Review

Gaining insight from data

Rocket science and astrophysicists

To some marketers, big data and analytics are still dubious concepts with little to offer the time-pressed digital marketer. For those who attended 'Gaining data from insight', an evening event run by MBN Solutions and hosted by IBM at its Southbank Client Centre in London on 17 September, the real-world examples from Boots and the All England Lawn Tennis Association (AELTA) showed those concepts in use right here, right now. For any further doubters who assume you need a rocket scientist to follow those examples, the Hartree Centre revealed that it offers a way for commercial organizations to get involved (although its data scientists are more focused on astrophysics than rockets).

Boots

Martin Squires, head of customer insight at Boots, explained that the health and beauty retailer has a brand statement, 'Let's feel good'. The retailer aims to be the champion of everyone's right to feel good, helping its customers to look and feel better than they thought possible. 'We need to understand what it is that Boots offers that feeds that brand promise', said Squires. In essence, his insight and analytics team is required to understand female consumers (or 'get women', as he put it) since 95 per cent of Boots Advantage cardholders are women, as are 80 per cent of the company's staff (although within his own department the male/female split is a more conventional 50/50).

Care over data inferences

Boots' big data resource is every receipt it has issued through its electronic point-of-sale terminals in the last 17 years. By associating transactional data with demographic data through the use of its loyalty card, this enables the company to investigate spending patterns and identify trends in behaviour. Squires stressed that Boots is careful with the inferences that can be made from its data, for example that a customer may have a partner because of the presence of male-oriented products in her shopping basket. 'We don't get too carried away with the data — we apply rules', he said.

All customers are online

Data is consolidated from the 18 million Advantage cardholders who visit the wide range of Boots retail outlets, from its main stores to opticians, pharmacies and online presence. On top of this, it runs a customer satisfaction survey across up to 20,000 consumers and a consumer panel of 60,000 shoppers to provide more attitudinal insights. Squires provided an example of the way consumer behaviour has changed and the impact it has on data and analytics. 'Ten years ago, we built a customer segmentation and had online shoppers as a specific group.

When we rebuilt it recently, they were no longer a distinctive segment — everybody is now online. All of our customers are now multi-channel', he said.

Near real-time optimization

This has created a new pressure on the business from its customers, who expect it to share data across all of its outlets and recognize them accordingly. With over 20 per cent of website views now coming from a mobile device, data use and personalization needs to take place rapidly. To support this, Boots has built an optimization process on top of its single customer view that selects customers, assigns them an optimal offer mix based on propensity scores, and then reviews the outcomes in order to improve the effect the next time around. 'It is not real-time, but it is faster than ever', said Squires, noting that the retailer has the full stack of IBM tools, from a DB2 data warehouse to SPSS modelling software, Coremetrics web analytics and Unica campaign management tools.

Anybody can be taught to code

Squires considered the current vogue for data scientists to handle the new complexity of data sources. 'The biggest thing is to get the right people. Do they need to know R, Javascript and Python, as well as having the ability to tell a story? I prefer to look for the four Cs — common sense, communication, creativity and curiosity. We can teach anybody to code, but we can't beat stupidity out of them', he said. Analysts in his department develop around 1,200 models across each sales period, with 100 permanent models in place for products that are always on sale at Boots.

Rules-based marketing

One effect that Squires is wary of is the over-optimization of marketing so that the top 10 per cent of customers, who demonstrate the highest propensity to purchase because they always buy from the store, receive a lot of offers, whereas the lowest 30 per cent of the customer base would get none. 'So you have to apply rules to optimisation, including rules from our suppliers, for example that their brand only appears in certain types of communications and not alongside certain other brands', he said. Out of the 500 possible offers that could be made to Advantage cardholders, the five best are selected. For each loyalty card statement, that produces 120 m possible variations. 'They are not fully individualised, but we get close to one-to-one marketing', said Squires.

Market research

Although large-scale data provides the ability to model out propensity in this way, he also noted that conventional market research is also important. 'It allows us to pick out the reaction to a product, not just its sales. A new fragrance may sell well at launch, but that doesn't mean our customers actually enjoyed it', he said.

All England Lawn Tennis Association

The All England Lawn Tennis Association faces a different problem when it comes to personalization. For the two weeks of the year when the Wimbledon tournament takes place, it welcomes 490,000 visitors to its grounds. But it has a global TV audience of 1 billion. 'It wants to drive fan engagement and help those viewers get closer to the tennis. For visitors to the website, it wants to create a digital experience that is the next best thing to being there', explained Chris Nott, CTO big data and analytics at IBM UK and Ireland. A team of 150 people work to capture the 3.2 million data points generated across the tournament. They include 48 specially trained tennis players working court-side in pairs whose task it is to capture the

speed of every service, its direction, the number of shots in a rally, where each shot was placed and how the point ended, all within one second. 'It is easier to train a tennis player in data entry than to train an IBMer on tennis and whether a shot was a forced or unforced error', said Nott.

