



Read more.

Publishing in the Era of Big Data

Publishing in the Era of Big Data

The mantra of the moment, in every field from retail to healthcare, is “Big Data” – defined as being data sets that are too large and complex to manipulate with standard methods or tools. Analyzing these data sets is quickly becoming the basis for competition, productivity and innovation; in fact, some predict Big Data will be as important to business and society as the Internet has become, and it is being used to predict where and when crimes will occur, flus will strike, where traffic will snarl – all very useful for deploying limited resources like police forces, health care professionals or traffic lights.

Big Data is ostensibly a driver of big business at retail – a predictor of what color will be in, what flavor will gain favor, what kinds of electronics will take off. By assessing past behaviour in fine detail and across huge numbers of instances (or people), trends become easier to predict and at an earlier stage such that you can know what your customer really, really wants almost before he does, and it all looks akin to magic. In a now-famous [example](#), Big Data allowed mass market retailer Target to know that a customer was pregnant – and so started sending sales flyers offering deals on baby furniture and maternity clothes – even before she told her family.

But what can Big Data do for publishers?

What IS Big Data for publishers anyway?

In the past, before digital reading, publishers had at hand the blunt instrument of units sold and could draw inferences by analyzing sales by region and broad demographics, and then anecdotally what people, or reviewers anyway, thought of the content.

With the onset of digital reading, though, it is now possible to know how a customer engages with the book itself – what books were left unopened, which were read to the very last word and how quickly.

An argument could be made that once a book is sold, who cares what readers thought of it? Cash in hand is worth a lot of speculation in the abstract.

But knowing what readers find engaging, and what they do not, can actually help publishers unlock previously hidden equity within their publishing lists and inform decisions on which authors and franchises to invest in, which ones are running out of steam, and where trouble may lie within any individual book, thus offering the opportunity to make a new edition a materially improved one.

We now have the opportunity to act on engagement, not just sales, which over time should lead to a stronger, more viable list overall.

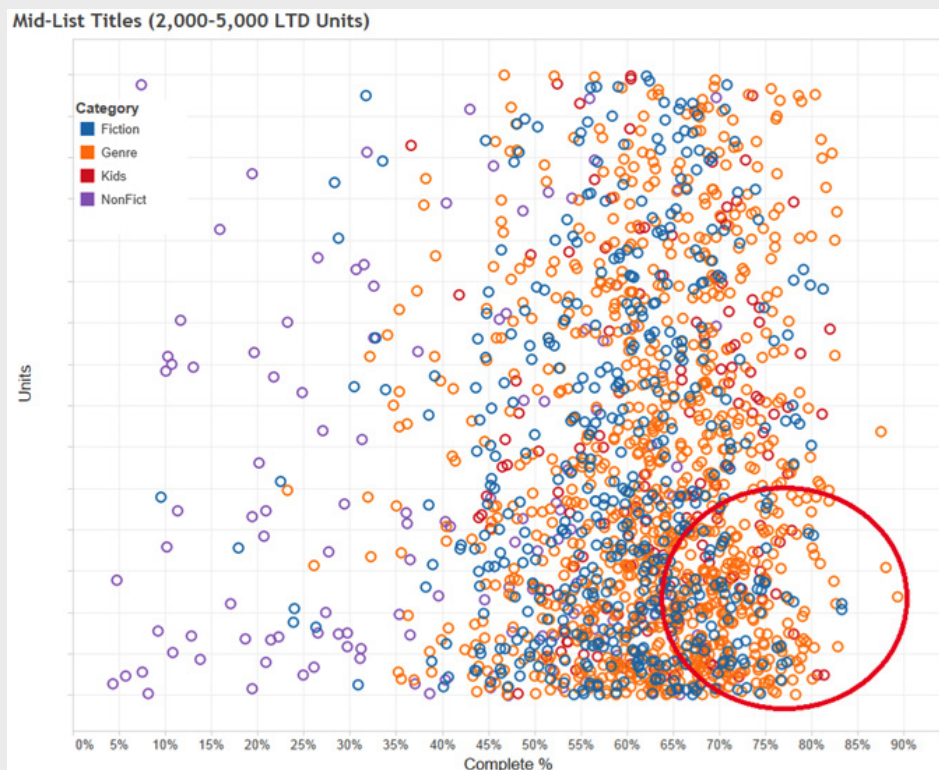
Discovering Hidden Equity

By looking at both sales and completion rates of titles, a publisher can gain an overall view of reader engagement.

