Guiding Business Innovation using Big Data and Data Analytics (Part 4)

1 April 2015 | **Business Innovation, Digital Transformation** | **Wolfgang Emmerich**
Reading time: 6 minutes

Business innovation in many of Sawhney et al's innovation dimensions requires a thorough understanding of a number of different concerns:

Where are new customer segments?

What do customers in these segments really want that a firm could provide?

How can the firm extract more value from existing customer relationships?

Where is the unmet demand that a new or existing product or service could meet?

Where is the right market for an existing or new product or service?

What is the market's response to a change in offering or branding?

The quality of the answers to these questions has a profound impact on the probability with which business innovation investments yield the desired returns. The answers should therefore be based on sound analytics of relevant data rather than gut feelings, anecdotal evidence or interactions with statistically insignificant focus groups. Unfortunately, that is currently often not the case. This is because providing a definitive answer is difficult. It requires analysis of sizeable datasets that have hitherto not been available and a firm grounding in statistics, in order to analyse these data sets. The "Big Data" phenomenon is rapidly changing this however. The key idea of Big Data is that storage of data on commodity hardware is very inexpensive and therefore there is no economic need not to keep data. Once it is kept it can be subjected to different forms of analyses which often cannot even be imagined at the time the data is collected.

Customers of firms leave a significant trail of data behind in both digital and real interactions with firms. Examples include point of sales data that is collected by retailers, details of e-commerce orders, flight bookings, hotel reservations, banking transactions, search histories, credit card records, records of phone calls and many more. In addition to data that firms keep themselves, they can obtain large data sets from the public domain. The Open Data Initiative in several jurisdictions obliges government departments and agencies to make the datasets they collate publicly available. In the UK this includes public health data,

prescription statistics, geographic data, traffic data and many other data sets. Moreover, data can be about offerings that their competitors have published electronically. For example, every online retailer needs to publish a catalogue of their portfolio together with pricing details, sales campaigns and goods' availability on the web. That data can be mined for competition and pricing analysis. There are large public data sets available, from which conclusions can be drawn about both customer behaviour and possible competitors. Customers might post content on social media, such as Twitter, Facebook, LinkedIn and the like which is publicly accessible. For example, search engine providers, such as Google, have collated large data sets, such as geographic information about the location of companies, their financials and other details that they have mined from the world-wide-web and that can be accessed through APIs.

These data sets contain authoritative answers to business innovation questions. For example, retailers can analyse point of sales data sets to profile and classify customers, and they can then use the insights gained to extract additional value about the customer relationship, for example through targeted advertisements. Telecom operators wishing to open a new retail point of presence can use position information in mobile phone records to search for optimal outlet locations in order to maximise customer footfall. Mobile telecom operators might also wish to innovate by offering payment services. They can use traces of position information which is available in phone records and overlay these with geographic information about retail outlets in order to see which outlets customers visit, in order to establish relationships with these retailers and get them to accept payments by phone.

The problem many businesses have is that conventional data analysis techniques, that involve extracting data, loading it into a data warehouse with a carefully designed schema and then writing business intelligence reports, do not really work in this settings for a number of reasons. Firstly, the data sets are far too heterogeneous as they might involve structured, semi-structured and unstructured data that cannot be squeezed into a relational data warehouse schema. Secondly, the questions that will be asked about the data are typically not known at the time when the data is collected so it is not possible to define a suitable schema to store the data. Thirdly, the time available to answer the relevant investment questions is typically an order or two shorter than the time that it takes to complete the required significant data warehouse initiative. Finally, the infrastructure and licensing costs for data warehouse environments are often prohibitive.

Fortunately, a useful eco system of tools and techniques to support the required management and analytics of big data sets has emerged from online advertisers, most notably Yahoo. These tools are made available as open source tools by the Apache Software Foundation that are bundled, packaged and supported by Hortonworks and Cloudera as

platforms for Big Data. These tools support the storage of vast heterogeneous data sets in so called "data lakes". Data in these lakes can then be transformed and analysed using distributed computing resources in a number of different manners in order to answer the required innovation questions.

But where do firms start? Do they build up the required skills and competence in data analytics, machine learning and management of big data infrastructures or do they partner with a provider who specialises in these techniques? Once they have hired their own resources or obtained skills from specialist partners, how do they organise the required programme of work? How will the collaboration between strategic innovation and data scientists work?

Analysis of Big Data requires cross-functional teams that combine skills in understanding innovation strategies with data science techniques, such as computational statistics and machine learning, distributed system programming and advanced data visualization of the insights that are obtained. In our experience, these teams work best in an agile manner in weekly or fortnightly sprints and generate answers to business innovation questions that inevitably will produce further questions.

I would welcome any feedback and examples about how some of these techniques described in this post have been used for business innovation.

You can view previous parts of this blog here: Part 1   Part 2   Part 3