

Executive summary

Consumers purchase goods and services using a complex set of shifting, individually important factors, whether it is access to the right selection of products, the date of an upcoming trip or event, or even the day's weather. The advent of "big data"—a growing mass of information consisting of public records, proprietary data, and collections available for purchase—has enabled merchants to pinpoint customer motivations more accurately than ever before. This white paper explores the state of the retail industry in the United States and explains how merchants can use available data to better address their markets.

By skillfully applying learnings from big-data analysis to decisions at every level, retailers can deliver a cost-effective, yet personalized, approach that is consistent for the buyer across both physical and digital channels. The goal is to yield a highly satisfactory interaction for both shopper and retailer. Unfortunately, customers are reporting a decreased level of satisfaction with some industry sectors and stores. Sales figures are mixed, with some retailers setting record profits and others reporting losses. Overall, according to U.S. Census Bureau estimates, the volume of online sales is rising steadily at 10 percent annually, with in-store purchases increasing more slowly, at 4 percent a year.¹

Though much of the information in the big "data lake" is unstructured, it can now be efficiently stored and analyzed to show patterns of thought and behavior that directly relate to sales and retail operations. Sellers can influence important purchase-related options, such as location, stock selection, pricing, packaging, promotions, payment schedule, delivery options, and store and shelf layout. Today's technologies and creative approaches make it easier to manage and experiment with these factors.

Merchants have an increased ability to provide excellent shopping experiences by seeking insights from this data lake about the things that matter to each customer. With today's powerful data storage, real-time analysis, and automation technologies, retailers can handle this volume of information more efficiently and access it as needed to inform operational and product decisions.



Retail transactions occur in stores and showrooms, at kiosks, via catalogs, and through mobile merchants of all sizes. They also increasingly happen online, on smartphones, and through automated debit and credit card transactions. Shopping experiences for those in the United States are often positive, but recent customer surveys reveal there is room for improvement.

To support and strengthen customer satisfaction and perception of value, more retailers are turning to big-data analysis. Merchants are combining their proprietary customer details with public and for-sale information to create data lakes that produce a single view of the customer across multiple communication channels. Often this data is then displayed through a business intelligence dashboard that uses machine learning and other techniques to present patterns and predictors of future customer actions.

Individual buyer information and group buying behavior can also help a company to personalize offers and better inform customers of special deals and other relevant news. This can lead to better customer experiences, whether that means a faster, easier process, a more effective product choice, or a more complete solution for the impetus behind the purchase.

With big data and linked solutions, it is easier than ever to make sense out of the masses of data and to take action to improve every step of the buyer's journey.







Decreasing consumer satisfaction

Consumers around the world are more experienced and informed than ever before and display higher purchasing standards across the board. They've come to expect an optimum overall experience as well as consistent value and reliability in brand, message, pricing, promotions, and loyalty and rewards programs. For their part, retailers want to provide excellent shopping experiences but have varying capacities in achieving that goal.

With today's broad information-sharing capabilities, consequences for failing to meet in-the-moment customer expectations can be sudden and far-reaching. For example, in 2015, major U.S. publicly owned department and discount stores saw an average 3.8 percent decrease in customer satisfaction, according to the University of Michigan's American Customer Satisfaction Index.² While there were bright spots, this was the lowest level of satisfaction since 2008.

Even industry leaders such as Home Depot and Whole Foods suffered large declines; Macy's experienced an 8 percent drop in satisfaction. The top complaints centered around brick-and-mortar stores: slow checkout lines and store cleanliness. All nondurable goods sectors showed a similar trend, including food and beverage, personal care, apparel, and athletic shoes.³

Sales trends show mixed results

Revenues and growth figures for retail industry leaders have varied widely over the last few years. For example, according to U.S. Census Bureau estimates, online retail sales have been increasing more than 10 percent annually, but total sales are rising more slowly, at about 4 percent a year.⁴ In the next five years, Internet-based sales are projected to increase at a higher rate than sales at brick-and-mortar stores. Trends like this are also echoed in other markets around the world.

In the offline space, retail payroll numbers are showing moderate progress, with a gain of nearly a quarter million new jobs between February 2015 and February 2016.⁵ Discount clothing chain T.J. Maxx was a retailer to watch last year, with a surprising 6 percent increase in sales.⁶ Only 1 percent of the overall T.J. Maxx revenue comes from online sales, although the company began an e-commerce initiative in 2013.

