

## Six Analytics Case Studies

### Q1: High School Data

**Intro:** 4 years ago, Facebook was considering building a product around high schools. At the time, we had 30M (global) users who had filled in a text field (no associated pages) with the supposed name of their current high school.

**Question:** How much of this data is real? We'd like to get an approximation with only Facebook data, so no trying to put together a listing of public high schools.

### High School Data Answer Guide

- Candidate should be able to decide what “real” means in a reasonable way on their own (I play fairly mum if asked and basically repeat why we care in product context — i.e. if this data is real we have a great knowledge base to bootstrap product off of, if fake we don't).
- The usual first step is to appeal to a list of schools. There are two problems with this approach you should raise:
  - No such list exists even really in the US, and certainly not internationally
  - We would rather use only FB data and need an answer quickly; we have no way to validate external lists.
- There are two paths people will usually go down here: 1) try to determine which schools are real, or 2) directly try to determine which associations are real. Both of these paths are reasonable. 1) is an easier problem, so if a candidate is spinning their wheels on 2), feel free to quickly redirect them to 1).
- One key moment in this interview usually involves looking at the distribution of schools by frequency. Get the candidate to draw a guess at the distribution and really drill into specifics here (e.g. right hand tail, how high does it go? what numbers do you expect on the x axis at the peak(s?)) and absolutely ding them for clarity of presentation, e.g. terrible axis labeling).
- Here's a list of some key features / insights / special cases. If they don't come up with these, try to nudge them towards them and get them to analyze / operationalize.
  - Friends (you're likely to have the same school as your friends)
  - Location (real schools will have the vast majority of attendees in same city)
  - Spelling / typos / standardization (great candidates will clean the data before doing anything else, but there are some tough things here, e.g. Montgomery vs Montgomery High vs Montgomery High School)
  - Hogwarts (a very large school that is fake) — how do we identify lower cardinality high schools?
  - Jefferson High School (a duplicated school name that will appear to have “too many” students) — how do we distinguish from Hogwarts?
- Try to leave 2-3 minutes to “pull it all together” and have the candidate operationalize insights into a number. If you don't feel you have enough to do this with at this time point, it is very likely a no-hire.
- Most candidates will come up with some subset of reasonable signals. Make sure they operationalize. For instance “my friends likely go to the same school” — how would you do this in practice? Raw n cutoff? Most common school? Proportions? These approaches will all be bad in some cases (respectively: user has few friends, user goes to small private school and most of friends go to large public school, user has 5000 friends and schools are smallish). Candidate should be able to have a pleasant discussion with you as you challenge this.

•

## Q2: Facebook Reactions

**Intro:** As you may know, one of our most commonly requested features is a Dislike button. We're not building that for a few reasons, but we're thinking of extending the like button to other lightweight reactions. The way this would work is you would long-press on Like, and it would bring up a menu with a few other reactions.

**Question 1:** This would have a big engineering cost. How would you figure out from existing data whether or not we should build it, and if so, which reactions we should build?

**Question 2:** Testing feed ranking is relatively easy: we can just roll out a new feed ranking to 5% of users. Testing this will be harder since you'd have to figure out how it looks if I try to react to your post and I'm in the 5% experiment and you're not. How might we go about testing this feature, given this problem?

