Consumer-facing Product Troubleshooting

Monitoring metrics is essential for determining the success of any product.

Once a key metric has been selected - one that accurately represents an overarching goal for the product - moving that metric in the right direction becomes the top priority.

Here we focus on how to determine the primary drivers of key metric changes and how to analyze them.

Metric changes are almost always due to one or more of the following factors:

data quality, product changes, seasonal factors, competition and other external factors, mix shift.

- **Data Quality Issue**

Data quality is often at the heart of <u>sudden</u> metric changes. The most common manifestation of a data quality problem is a sudden and drastic change that cannot otherwise be easily explained. By better understanding the underlying sources of data quality issues, we can develop action plans to address them.

Discrepancies in data are often due to errors in the way the data is recorded. (missing/duplicate/incorrect records)

While "transforming" raw data makes it more usable, it may also introduce errors. (data transformation issues)

Data quality issues often stem from logging issues. Identifying missing data, duplicate logging or logical errors will help diagnose this problem. The process of transforming raw data can also lead to errors. (**Takeaway**)

- **Effect of Product Changes**

There has been an explosion in the development of new products in recent years with the rise of experimental product iteration (aka, A/B testing).

When a product changes, a reaction almost always follows causing a shift in the key metrics. Therefore, it is critical to understand the impact of shipping a product/launching a feature, and ensure that the right outcome is achieved (Causal Inference).

Source: App Annie

Figure 1
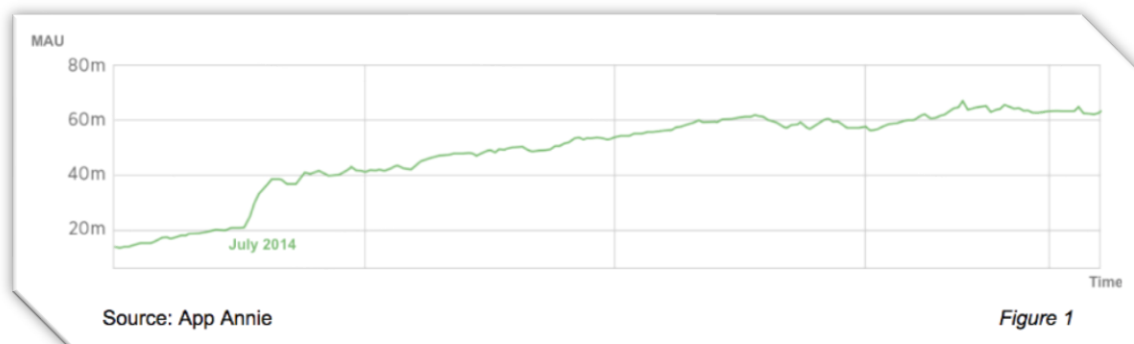
Example: in July 2014, when Facebook began notifying users that they would no longer have the option of messaging within its core mobile app, nearly 20 million iOS users in the US downloaded the separate Messenger app within a one-month period.

Product changes aren't always intentional. Therefore, carefully tracking your key metrics - particularly after a major product change or app store update - can help you detect, understand and limit the damage bugs cause.

The best way to evaluate the impact of a change to your product is A/B testing. In A/B testing, two or more variations are shown at random to users and statistical analysis is used to determine which variation performs better against a given goal.

It is especially important to use A/B testing when <u>a change is relatively small</u>, which is most often the case: A large change, such as the Facebook example above, can sometimes be 'eyeballed'. Usually, though, it is nearly

impossible to effectively detect and measure impact without an A/B test.

A product change - whether a new product/feature rollout/shipping/launching, a changing notification strategy or an unintended bug - will invariably result in a shift in metrics. (**Takeaway**)

- **Seasonal Impact on Product**

We will explore seasonal factors, which often cause of dramatic fluctuations in key metrics. People behave differently based on time and circumstance, and seasonality is one of the most prominent influences that can be analyzed using data. Seasonality is a broad label for periodic changes in user behavior. Examples of seasonality include behavioral changes based on time of day, day of the week, season of the year, etc.

