The Holy Trinity: Big Data, Analytics & Real Estate *(Part 1 of 2)*

APRIL 2015

AUTHORS:
ALEKS VELKOSKI, DATA SCIENTIST, DATA ANALYTICS GROUP, NAR
TODD CARPENTER, MANAGING DIRECTOR, DATA ANALYTICS GROUP, NAR

AUDIENCE

- **REALTORS®, Brokers, and Associations** who are interested in understanding Big Data and its practical applications for real estate.
- **Brokers** who are interested in tools available to their agents and their brokerages.

ABSTRACT

This paper provides a primer on big data—what it is, how it's collected, and how it's analyzed—offers examples of real-life applications, and provides advice and tips on how REALTOR® associations and real estate professionals can apply big data practices to their business.

WHAT IS BIG DATA, & WHY SHOULD I CARE?

Big data promises some exciting applications for real estate. However, before you can begin to envision ways to use big data in your business, it's important to understand the concept. Although there's been plenty of media hype about big data's promise to transform businesses and upend traditional players in various industries, big data is widely misunderstood.

Even experts have a hard time agreeing on a definition. Among all the definitions, though, are three common elements: size, complexity, and speed.

In big-data vernacular, these characteristics are known as the "three V's":

- Volume
- Variety
- Velocity

These are quasi-tangible concepts: We know what they represent, but we can't precisely quantify them. In other words, there's no size, complexity, or speed beyond which data is considered big. What we *can* say is that data is big when its size and complexity—and the speed needed to acquire and process it—make it difficult to manage and extract information using traditional systems, technologies, and techniques (e.g., desktop analysis applications, relational databases, and extract, transform, and load processing).

DATA VOLUME: ALWAYS GROWING

Data volume is often measured in bytes, with one byte roughly equivalent to one character of text and one gigabyte—or one billion bytes—equivalent to eight minutes of high-definition video. The world's ability to store data has grown exponentially, and continues to grow. In 1986, for example, world capacity was estimated at 2.64 billion gigabytes. By 2007, it was 112 times that, or 294.98 billion gigabytes.

This exponential increase in the size of data and data storage capacity is what is meant by *data volume* in the big data context.

DATA VARIETY: ENDLESS

Organizations today collect more types of data than in the past. For example, 10 or 15 years ago, a retailer likely tracked consumer transactions—that is, data about a product purchased and its purchase price. Today, however, that same retailer may track everything from purchases to customers' social network activity to the GPS signals from customers' cell phones.

This increased diversity of data is the idea behind *data variety* in the big data context.

DATA VELOCITY: REAL TIME

As data volume and variety have increased, so has the speed of acquiring data and the need to process data quickly. Traditionally, organizations acquired data at discrete time intervals, like once per day, through a batch process or some other ad hoc method. Today, in addition to batch or ad hoc data acquisition and processing, some organizations are challenged with managing streaming data or data that is acquired and processed in real- or near real-time.

This increase in the speed of data acquisition and processing is what is meant by *data velocity* in the big data context.

SOURCES OF BIG DATA

Big data can derive from many sources. Some of the most prominent sources include

- Mobile devices
- User generated content on social networks and other web applications
- Sensors

MOBILE DEVICES: WALKING MONITORS

Mobile devices generate data 24 hours a day, seven days a week, tracking everything from our location to our texts, phone calls, voicemail messages, pictures, videos, personal health metrics, website visits, and application usage. The most common mobile device is the phone. In the United States alone, 91 percent of adults own at least one mobile phone, with the number of wireless devices (smartphones, tablets, and wireless cards) exceeding the population by over 10 million, according to a cellular industry trade group. In emerging nations, 83% of adults own at least one mobile phone. People commonly use their phones not just to call but also to send text messages and take photos and video. Add to that the data captured via laptops, tablets, and other mobile devices, and it's easy to conceptualize the depth of impact mobile devices have on big data.

USER GENERATED CONTENT: POSTS, SEARCHES, & SHARES

User generated content has become one of the most important drivers of big data. If you have conducted a search on Google, updated your status on Facebook, rated a book on Amazon, uploaded an image on Instagram, commented on a video on YouTube, or rated a movie on Netflix, you have contributed to big data. As we share more of our information online, and as we engage more frequently with applications that help us generate content, we collectively expand our role in crafting big data.

SENSORS: HERE, THERE, & EVERYWHERE

Sensors are likely to become one of the most influential sources of big data in the future. Today, sensors already capture mountains of data from vehicles, assembly plants, jet engines, household appliances, city streets, farm machinery, and medical devices, among many other items. Organizations are even beginning to use sensors to capture data on workplace behavior. New sensor technologies like iBeacons—low-powered, low-cost sensors that can be used to interact with devices (like mobile phones) in close proximity—are making it even easier for sensors to become a behemoth in the big data landscape.