**Question 3:** How would you assess the launch as far as whether we're providing value to our users?

## Facebook Reactions Answer Guide

- **Part 1: This would have a big engineering cost. How would you figure out from existing data whether or not we should build it, and if so, which reactions we should build?**
- Some answers:
  - Comment text is the logical first place most people start — press for details if the candidate doesn't volunteer them. Some good ideas:
    - Sentiment analysis (press for details, not implementing sentiment analysis, but talking about which sentiments to look for, how to group similar things together, etc.)
    - Short comments versus long comments (short comments are more lightweight reactions)
    - Comments on posts with no likes (suggesting that a like isn't sufficient)
    - Comments where the commenter only commented once on that post (a reactive comment instead of a conversational one)
  - Stickers (similar for emoji)
    - How do you extract a sentiment from a sticker? (the right answer is actually to pull most common stickers and just look at them and reason, though there are data approaches)
    - How do you “group” stickers together into sentiment categories? (a good answer here is the ones that appear in the same reactive thread should be similar)
  - Long view duration without taking an action
- **Part 2: Testing feed ranking is relatively easy: we can just roll out a new feed ranking to 5% of users. Testing this will be harder since what does this look like if I try to react to your post and I'm in the 5% experiment and you're not?**
- Candidate should be expected to address both how to get data that's informative, and how to design experiment for decent user experience. This is a hard problem and there is no right answer, so expect this to be conversational.
  - Only show if both user and target are in the experiment. This is pretty poor as your 5% experiment is now 1 in 400, and also will confuse users.
  - Do a market test. This is good — New Zealand to the rescue! You'll still want a small holdout there though, or you can compare to trend lines in a similar “paired” country.
  - Run experiment on 5% of posters — weird user experience to be able to react on some posts and not others.
  - Run experiment on 5% of users — converting reactions to likes or comments if the poster doesn't have it. Candidate should be able to talk through pros/cons here (major con is that poster will receive different experience that the user may not expect).
  - Run experiment on 5% of posts — super weird.

- Part 3: **How would you assess the launch as far as whether we're providing value to our users?**
- Good answers would talk about cannibalization of likes and comments and be able to reason (again, no perfect answer) about how many likes a reaction might be worth. Bonus points for thinking about cannibalization of different types of comments. Another good thing is talking about participation rate of any engagement on a post, which is similar, but can be brought back to actually connecting. Thinking about overall Facebook metrics (e.g. time spent) is a good feature here as well. Another bonus point is thinking about how receiving reactions affects posting behavior.
- 3a: **Optional followup: Suppose all this experiment did was change a bunch of likes to reactions. Would this be a success?**
- Answer here is unclear; probably no because we spent a lot of effort building it. You could make an argument for yes though — haven't hurt metrics and users are using the new feature so they must be deriving value from it.
- 3b: **Optional followup: Suppose all this experiment did was change a bunch of comments to reactions. Would this be a success?**
- - Answer here is also unclear; probably no. Best answer is what type of comments were converted? If it's short ones, you just made life easier for people (though if you're not doing it more, it's not clear that's true). You could slice here to see if some people love reactions and some people are confused by the complication (e.g. by age).
- 

### Q3: Fraud Prevention

**Intro:** Facebook has a lot of transactions, some of which are fraudulent. Unfortunately, if a transaction is fraudulent, we don't always learn this right away (it can be any time from immediate to up to six months later).

**Question:** Given the new transactions that happened every month, and the fraudulent transactions that were identified in each month (these can be mapped to the original transactions), what's the best way to assess the effectiveness of our fraud detection system without having to wait six months?

### Fraud Prevention Answer Guide

- One naive answer is to define the loss rate as  $\frac{\text{\$fraud\_amount\_identified\_this\_month}}{\text{\$new\_transaction\_amount\_this\_month}}$ . But since the fraudulent transactions could have happened in previous months, this doesn't really make a lot of sense. This is only true if the transaction volume and fraud rate were flat month over month.
- Give candidate some hints to plot the fraud arrival curve if they haven't done it already - % fraud received vs. number of months elapsed. This would show that for a typical cohort of transactions, % of fraud received (cumulatively) over time (e.g., 10% fraud in the first month, 40% in the first 2 months, 70% in the first 3 months, etc.)
- Then calculate the estimated fraud rate based on the arrival curve -  $\frac{\text{\$estimated\_fraud\_amount\_for\_transactions\_this\_month}}{\text{\$transaction\_amount\_this\_month}}$ . Update the previous months' fraud rate using the newly identified fraudulent transactions.