Seasonality is often the root cause of changes in key metrics. To properly analyze any behavioral change, you must understand your product's overall ecosystem. For example, it might be important to know what young people do during the summer, how middle-aged women shop, how people using Android behave compared to those using iOS or who is likely to be an early adopter of your product.
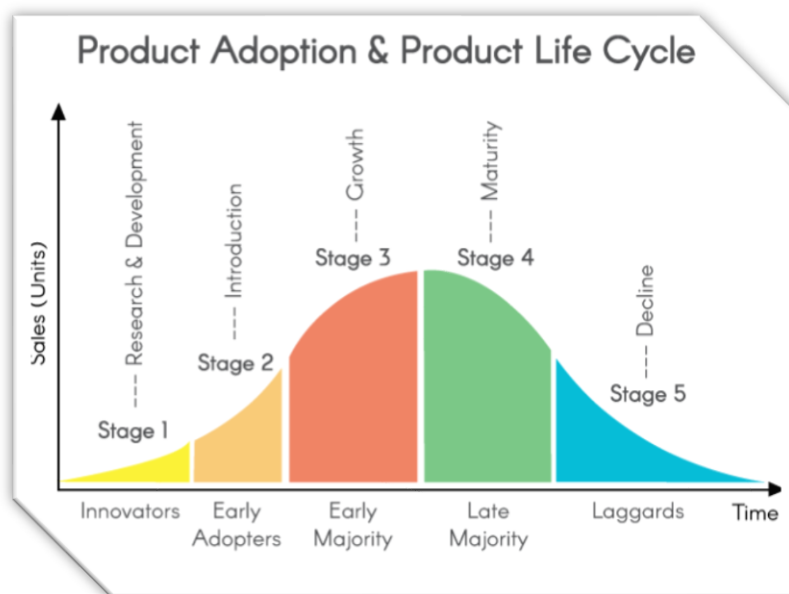
Understanding this ecosystem allows you to ask good questions and develop good hypotheses, which is how effective analysis of behavioral change almost always begins. Once you have a comprehensive list of hypotheses for why the key metric might have changed, you're ready to investigate further.

If you are testing a hypothesis that seasonality is a factor in the change, you may take several approaches:

Week-over-week movement: Most products exhibit a very strong day-of-the-week effect, because people behave differently on weekdays than they do on weekends. To remove this effect, use seven-day (or rolling seven-day) averaging of the metric of interest.

Example shows strong weekly patterns in usage of Yelp, with increases on weekends. Without 7-day averaging, any week-to-week changes would be lost in day-to-day fluctuations.

Note social apps often see similar intraday effects, because most people spend far more time on social apps in the evenings than they do during the day. If the lowest level of granularity you studied for such an app was daily use, you would miss that signal completely.

Day-over-day movement: If you are interested in examining day-to-day changes, try overlaying each day of the week as seen below. This approach will give you more granular information; for example, this shows us that week 3 has higher highs and lower lows compared to other weeks.
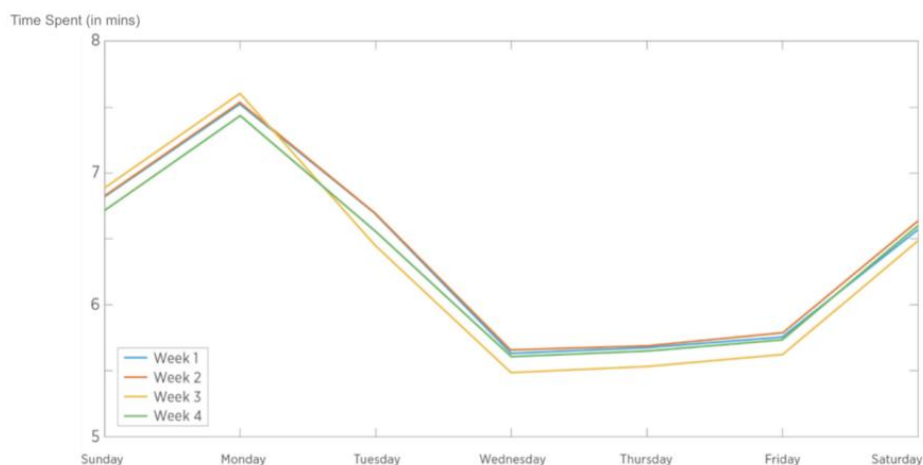

Year-over-year changes: Observing annual trends in your data is one of the most powerful ways to identify the effects of seasonality. This method will illuminate the effects of holidays, the back-to-school season, festivals, cold weather, etc. and help you determine whether those impacts on your key metrics are amplifying or diminishing over time.