In essence, big data is being generated and captured all the time by both machines and people. In and of itself, all that data isn't very useful. What makes it useful is the ability to learn from it and then leverage those learnings to solve problems. That's where *analytics* come into the picture.

THE ROLE OF ANALYTICS

"Big data" and "analytics" are often used interchangeably. Although related, they are not the same. Big data is about the features and characteristics of the data itself and how those characteristics (among them, volume, variety, and velocity) impact our ability to work with the data. Analytics is the systematic use of data and related business insights to drive fact-based decision making. Organizations use analytics of data—big and small—to better understand their business. Analytics is typically categorized as *descriptive*, *predictive*, and *prescriptive*.

DESCRIPTIVE ANALYTICS: MEASURING THE PAST

The first phase of analysis and the foundation for fact-based learning, descriptive analytics is about leveraging data to answer the question ***"What has happened?"*** Most organizations use some form of descriptive analytics to understand how their business has performed in the past. Dashboards that plot various key performance indicators (KPIs)—such as historical sales volume and the number of active customers—are a practical implementation of descriptive analytics.

PREDICTIVE ANALYTICS: FORECASTING THE FUTURE

The second phase of analysis, predictive analytics, sets the stage for understanding possible future outcomes. It focuses on answering the question ***"What could happen?"*** Predictive analytics is more complicated and less common than descriptive analytics, but its popularity is growing rapidly. A statistical model that describes the drivers of sales volume and estimates future sales volume based on those drivers is one implementation of predictive analytics.

PRESCRIPTIVE ANALYTICS: WHICH PATH FORWARD?

The final phase of analysis, prescriptive analytics, focuses on using data to answer the question, ***"What should we do?"*** It involves accounting for the potential benefits and risks of different decisions in order to make decisions that maximize performance while minimizing risk. In the example of estimating future sales volume, a prescriptive analysis might extend a predictive analysis by examining how the forecast might change given the various actions an organization could take to influence sales.

Analytics—descriptive, predictive, and prescriptive—is the mechanism organizations use to extract insights from big data.

Your Role in the Data Deluge

As the sources of data relating to real property become more diverse, the share of data that is created by real estate professionals will shrink. Your ability to manage and evaluate large data sets will become a more important part of your value proposition, while the role of "data gatekeeper" will become less meaningful. Here are some basic strategies you can use to stay current with big data in real estate:

- Focus on innovative tools being created and provided by leading consumer real estate portals, like realtor.com®. As these sites add additional data about crime, traffic, schools, and the like, and as more of that data is accessible to the average home buyer, you will be responsible for understanding and synthesizing this data, and making it useful for your customers.
- Follow the progress being made by local municipalities in providing data to its citizens. Large, medium, and even small communities participate. Visit https://www.data.gov to better understand the types of data that are openly available at the national level, and https://www.data.gov/open-gov/ for data available at the local level.
- Encourage local boards and MLS's to investigate new sources of data that can integrate with member-only and public facing real estate databases.
- Encourage local boards and MLS's to evaluate analytical tools and processes that can help members become the local expert on big data in real estate.
- When evaluating third party vendors who offer big data services, ask them to explain their methods. Use this white paper to develop appropriate questions.
- Evaluate new mobile tools as they come online. Consider how they may be used as sensor devices, which can capture data about special interactions between physical objects and the world around them. In some cases, this data can be leveraged to provide significant value.
- Follow emerging trends that may affect real estate. These include hot topics like *smart homes*, *the Internet of everything*, *user-generated content*, and *crowdsourced content*.
- Learn more on big data and analytics—and how they are applied in practice—by reading one or more of these popular books:
    - Big Data: A Revolution That Will Transform How We Live, Work, and Think by Viktor Mayer-Schönberger and Kenneth Cukier
    - Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die by Eric Siegel
    - Competing on Analytics: The New Science of Winning by Jeanne G. Harris and Thomas H. Davenport
    - *The Human Face of Big Data* by Rick Smolan and Jennifer Erwitt.

COMMON APPLICATIONS OF BIG DATA & ANALYTICS

Web companies, brick-and-mortar companies, and governments are all using big data in combinations with analytics to make business decisions.

E-COMMERCE

One of the most prolific uses of big data and analytics comes from the online shopping world, where top players such as Amazon, Netflix, and Yahoo! have developed algorithms, or formulas, that use customers' online purchasing habits to provide them with product recommendations. Recommendations may lead people to discover and subsequently purchase products they didn't initially intend to buy. Thus, recommendations have the potential to improve customers' shopping experience while increasing company revenue.