#### Q4: Mentions

**Intro:** We launched a mobile app called Mentions, which is built specifically for celebrities. The app allows them to quickly share content, read fan posts about themselves, and interact with fans.

**Question:** How do you determine whether the app is successful? What metrics would you use, and what other factors would you consider?

#### Mentions Answer Guide

You should be able to get much of the basic usage stats, and recognize that those stats need some sort of context before we can determine success. e.g. what effect does usage of Mentions have on the fb ecosystem, and how big is the market size for the app?

- Let's start with: is the app being used, and is its usage sustainable? Look at things like DAP, retention, time spent, sessions, posts, likes/comments.
- An important observation is that the second order effects from the app are what make it valuable. The goal is to get more/better content from celebrities for the benefit of everyone else using Facebook.
- With that in mind, we want to zoom in on posting behavior and see things like: how many people they are reaching, and how many feed stories they are generating across Facebook. Main point is that we want to try to measure the benefit these celebrity posts provide to the broader ecosystem on Facebook. If you can tie this back to engagement of the general userbase, or even a step beyond that to revenue, it becomes easier to value Mentions relative to other products we build and determine success.
- We want to make sure their Mentions usage is additive and not just cannibalizing their posts on other interfaces. Can we do an a/b test? Maybe, but might be tricky - would we just block a bunch of super famous people from downloading the app? Doesn't seem like a good idea. Maybe we can have one set of celebrities who we aggressively try to get to install the app, and another that \*can\* install it but we won't push actively. We could also try a correlational approach with a cohort analysis when people install the app. In any case we will probably be dealing with small sample sizes and noisy results. The candidate should be able to identify some of these approaches and explain drawbacks of each. If they bring up the cohort analysis, push for details until you're confident they could actually execute such an analysis.
- Mentions usage numbers need context. Is 1000 DAP a lot or a little? We could come up with a list of the X thousand most important celebrities and see what fraction are on Mentions. Compare vs the fraction on FB overall, on Instagram, Twitter. Also compare posting behavior on each network for this set.

#### Q5: News Feed Health

**Intro:** We have a product called News Feed that is the first thing that you see when you go into Facebook. This product contains many different types of stories. You could see a friends status update about a cup of coffee followed by an shared article about a serious current event.

**Question 1:** If you were the first data scientist on this product, how would you go about understanding the performance of the product?"

**Question 2:** Ads PM wants to double the number of ads in Feed, how do you make sure that is a good/bad decision for the product?

#### News Feed Health Answer Guide

- **[Part A] How do you measure News Feed health?**

- Set the question up so that the person answering will understand the complexity of the News Feed product regardless of how often they use the product. For example, “we have a product called News Feed that is the first thing that you see when you go into Facebook. This product contains many different types of stories. You could see a friends status update about a cup of coffee followed by an shared article about a serious current event. If you were the first data scientist on this product, how would you go about understanding the performance of the product?”
- Superficial answer is an unorganized list of high level metrics.
  - If the person is going in this direction, try to pick a couple metrics and dive into why they are important, how you would normalize them (dap/map/wap/impressions/time spent etc).
  - The other direction you could take here is give an example of when some of the metrics wont tell the right story. For example, if they say clicks is an important metric... *you can give the example of a natural event occurring, so everyone's feed is suddenly filled with articles. Because there is much more to click on, clicks go up. BUT how do we know if we are serving better content vs more clicky content?* Diving into metrics like this with examples makes it easy for the person getting interview to understand the problem and it gets at the complex product problem.
- Better answers separate out different objectives of consumption, feedback, driving traffic, driving reshares.
  - This can be done in many different ways. Some people separate out passive vs active engagement to measure. Some people think about the funnel of engagement. Others think about growth/retention and then engagement. There is not a right and wrong answer here, but look for a logical thought process.
  - The other thing that is really important to look for here, is if they are thinking about the unique challenges that come with evaluating News Feed. They should think about different types of content and how that plays into the health of the product. Even better, they should think about the network and inventory. Here again, if they don't get these right away, prompt them with examples!
- **[Part 2] An ads product manager wants to double the number of ads in Feed--how would you decide if this is a good or bad decision for the product?**
  - Hopefully they jump to an AB test. From here, you want to get at what their product intuition is. I like to tell them there are a bazillion metrics you can look into for an AB test, for this test, what you be your highest pri metrics to look into?
    - A good answer would not only look at this from the Ads perspective, but also from the entire product perspective.
    - Lots of times people want to look at if the ave CTR went up. You can question this a bit if you have time. Since we should you the best ads first, ave CTR should actually decrease.
  - Assume we run the AB test. Put some metrics on the board going in different directions. They are all stat sig. It is best to use metrics that they have come up with and that are hard to tell which is more valuable at a glance. Ask them if you should roll out the test based on those results.
    - A good answer here would look at the scale of the actions, not just the change.
    - A better answer would get a the value of the actions. This is a great thing to push people to explain :)
  - *Variation of this question: let's go back in time to before there were ads in news feed*
    - The world is going mobile, which is not monetized. We are considering adding ads to news feed as a solution, but want to be careful that we understand and

