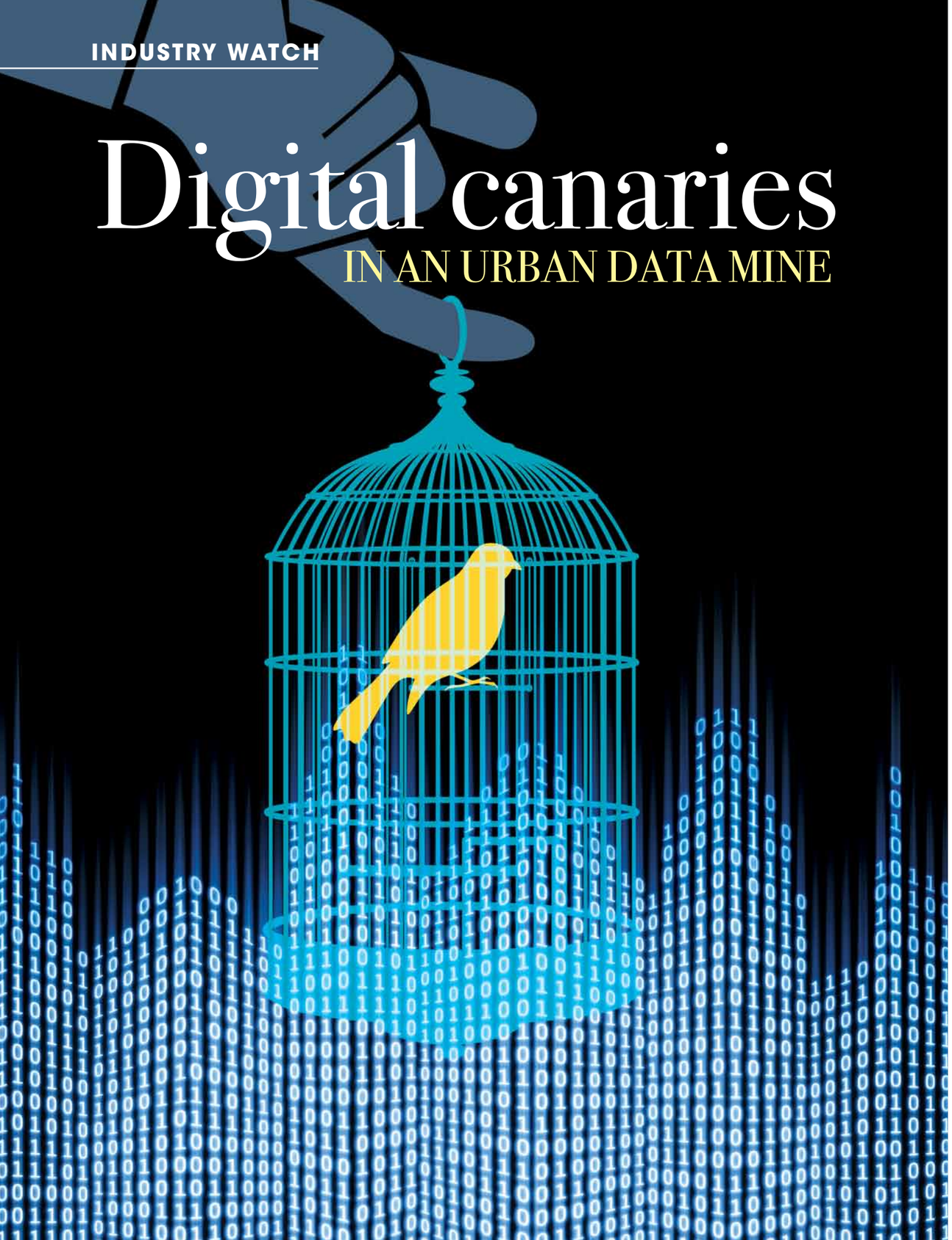
5-2014

Canaries in the Urban Data Mine: Analytics for Smarter City Life

Robert J. KAUFFMAN
Singapore Management University, rkauffman@smu.edu.sg

Digital canaries
IN AN URBAN DATA MINE



Technology-based sensors and data analytics have created unprecedented levels of informedness for consumers, corporations, the public and government agencies. The information they generate provides a basis for smarter cities and more sustainable urban living.

