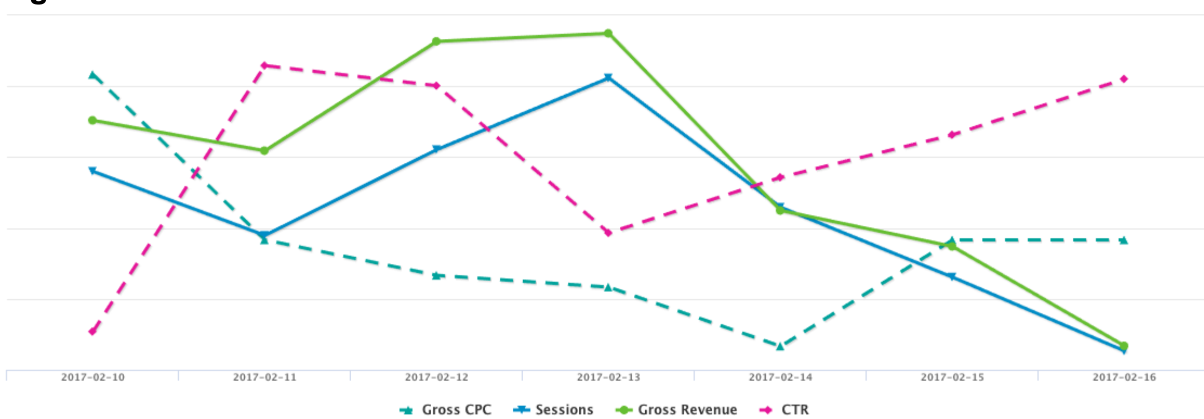
Extracting Revenue Impact from Business Metrics

Identifying revenue and optimization relationship

When I joined System1, I wanted to make an impact as quickly as possible. The first change I pushed into our system was designed to optimize our click-thru rate (CTR) model. I was hoping to see a clear uptick in user engagement, but instead I saw something similar to the graph below (**Fig. 1**)

Figure 1



I *did* see that the CTR (pink line) had increased following my push on 2/13, but unfortunately Gross Revenue (green line) was down. In this case,

$$Gross\ Revenue\ =\ Sessions \times CTR \times Gross\ CPC$$

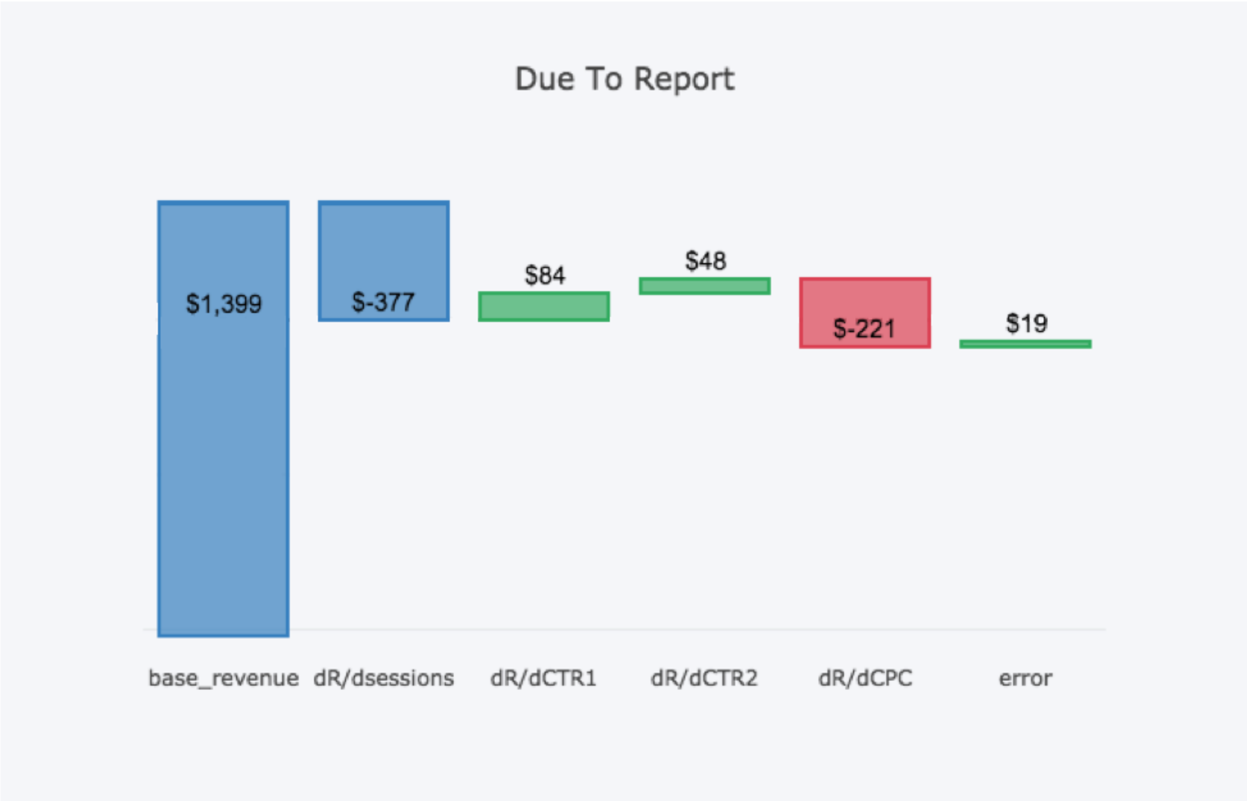
If CTR was up but Gross Revenue was down, then the equation above requires that at least one of Sessions or Gross CPC must have gone down. However, Sessions and Gross CPC were **both** down, so it was hard to tell at a glance whether decreased Sessions was mostly responsible for the decrease in Gross Revenue, or whether it was due to decreased Gross CPC, or whether they were both equally responsible.