are comfortable with any tradeoffs. You are in charge of recommending whether we ship ads in news feed. What questions would you want to ask and what data would you want to see before you were comfortable making a recommendation?

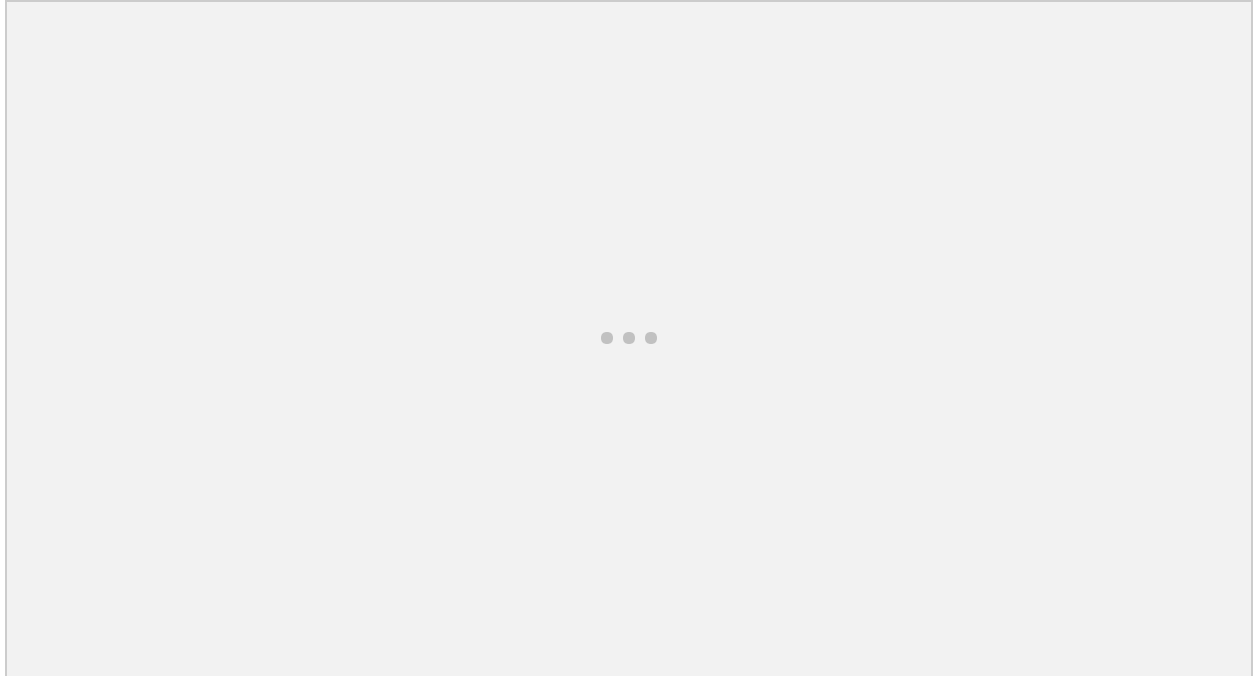
- Candidate should say something about an a/b test, experiment, control group, etc.
- They should talk about engagement metrics such as time spent, likes/comments/shares, feed story views, content production, etc. They might also talk about negative engagement (x-outs/hides, mark-as-spams). It's good if they note that this doesn't have to be a binary tradeoff; you can test different levels of ads in feed and they will have different tradeoffs.
- They should also talk about advertising metrics; specifically, how much did revenue go up? They might talk about seeing how feed ads perform relative to existing right hand column ads (in terms of CTR, CPM, etc). They might ask how many ads we are showing (% of feed).
- Candidate should explicitly call out that we may likely have a trade-off scenerio between engagement and revenue. How can we make a decision? Ideally we can translate the engagement loss into a dollar value that we can compare.
- Scenario: Revenue goes up by 10%, likes down 4%. Should we ship this? Answer is we probably need more information. What are the absolute values - if we make very little revenue today, 10% might not be much. What about other engagement metrics? If comments and time spent are flat, how could we interpret this?

