

# Intercom Analyst Screener

Shaun McGirr

February 26, 2016

## 1 Products & Frameworks

### 1.1 Google's HEART

*Evaluate Google's HEART framework and list out the main strengths and weaknesses of it, with rationale for your opinion.*

Google's HEART framework<sup>1</sup> suggests one approach to measuring the user experience and generating data to inform product decisions. It grew from a desire to go beyond blunt measures like 'page hits', which are too generic to be informative about what to build next. In the worst case, chasing page hits leads to endless A/B tests with goals disconnected from user experience.

The authors sell HEART as an improvement over traditional small-scale user experience research (observing users complete tasks is costly) and technical metrics under the PULSE framework. It aims to measure user experience as directly as traditional methods, but leverage the wealth of behavioural data generated by web products to measure more with less.

#### Strengths

- The separation of the goals of the business/product/feature from their measurement is a genuine strength. One of the toughest challenges in analytics is to discover the underlying motivation of our stakeholders. As with most technical fields there is a constant urge to choose a solution (eg 'maximise weekly active users') before the problem is even well-defined. Beginning the search for metrics by being clear about the goals under each category helps to narrow the field of useful metrics, and makes the eventual choice more trusted.
- The inclusion of the 'signals' step between definition of goals and metrics is also useful, for a different reason. Common goals for a business or product, like 'grow our active user base', lead quite directly to a set of useful metrics, like 'change in active users'. The signals step forces an important check: what would success or failure of the goal actually look

---

<sup>1</sup>Available at <http://research.google.com/pubs/pub36299.html>

like? Two common pitfalls of analytics that this can help remedy are: a) trying to build metrics when the relevant data are not actually available, and b) assuming that various stakeholders have a common definition of success.

- Some of the technical challenges to achieving good coverage of metrics are fading, as the world becomes more web-page centric and less obsessed with mobile apps or Flash. This means those who find the framework useful for web products should find it more useful over time!
- It is built from the experiences of many different product teams at one of the most important web companies, so the expertise it contains is battle-hardened and likely generalisable across many types of products.

### Weaknesses

- Despite being linked by the word HEART, the way the categories in the framework are described suggests they emerged as a way of ‘binning together’ metrics that were *already in use*. If that is true then the development of this framework did not follow its own key strength described above: no goal was defined up front for the framework itself. Beyond ‘we need more and different measures’ the authors don’t present a clear case for why these categories and not others. This leads to a deeper problem. Although Happiness, Engagement, Adoption, Retention and Task Success can all be conceptualized and measured, and may even correlate with a generic definition of ‘good user experience’, do they actually represent user experience? In fact, beyond Happiness and Task Success, these categories have little relevance to the way a user experiences a product: Engagement, Adoption and Retention are goals of the product team. Google’s (many) decisions to kill off products that were presumably doing OK across these goal categories suggests that the underlying definition of user experience in this framework is undercooked. **This weakness could be remedied by making a clear case for why these goal categories matter more than any other plausible set. Making a convincing case would rely on having a clear model of how users of a given product make their decisions: in this model, Happiness and Task Success would be inputs, while Engagement, Adoption and Retention are outputs.**
- A framework for measuring improvement in user experience should at minimum give some idea of overall progress on achieving a better experience, and make clear the tradeoffs between different goals. As presented, HEART does neither. A set of goals could be defined under each category and neatly cascaded in to signals and metrics, and these might all look good on a dashboard. It would be easy for stakeholders to assume that as long as all these metrics head in the right direction, user experience is improving overall. This would be an easy mistake to make because metrics have a way of becoming the focus of decisions, rather than one source of information that feeds decisions. Having said that, improvement across

the board is quite unlikely, more common is that some metrics improve while others decline. In this case the optimal decision is unclear because doing something to arrest decline in metric A could harm progress on metric B. Add to this the standard resource constraints of any team, and the next-best-action is unlikely to be clear. **What a better user experience means depends on the product, and the HEART framework is a useful step to clarifying this, but not the end. Information from outside the metrics must inform the measurement of overall progress. The framework could be improved by an additional column mapping dependencies and interactions between goals, to make clear the tradeoffs.**