An opportunity lies among those books that have high completion rates yet suffer low sales. Clearly those readers who have stumbled upon these books have loved them – so while the marketing team or editorial department may not have seen a winner among these titles, the reader did; it may well be worth experimenting with marketing efforts to see if the books catch on with a wider audience once they've been surfaced and attention is drawn to them.

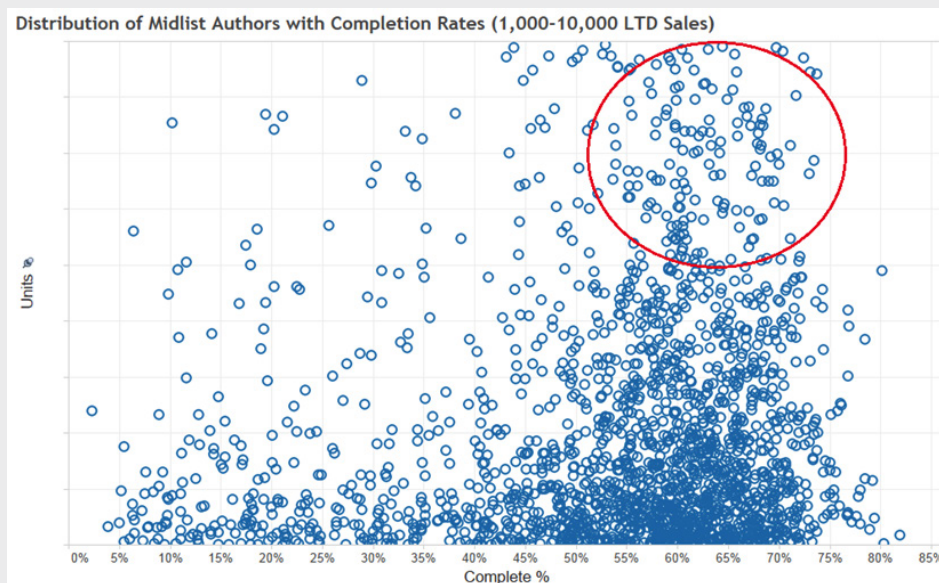
Over all, being able to identify true reader engagement allows for more targeted allotment of marketing and publicity resources. For single titles that are over-performing in terms of engagement versus sales, is there potential to create more awareness, a follow-up title, or turn the title into a series of books?

High-engagement, low-sales titles



Identifying the next Dan Brown

Similarly, the same metrics can be used to identify your next big author who currently may be hidden behind modest sales numbers. How? By looking at how readers are actually engaging with books written by those midlist authors, and using this information to identify which authors are outperforming industry and in-house benchmarks. Engagement metrics over time can also help identify when it is prudent to invest further in an author and what the size and shape that investment should take.



Recently, a publisher had a pricey deal on the table. One of its stable of tried-and-true bestselling authors was asking for a major six-figure, three book deal. The advance was considerably higher than the publisher had committed to in the past – the author was successful, with a number of previous bestsellers; nonetheless, the advance was sizeable and potentially a risk if sales plateaued or declined.

