September 2006), we see that people who joined FB as **tweens have the highest long term retention. This is true from confirmation year 2007 to 2016.**
    - E.g., people born in 1995 and confirmed in 2007, approximately 12 y/o at confirmation, have **92%** MAP retention on Day 3,652!
    - E.g., people born in 2005 and confirmed in 2016, approximately 11 y/o at confirmation, have **87%** MAP retention on Day 365
- Users joining as young teens used to have high LTR but that's declined drastically in more recent confirmation years
    - E.g., people born in 1996 and confirmed in 2010 (approx 14 y/o at confirmation), have 58% MAP retention at Day 2,000
    - But for people born in 1997 and confirmed in 2011 (approx 14 y/o at confirmation) have dropped to 41% at Day 2,000. Even on the same date (Day 1,635), for those born in 1997 and confirmed in 2011, their MAP retention was 42%
    - And even more recently, for people born in 2002 and confirmed in 2016 (approx 14 y/o at confirmation) have dropped to 40% as early as Day 365 and projected to drop to around 25% by Day 2,000

1961 to 1980

- About 32% of MAP is represented by birth years 1961 to 1980
-

1940 to 1960

- About 17% of MAP is represented by birth years 1940 to 1960

CONFIDENTIAL
CONFIDENTIAL

Robert Chen (PM) edited *Nov 29, 2018 at 8:38 pm*

(Caption Image: Chart 1: Sample Chart for Confirmed Year 2013)
We will start by looking at the big picture, birth years 1940 to 2006 and confirmed years 2004 to 2016.

(Caption Image: Chart 2: All Birth Years and All Confirmed Years)
Start with overall

We time travel At a high level, we observe the following:

- The most retentive cohorts (the most red) are:
  - **Tweens (10-12 years old)**
  - Users from early confirmation years (2004 to 2009/2010)
  - See TL;DR section above for further discussion
- The least retentive cohorts (the most green) are:
  - **Late teens/young adults**—we see this pattern emerging as early as 2007/2008!
    - **Implication**: Social product for young adults are unlikely to be retentive
  - Over time, the band of less retentive users expands in both directions
    - Younger and younger teens become less and less retentive. Hypothesis here is that people are joining social media earlier and earlier in life and habits are established early on
    - Older and older adults become less and less retentive. Hypothesis here is that most of the young and middle aged adults already joined a social media service earlier on so the people who are joining Facebook in recent years are more "marginal," either because they are already on an alternative service (e.g., Snapchat, Instagram) or because they naturally have less affinity toward social media
- MAP_28 long term retention tends to be stable over time.
  - For example, if you start "red," you tend to remain red
- Long term retention can be very high despite small cohort sizes
  - This is the case with the early cohorts from 2004-2005 as well as the small cohorts for the very old and very young in almost all the confirmed years
- Facebook's largest influx of users (cohort size) occurred in 2009, 2016, 2010, 2015, 2011
  - However, the users from earlier confirmed years are much more retentive
  - Therefore, the biggest confirmations years by proportion of current MAP are 2009 (22%), 2010 (14%), and 2008 (11%)
  - For the large number of confirmed users in the more recent years, we wonder if a significant portion of those are spam/fake accounts
- Difference in the shape of the curve between the Cohort Size (middle charts) versus Approx MAP_28 on Latest Day (bottom charts) shows you which cohorts are relatively more retentive
- The biggest confirmations years by proportion of current MAP are 2009 (22%), 2010 (14%), and 2008 (11%)

PX-01325, Page 31 of 35

Methodology

Santosh could you write up a short methodology section?

Appendix

Robert Chen (PM) edited *Nov 29, 2018 at 6:02 am*

The Big Kahuna: Long Term Retention in the US since 2004 by Year of Birth

Start with overall
Then zoom in on particular years

[Inset "The Big Kahuna" here]

- Difference in the shape of the curve between the Cohort Size (middle charts) versus Approx MAP_28 on Latest Day (bottom charts) shows you which cohorts are relatively more retentive
- The biggest confirmations years by proportion of current MAP are 2009 (22%), 2010 (14%), and 2008 (11%)

Robert Chen (PM) *Nov 28, 2018 at 1:28 am*

Replace image with "The Big Kahuna"

Robert Chen (PM) renamed The Young Ones are the Best Ones: Long Term Retention in the United States to Long Term Retention: The Young Ones are the Best Ones and Other Learnings *Nov 27, 2018 at 10:31 pm*

Robert Chen (PM) edited *Nov 28, 2018 at 2:08 am*

PX-01325, Page 32 of 35

Long Term Retention: The Young Ones are the Best Ones - But Oldies are Decent Too and Other Learnings

TL;DR

Observations

We observe very stark differences in **long term retention** in the United States depending on **(1) Age at which you join Facebook** and **(2) Year in which you join Facebook**.