Some old-line retailers, like J.C. Penney and Barnes & Noble, are seeing success at last with an omnichannel marketing approach. J.C. Penney is running an integrated social media and in-store campaign with private-label items on sale for a penny, promoted online using the hashtag #SoWorthlt.⁷ This reinforces the message that every trip to a J.C. Penney store is worth the time, money, and effort spent. J.C. Penney is expected to report a profit in 2016 for the first time in five years. Barnes & Noble continues to rely on in-person events for steady sales.

Among specialty fashion retailers, newer operations, such as Zara and H&M, are often outperforming standard department stores.⁸ Both H&M and Zara are less expensive and more responsive to fashion than traditional retailers and can bring new designs to market much more quickly.⁹ Zara can manufacture more of an existing style in two weeks and can produce a new one in as little as five.¹⁰ This means that this retailer can manage inventory more efficiently than other companies, reducing the risk that items will remain unsold.











Drivers of online sales growth

Currently, e-commerce is steadily increasing its contribution to total retail sales. As an example, by 2020, e commerce is expected to account for a full tenth of all U.S. retail sales. What is causing this steady increase? More than ever, customers are accessing the Internet through smartphones and tablet devices, adding hours per week to the time that they spend online.

Consumers can now research products, find out what is in stock, and get the best prices without ever speaking to a sales representative. Frequently, retailers are making it possible for a shopper to complete their entire purchase online and have their order delivered to their door.

Another factor contributing to this growth is increased investment by traditional retailers in online sales divisions. Amazon dominates in the online space, and a number of other retailers, like Target, have flourishing online businesses. Target ran a holiday free-shipping promotion that helped it outpace Amazon during the fourth quarter of 2015, with online sales increasing by 34 percent compared with Amazon's 25 percent.¹²

Large retailers are concentrating on presenting more omnichannel retail experiences and exploring additional ways to fulfill customer needs. Common scenarios include allowing online shoppers to pick up their order at a brick-and-mortar store, with sales associates offering to order out-of-stock items through an in-store online portal. When the customer has multiple ways to feel connected to the retailer, the relationship is strong enough that the concept of channel starts to disappear.¹³

How customers experience shopping



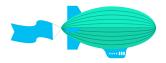
Gather information through other people

- Recommendations from friends
- Online reviews
- Social media
- News



Make personal contact with a retailer

- Sampling, trade shows, events
- Stores and showrooms
- Kiosks and event booths
- Through mobile merchants
- Nurture and follow-up contacts



Learn about a brand through media, advertising, and community participation

- Radio, TV ads
- Product placement
- Sponsorships
- Digital advertising: online, apps, social media



No personal contact with a retailer

- Online research about purchase options and product details using a cell phone, smartphone, laptop computer, tablet device, or digital kiosk/station
- Self-checkout transactions
- Retail transactions performed online and on cell phones
- Previously authorized credit and debit charges
- Purchases through partnerships and affiliates







Volumes of information

Every day, consumer actions generate terabytes of new information about customer intent, actions, and preferences. This data is often relevant to decision-makers in the retail industry and comes in many forms, from the quantitative details of a transaction and account history, to free-form information such as security camera videos, local events, and weather patterns. Last November, Gartner analysts predicted that 6.4 billion connected things will be in use worldwide in 2016, up 30 percent from 2015.¹

Retailers can use the data to form and test reasonable hypotheses that impact operational decisions. For example, Lowe's analyzed big data to help design its Manhattan store that opened in 2015. Rather than the typical mix of inventory, Lowe's decided to stock small appliances and other apartment-scale offerings, a decision that is paying off for the merchant and the community.

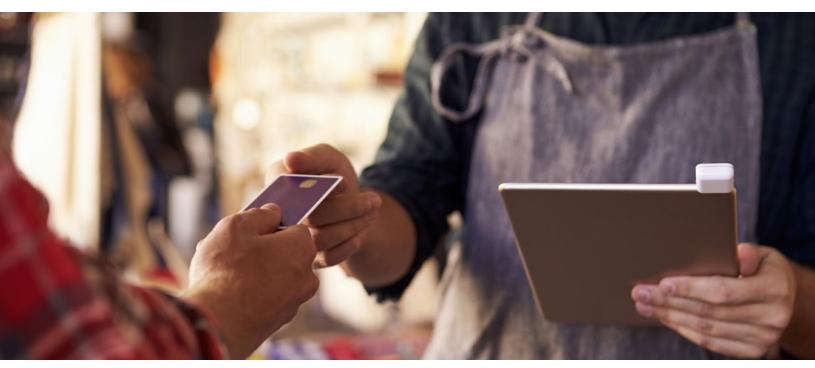


Interactions between consumers and retailers

People perform a wide variety of shopping activities, and the shopping experience can be simple or complex. From impulse buying to extended research on a major purchase, the purchasing journey can take seconds or years, with many opportunities for retailers to connect with potential customers in their unique markets.