For example, the chart below on the left shows a seven-day rolling average of revenue for an e-commerce company that is clearly rising. In some cases, where changes are less obvious, a longer rolling average such as 28 days (see the chart below on the right) can help demonstrate how a metric is **trending over time**.
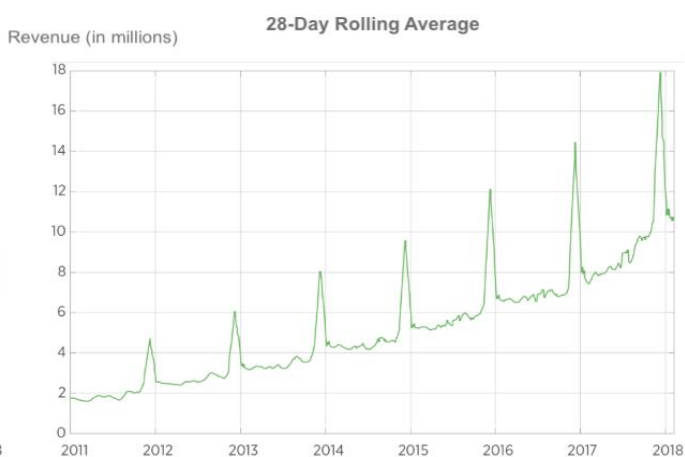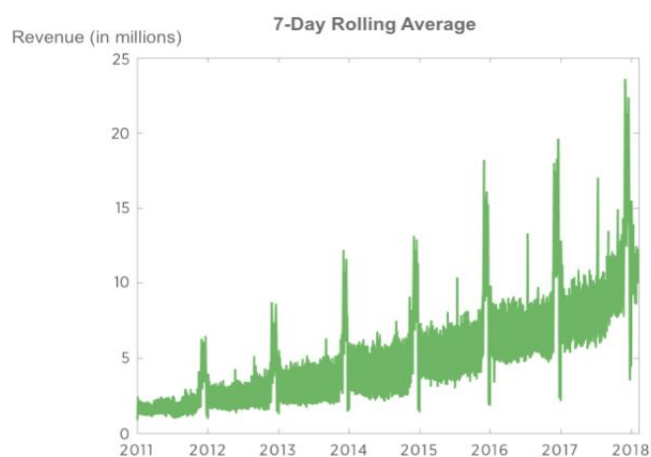

We should look at the year-over-year metric (see below left), which will tell us much more about **seasonal patterns**. The below chart on the right shows the same 28-day rolling average pattern year-over-year; the changes within the year are much more apparent. Understanding the pattern will help you plan for the future.

