KNOWLEDGE@WHARTON

Clumpiness and Customer Lifetime Value

MARKETING

Are Your Customers 'Clumpy'? What Binge-buying Means for Marketers

Dec 17, 2014   📍 North America   🏷 Research, Video

If you've ever spent a weekend binge-watching a season of "Breaking Bad" or some other series on streaming video, you can take comfort in the fact that you're certainly not alone. Whole sections of the population are consuming digital products and services in a "clumping" pattern that features extended periods of inactivity punctuated by short, intense buying bursts.

Once marketers realize this — and they already have the data in hand to track it — they could mine a rich, new digital behavior vein, says Wharton marketing professor Eric Bradlow. He notes that a basic customer value measure for decades has been RFM segmentation (recency, frequency and monetary value). "My research ... says that's not a complete characterization...," he adds. One more letter is needed: "I call that 'C,' which [stands for] clumpiness." Bradlow explains that concept in this Knowledge@Wharton interview, based on a research paper titled "New Measures of Clumpiness for Incidence Data," which he co-authored with Yao Zhang, an associate at Credit Suisse, and Dylan Small, a Wharton statistics professor.

Edited excerpts from the interview follow.

What is "clumpiness" in customer data, and why it matters:

One of the most established practices in the field of marketing and customer valuation is to summarize a customer using what's called RFM segmentation — recency, frequency, monetary value — which means I take everything I know about my customer and I compute just three simple numbers: How recently did they buy? How frequently do they buy? And when they buy, how much money do they spend? ... It's the basis on which most companies decide who are the valuable customers and who are the non-valuable customers.

My research ... says that's not a complete characterization of customers. You have to add one more letter to RFM, and I call that "C," which [stands for] clumpiness , which means some customers do buy in a regular pattern. Historically, if you bought orange juice, if you bought diapers, you bought things in a regular pattern. But clumpiness refers to the fact that people buy in bursts. And those burst periods indicate something very different about the customer and that those customers could be extremely valuable.

On the key takeaways:

The key takeaways of my research are very simple. Let's imagine you want ... to predict who are going to be the valuable customers in the future. And you have four things you can use to predict it. As I mentioned: recency, frequency, monetary value and let's say the marketing spend towards the customer. Those are the classic ways in which companies build what are called scoring models. I'm claiming you need to add one more number, and that's C — how

clumpy the customer is. This is no more difficult to compute than R, F and M. You can do it in Excel. It's very quick to compute. You can compute it for literally 100 million customers in a second.

Burst periods indicate something very different about the customer and that those customers could be extremely valuable.

And the findings of my research suggest that higher clumpy customers are worth more out of sample, meaning in their future value, even after controlling for RFM and marketing expenditure — which means we have found another variable that firms should track [concerning] our customers and use it to predict their worth in the future.

The most surprising conclusions:

Two things surprised me about my conclusions. One is I just figured that this RFM based segmentation, which had been around for so long and is used by so many firms, had been validated in the sense that there wasn't anything else simple out there that could help explain customer value. You can do all kinds of fancy web-scraping and all kinds of other variable construction, but clumpiness is so simple.

So, first, I was surprised that that had been missed — that, in other words, hot and cold periods are indicative of something about the customer. I think the second part that surprised me, at least in the data sets I've analyzed, [is that] it's true for digital and online consumption goods, but it's not true for regular consumer package goods. In other words, with historical models I can see why they fit fine, because you buy toilet paper in a regular pattern; you buy orange juice in a regular pattern. But you don't consume Hulu in a regular pattern. You don't bid on auctions at eBay in a regular pattern. You don't buy books at Amazon on a regular pattern.

If you look at historically purchased goods, clumpiness really isn't there. But if you look in the new wave, the new economy, clumpiness is pervasive in every data set I've analyzed.

On the practical implications:

I think of all the research I've done over my … 20-year career, this is probably the most practical thing I've done. The work I do tends to be what I call fancy, complex statistical modeling. And this isn't about statistical modeling. This is about a number — clumpiness — that firms can actually compute today. They don't need to collect any additional data. It's the same data they're using to compute R, F and M and customer lifetime value. And they can figure out how much value it adds to predicting customer value. Your rank ordering of customers will change. Your decisions about which customers are valuable to reactivate — imagine customers have churned — well, which ones are valuable to reactivate?

My claim is the clumpy ones, even though they've churned … are the ones to reactivate. If you reactivate them, they'll come back and be clumpy again, and do a lot of stuff in the future. So, I think it has huge practical value. And the beauty of it is, if you go to my website I have an Excel sheet there that has worked out examples. It actually has an Excel sheet that you can just download and you can start using clumpiness today.

What new rules, procedural changes or strategies would you suggest as a result of your research?

A lot of people today talk about big data. I love big data, but I'll tell you what I love even more than big data. I love data compression. And what I mean by data compression is, you can collect thousands and thousands of variables on people now. You can track where they are and you can track what they bought, what web pages they looked at. But that's not science. That's data collection. Now a question is, which of that information is actually useful for the business problem at hand? And that's what I call data compression.

So, the way I view clumpiness is as an addition to traditional variables like RFM, marketing activity and stuff like that — I view it as a form of, let's call it increased data compression. I'm just telling you that you need to keep a little bit more data. You can't compress things down to three numbers. You've got to compress it down to four.