By Robert J. Kauffman

Two centuries ago in Great Britain, coal miners took caged canaries down into a mineshaft, along with equipment, gas lamps, food and water. The canaries acted as environmental sensors for threats from poison gases. They reacted quickly to adverse conditions. And if a canary fell from its perch, this signalled it was time for the miners to get out.

Today, technology-based sensors, acting as "digital canaries", make us more informed than ever before. We learn about social sentiments from social media. We obtain online news via mobile devices. We get frequent micro-locational weather reports. And we receive updates on traffic and transit delays, stream flooding, and other problems. We also obtain news of air pollution through environmental sensing and crime downtown through public safety monitoring. Business managers, meanwhile, learn about and leverage data streams that cover logistics, supply chain management and product movement from store shelves to consumers, while government agencies and officials track political opinions of citizens and social programme effectiveness. At the 2013 Wimbledon Tennis Championships, the organisers even appointed a "Twitter Chief" to inform people who were queueing for stadium entry about when they would no longer be able to see the men's final.

There are sensors everywhere now. Through them and the informedness they create, it is possible to do better planning, risk management and public communication, and pre-empt problems-in-the-making.

Digital canaries

Having data to create relevant information supports an understanding of the rich tapestry of urban life in ways never before possible. Sensors in the city act as "digital canaries in an urban data mine", with so much information to be revealed and leveraged for smarter living. Urban data analytics tell us things we need to know so sustainability becomes possible—for consumers, corporations, the public and government agencies—in contemporary city life.

As large-scale data have become available, so have emerging analytics methods based on multidisciplinary knowledge: machine learning, text mining, Bayesian statistics, randomised closed-loop network experiments, social networks methods, and more. These are elements of a new *computational social science* coming together for an understanding of how urban activities can be understood. This will support the emergence of a new *e-social science*.

Data to information

Information is valuable. However, *data* captured in smart city analytics processes only have potential value. Their value will be unlocked though a process of *transformation*—the mapping of data facts to useful information for decisions and actions.

The *value of information* for data analytics applied in urban settings is the value of a decision made in the presence of the information minus the value of the same decision made without it—adjusted for the acquisition and analytics process costs. Fortunately, the costs of its acquisition have dropped dramatically of late.

The *value of information* for data analytics applied in urban settings is the value of a decision made in the presence of the information minus the value of the same decision made without it—adjusted for the acquisition and analytics process costs.

Working with big data supports efforts to achieve smart city life. We can capture data in large *stocks,* based on their storage in data warehouses as transactions, sales records, inquiries, taxi pick-ups, inventory levels, water drain gauge levels, airline fares, and housing starts. We also can acquire data in near-to and real-time *flows:* social sentiments, financial market prices, and seller inventories in transparent Internet markets. Flow data are used for electronic road pricing for highway access, and online reservations at our favourite restaurants.

Contextual information is important too. Examples are stock prices and parking in different areas of the city. Individual data points only have meaning when we know their means, variances, maxima and minima, and the stability of the readings we obtain about them over time. We want to know what causes them to change, for example, announcements of changes in interest rates or big-draw sporting events.

This is true for collections of large time-series data sets as well. For example, the social sentiments that emerge in different issue areas involving quality of urban life-related issues—the availability of housing for young families, the effectiveness of healthcare outreach and services for senior citizens, and the reliability of public transit services. Are the social sentiments trending up or down? Influenced by recent governmental actions? Or stable?

Obtaining useful information requires knowledge of data in a context before it can be used to formulate strategies and actions to make desirable impacts. Information makes people and decision-makers informed. This can be about whether it makes sense for rail transit planners to experiment with differential prices during peak hours. Or whether the timing is right for a young couple to buy a home. It also includes something so simple as gauging if it makes sense for a commuter to carry an umbrella to work.

Informedness is key to the operation of smart future cities. It arises as: *consumer informedness, firm informedness, citizen or social informedness, and government informedness.* In a smart city, informedness has diverse facets—as an understanding of consumer choices, knowledge of regulatory impacts, awareness of service effectiveness, and assessment of how actions to promote change need to be implemented for impact. So it is critical to identify the kinds of data to acquire, to provide the greatest fidelity for the analysis of the problems in the bull's eye.