We can now clearly see that revenue between May and July is influenced by seasonality. Revenue decreases throughout the month of May, begins to sharply increase in early June, then decreases again until around July 4, at which point we see a large bump in both 2016 and 2017, likely due to a sale event.
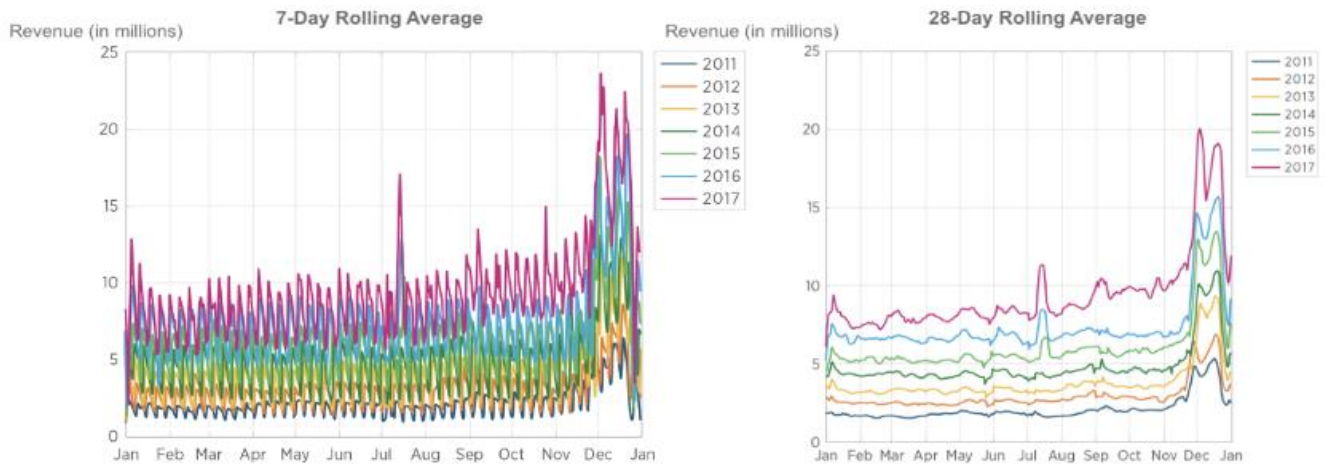
Year-over-year changes can be even better understood by **plotting the ratio of revenue from one year to the next. Indexing to 1** is another powerful method to see how things change from the beginning of a year.
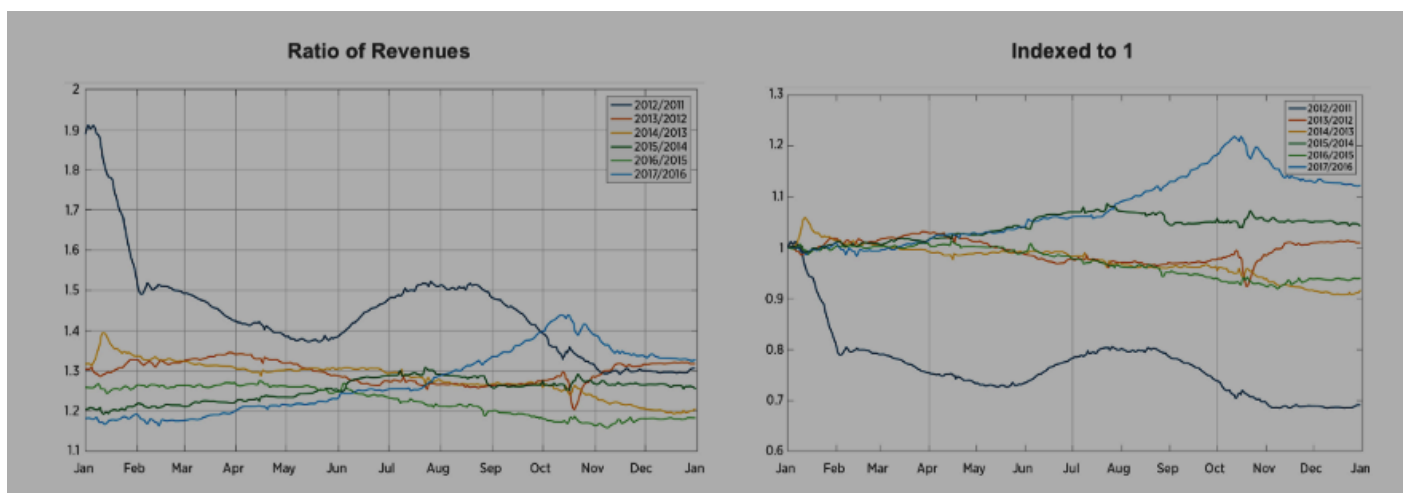
Left, revenue growth declined year-over-year throughout 2016, but in 2017, things started to improve significantly.

Right, it shows 2017 was the best year from a year-over-year perspective, as the company managed to turn things around.

Comparison: You can also analyze seasonality by comparing your key metrics to other apps with similar demographics. In this analysis, it is best <u>to average over at least seven days and index to 1</u>, in order to identify **macro trends**. Comparing your metric to those of apps in <u>different locations</u> (and with different school calendars) can also help identify this effect.

For example, if your app targets younger populations and your key metrics shift in August and September, try to obtain data from another app with a similar demographic, such as Snapchat, to see if it <u>experienced a similar shift</u>. If so, the back-to-school effect likely caused your shift.

Behavioral changes based on seasonality often abruptly change key metrics. Multiple techniques can be used to isolate the impact of seasonality. (**Takeaway**)

- **Competitive Pressure on Product**

User behavior is also affected by **competition and other external events**, such as government actions, new products and social media campaigns. These factors can dramatically change how your users engage with your product.