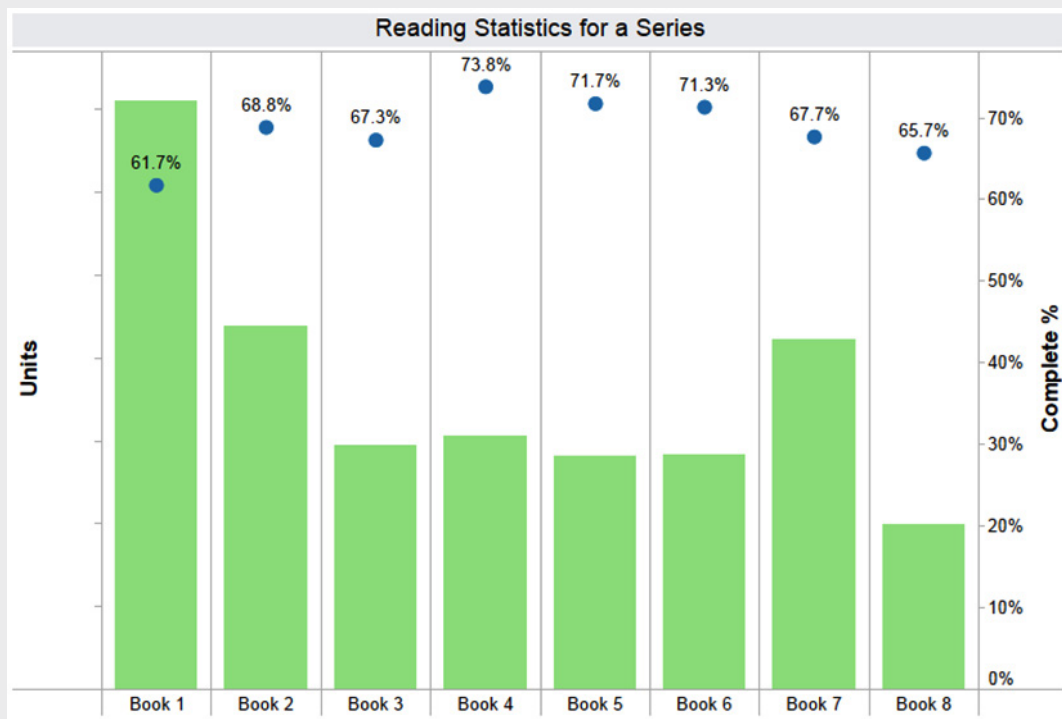
A look at the reading data, though, showed an encouraging story. Over the course of 12 books, engagement was actually increasing – more people were finishing books, and finishing them faster, as new books came out. The publisher had been investing in the author since the beginning, and determined that leaving the deal on the table meant they had built up equity for another publisher to take advantage of; as a result of looking at not only sales but engagement, a deal was struck in which the publisher was confident and which was borne out in sales.

In short, engagement data can be useful in plotting marketing and publicity strategies – the task shifts from hoping a book takes off to knowing that it will once people know about it. As well, past sales combined with rates of engagement can help arm publishers in negotiating advances and in making decisions on signing multi-book deals.

Betting on a Series

Metrics can also be put to work to identify the general health and wellbeing of a series, which in turn can inform either editorial, marketing or overall publishing decisions.

Take the following graph showing sales versus completion rates for books one through eight of a series:



In this example, completion rates vary considerably. The author seems to have hit a peak in book four, which has the highest engagement rate (meaning the highest number of people who finished the book once they started it) and remained steady through books five and six. From an editorial point of view, what was different about these books than the previous?