- There is an undertone in the authors' presentation of the framework that data from web logs is 'behavioural' simply because it represents actions taken by users in a natural, non-laboratory setting. While this kind of data does represent a user action (eg 'user clicked back button in browser'), it gives only a very limited view of the underlying behaviour (eg 'user is frustrated by workflow organisation in the product'). While a strength of the framework is its encouragement to think carefully about goals and their relevant signals, this view of data as 'unfiltered' can be dangerous: no data is generated in isolation of important decisions by the product team over what to log and how. The danger is that one random decision by an engineer some time ago leads a particular set of data points to become *the* relevant metric everyone focuses on, even if it is a poor proxy for the part of the user experience we care about improving. **This weakness can be guarded against by defining metrics broadly, in terms of the experience of the user that generated the data, rather than narrowly in terms of 'we already pulled that out of the web logs'.**

## 1.2 Shaun at Google

*If you worked at Google, which of their products would you like to work on most? Explain your answer.*

Google+, because I like a challenge! The promise of this product for me was always as an integration layer between all the Google services I already use. The failure of this product is a consequence, in my view, of trying to make it a mainstream social media channel, of which there were already too many. Despite several relaunches and at least two Executives in charge, Plus is still a bit confused: every time I visit [plus.google.com](https://plus.google.com) it looks different. Now it includes a chat/Hangouts sidebar, and still shows the 'Circles' I diligently defined years ago in anticipation of its promise to 'share what you want with whom you want' (even though I've never used them).

As someone interested in developing my product management skills further, it would be a fascinating task to understand the intended role of Google+ within their ecosystem, test whether that role meets some unmet need of users, and

proceed from that point of validated assumptions. Of course, I would have to enter that task willing to entertain the possibility that there is no need for the product at all!

## 2 Data, Data, Data

If you had access to any dataset from any company or organization in the world...

### 2.1 What would the company/dataset be?

One dataset by itself is usually not as interesting as two datasets joined together. For this reason I'd like to cheat a little and give two answers:

- All New Zealand property valuations and sale prices held by QV.co.nz, a government-owned company that charges steeply for access to data on individual properties. They charge even more steeply for bulk access, under a restrictive license.
- Ownership data on properties in New Zealand, which is available free-of-charge but under a not-quite-open license. This supposedly protects individuals' privacy, but also limits users' ability to publish their analyses.

### 2.2 What would you like to explore within the dataset?

I would use the data described above to build a comprehensive picture of what is actually happening in the property market. Citizens, journalists, and policy-makers could all trust the data and have more informed conversations: at the moment the scale of the problems are difficult to estimate and easy to dismiss.

In particular I would build:

- An underlying dataset that tracks changes in valuations and sale prices over time, at small geographic areas.
- Another dataset tracking changes in ownership patterns over time, with the same geographic detail.
- A tool allowing users to select various geographic areas and see the *distribution* of the key metrics and how these change over time. At the moment the media quote median or (even worse) mean sale prices, obscuring important information, especially about affordability at the cheaper end of the market.
- The key question I would want to answer is 'what proportion of the housing stock is affordable for middle- and lower-income New Zealanders?'

## 2.3 Why this dataset/this exploration?

The property market in New Zealand has several of the characteristics of Ireland's in 2006-07, right before the bursting of that property bubble, which in turn triggered a banking crisis: real economic growth driven by (now-depressed) dairy price increases, restrictive town planning policies that make new building very expensive, and significant tax incentives for individuals to own more and more real estate.

Add to this the size of New Zealand's population (about the same as Ireland's) and the fact the problem is worst in our largest city of Auckland (about the same size as Dublin, with the next-largest city much smaller) and the parallels are uncomfortable. Of course, a good analyst never jumps to conclusions...