Generally, the larger the change in a key metric and the shorter the time frame in which it changes, the easier it is to identify the cause. This is particularly true with many external events, although the effects of competition are often difficult to quantify.
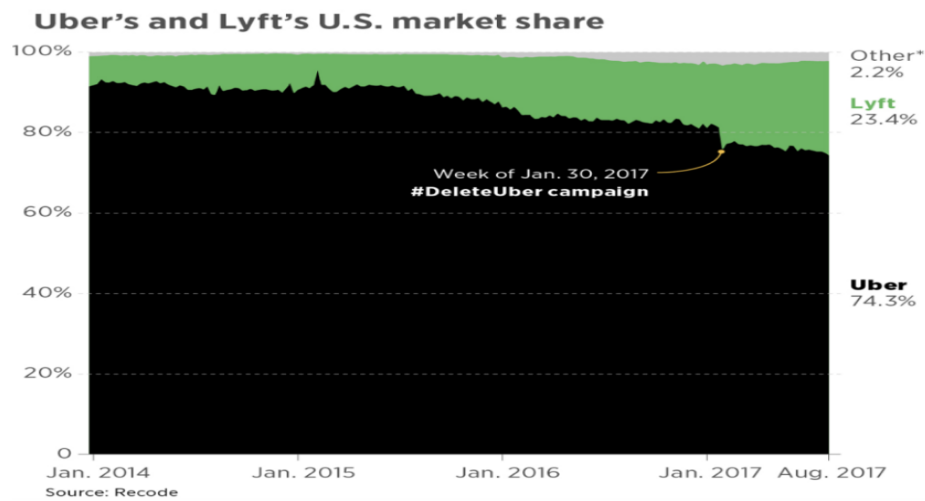
Uber's and Lyft's U.S. market share

Week of Jan. 30, 2017
#DeleteUber campaign

Other* 2.2%
Lyft 23.4%
Uber 74.3%

Jan. 2014   Jan. 2015   Jan. 2016   Jan. 2017   Aug. 2017
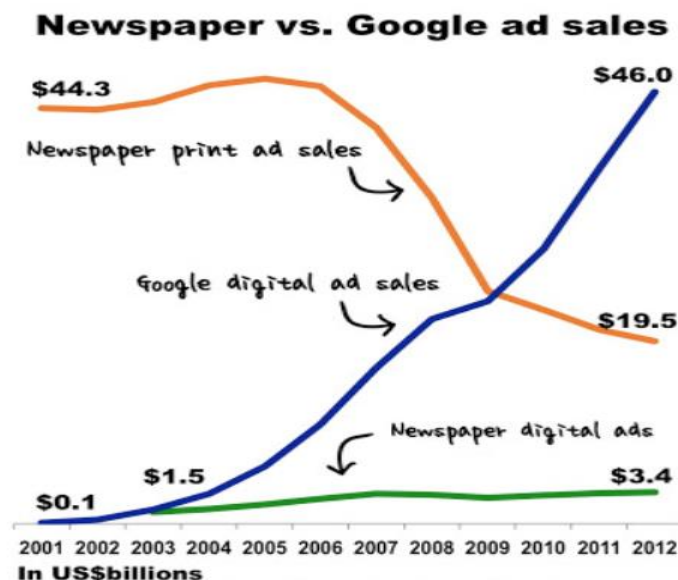Source: Recode

Figure 1

Example, Uber saw a dramatic drop in usage and overall market share due to the #DeleteUber campaign.

The impact of competition on your product can be subtle and difficult to identify — and even harder to fix. This is because competition generally manifests as churn; while we can identify that people are churning, it is difficult if not impossible to know whether people are leaving for a competitor without user experience (UX) research. Leveraging **third-party datasets** that cover your industry (*Sensor Tower/ App Annie*) and frequently commissioning **user surveys** will help you understand how competition is *cannibalizing your user base*. (understanding **competition, cannibalization and incrementality**.)
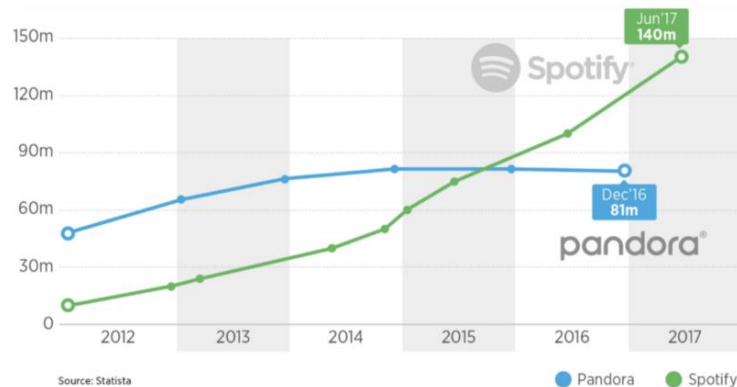
Example, Pandora was growing steadily until Spotify's growth started to pick up speed. It is likely that users began moving from Pandora to Spotify, further accelerating Spotify's growth and flattening Pandora's. However, it is difficult to quantify this shift precisely.

