
How Online Product Reviews Affect Market Share



By Hallie Cho , Vanderbilt University; Manuel Sosa and Sameer Hasija , INSEAD

To find out how user-generated content affects your bottom line, you need a fine-tuned strategy for data analysis.

Intuitively, we know that positive online buzz about a product or service is better for sales than the negative kind. Measuring that bump in demand, however, is a much more complicated matter. Precise quantification would require knowing how the different forms of online data – for example, quantitative and qualitative, or numbers and text – affect consumers’ decision making as they research a purchase.

It’s a challenging task, but not an impossible one. For our recent [working paper](#) focused on the US automobile industry, we were able to attribute changes in market share directly to both the quantitative and qualitative portions of online product reviews.

Analysing online reviews

We looked at the data on 416 car models during the years 2002-2013. For each, we captured the most relevant considerations from the consumer

standpoint, including price, horsepower, size and miles per dollar. We then simulated customer decision making by plugging each characteristic into an econometric model. That way, we could account for any features of the automobiles themselves that might explain rising or falling market share. We measured market share for the vehicles in our dataset by dividing the number of units sold per year by the total market size (i.e. the number of US households that year – as every household needs a mode of transportation).

We also accessed publicly available consumer reviews from a leading industry website. Each model had an average of 45 associated reviews during its year of release; the most popular model had nearly 600.

Like most product reviews, the ones we studied consisted of both a star rating (one to five stars) and text component (which the reviewer could leave blank). Instead of treating both parts of the review as identical, we subjected the text component to a machine-learning (ML) sentiment analysis. The ML model was trained with the help of humans recruited through Amazon's Mechanical Turk, who scored a sample set of reviews for positive or negative sentiment.

The bifurcated nature of consumer reviews turned out to be pivotal. Looked at in isolation, the impact of star ratings on a car's market share exhibited decreasing returns on aggregated rating, meaning that a higher rating had negligible effect on market share when the car ratings were already high. Hence, a five-star car would see less of a demand bump from its online buzz than a four-star one, which doesn't make intuitive sense.

However, when we examined the interaction of star and sentiment ratings, the picture came into sharper focus. The decreasing-return impact curve became a steep upward line – but only for those models with high sentiment scores. In other words, a high star rating's impact on market share was moderated by the overall sentiment of the written reviews. Great star ratings meant less to consumers without a corpus of text recommendations behind them (see 3D plot below).