However, people are also increasingly distracted by life outside the shopping experience. In this pressured environment, retailers and brands have to work doubly hard to increase the level of customer attraction, engagement, and retention. While shopping behavior may result in a real-time monetary exchange, all the contacts made prior to and after that moment are equally important to delivering the best shopping experience.

On desktops and laptops, Facebook is still the number one site that people access on the Internet, falling second on mobile devices to Google Maps. ¹⁴ In the second quarter of 2015, Facebook reported 399 million mobile-only users—that is, users who never use Facebook on a desktop computer. ¹⁵ That was up 180 million from the prior year. Moreover, a broader category of mobile users—who aren't necessarily mobile-only users—increased even more, by 251 million to 1.1 billion. The other three top applications, in order, are YouTube, Google Search, and Google Maps.



 $1.\ http://www.information-age.com/technology/mobile-and-networking/123458905/gartners-internet-things-predictions and the property of the p$







Structured or unstructured

Traditionally, information technology was created to utilize structured data, such as with relational databases. However, more than 85 percent of all data available today is considered "unstructured" and requires extra organization, careful validation, and human insight¹⁶ to be used effectively.¹⁷

Types of unstructured data



Social media comments and behavior



Logs of web browsing data with page titles, content, and names of video/audio/image files



Telephone logs, including messaging, media-sharing, and apps used



Documents, videos, audio files, and presentations



Email titles, delivery information, and message content

Personally identifiable information and anonymous data

Most merchants collect and store specific information about their customers, known as personally identifiable information, or PII. Consumers trust merchants with everything from their credit card and banking information to their physical locations and home addresses. They also share wish lists, life goals, vacation plans, and details about their families. Some of the most useful data in this category also requires the most security.

When PII data is aggregated and stripped of information that connects it with a single person, it becomes anonymous data. For example, bundling the wish lists of thousands of people yields trends without risking any individual's privacy. Anonymous data has been used for decades in retail purchasing, and with current processing methodologies it can be accessed in real-time rather than waiting weeks or months for analysis.





Proprietary data

The unique information retailers collect is called proprietary data, and the retailer owns the data. It is the retailer's responsibility to keep it safe, which often requires special security and privacy measures. The data is stored in manual or automated on-premises systems and, increasingly, in the cloud.

Proprietary data can include purchase transaction details, such as basket selection, geographic information, age, gender, time of purchase, and the length of time shopping. Common sources of cloud-based data are email and chat correspondence records, website visit behavior, and payment-processing information. Merchants can also record derivative details about consumers, such as using security camera videos to make guesses about the age and gender of store visitors.

Open data

A major unstructured data source that is growing exponentially is open data, also known as public data. This information, often released by the government and science sectors, is freely available to use and republish, without restriction. It is usually stripped of PII and may require extensive programming to match it with other data sets. Examples include economic indicators, public health information, and use of government services.

Internet of things

The next wave of data is coming from the Internet of Things (IoT). The information, including locations and operational data, is coming from cell phones, traffic monitors, small appliances, utilities, records, and more. Tiny sensors are being used in consumer wearables, such as headphones. Radio-frequency identification (RFID) chips, beacons, sensors, digital interfaces, and other automated devices all fall into this category. Retail applications of this technology include store traffic counters and improvements in inventory tracking in the form of smart bins, hangers, and shelves. The technology analyst firm Gartner estimates 26 billion of these types of units will be in use by the year 2020.¹⁸

Data for purchase

Mailing lists have been available to merchants for decades. Now, thousands of entities are packaging their proprietary data and selling it to businesses and interested parties. The U.S. Census, for example, provides a lot of information without charge, but a set of unpublished material is available for purchase. Even major credit card companies like MasterCard and American Express sell their non-PII data for the right price.¹⁹



Breakthroughs in merchandising strategy

Merchants can communicate with customers through personal contacts as well as newer information-sharing channels, such as websites and email newsletters. The overall goal is to enable people to find personally relevant details about brands, products, and purchase options at the right time for them. Consumers are indeed using this wealth of information to better control the buying process and their experience, but gaps still exist between shopping expectations and reality.