Similar to behavioral changes due to competition, **long-term behavioral trends** are challenging to detect but can be guided by UX research. These macro changes are often driven by external events. Of course, the impacts of such macro changes are not all positive.

Newspaper vs. Google ad sales

$44.3 — Newspaper print ad sales
$46.0
Google digital ad sales
$19.5
Newspaper digital ads
$1.5 ... $3.4
$0.1

2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
In US$billions

Example, while Google's ad sales dramatically improved, US newspapers saw print circulation decline quickly, which later manifested in significantly lower revenue from ad sales.
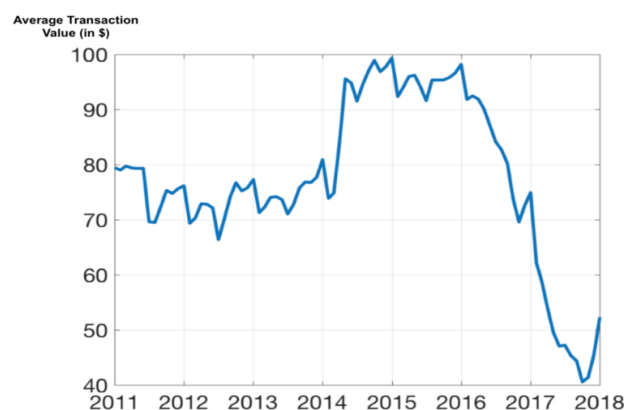
External factors can dramatically affect the performance of your product, and most - aside from macro trends - are difficult to predict. However, if you are disciplined in monitoring key metrics and ensure your company has strong analytical DNA, you will be <u>better positioned to take advantage of opportunities and mitigate the negative impacts of external events</u>.

Effects from competition are difficult to measure and address but *industry data sources can offer direction*.

Singular external events usually result in abrupt changes to metrics. Macro trends may dramatically affect metrics for your product, but will usually do so over an extended period of time. (**Takeaway**)

- **Effect of Behavioral Changes on Product (Mix Shift)**

"Mix" has multiple meanings, and is sometimes known as **Simpson's Paradox**. "user population mix" is the proportion of a specific user base (e.g., users from a given country) relative to the overall user base. A change in a mix over time is known as "mix shift."

Average Transaction Value (in $)

100
90
80
70
60
50
40
2011 2012 2013 2014 2015 2016 2017 2018

Example, the above shows how the <u>average transaction value on Prime</u> has changed over time, which is due to both price adjustments and mix shift of various groups. Amazon Prime offers different prices for regular, student and low-income customers, as well as annual and monthly plans and occasional sales. The drop in average price per customer over 2016 and 2017 is almost entirely attributable to greater adoption among low-income, monthly and student members, all of whom pay lower prices than other customers. The pricing of each product stayed the same during this period; it was the mix of customers that changed.

Pure Mix Shift
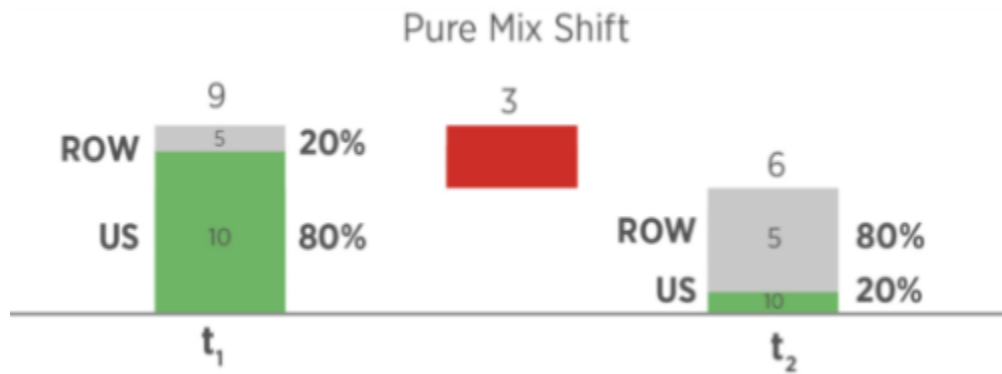
Example, at time t1, TS/DAU for U.S. users is 10 minutes per day; for ROW, it is 5 minutes per day. 80 percent of users are from the U.S., and overall TS/DAU is therefore 9 minutes per user. At time t2, TS/DAU remains unchanged t1 for both the U.S. and ROW. However, the mix of users has flipped: the U.S. now has a 20 percent share, while ROW has the other 80 percent. The new overall TS/DAU is therefore 6 minutes per user — a decrease of 3 minutes per user that is due entirely to mix shift.