Strategic data analytics process

Researchers at the Living Analytics Research Centre (LARC) at Singapore Management University are studying settings in business, consumer and social insights research that involve the "digital traces" of urban living. Data in corporate, public and social domains can be tracked, mapped and understood as a basis for improving strategy and taking appropriate actions. For example, investigating urban rail transit and hospitality services, retail telecoms and digital entertainment, financial services and e-commerce, and public safety and election campaigns have all become commonplace. Some data occur as sentiments expressed in social media, others as transactions in ATM networks and bank branches, and still others as consumer purchases in shops, restaurants, and online.

Considering these settings—and the role of data, the costs to transform it into meaningful information, and the value it produces for organisational strategy and informed action—I see a common *strategic data analytics process.* It involves data, information, strategy, action and evaluation, and can be iteratively implemented, so the most value can be achieved by organisations that use this process (refer to Figure 1).

Dr. Jamshid Vayghan, chief technology officer for IBM's Global Business Services in Latin America, echoed this view in a speech at Singapore Management University in August 2012. Data analytics, he argued, should be managed as a business process—and one that reflects repeatable best practices. Appropriate requirements analysis, functionality and usability design, cost-effective and action-relevant data capture, and maintenance and updating of data and algorithms that constitute the data analytics must be undertaken. Building strong data analytics capabilities requires human, process and technology capital investments. And they all need to be focused on building best practices for the organisational processes that produce strategic business, consumer and social insights. This allows goals to be targeted so that current products and services can be refined, new ones can be designed, adjusted business and social processes can be introduced, and new programs can be launched. Better strategy formulation is the main driver of data analytics.

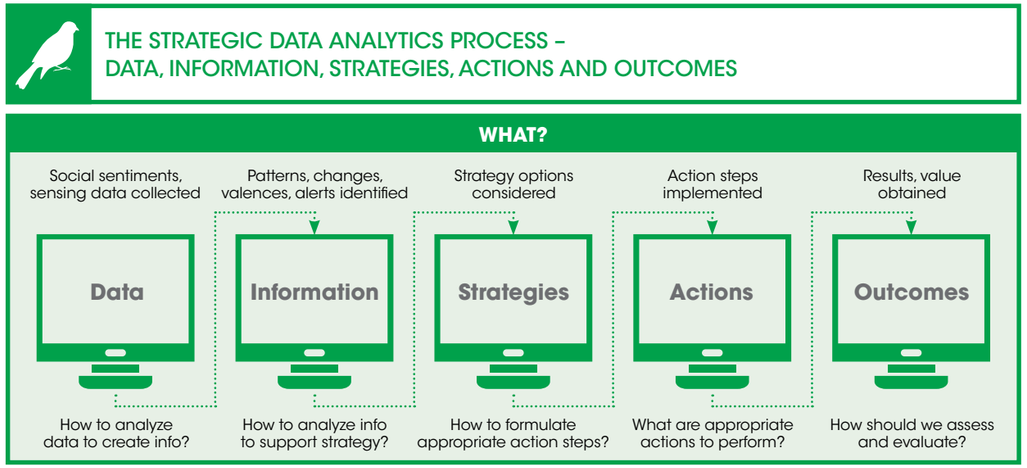


FIGURE 1 *Source: Kauffman and Lim-Wavde (2013)*

Formulating effective data analytics strategies requires an understanding of how value is created for a targeted set of stakeholders. Value takes on many guises in different urban contexts: as healthy profit margins achieved in business by local firms; as effective risk management protocols for contagious disease control in public health; as citizens' well-being with clean indoor and outdoor air; and as knowledge-based trust and understanding through the public's informedness regarding government policy actions. This goes beyond business value to citizen and social welfare.

Strategies to actions

Big data analytics support organisational programmes for a range of strategic objectives. An example is *observational micro-segmentation of consumers.* No longer is it necessary for organisations to employ high-priced consultants and generic market-wide segmentation templates to establish their market segments.

In LARC, my research group has analysed societal-scale data related to the density of household cable TV viewing in different segments. The analysis work showed enormous flexibility in identifying channel and programme viewership micro-segments based on 30-second data from two-way TV set-top boxes. We also explored the extent to which households efficiently use the spectrum of channels to which they subscribe, and learned that viewing time is a key determinant of the outcome (refer to Figures 2 and 3).

The formulation of effective data analytics strategies requires an understanding of how value is created for the targeted set of stakeholders.