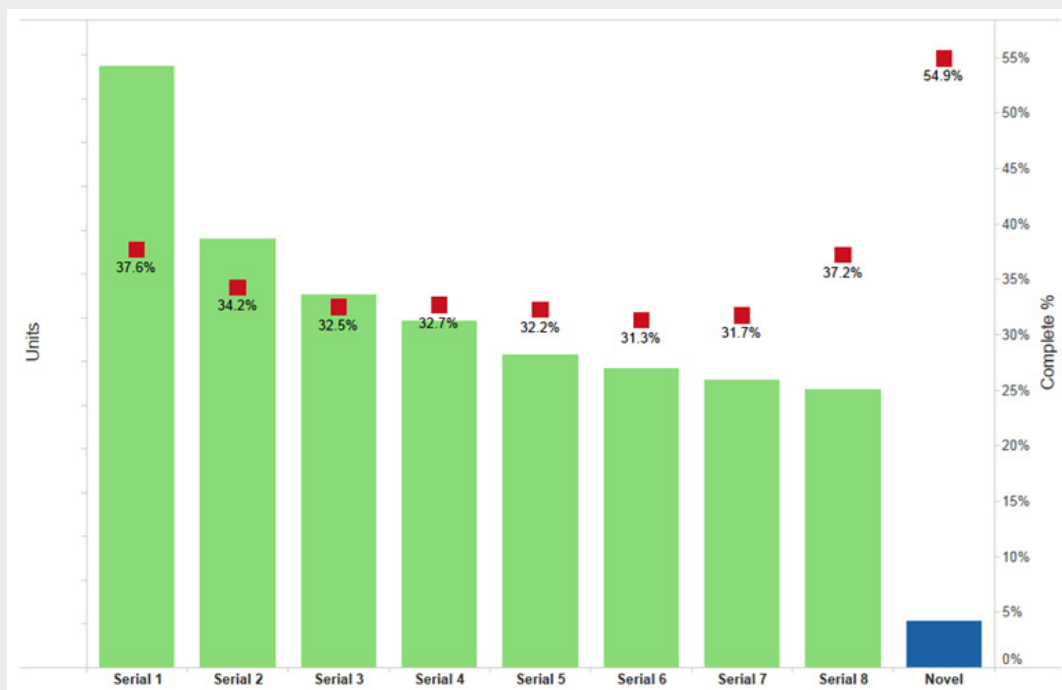
Meanwhile, completion rates began to drop off with book seven with a further decline in book eight.

This opens up a few questions if not answers: is it time to evaluate whether the series should wrap? Or whether to continue to invest in the author? From an editorial point of view, was there a change in books seven and eight which warrants a waning of interest? Can that change be corrected?

Serialization and the Viability of a Book

Serialization – breaking a book up into sections – is not only a great way to experiment with sales and possibly pick up incremental revenue, but it’s also a useful guide to gauging audience engagement with the whole.

Without access to reading data, the following serialization would appear to have been successful – sales were respectable and that would seem to predict strong sales once the full-length novel was released.

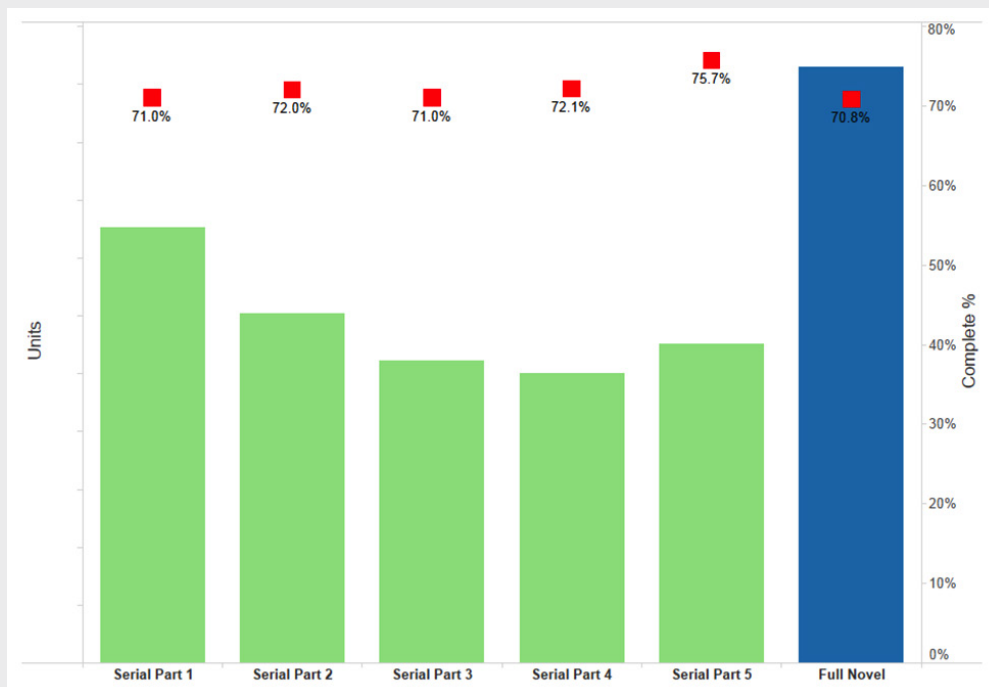


Alas, the novel sold poorly. Absent of engagement information, the publisher could easily blame serialization for the dismal performance, inferring that the serials cannibalized sales of the full-length novel.