No Mix Shift
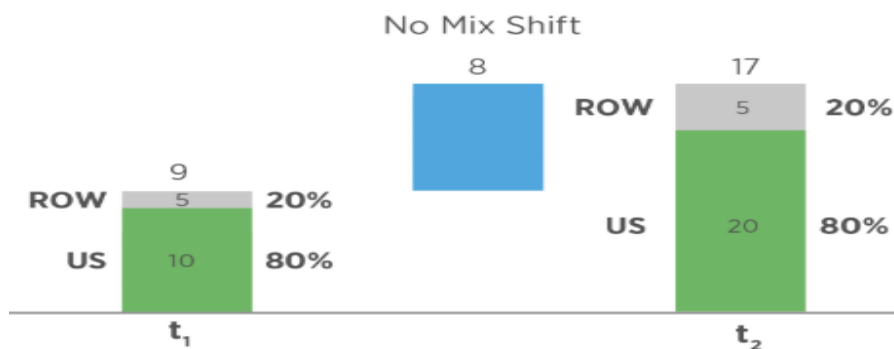
Example, at time t1, TS/DAU is again 10 minutes per day for U.S. users and 5 minutes per day for ROW. The U.S. again has an 80 percent share of users, for an overall TS/DAU of 9 minutes per user. But at time t2, TS/DAU changes for the U.S., increasing to 20 minutes per day, while it remains the same for ROW. The share also remains exactly the same. The new overall TS/DAU has therefore changed to 17 minutes per user — an increase of 8 minutes completely attributable to the increase in U.S. user engagement with no mix effect.

Mix shift effects can be analyzed <u>across multiple dimensions</u>: country, region, platform, age, gender, connectivity class, device class, etc. For the most effective mix shift analysis, you should first carefully examine the problem and develop a hypothesis regarding mix effect.

Analyzing mix shift can help you identify the effects of changes in the population mix versus inherent changes in user engagement. (**Takeaway**)

- **Action Plan**

Once you have confirmed that there is indeed a change in metric worth investigating, you need to develop a systematic and structured approach to identifying and attempting to eliminate each possible cause.

The first step is selecting two points in time that best represent the change in metric you are investigating.

Next, you should ask lots of questions about what could have caused the change in your key metric. Once you have a comprehensive list of hypotheses, eliminate or investigate factors one by one.

<u>Data Quality</u>

Investigate issues with data quality first, as they may be the easiest to identify. Look for logging issues related to product changes — for example, a bug that incorrectly records DAU (daily active users) for a certain locale, language, country, device, etc.

- ✓ To localize such an issue, investigate whether the change is systemic across all dimensions or specific to some dimensions.
- ✓ Examine other correlated metrics for similar changes. For example, if number of sessions is correlated with DAU, and you see a change in DAU but not number of sessions, a logging bug may be the problem.

Product Changes

- ✓ List changes made to the product in the given timeframe.
- ✓ If you have an experimentation framework (A/B testing), quantify the impact each product change would have on your key metric.
- ✓ Look for behavioral changes due to product changes. Examine behavioral changes by group (country, device, etc.) to determine whether the changes are localized, then examine the time at which you saw the metric change. If the metric change happened outside of the time period you would expect based on the timing of the product change, it is unlikely the latter caused the former. Remember, too, that network effects can sometimes carry the impact of an issue beyond the population that is primarily affected; for example, if a bug prevents people in US from using a communication platform, this could lower engagement of people in other countries, as well.

Competition and Other External Factors

Leveraging third-party datasets that cover your industry (Sensor Tower/ App Annie) and frequently commissioning user surveys are some useful means to investigate external factors

Behavioral Changes:

Seasonality is the generally the largest contributor to behavioral change, though external events and competition may also influence this factor.

Mix Shift:

The first step to diagnose mix shift is hypothesizing the dimension in which you expect it to occur. Note mix shift may be a strong factor in long-term changes, but likely will not be the source of changes week over week.