In order to please customers, retailers can choose from a variety of methods. One of the most cost-effective is matching different data sets to yield new insights. Using currently available data-management solutions, retailers can efficiently combine and utilize data from disparate sources to draw more accurate and timely insights about customers.

For example, knowing what is being said about a product or company on Facebook and on review sites such as Yelp, TripAdvisor, and Angie's List can be a great help in maintaining a competitive advantage and a healthy connection with customers. To make this task easier, many companies are analyzing customer sentiment on these social media platforms, based on natural language-processing algorithms. This analysis can be done in real-time in order to proactively respond to current public opinion, adding the retailer's voice to the conversation.





"[According to a Nielsen survey] while 51 percent of respondents said they browse products in stores before ultimately buying them on the Web, an even larger number said they do the opposite."

- Sarah Halzack, "One way online shopping is actually helping brick-and-mortar retailers," Washington Post, Sept. 3, 2014



Factors that make a difference to customers

People purchase goods and services based on multiple, individually important factors, some of which a merchant can control (see Fig. 1). Analyzing the masses of available data can help to measure and understand the things that are controllable and incorporate strategies to address the ones that aren't.

Factors that are not in a retailer's control include the weather, sporting events, holidays, and surrounding traffic. Important purchase-related options that a seller can influence are location, stock selection, pricing, packaging, promotions, payment schedule, delivery options, and store and shelf layout. Today's technologies and creative approaches make it easier to manage and experiment with these factors.

Fig. 1. Consumers' purchase influencers when shopping online

Percent of respondents indicating what influenced their online shopping purchase decisions.



One thing that hasn't changed is that trying, seeing, and touching the product is still centrally important to consumers. The rise of showrooming, which is the practice of purchasing goods or services online after in-person research, has benefited Amazon, eBay, and other web-based retailers. At the same time, the opposite practice, webrooming, is becoming more prevalent. This happens when a shopper researches a purchase online and then buys in-person, sometimes even after checking details like local stock levels and sales. Retailers are investing in online, in-person, and hybrid interfaces to make it easier for people to move seamlessly between these channels.







Using big data learnings wisely

Finding relevant patterns in this big-data lake can help retailers make better decisions about store design, shelf layout, staffing, product planning, and stock selection, levels, and locations. It is now possible to get a single view of the customer across multiple communication channels: online, in-store, and social media. Retailers can combine this information into a dashboard using machine learning and other tools to display patterns and predictors of future customer actions.

However, inappropriate or unwelcome use of this information can harm the relationship between a customer and a company or technology. Recently, advertising firm Titan had to halt a deployment and remove Bluetooth-responsive beacons from hundreds of pay telephone booths in New York City.²⁰ Residents and privacy advocates protested after finding that the beacons had been sensing nearby cell phones and recording web-browsing histories and other data about device users, and then pushing ads and offers for apps to their phones.

Many retailers are investing in big data as a way to better understand the desires of current customers in order to implement more tailored and successful marketing. As Andrew Laudato, senior vice president and CIO of Pier 1 Imports, stated, "The retail industry is going through a paradigm shift. Customers are shopping more online and they have more choices. In response, our goal is to ensure wherever she engages with the brand, whether in the store, on her home computer, on her mobile device—it's a very personal and consistent message."²¹









Better ways to engage customers

It can be a challenge to provide a seamless shopping experience across channels. With big data, it is much easier to understand and keep track of the customer base—as a group, in segments, and as individuals. Real-time processing can help a retailer immediately orient to the individual characteristics of each customer to decide, for example, whether to authorize a credit purchase.

When analyzing data from multiple sources, new patterns arise that help determine what offers will have value to the customer. For example, matching information in different data sets, such as local shopping activity levels and the times of highest use of area parking lots, might inspire a pilot marketing promotion offering free valet parking to preferred shoppers when they most want to come in.

Personalization

The advances in knowledge management have led to more customized buyer experiences and deals based on individual needs and wants. Retailers now can create recommendation engines and use machine-learning capabilities to automate their delivery. With data-based insight, retailers can deliver highly satisfying shopping experiences at a much faster pace and lower cost than ever before.

By creating a personalized customer journey and helping shoppers to more effectively achieve their goals, retailers build customer appreciation and loyalty. If customers know that they will be rewarded for their patronage with a unique, enjoyable experience, they are more likely to shop with the same merchant again.