A look at the completion rates or engagement with each installment, though, shows that there is more to the story. Regardless of sales, less than 40% of readers were able to finish the first installment. However short that first taste was, fewer than half of those who started reading it got through to the end. The readers who sampled the rest of the serials fared no better. However compelling the premise of the story may have seemed, it held few people’s attention.

The fact that the novel didn't attract many sales shouldn't have been surprising. It simply wasn't a very good book from a reader's point of view.

Compare those results to a similar effort by another author working within the same genre and same price points for both the serialization and full-length novel.



In this second example, completion rates are significantly higher – more than 70% of readers completed the first installment of the book, a number that held steady or improved over the course of the serialization. When the full-length novel was finally released, sales soared – exceeding results of even the first serial. While this is very good news, it should not have been a total surprise given how engaged readers were with the component parts.

Serialization is a magic bullet for publishers. Not only does it offer the opportunity to increase overall sales, but it also can be a useful tool for predicting future success of the novel. For publishers who carefully watch and respond to engagement numbers, a serialization will also point out if and where readers are losing interest in a story. If there is a willingness to experiment and bring engagement data back to the editorial department, reading statistics can help identify pain points and areas where a story can be “fixed.”

The Best among Equals – Benchmarking Imprints and Genres

It's a brand new way of looking at the publishing business, but benchmarking engagement rates across genres and imprints can help a publisher drive performance, identify and fix trouble spots, and find more of that untapped equity hiding behind poor sales performance.

Here's an example.

The industry benchmark for completion in mystery books is 62%. For the most part, the following imprints fall in and around this mark; three of these imprints actually exceed the average with a 66% completion rate. But Imprint 2 falls behind with only 56% completion. What is happening in this imprint that differs from the others?

		Imprints with Completion Rates					
		Open %	Complete %	Avg. Time (hrs)	Avg. avgsessions	Avg. Session Length	
Mystery	Imprint 1	46.4%	62.1%	9.1	24.9	30.3	
	Imprint 2	49.5%	56.4%	7.8	21.1	31.6	
	Imprint 3	51.0%	66.0%	9.3	24.7	34.1	
	Imprint 4	48.2%	60.5%	8.3	21.9	33.2	
	Imprint 5	45.5%	66.0%	7.2	20.7	29.3	
	Imprint 6	49.3%	66.1%	9.2	26.3	31.8	
	Imprint 7	43.9%	61.9%	5.9	12.1	17.9	

Imprint 3 is most likely to be opened (51%) and one of the top imprints for completion; it is also the imprint that keeps readers engaged for longer – the session lengths, or the amount of consecutive time spent reading, is longer than the rest. Differences in open rates can be tied back to cover image selection, author branding and marketing or publicity efforts. Whatever it is, there is something about this imprint in particular that is working well and which may point to best practices worth adopting across the board.

Meanwhile, the same data can be used to identify where easy improvements may be able to drive sales. Imprint 5 is a good example of this – it has a low open rate (45%), but one of the higher completion rates (66%) – what is it about this imprint versus the others that doesn't have readers as excited to open books? Is it a matter of the title, the cover design or some other element?

Betting the Business

Basing business decisions on unit sales alone is only taking into consideration one part of the puzzle. Engagement data in the form of completion rates should take at least some of the guess-work and risk out of the equation.