## Q6: Social Graph Equality

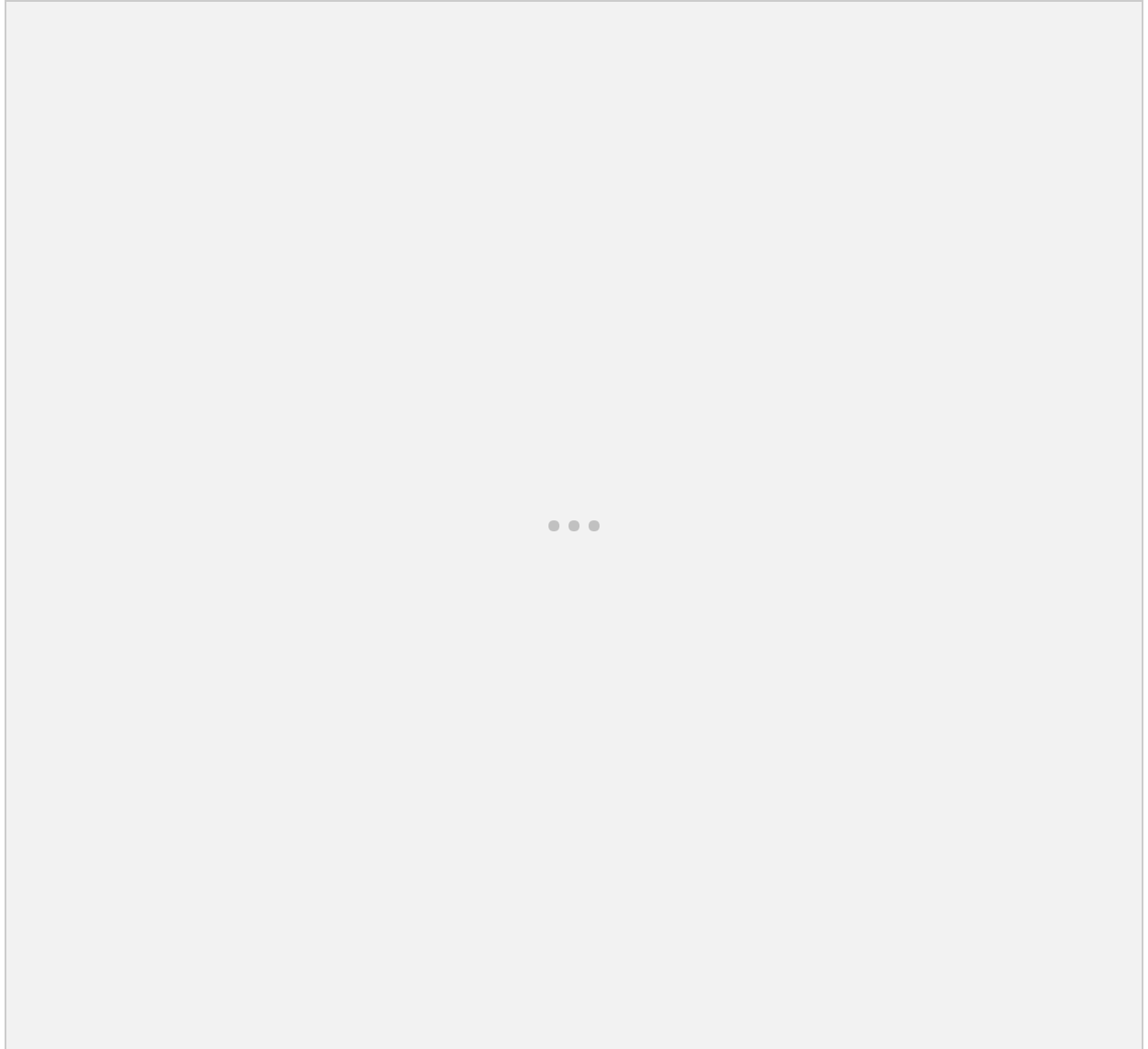
**Question: If we model Facebook and Twitter with users being nodes and friendships being edges, what's the key difference in terms of graph structure?**