THE STRATEGIC DATA ANALYTICS PROCESS – DATA, INFORMATION, STRATEGIES, ACTIONS AND OUTCOMES

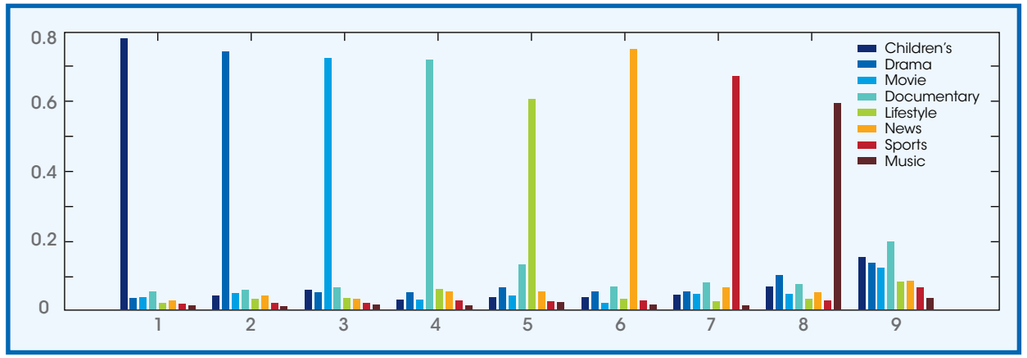


FIGURE 2

Source: Living Analytics Research Centre, 2014

SPLINE REGRESSION FOR CHANNEL VIEWING EFFICIENCY

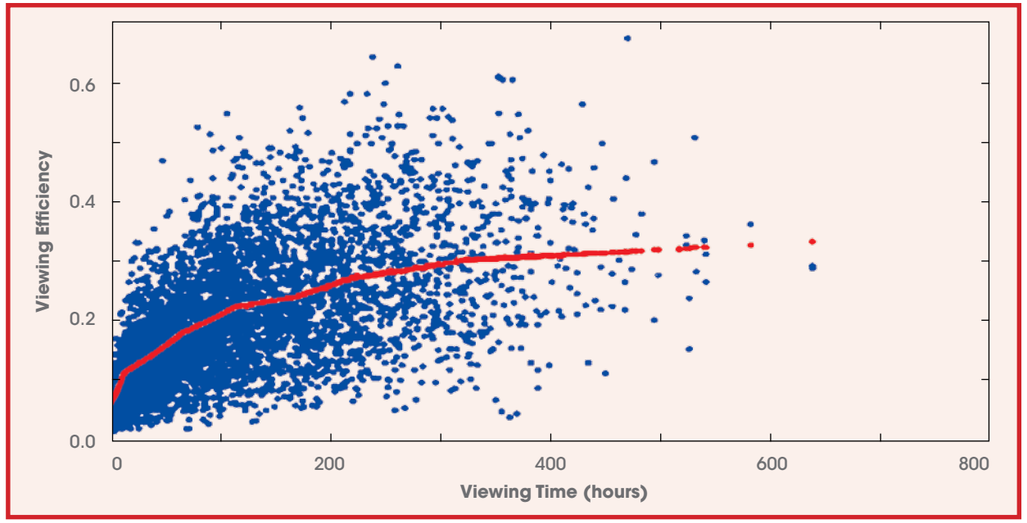


FIGURE 3

Source: Living Analytics Research Centre, 2014

Public information and communication. Another context for the application of urban data analytics is *strategic public information and communication enhancement* in social media. Analytics help public organisation and government agency communications to be effectively constructed, targeted and timed to have beneficial effects. In joint research that I have been conducting with computational and behavioural science research scientists at Singapore's Agency for Science, Technology and Research (A*STAR), we are studying the tie-in between tweet volumes, and positive and negative social sentiments from Twitter users related to the city's MRT (Mass Rapid Transit), and other current social issues (refer to Figures 4, 5 and 6).