At the moment, the real estate industry is the only party with such comprehensive access, and they have little interest in revealing the severity of our property bubble. The media publishes press releases from the industry with median house prices by region, but these over-summarise the raw data and obscure potentially important patterns.

## 3 Technical Analytics Task

### 3.1 Context of the data

Intercom launched the "Inbox Insights" feature in 2015 to help its customers, in particular support managers, understand their team's workload. These managers are responsible for ensuring a good customer experience, which requires constant and close attention to the distribution of work across team members and to changes in patterns of customer demand across time.<sup>2</sup>

The product team is interested in whether this feature has been successful, and whether there is scope for growing the business with the feature. They have provided data about customer signups to Intercom from Feb-Sep 2015, and usage of the Inbox Insights feature from May-Sep 2015.

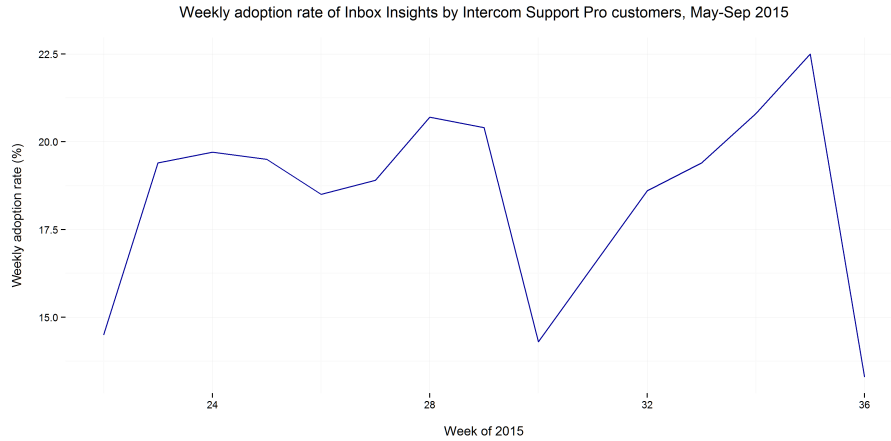
### 3.2 Questions

#### 3.2.1 What is the weekly adoption rate among Support Pro customers?

**Answer:** See graph below. The adoption rate varies between a low of 14% in weeks 22, 30 and 36, and a high of 22.5% in week 35.

---

<sup>2</sup>See <https://blog.intercom.io/design-principles-insights-are-not-about-analytics/>



### Method:

1. Define a customer as a unique *app id* in the ‘apps on support’ file.
2. As we are interested in only Support Pro customers, and in a weekly rate, subset the data to one row per Support Pro customer per week; the count of these will become our denominator. There are 2853 unique customers over 28 weeks of data.
3. Define adoption as occurring when any user within a Support Pro customer loads Inbox Insights at least once in a given week.<sup>3</sup> A better threshold might be ‘more than once’: calculating ‘loads per week’ showed about 32% of user-week observations only involved one load of Inbox Insights, which is a very low bar for ‘adoption’. This would obviously require discussion with the product team. I would also clarify whether they consider this metric to be relevant at the customer (ie *app id*) level, or whether they care more about individual users within a customer. Both are possible to calculate but imply different concepts of adoption.
4. There were no adoptions before May 29th so assume this was the launch week, week 22 of 2015. Also note there was no usage data for week 31 of 2015 (both worth checking with product team!)

### 3.2.2 What kind of retention rates are we seeing among customers using the product?

**Answer:** See the table below. Intercom’s retention rate for Support Pro customers over this period seems pretty solid at 74% even for those who never used Inbox Insights. Retention rates increase steadily with more weeks using the

<sup>3</sup>A purist definition of adoption would probably require it to be a ‘terminal’ state: ie once a user has adopted the feature, they fall out of the dataset for the remaining time periods. For time’s sake, I leave to one side and interpret adoption as ‘usage’.

feature, which is what the product team would like to see! Of course, it could be that being likely to retain (for other reasons) is correlated with trying out Inbox Insights, so a more sophisticated analysis that rules out other plausible causes is needed before breaking out the beers.