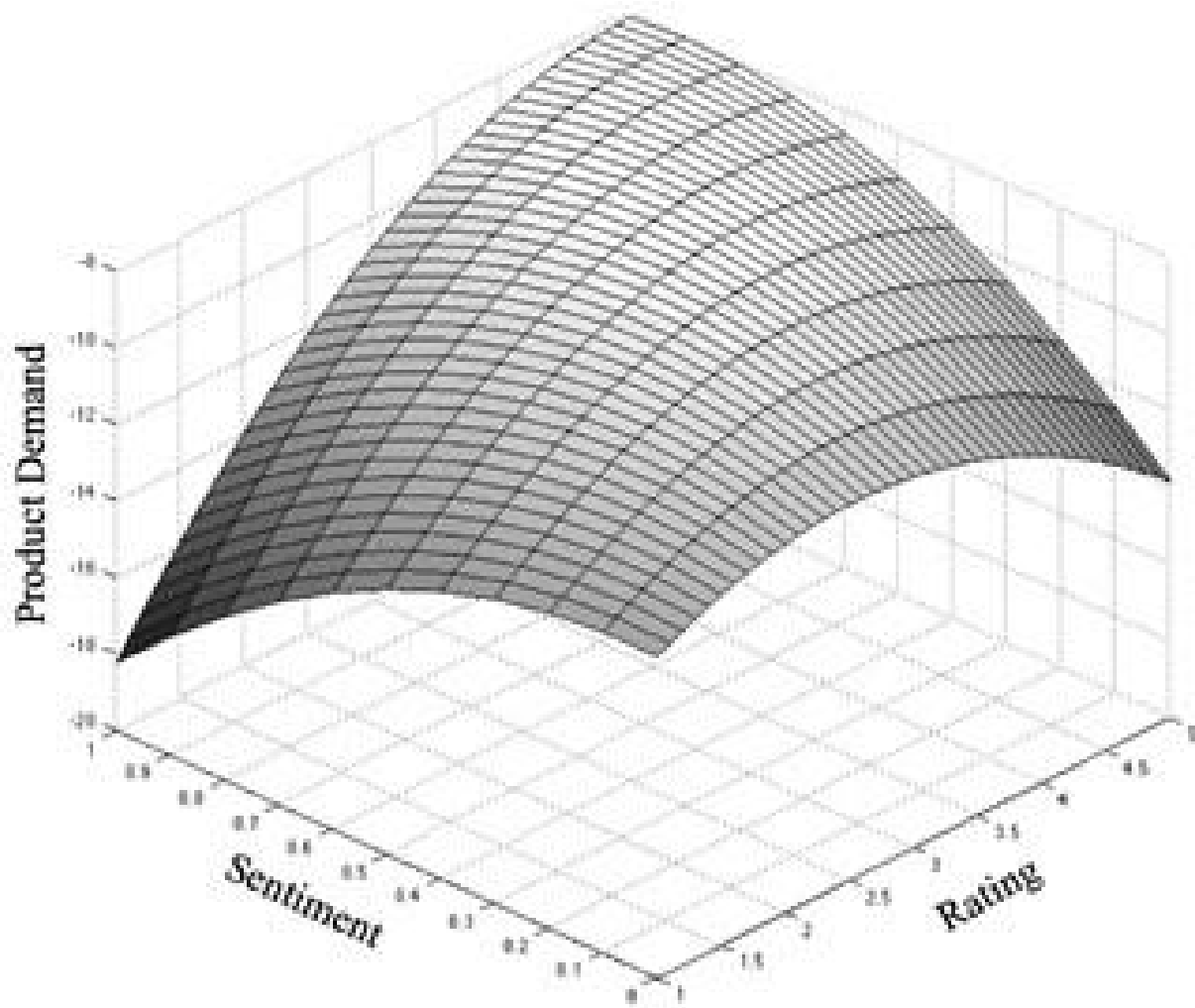


Figure 1: Joint Effect of Review Rating and Sentiment on Product Demand

Dual-system thinking

We liken the joint effect of sentiment and star rating to Daniel Kahneman’s “System 1” and “System 2” framework of human cognition, popularised in his best-selling book *Thinking, Fast and Slow*. System 1 (rapid and intuitive) thinking corresponds to the aggregated star rating, which gives a quick-hit crowdsourced impression of product quality. System 2 (deliberate and rational) covers the written reviews, which demand mental effort on the part of both writer and reader. In Kahneman’s framework, System 2 is responsible for endorsing or re-evaluating System 1’s automatic impressions before converting them into beliefs and actions.

When we see a star rating that is too close to perfect, our scepticism directs us to the written reviews for corroboration. If the enthusiasm in the text is

less impressive than the rating, we may retreat a step or two from the brink of decision and consider other options.

We found empirical evidence supporting this explanation after ruling out several alternatives. For example, we controlled for the possible influence of past reviews on the present – i.e. the possibility that the popularity of previous models of the same car could colour review content for a current model, or that reviews posted early could set the tone for those that followed. When we gave less weight to those earlier entries in our analysis, the findings still held. Our analysis is also robust when considering different reading behaviours (e.g. when people read only negative reviews, or only the most recent reviews).

Lessons for data strategists

Even though our study was conducted in one particular industry, universal lessons about data-driven strategies can be drawn from it. First, online data come in multiple formats, both structured (e.g. star ratings) and unstructured (e.g. written reviews). A genuinely comprehensive data strategy will not only encompass all formats but also examine the ways in which they may interact. This is important because when these various formats are packaged together, as in online product reviews, they may send different signals to consumers. Analysing only one format may yield a skewed interpretation of the overall data, while a holistic view is more likely to generate clues as to what drives consumer behaviour.

Second, data-driven strategies should be context-dependent. Before you can draw firm conclusions about the business impact of your data, you need to know what moves the needle for consumers in your industry and design experiments accordingly. For example, we would not have been able to produce a clear picture based on the data we had, had we not first replicated the typical buying process in the industry we studied with a validated econometric model. A plug-and-play approach to data science won't get you where you need to go.

Third, our study is a fantastic example of the sort of scalable data analysis in which all companies can now engage regardless of their size, thanks to advancements in artificial intelligence (AI). Tools such as sentiment analysis have developed by leaps and bounds in terms of both technological refinement and accessibility. Consequently, there's no reason small-medium enterprises (SMEs) can't get in on the action once reserved for large

established firms.

No matter your industry or degree of data sophistication, your best strategic bet is merging AI-based capabilities with comprehensive data collection and nuanced industry knowledge.

Find article at

<https://knowledge.insead.edu/operations/how-online-product-reviews-affect-market-share>

About the author(s)

Hallie Cho is an Assistant Professor of Operations Management at Vanderbilt University.

Manuel Sosa is a Professor of Technology and Operations Management and the Director of the Heinrich and Esther Baumann-Steiner Fund for Creativity and Business at INSEAD. He also directs the [Design Thinking and Creativity for Business](#) programme at INSEAD.

Sameer Hasija is Dean of Executive Education and Professor of Technology and Operations Management at INSEAD.