This problem comes up whenever we are trying to understand a change in a metric that is the multiplicative product of several other metrics. The more variables we are multiplying, the harder it is to untangle the individual contributions!

Working with one of my colleagues, we put together a new report type that helps identify the impact of multiplicative factors on a metric. We eventually extended this report into a tool that the whole company now uses, called the “Due-To Report.” A typical use for this tool is to compare a metric (e.g. Revenue) over two time periods (e.g. week over week).

In the example below (**Fig. 2**), Revenue is the product of *four* factors (Sessions, CTR1, CTR2, & CPC), so the situation is even more complex than in the example above. However, the Due-To Report makes it clear that an increase in CTR1 contributed $84 to total revenue this week compared to the week before. We can also see that a decrease in CPC took away $221 from total revenue.

Figure 2



The graph above is called a waterfall graph because each bar starts where the last one left off. The “base_revenue” bar is the revenue for last week and adding all the bars together gives us this week’s revenue. The blue bars are not controlled by model optimizations. The green bars indicated positive contributions and the red bar indicates a negative contribution. To find the new revenue this week you would do the following:

$$\$1399 - \$377 + \$84 + \$48 - \$221 + \$19 = \$952$$

You can see that CTR1 made $84 but that gain was offset by the drop in CPC of $221, explaining why our overall revenue was down. This report allows us to pick apart some of the causality not visible in **Fig. 1** and gives us actionable items that we can further investigate or improve upon to continue to drive revenue.

How it works

This week's Revenue (R_{i+1}) and last week's Revenue (R_i) are both known and can be connected using a 1st-degree Taylor series approximation:

$$R_{i+1} = R_i + \Delta t\left(\frac{\partial R}{\partial t}\right)_i + \epsilon \qquad (1)$$

To decompose$(\frac{\partial R}{\partial t})_i$, we use the Revenue equation:

$$R = Sessions \times CTR1 \times CTR2 \times CPC \qquad (2)$$

combined with the product rule from calculus to get:

$$\begin{aligned}\frac{\partial R}{\partial t} = &\left(\frac{\partial Sessions}{\partial t} \times CTR1 \times CTR2 \times CPC\right) + \\ &\left(Sessions \times \frac{\partial CTR1}{\partial t} \times CTR2 \times CPC\right) + \\ &\left(Sessions \times CTR1 \times \frac{\partial CTR2}{\partial t} \times CPC\right) + \\ &\left(Sessions \times CTR1 \times CTR2 \times \frac{\partial CPC}{\partial t}\right)\end{aligned} \qquad (3)$$

and approximate each of $\frac{\partial Sessions}{\partial t}$, $\frac{\partial CTR1}{\partial t}$, $\frac{\partial CTR2}{\partial t}$,and $\frac{\partial CPC}{\partial t}$using the forward finite difference:

$$\left(\frac{\partial u}{\partial t}\right)_i \approx \frac{u_{i+1} - u_i}{\triangle t} \qquad (4)$$

where u_i is last week's value and u_{i+1} is this week's value for a given metric.

Each of the four summands in (3) represents the additive contribution to R of one of the multiplicative factors from (2). For instance, the equation:

$$\frac{\partial Sessions}{\partial t} \times CTR1 \times CTR2 \times CPC = (Sessions_{i+1} - Sessions_i) \times CTR1_i \times CTR2_i \times CPC_i$$

is used to create the waterfall column informally labelled $dR/dSessions = -\$377$ above.

We then solve for ϵ in equation (1), since all other variables are known, to get the final column of the waterfall graph.

Written by Jeff Roach
Data Scientist, System1