Advances in knowledge management and sheer creativity have led to some truly amazing personal experiences for today's customers. To inspire customers, Lowe's invested in digital display technologies, such as the Holoroom, a virtual representation of a room being remodeled.²² If the customer likes the image, they can print it, along with a list of materials necessary for the project.

Hyperlocal marketing

Some of the aspects of a shopping experience are specific to the location. For example, a consumer may be looking for a quick meal in a nearby neighborhood. With the customer's permission, rating websites like Yelp can present a real-time list of the top restaurants in the area, even showing only establishments that are open at that time. Hyperlocal marketing can also involve targeted special deals, such as shopper-relevant coupons on the back of a grocery receipt. By making the customer an offer unique to their need, location, and timing, a retailer is much more likely to make a positive connection with the shopper.²³

Online to offline and back

While a consumer may have slightly different behaviors in online and offline settings, it is easier if the retailer maintains a level of consistency. For example, pricing should be the same on a store's website as it is on the shelf, or the customer will want a reasonable explanation for the difference. Another common scenario is an online order that requires a return. This should be a seamless experience for a customer, but often one part of the retail operation has different information than another, forcing the shopper to explain themselves and provide order details before being helped.

With enough relevant data, the retailer can form a hypothesis about when and where that shopper would find it easiest to return the item. Are there any retail outlets near that person's house? Are they an executive and likely to have limited time for errands, or even have a personal assistant who might return the item? By using data to personalize and shape the process, the retailer can influence the shopper to use the most effective and engaging method, increasing satisfaction and profit.

Other processes should be congruent across channels as well. For example, is it possible for a shopper to go to a store after placing an online order and cancel that order in favor of an in-person purchase? For most retailers, this scenario would cause issues. Big data can quickly illuminate places where customers get stuck.





Conclusion

Combining direct feedback with insights from big data can inform merchandising strategy and produce a higher level of customer satisfaction for shoppers. By detecting meaningful patterns in the data lake, retailers can improve their ability to provide an optimum customer experience.

In summary, retailers have the opportunity to learn more about their customers than ever before. The first step is to combine proprietary data with purchased and public information to create a data lake that can be quickly and efficiently analyzed on demand. This enables merchants to build a single view of each customer across multiple communication channels. This may take the form of a dashboard that uses machine learning and other state-of-the-art solutions to make sense out of the masses of data available.

These insights can transform multiple aspects of retailing. The results can be used to make relevant, personalized offers, select stock, optimize pricing, and match investments in operations to customer demand. With this data-based insight, retailers can deliver highly satisfying shopping experiences at a much faster pace and a lower cost than ever before.



Use cases

Pier 1 Imports

Pier 1 has a unified vision to create special shopping experiences for customers, whether in-store or online. They have invested in Microsoft Azure HDInsight to help them consolidate a high volume of multistructured data and process it in a format where it can be queried through batch, interactive, and real-time analytics. They now use this system to predict which products might draw customers back to the store or website for another purchase.

View Case Study

Ziosk

Ziosk is a device and software offering for U.S. casual-dining restaurants. Tablets customized by Ziosk sitting on restaurant tables offer menus, and the ability to digitally order and pay for a meal, news, entertainment, and video games. The company started using Microsoft Azure HDInsight to aggregate a broad range of user-experience data, from the content guests use, to the parts of the user interface they click on the most. With this understanding, Ziosk can better predict what offers, content, and user-interface elements provide value to customers.

View Case Study

zulily

zulily offers thousands of unique, fresh products on its e-commerce website, for more than \$1 billion in sales during 2014. To help its merchandising professionals create the best, most personal user experience for shoppers, zulily decided to invest in a solution that could integrate and analyze both its structured and unstructured data while lowering development costs. The company used the Hortonworks Data Platform (now part of Microsoft Azure HDInsight) to create advanced analytics solutions, ultimately presenting the most relevant products and offers to zulily customers. View Case Study

"The retail industry is going through a paradigm shift."

- Andrew Laudato, senior vice president and CIO, Pier 1 Imports





Hortonworks and Microsoft can help retailers optimize their insights from data

Hortonworks and Microsoft have come together to transform data within your organization into intelligent action. Over the past decade, we've learned that the characteristics of data used to drive decisions have fundamentally changed. It is no longer sufficient to only use the relational data captured from your line of business (LOB) applications or data from your enterprise resource planning and CRM; nonrelational data needs to be captured from social feeds, click streams, RFID, weather feeds, and devices, which can unlock profound insights to the business. It is no longer just about storing your data-at-rest in a permanent store; data-in-motion needs to be captured—particularly as smart devices are wired up in the IoT scenarios. Finally, it's not just about data that lives on-premises in your own datacenter; the cloud needs to be incorporated—because it provides virtually limitless scale while lowering the costs of your IT operations.