TWEETS FOR "MRT" JANUARY 1 TO FEBRUARY 7, 2013

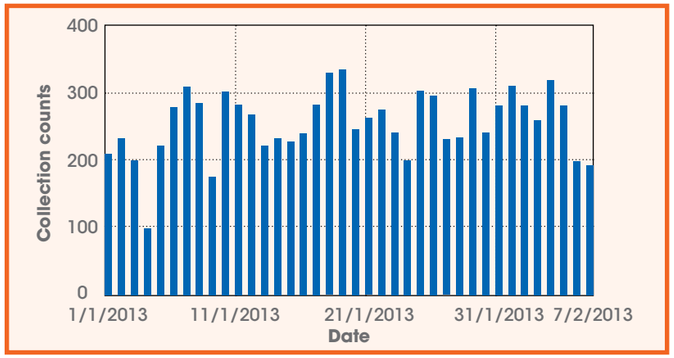


FIGURE 4

Source: Wang, Kauffman, Goh, Yang and Chin (2014)

POSITIVE AND NEGATIVE TWEET DENSITIES

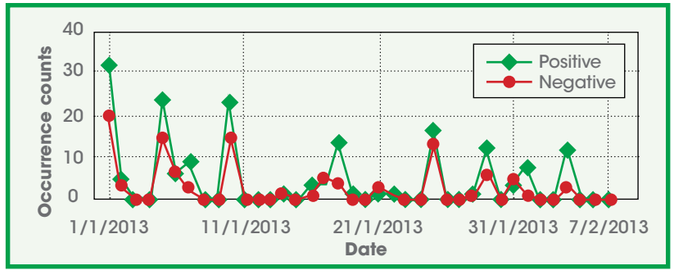


FIGURE 5

Source: Wang, Kauffman, Goh, Yang and Chin (2014)

POSITIVE AND NEGATIVE TWEET TRENDS

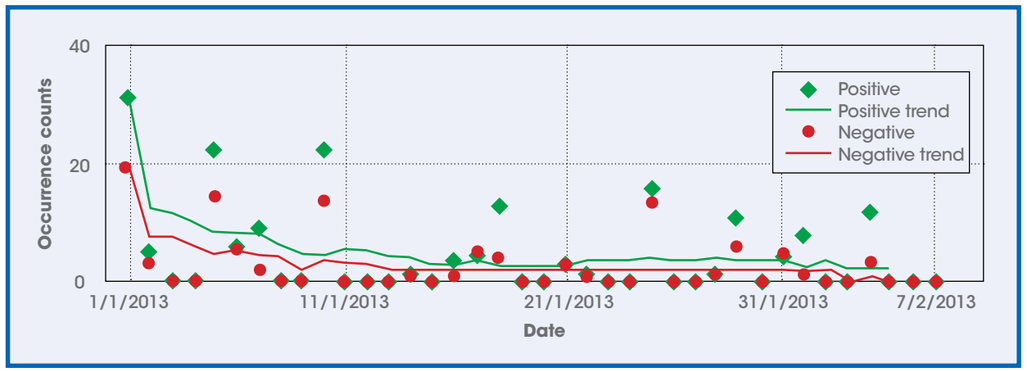


FIGURE 6

Source: Wang, Kauffman, Goh, Yang and Chin (2014)

We have attempted to connect the social sentiments to news in Singapore's *The Straits Times*, as well as social media news and announcements that act as stimuli for social reactions. We use various approaches to identify the evolving density of social expression, while controlling for prevailing baseline levels of tweet activities, and recognising the sentiment valence patterns (positive, neutral, negative), and the sub-negative categories of anger, anxiety and disappointment.

Geolocation and contextual awareness. Another application area for city analytics is the use of mobile phone tracking, GPS, and location-based marketing, to maximise the benefits consumers experience while they shop. The goal is to create *contextual awareness* for sales and service that goes beyond what a consumer knows herself.

In the past, many different kinds of stores–department stores and even bookstores–were "information deserts" for the consumer. Like water in the desert, little information was available to a customer in most stores. If you happened to have a mobile phone, you could obtain more product information than most store employees typically knew, and make well-informed price comparisons and purchases. New in-store and in-mall applications have emerged, making it possible for consumers to receive micro-location-based recommendations, coupons, discounts, and promotions.

Key sensors of consumer behaviour in the city are the many credit and debit cards that we carry in our purses and wallets. Card products from financial services firms are ubiquitous and soon our mobile phones will support card-free, contactless mobile payments. They all have the capability to track what we buy and where we buy it. Our daily transactions will become the digital traces of our lives as consumers.

The cards offer "15% off at ESPRIT", for example, co-branding incentives with airlines, malls and merchants, and micro-segment focused loyalty vehicles. This might be successful and independent women, as is the case with DBS Bank's recent "DBS Women's Mastercard". Another is Standard Chartered Bank's card use-focused mobile recommendation application, "Breeze", for location-based promotions, categories and merchant recommendations.