TRADITIONAL RETAIL

Have you ever wondered why it's so easy to find canned pumpkin at the grocery store in the weeks leading-up to Thanksgiving? Using big data and analytics, retailers are able to pinpoint exactly when, where, and how to display products in their store in order to maximize sales. Retailers can predict purchase decisions with outstanding accuracy, sometimes with unintended consequences. Target, for example, used data to predict which customers were likely to be pregnant and sent those customers promotions for products that would interest pregnant people. Many consumers found this practice to be a violation of their privacy. Target is now subtler in its approach but still uses big data and analytics to improve sales.

MUNICIPALITIES

Leading cities around the world are leveraging big data and analytics to improve performance. For example, the City of Chicago employs a team of data scientists who analyze crime, environmental, and health data in order to predict everything from acts of violence to rat infestations and restaurants at risk of failing health inspections. The city is even installing sensors throughout the central business district to monitor weather, noise, foot and street traffic, and other metrics to quantify an assortment of block-level trends. With big data and analytics, the City of Chicago is improving the services it provides to its residents.

Politicians are using big data and analytics to target voters who they believe can sway an election in their favor . . .

Health organizations use big data and analytics to predict viral outbreaks . . .

Corporations are using big data to predict the "flight risk" of their own employees . . .

Big data and analytics are transforming industries, and real estate is no exception.

BIG DATA & ANALYTICS IN REAL ESTATE

We said earlier that data is big when its size and complexity—and the speed needed to acquire and process it—make it difficult to manage and extract information using traditional systems, technologies, and techniques. As the tools and processes to manage and analyze data evolve, data that was once "big" will become small.

Organized real estate provides a great example of this evolutionary process. More than a century ago, real estate agents began to warehouse the combined knowledge of available local listings. Although agents may not have known it at that time, they were amassing big data—files of historical market activity, including recent transactions. The MLS

book was a revolutionary big data management platform. Today, MLS systems aggregate far more data, including property, school, demographic, tax, and traffic data, which makes data from the original MLS book seem small. Ten years from now, the amount of real estate data will far surpass that of today.

Recent start-ups are already layering on data in ways that will change the industry. Here are two examples.

SMARTZIP

Leveraging patent-pending home intelligence, predictive analytics, and automated marketing techniques, SmartZip aims to help real estate agents connect with those homeowners who have the highest propensity to sell their homes. The company was founded in 2008 with the idea of bringing Morningstar-style analytics to the real estate world. "In the financial services world, a Morningstar Rating uses data and analytics to predict which stock or fund is a better asset to buy and hold," says Avi Gupta, president and CEO of SmartZip. "A similar rating didn't exist for homes, which is the largest purchase most consumers make in their lifetime. So we decided to build that rating. We wanted to help the real estate industry make better decisions about the home buying process." As one of America's fastest growing private companies, SmartZip is leveraging big data and analytics to optimize customer acquisition.

REVALUATE

"According to International Data Corporation (IDC), 23% of the digital universe would be useful if tagged and analyzed—yet only 3% is currently tagged and only 0.5% is being analyzed." says Revaluate cofounder Chris Drayer. "The opportunity we see is in closing that 22.5% gap for data that relates to housing and neighborhood." Drayer and cofounders Max Galka and Tim Segraves recently launched Revaluate with a focus on the Manhattan real estate market; the company provide a glimpse at the level of data that will likely be aggregated nationally someday. Revaluate gathers data about rat reports, bed bug complaints, elevator maintenance, nearby construction projects, and hundreds of other metrics to give consumers a better idea of what it's like to live in a building. "Currently, our database consists of about 35 million records (roughly 1,000 for each residential address in Manhattan), coming from about 2,000 different sources, public and private," Drayer says.

Cities like New York and Chicago have considerable data resources, many of which are available to the public for use. But the richness of big data is increasing everywhere, thanks to the widespread use of smartphones, web-connected thermostats and appliances, and user-generated content.

Smartphones—commonly seen today as consumption devices—are beginning to function as powerful sensors that can inform the real estate industry about everything from commute times to neighborhood walk-ability. Navigation applications like Waze, for example, guide commuters through rush hour by using the GPS technology in users' phones to measure traffic patterns and adjust routes so that users can avoid high-traffic areas and reduce commute times. Smart home appliances are participating in this innovation as well by collecting data about energy efficiency and connecting to smartphones. Data that once was considered private is now being shared willingly. And all this recorded information enables data collectors to generate in-depth consumer profiles for specific neighborhoods—layer upon layer of "lifestyle data" that can be leveraged by real estate sellers, buyers, and professionals.