Aggressive play algorithm

The data set is relied on by commentators and media, as well as being provided to players (along with edited video) within 20 minutes of a match finishing for coaching purposes. Each day at Wimbledon, IBM provides the BBC with 1,000 TV graphics and advises on selecting the ones that best illustrate the run of play. An innovation for 2015 was an indicator of aggressive play using an algorithm looking at how hard the ball has been hit, where it landed and how far the opposing player had to run. It also created a Championship Tracker that identified the top three things each player needed to do during a match to improve their chances. Fans get to access much of this data via the AELTA website and social media feeds.

Leyton Hewitt and Sam Groth

'That had an impact on Day One when Leyton Hewitt was playing his final championship match. We notified the Communication Centre when he was on 1,497 winners so they could prepare digital content about him being fourth on the list of all-time grass court winners. When he hit his 1,500th, they pushed that out,' said Nott. 'We did something similar on Day Six when Sam Groth hit a 147mph serve, just 1mph off the record. We tweeted a poll to fans on whether he would beat the record or not.' (For the record, most fans thought he would, but he didn't.)

IBM Watson

Nott pointed out that 80 per cent of website users come via the desktop, since many have a browser screen open while they are at work, but conversely 55 per cent of page views come via mobile devices. To deepen fan engagement, IBM also deployed its Watson artificial intelligence and natural language processing to provide answers to questions sent in, as well as to deliver insights to media. This might mean a simple data query, such as how many French players had reached the third round, or a more semantically challenging one, such as 'What is love?' (Watson correctly responded with an answer about the scoring system used in tennis.) To maintain availability, IBM applied its own predictive analytics to forecasting the load on the system and ensuring sufficient capacity was available across its triple-centred data processing operations. Maintaining data security is another key challenge; according to Nott, there were more Dedicated Denial of Service attacks during Wimbledon 2015 fortnight than for the whole year of 2014.

(Readers can learn more about the marketing activity driven by AELTA and IBM through 'The Punnet' communications centre at the 2014 Wimbledon Tennis Championships in this case study: www.palgrave-journals.com/dddmp/journal/v16/n4/full/dddmp201523a.html)

Science and Technology Facilities Council

Probably the most eye-opening presentation of the evening was given by David Moss, advanced technology solutions manager at the Science and Technology Facilities Council. STFC is one of seven research councils operated by the Department for Business Industry and Skills and it has the task of funding, designing and building world-class scientific infrastructure. 'We built the sensors and detectors for the Large Hadron Collider. Proving graphene was done using one of our centres', said Moss.

Square Kilometre Array

The scale of the data processing and management facilities operated by STFC is truly staggering — its Blue Joule petabyte system built three years ago can run a quadrillion calculations every second (this would take 19 years on an iPad). This capability is essential to support the new Square Kilometre Array radio telescope when it comes onstream, since it will involve 268,000 petabytes of correlation output. For STFC, a more 'normal' scale of facility is its big data analytics cluster, which has 288 terabytes of capacity.

The Hartree Centre

'We also employ a lot of wonderful people,' added Moss. This is the most interesting and critical aspect of the Council's role, focused on the Hartree Centre (which IBM is closely involved with). Moss explained, 'Our people are why organizations come to us, because we speak the language of industry. We collaborate on research and development projects as part of the Government's move away from the old model of just spending on science — now we have to show the value it returns, intellectual property it generates and our engagement with industry'.

GSK and Democrata

The Hartree Centre has worked with GSK on a bio-medical project, running ten years of its research data through the centre's unique data visualization facilities to help researchers to identify insights. Similarly, it worked on a project with Democrata using open data sources to map potential archaeological sites. Given that 3 per cent of the costs of major infrastructure projects go on archaeological discovery and excavation, being able to predict where construction might intersect with a site of historical interest can help to save money. 'We modelled where we might find archaeology using York University's archive of 40,000 records and Watson's content analytics to build a map of sites of interest', said Moss.

Knowhow and Hampshire County Council

Another project saw the centre work with Knowhow and Hampshire County Council to predict where water-related emergencies might occur. This has real-world implications for where emergency response vehicles should be stationed. 'Flooding accounts for between 8 and 12 per cent of road accidents and 32 per cent of flood-related deaths are in vehicles', said Moss. 'The arrival of an emergency vehicle at 10 minutes compared to 5 minutes of an incident leads to a four-fold increase in the risk of a fatality'. By combining a range of data on flood risk and accidents, the project was able to predict to 94.2 per cent accuracy where flooding would occur.

Commercial opportunities

The Hartree Centre offers this kind of big data analytical support to commercial organizations of all descriptions as part of its mission to engage with industry. With the competition to recruit data scientists now very intense and the start-up costs of big data projects sometimes prohibitive, this is a significant opportunity for any business trying to unpick a complex problem or looking for fresh insights to benefit from a publicly funded resource. This aspect of the centre is not currently well-known outside of a limited sphere, so now is the time to join the queue.

David Reed FIDM