### Social Graph Equality

- Comparing Facebook and Twitter (or Instagram), what's the key difference in terms of social graph? If the candidate isn't familiar with the diff, we can easily explain. But the key is to lead the candidates to bi-directional relations (FB) vs directional relations (TWTR/IG). Ignore the "following" feature on FB. Once the candidates get this, I ask what do we expect the graph look differently knowing this key difference. Most people will be able to guess TW/IG graph having more central nodes (e.g. celebrities), and FB graph is more inter-connected. If they don't, I usually draw two simply diagram on white board, and ask them to identify which is FB which is TW/IG. Most people will identify the left one as TWTR and the right one as FB. I ask them why. Some people will say it is the friends vs followers, but I then remind them that I didn't put directions on the graph, but they still identify correctly, so it must be something else. Hopefully, candidates realize it the existence of the central hub (A).



- 
- 
- Now enters the key part of interview. I ask them this “equality” concept is what we want to measure and monitor, and let them choose a metric to quantify this for the whole graph (not per user).
  - Weak candidates will likely say something like average number of connections, or average of 2nd degree connections. If they are smarter, they will use the two diagrams to validate their answer, and find out the two diagrams have the same average number of connections, 2nd degree connections.
  - Better candidates will realize it is not the average that matters, but the distribution of connections. Some people will throw in things like standard deviation of connections, etc. This is on the right track. I then ask them to choose something that is more concrete. Sometimes, I give a hint of using Gini-coefficient to quantify social wealth distribution.
  - Up to here, good candidates should be able to come up with a metric like: percent of total connections accounted by top 10% of people, or something similar.
- I then ask them to draw the full curves of FB and TWTR on white board, and the curves should be something like below.



- - Then I ask them to explain the 45 degree line, what is the meaning, etc. Why is FB or TWTR more skewed, etc. A lot of signals can be obtained here to see a person can interpret charts well, can virtualize graph well, can explain things in plain language well, etc.
  - I then ask candidates to draw the TWTR line from a followers perspective and a followings perspective, and ask which one is more skewed, and why they think so.
  - I then usually ask some product sense questions, for example, what product changes can cause the line go more skewed. What's the pros/cons of a more-skewed or less-skewed network for users, for us, for advertisers, etc. If we want to put effort into making the graph more/less skewed, what can we do, what does that mean for users, for different use cases, etc.
-