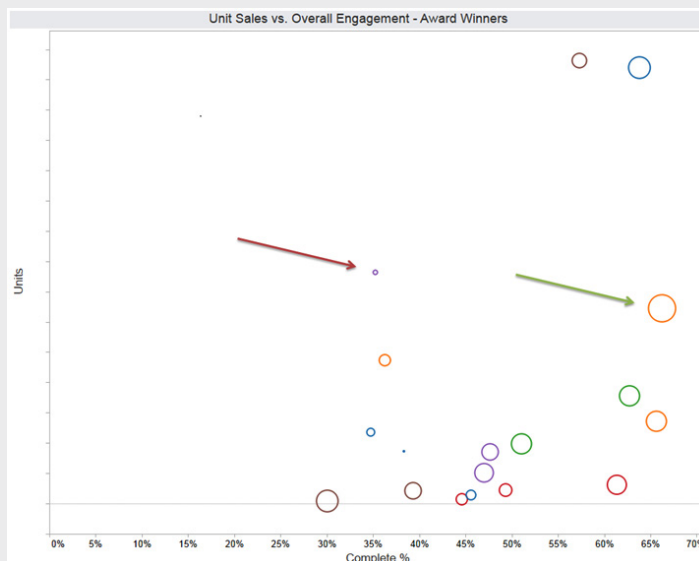
As frequently happens with print books, eBooks are often purchased with the best of intentions but end up sitting on the 'shelf' – readers do not always read their purchases, and in some cases, they don't even open them.

If Book A by an author sells X number of books – but only 40% of those who purchased it even opened the book, should you include the remaining 60% of readers in your projections or profit and loss statement? What if of the 40% of people who opened the book, 70% abandoned the book without finishing it? Predicting future sales of a new work by the same author will be risky if a publisher factors in only previous sales. Attention and engagement with the author's work is slight, and it is unlikely a large number of fans will buy another book to sit on the shelf, unread.

Another beguiling element in the business of publishing is the award-winner. Winning an award can lend a title an unexpected boost in sales, similar to the 'Oprah' effect.

But how much can you rely on the accolades?

Take the following literary award-winning titles:



All of the titles included in the graph above have won at least one literary award within the last 3 years. Similar to what we've demonstrated earlier, we measured both sales and completion rates. Added to this graph, as represented by the size of the bubble, are the open rates for each title.

Unit Sales vs. Overall Engagement - Award Winners		
Title	Open %	Complete %
Award Winner 1	95.3%	66.2%
Award Winner 2	25.7%	35.2%

Titles 1 and 2 both sold a comparable number of units – but their performance is actually drastically different. If you were a publisher with both of these books on your roster, it would be useful to know the differences between them when planning for new work by the same authors.

In the example of Award Winner 1, not only is this author an award winner, but has written a book that people actually want to read. An open rate of 95% is extremely high for any title. What was it about the award itself or more likely the marketing and publicity surrounding this book that got readers excited? What was it about this cover that begged readers to open it immediately? Knowing these things worked well, what can you do to duplicate these results in the future? Furthermore, a strong number of those who opened the book and started reading it got to the very end. Does this author have a backlist which can be promoted more aggressively? A new book in the works? Clearly this author is someone to invest in.

As for Award Winner 2, while sales themselves are comparable to Winner 1, a significant number of readers didn't want to open the book, and of those who did, only a smattering hung in to finish it. Why it was so poorly received after sale is a mystery worth considering when investing further in the author or the book.

The Bottom Line

We are at the very earliest stages of the possible when it comes to applying Big Data to the publishing world but even with these relatively simple tools, much can be learned to benefit overall business. Digital reading can help inform the business itself in a revolutionary way – understanding how readers experience your work can help in decision-making, uncover what would have been lost sales, and protect against costly errors in judgment.

In the near future we will have access to even more reading data. From this we will be able to gain insight into what “discoverability” really means for readers – where and how they find books, what triggers a purchase, what kinds of reading habits those bargain shoppers have versus those who buy at full price and the relative value of each. We will be able to track exactly where in a book interest begins to wane, which can be used to make editorial decisions and ultimately create a stronger title.

Digital books and eReading, therefore, are a powerful tool for building revenue and margin, and soon will help generate a deeper understanding of what readers are passionate about, offering insight into what ideas, stories, even what language is truly engaging. Publishers have the power to both create and define culture; the books they create are the stories we tell ourselves about our time, history and the future. Big Data is a window into what truly resonates – the only limit to what can be done with that knowledge is imagination itself.