1. Tweens (approximate age 10-12) have the highest long term retention on an absolute basis.
    a. As an example, for people who joined Facebook in 2016, 12 year-olds (born in 2004) have long term retention MAP_28 (LTR) of 72% at an FB age of 365. 32 year-olds (born in 1984) have long term retention of 25% at the same FB age. This is a **3x difference** for a 12 y/o versus a 32 y/o! The difference is similar in magnitude across multiple confirmation years.
    b. You can also see this very clearly in the Chart 2 below. There is a red band (indicating high retention) on the right in every confirmation year. The exceptions are in the very early days of Facebook (2004 and 2005) when high schoolers did not have access to Facebook.
2. The earlier the confirmation year (the year in which a user joined Facebook), the higher the long term retention.
    a. For example, for somebody who was born in 1953, this user's long term retention markedly higher the earlier the user joined Facebook
        i. Joined 2009: 76% at Day 1,800
        ii. Joined 2011: 54% at Day 1,800
        iii. Joined 2013: Projected ~47% at Day 1,800 (48% at Day 1,400)
        iv. Joined 2015: Projected ~35% at Day 1,800 (42% at Day 700)
    b. You can also see this clearly in the Chart 2 below. The early confirmation years are much redder (LTR closer to 100%) than the latter confirmation years despite the early confirmation years despite having users who are of much greater Facebook age.

Implications

1. There is something about tweens (age 10-12) that makes them special.
    a. Why is their long term retention multiple times larger than adults?
    b. Perhaps Facebook as a whole should invest more heavily in bringing in larger volumes of tweens
2. Facebook's long term retention deteriorates with each passing year.
    a. Why? Some hypotheses:
        i. Competition: Other social services are better and better

PX-01325, Page 33 of 35

ii. Nature of underlying population: The people who join Facebook later on are more marginal. Perhaps they have worse connectivity? Or perhaps they have negative sentiment toward Facebook?

iii. Changes in Facebook's product/the code and pixels: Our product has changed in a way that does not encourage strong long term retention

iv. Changes in the friend graph on Facebook: People who joined earlier on have a different graph than those who joined later on. The graph of those who joined earlier on lends itself to better long term retention

b. Should we invest more in keeping the people that we have or acquiring new people?

Context and Definitions

Santosh and I initially looked into long term retention in order to calculate the lifetime value for teens. What we found was a dataset that was interesting in of itself.

We define **long term retention** as the continued use of Facebook over very long periods of time, hundreds or even thousands of days. For the purposes of this note, we looked the percent of users who were MAP_28 after n number of days. It is possible to use other xAP metrics, such as DAP, or even other metrics, such as original broadcast sharing posts (OBPS), if properly defined for the retention context.

Quick Primer on Charts

These LTR color charts are similar in format to triangle charts. You can skim this section if familiar.

The color chart represents 3 dimensions of data.

- On the X Axis is the year the user was born, the date of birth year
- On the Y Axis is the Facebook age—the number of days after the user joins Facebook
- On the "Z Axis" (the color of the data point) is the long term retention percent
  - 0%, the lower bound, is green
  - 50%, the midpoint, is yellow
  - 100%, the upper bound, is red
  - Red = "good" and green = "bad"
- The example below is the 2013 Confirm Year color chart
  - As evidenced by the red band on the right, young teens/tweens have the highest long term retention
  - The green blob in the middle shows that older teens and younger adults tend to have the lowest long term retention
  - The yellow band on the left shows that much older adults tend to have reasonably good long term retention (in between the tweens and young adults)
  - We also see columns that stand out of place for 1940, 1950, 1960, 1970, 1980, 1990, and 2000. This could be a reflection of human psychology when creating "fake" accounts per Spencer Burns
- In some of the color charts, the data is blank for the earlier Facebook ages. This is because our current dataset does not go that far back in time. E.g., we didn't have dim_all_users in 2005.

CONFIDENTIAL
CONFIDENTIAL

The Big Kahuna: Long Term Retention in the US since 2004 by Year of Birth

[Inset "The Big Kahuna" here]

We time travel

- MAP_28 long term retention tends to be stable over time
- Long term retention can be very high despite small cohort sizes
-

Robert Chen (PM) renamed The Young Ones are the Best Ones - But Oldies are Decent Too to The Young Ones are the Best Ones: Long Term Retention in the United States *Nov 27, 2018 at 7:41 pm*

Robert Chen (PM) added Santosh Kumar Panda to the thread *Nov 21, 2018 at 10:21 pm*

Robert Chen (PM) added you to a document *Nov 21, 2018 at 10:16 pm*

Robert Chen (PM) edited *Nov 21, 2018 at 10:16 pm*

The Young Ones are the Best Ones - But Oldies are Decent Too

By Santosh Kumar Panda and Robert Chen

Robert Chen (PM) created the document *Nov 21, 2018 at 10:21 pm*

PX-01325, Page 35 of 35