I've done a lot of work on clumpiness, I know it exists across industries. I know it can be of predictive value. Here's what I don't know: what causes it…. I've related marketing activity to clumpiness. Firms can try to make you clumpy by sending you an e-mail, by sending you a catalog, by targeting you, etc.

But I haven't really studied yet what's the optimal way in which firms should target you, knowing that clumpiness exists. I haven't looked at, for example, do you consume more clumpy content if it's a series? Imagine watching "Breaking Bad" or "Mad Men" or something like that. Or imagine you're a firm and you're trying to sell a suite of products like a facial care line and a moisturizer line, and all this other stuff. Should you package it together and make it seem like people are progressing towards a goal?

If you look at historically purchased goods, clumpiness really isn't there. But if you look in the new wave, the new economy, clumpiness is pervasive in every data set I've analyzed.

So … I know mathematically how to compute it. I know it's trivial for firms to do. I know it's predictive. But the part that's left unknown to me is the psychology of why, which is why I'm partnering right now with a lot of my more consumer psychology-oriented colleagues. We're going to start running a lot of behavioral experiments in the lab to try to get to the underlying psychology of why people behave in a clumpy fashion.

On "Clumpy" vs. "Bingeing"

I like the word "clumpiness." Other people like the word "bingeing." The reason I like clumpiness is that it refers to the opposite, which is non-clumpy, which is kind of equally spaced arrivals or equally spaced purchases. I don't think I've seen a story about clumpiness, but any time you see a story about people bingeing [on] content or people consuming things – "a student sat up for 18 hours watching this" — it applies. And the concept is so pervasive: Every time I talk to managers, students or academics about it, everyone believes it exists.

Dispelled misperceptions:

My research dispels the idea that in some sense customers can just be categorized by a simple set of numbers. You need to go a little bit beyond that. You need to go a little bit beyond what I would call simple theories of how people behave. If you look at recency, frequency, monetary value, which is kind of the historical basis of consumer behavior, it basically ignores what I call

the inter-arrival times. It basically says, I can take all the data — like it was a two-day window and then a four-day window and then a three-day window, then a six-day window — I can throw all of that away and all I need to know is when's the last time you came and how many times had you come?"

What this dispels is [the notion] that the arrival pattern of people is uninformative. It's very informative. People [who] come in bursts, then go away and then come back in bursts and then go away … those people are fundamentally different. I personally believe there are clumpy-type people and non-clumpy-type people.

What we've also shown is, it varies by product categories. So, we've found, for example, that women tend to be more clumpy than men. We've found that younger people tend to be more clumpy in their consumption than older people. So, I think [one of] the myths that we're going to dispel is that not only are all people created equal, but that there are simple ways to just categorize all people into a certain type.

How the study stands apart:

There's a whole class of mathematical models that have been popularized — although they've been around for 50 years — over the last 10 years called "hidden Markov models." … Let's imagine there are two states of the world. You're in a hot state or a cold state, and you rotate back and forth between a hot and a cold state. That mathematical model is clumpiness. You're hot, you do a lot of stuff. You're cold, you don't. Hot, cold, hot, cold…. What I wanted to do was to bring to the practitioner a way that they could compute a simple number. It's a statistic. It's not a statistics paper. It's a paper about a number, a statistic, as we call it.

You just compute the number and then do what you want with it. You could try to use it to predict customer value. You could use it to see whether men are more clumpy than women. You could use it to segment people. That's what typifies and separates this work — it's a simple metric-based approach that practitioners can use. It's not a fancy modeling based approach. But they're both trying to cover the same problem.

The work I'm doing isn't ivory tower mathematics. It's a simple number that someone can compute.

An example of how "clumpiness" might be used:

What we've studied so far with reaching clumpy customers is whether e-mail, catalogs, different types of marketing channels are more effective. What we found, not surprisingly, is e-mail has more of a short-term effect, as you would expect. [A] catalog has more of a longer-term effect.

What we've yet to really understand is — are there certain words in an e-mail or a catalog or a video campaign that will engage or, if you'd like, cause people to be more clumpy? Are there certain topics that are more clumpy, some product categories that will necessarily be more clumpy? All we've done so far is to establish that the phenomenon exists. I know it exists across lots of industries. I know certain types of people tend to be more clumpy.

The part that I haven't done, which is shocking because I'm a professor of marketing, is talk about the marketing implications of it yet. That's going to require bigger and newer data sets that allow me to link things about marketing campaigns to people's clumpy behavior. I know how to do it. I just need richer and better data to do it.

What's next?

I'm thinking of about three different streams to follow up this research. First of all, I'd be thrilled to just analyze more data sets and prove how pervasive the clumpiness measure is. I've analyzed data sets from Amazon, from CDNow, eBay, Hulu, YouTube and also from some traditional consumer package goods companies. [Now,] I just want to apply it to new data sets.

The second is I want to understand the psychological processes. Why are people behaving in a clumpy fashion? The third and final piece is I want to relate marketing activity to clumpiness. Now, that's going to require not just people's behaviors — like what did they do? What web sites did they visit? What did they purchase? [It will also require] information about the

marketing campaigns themselves, possibly even the copy of the marketing campaign, which channels they were sent through — and that's going to allow me to come up with optimization — ways for firms to optimize their marketing campaigns to activate clumpiness.

All materials copyright of the Wharton School (http://www.wharton.upenn.edu/) of the University of Pennsylvania (http://www.upenn.edu/).