CONCLUSION

Big data is everywhere. Analytics leverages data to derive meaningful insights. One of the most critical drivers of success for tomorrow's real estate professional will be acknowledging, and capitalizing on, big data and analytics in order to find new and innovative ways to think about and serve consumers and communities. With more and better data at your fingertips, you'll make better business decisions and more easily demonstrate your primacy to consumers.

BIBLIOGRAPHY

Bleiberg, J. (2014). What Does Big Data Actually Mean? http://www.brookings.edu/blogs/techtank/posts/2014/09/11-big-data-definition.

Cody, S., & Asher, A. (2014). Smarter, Better, Faster: The Potential for Predictive Analytics and Rapid-Cycle Evaluation to Improve Program Development and Outcomes. http://www.brookings.edu/~/media/research/files/papers/2014/06/19_hamilton_policies_addressing_poverty/predictive_analytics_rapid_cycle_evaluation_cody_asher.pdf.

Cortada, J.W., Gordon, D., & Lenihan, B. (2012). The Value of Analytics in Healthcare. http://www.ehidc.org/resource-center/publications/doc_download/229-white-paper-the-value-of-analytics-in-healthcare-from-insights-to-outcomes-data-and-analytics.

Davenport, T. (2012). The three '..tives' of business analytics; predictive, prescriptive, and descriptive. http://www.enterprisecioforum.com/en/video/three-tives-business-analytics-predictiv.

Elmasri, R., & Navathe, S.B. (2007). XML: Extensible Markup Language [PowerPoint slides]. https://www.cs.purdue.edu/homes/ake/cs348/Chapter27.ppt.

Eltabakh, M. (2013). CS525: Special Topics in DBs Large-Scale Data Management [PowerPoint slides]. http://web.cs.wpi.edu/~cs525/s13-MYE/readings.html.

HIll, K. (2012). How Target figured out a teen girl was pregnant before her father did. http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

IBM Software (2013). Descriptive, predictive, prescriptive: Transforming asset and facilities management with analytics. http://docs.caba.org/documents/IS/IS-2014-49.pdf.

Kang, C. (2011). Number of cell phones exceeds U.S. population: CTIA trade group. http://www.washingtonpost.com/blogs/post-tech/post/number-of-cell-phones-exceeds-us-population-ctia-trade-group/2011/10/11/gIQARNcEcL_blog.html.

Johnson, R. (2011). Semistructured Data, XML, DTDs [PowerPoint slides]. www.cs.toronto.edu/~ryanjohn/teaching/cscc43-s11/c43-xml-v02.pdf.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Marcus, G. (2013). Steamrolled by Big Data. http://www.newyorker.com/tech/elements/steamrolled-by-big-data.

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. http://www.buyukverienstitusu.com/s/1870/i/Big_Data_2.pdf.

Mi, C.M., Shan, X.F., & Ma, J. (2014). Improved Network-Based Recommendation Algorithm. iBusiness, 6, 109-116. http://dx.doi.org/10.4236/ib.2014.63012.

Rainie, L. (2013). Cell phone ownership hits 91% of adults. http://www.pewresearch.org/fact-tank/2013/06/06/cell-phone-ownership-hits-91-of-adults/.

Rivera, J., & Meulen, R. (2013). Gartner Survey Reveals That 64 Per Cent of Organizations Have Invested

or Plan to Invest in Big Data in 2013. http://www.gartner.com/newsroom/id/2593815.

Short, J.E., Bohn, R.E., & Baru, C. (2011). How Much Information? 2010 Report on Enterprise Server Information. http://hmi.ucsd.edu/pdf/HMI_2010_EnterpriseReport_Jan_2011.pdf.

Silverman, R. E. (2013). Tracking sensors invade the workplace. http://online.wsj.com/news/articles/SB10001424127887324034804578344303429080678.

Sun, H., & Heller, P. (2012). Oracle Information Architecture: An Architect's Guide to Big Data. http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf.

Vastag, B. (2011). Exabytes: Documenting the Digital Age and Huge Growth in Computing Capacity. http://www.washingtonpost.com/wp-dyn/content/article/2011/02/10/AR2011021004916.html.

Watson, M., Nelson, D., & Cacioppi, P. (2014). Managerial Analytics: An Applied Guide to Principles, Methods, Tools, and Best Practices. New York, NY: Pearson FT Press.

Wike, R., Simmons, K., Poushter, J., Ponce, A., & Devlin, K. (2014). Emerging nations embrace Internet, mobile technology. Cell phones nearly ubiquitous in many countries. http://www.pewglobal.org/files/2014/02/Pew-Research-Center-Global-Attitudes-Project-Technology-Report-FINAL-February-13-20146.pdf.