Weeks using	Customers not retained	Customers retained	Retention rate
0	282	814	74
1	85	468	85
2	27	270	91
3	10	169	94
4	6	96	94
5	1	62	98
6	0	56	100
7	0	34	100
8	0	23	100
9	1	31	97
10	0	22	100
11	0	13	100
12	0	16	100
13	0	16	100
14	0	35	100

#### Method:

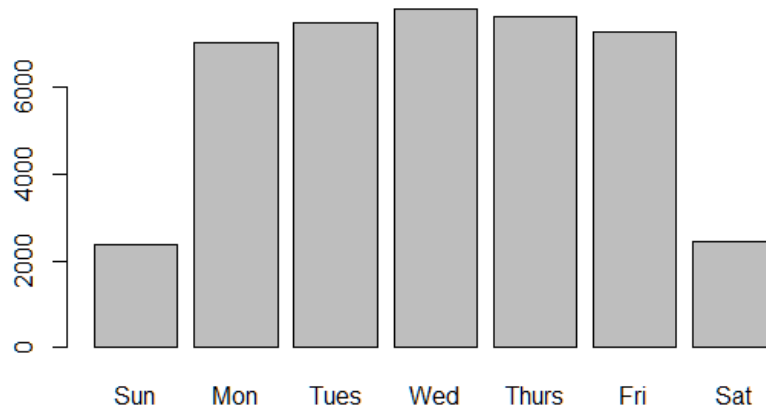
1. Define the retention rate as the percentage of Support Pro customers in a given window who are still Support Pro customers at the end of that window. For this exercise my window will be from week 22 to 36, as this is when the feature was live.<sup>4</sup>
2. Define ‘usage’ as the number of weeks within the window where the customer used Inbox Insights.
3. Merge these two tables (customers who retained/didn’t; customers who used/didn’t) and calculate the retention rate.

#### 3.2.3 Can you tell us anything about usage patterns? Do customers use the feature every day of the week, or every week of the month? Or is it a feature that is only used occasionally?

As I expected, usage is much heavier during the week than on weekends.

<sup>4</sup>This groups together users who were customers from the beginning of the window, with those joining later. A more sophisticated analysis could use survival modelling to estimate the retention effect of Inbox Insights usage when the window is a single week. Here I exclude customers with only one week spent as Support Pro customers: this is too few weeks to reliably calculate a retention metric, especially as many of these customers only appear in the last week of the window.

### All Inbox Insights usage by day of week, May-Sep 2015



There are no significant week-of-month patterns in these data. Whether this was to be expected or not depends on the support model of these Intercom customers; I suspect they are much more concerned with a weekly ‘run rate’ than meeting end-of-month deadlines for closed ticket metrics.

#### 3.2.4 Is there anything else interesting you noticed about the data?

Yes, an overwhelming majority of users (89%) opt for the ‘day’ time grouping of Inbox Insights, rather than ‘hour of day’ or ‘day of week’. While this is the first option selected in the online help picture, and so is likely the default option, it is still very sticky! It could mean that the majority of Intercom’s customers simply care about conversation responses/closes within a day (rather than workforce balancing), or it could mean they haven’t explored the other options enough.

#### 3.2.5 Follow-up questions to analytics/product teams

As elaborated in my commentary above:

1. How high should the bar be to consider this feature ‘adopted’? Is one use per week enough?
2. What is more important, adoption by the customer account, or by the users within the account?
3. What on earth happened to the data from week 31 of 2015?!?



### **3.2.6 30 seconds with Paul Adams**

Paul, Inbox Insights has launched successfully. Usage is stable at about 20% of our Support Pro customers. While about three quarters of these customers retain anyway, without using Inbox Insights, usage is correlated with retention: customers using the feature at least every second week are showing almost no attrition. Of course, this could just be because they are fantastic customers! But the evidence so far gives us a good base to gather qualitative information about adoption of this feature and use that to inform development of further insight features.