All these provide searchstream and consumer purchase data, with time, merchant and location stamps, and spend amounts. These create the basis for new and valuable insights on *consumer spatial transaction trajectory* over time for the firms that effectively data-mine the relevant digital traces. This will yield business value that the merchant and consumer can share: higher firm profitability via greater consumer willingness-to-pay levels. Value-sharing in all kinds of social settings is desirable.

Data analytics: limits to value?

Tropical rainforests showcase the complexity and interrelatedness of ecosystem life, with many species competing for the available resources. Cities are like that too. The challenge of making life in cities comfortable for people is a matter of managing the complexities and the interdependencies among the inhabitants.

Integrated data analytics that bring together the point-of-service or the point-of-sale with an organisation's data warehouse and operational systems can create enormous value for consumers, business partners, and citizens and public services providers. They can build new levels of informedness to support customer and user-focused product and service delivery. This is true for the operation of taxi fleets, city buses, rail transit, tourist attractions, and other urban services. Hospital emergency rooms, and water and electrical utility services are all subject to high demand at different times too.

Data analytics, no matter their level of sophistication, will be based on economic models of human behaviour in the urban environment. Nobel Prize-winning economist, Herbert Simon, has reminded us: "Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behaviour over time is largely a reflection of the complexity of the environment in which we find ourselves."[i]

Indeed, cities are the ultimate in complex human systems. It will be challenging for data scientists, business analysts and policy-makers to address the problems that arise due to the human complexities in city life. The decisions that city dwellers make–where they shop for food, whether they own a car, how much energy they use, when they commute to work, to what entertainment services they consume–can be modelled (if imperfectly) with data analytics that support predictions.

Difficulties will arise, of course: not all consumers make rational choices–even with "perfect" information about products and services. What's more likely? That decisions based on bounded rationality still will abound. And although such decisions may be "good enough" in some sense–for the individual perhaps, they still may not truly be good in the social sense. In many settings, "one person's music is another person's noise".

Thus, one can expect that business and social systems built on data analytics still will have "hits and misses" in their predictions about consumers, firms, citizens, government agencies, and societal processes. But the smart money should be on their having impacts that involve greater pleasure for most, with a modicum of pain for some.

So it makes sense to raise a caution flag... at least to half-mast height. What are the potential limits to value for data analytics in support of smarter city life? Can the "digital canaries" be our sentinels and saviours in different issue and policy areas? Will they help us to organise systems that make sustainable urban living more achievable?

Due to the burgeoning populations, limited land and increasing industrialisation of Asian cities, we must assess the issue areas in which the new sensor mechanisms will provide the greatest societal return on investment. Will they become strategic necessities, but still not sufficient to deliver desired levels of social welfare in different settings? In what settings will they offer more complete solutions: mass transit, healthcare and housing? Business or social services?

Herbert Simon offered some hints in a newspaper interview in 2000, "I don't care how big and fast computers are, they're not as big and fast as the world."[ii] Cities, especially Asian cities are ever bigger, ever faster-paced, and ever more complex. Yet other sage advice from the British statistician, George Box, suggested that "some models are useful" for many problems. He admonished us: we must "[r]emember that all models are wrong; the practical question is how wrong do they have to be to not be useful?" [iii] Still, data analytics for business, consumer and social insights have a bright future. They have the potential to improve urban social welfare through the greater informedness they create, and smarter cities will be better places to live in as a result.

> "Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behaviour over time is largely a reflection of the complexity of the environment in which we find ourselves."
> – Herbert Simon, Nobel Prize-winning economist.

Robert J. Kauffman
is associate dean (Research, School of IS), Deputy Director (Living Analytics Research Centre), and Professor of IS at Singapore Management University. He was the Lee Kuan Yew Faculty Fellow for Research Excellence in 2012-2013

Reference

[i] Herbert Simon (1996) "The Sciences of the artificial", 3rd edition, Cambridge: MIT Press.
[ii] Spice, B., "CMU's Simon reflects on how computers will continue to shape the world", Post-Gazette News, October 16, 2000.
[iii] Box, George E.P., Norman R. Draper, 1987, "Empirical model-building and response surfaces", Wiley.