Hortonworks and Microsoft do all of this together through our joint cloud offerings. Microsoft Azure HDInsight, built on the Hortonworks Data Platform, is the premier offering for customers pursuing big data and analytics use cases in the cloud. Microsoft Azure HDInsight also offers a seamless extension of Hortonworks Data Platform from a customer's data center to the cloud, as Hortonworks' premier Connected Data Platforms cloud solution.

On-premises, Microsoft provides SQL Server, the world's most widely deployed relational database management system in the world, while Hortonworks provides Hortonworks Data Platform and Data Flow, one of the world's most widely deployed Hadoop distributions. In the cloud, Microsoft offers IoT Suite and Cortana Intelligence Suite, which showcases Azure HDInsight. And to unify the solutions, Microsoft offers PolyBase on-premises and in the cloud, which allows you to federate and distribute your queries across relational and nonrelational sources.

Microsoft and Hortonworks have been pioneering data solutions jointly for our customers for over four years. Together, we help organizations realize the advantages of doing analytics on all of your data to drive intelligent action.

Learn more about Azure HDInsight at www.hortonworks.com/HDInsight or email HDI@hortonworks.com to schedule a consultation.

- U.S. Census Monthly & Annual Retail Trade page, http://www.census.gov/retail
- American Customer Satisfaction Index, http://www.theacsi.org Open Data Handbook, http://www.opendatahandbook.org

- 1. U.S. Census Retail page, http://www.census.gov/retail/index.html.
 2. 2015 ACSI Market Retail Report, https://atlas.qz.com/charts/4y/iy/m_sl.
 3. Nondurable Products and National ACSI 0.3 2015, press release, http://www.theacsi.org/news-and-resources/press-releases/press-2015/press-release-nondurable-products-and-national-acsi-q3-2015.
 4. U.S. Census Retail page, http://www.census.gov/retail/index.html
 5. "Payroll Shows Pace of Retail Economy Is Steady and Improving," National Retail Federation, March 4, 2016.
 6. "The TIX Companies, Inc. Reports Above-Plan Q4 and FYI6 Results," press release, Feb. 24, 2016.
 7. "I.C. Penney Emphasizes Customer Value with Launch of 'Get Your Penney's Worth," press release, Feb. 25, 2016.

- 8. "U.S. sales growth over the last two years," Quartz.com chart on retail industry, https://atlas.gz.com/charts/
- Ekf-wcKsg.

 9. "Where Have I Seen You Before? At Zara, in Midtown, It's All a Tribute," New York Times, March 27, 2012.

- T2."Mannequins, home decor kept Target turnaround successful in Q4," USA Today, Feb. 24, 2016.

 13. A Simple Explanation of 'The Internet of Things;" Forbes, May 13, 2014.

 14. "Facebook is making massive amounts of money on sky-rocketing mobile usage," Quartz, July 23, 2014.
- 13. Iou.

 15. "Improving the Effectiveness of Customer Sentiment Analysis," *Data Informed*, Sept. 13, 2012.

 17. "The biggest sources of unstructured data in the enterprise," ComputerWeekly.com, March 2007.

 18. "Gartner says the Internet of Things Installed Base Will Grow to 26 Billion Units by 2020," press release, Gartner.com, Dec. 12, 2013.

- 19: Mastercard, AmEx Quietly Feed Data to Advertisers," Advertising Age, April 15, 2013.
 20. "Hundreds of Bluetooth Ad Beacons Uncovered in NYC Payphones," Escapist, Oct. 6, 2014.
 21. "Pier I Imports Finds a Better Connection with Customers Through Cloud Machine Learning," case study,
- 22. "Lowe's 3-D room makes remodeling a (virtual) reality," CNBC.com, June 11, 2014. 23. "3 Ways Location-Based Marketing Is Shaping Retail in 2016," Forbes, March 3, 2016.
- © 2016 Microsoft Corporation. All rights reserved. This document is provided "as-is." Information and views expressed in this document, including URL and other Internet website references, may change without notice. You bear the risk of using it.





This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal reference purposes