Importance of Big Data and Data Analytics in E- Commerce in India

Nithya S. M. [1] Kusuma M. [2] Murugaiah V. [3]
[1,2,3] Institute of Management Studies, Davangere University, Tholahunase, India.
Corresponding authour: nithyasmk@gmail.com

OIDA International Journal of Sustainable Development, Ontario International Development Agency, Canada.
ISSN 1923-6654 (print) ISSN 1923-6662 (online) www.oidaijsd.com
Also available at http://www.ssrn.com/link/OIDA-Intl-Journal-Sustainable-Dev.html

Abstract: Earlier, it was 5 M's considered as resources needed for successfully running any business, Money, Machine, Method, Material and Money. But now most of the organizations use Data as one of the major resources for any business. Some of the complicated and time-consuming tasks can be completed in seconds. E-commerce is one of the emerging industries where we see the innovation in technology, products, trends, websites, Blogs, social marketing happens constantly. In this paper the attempt has been made to analyze the importance of Data and Data analytics in E commerce industry in India. The detailed analysis has been made about the usage of data and application of data analytics in their business to enhance their performance. This paper is based on literature review and analysis of reports from NASSCOM and McKinsey Global Institute. Based on case study on Amazon company the application of Big data and data analytics as been discussed.

Keywords: Data analytics, predictive analysis and behavioral pattern.

Introduction

In India the online business market is well developed. Ecommerce industry provides a major contribution to its development. Even so, few understand the intricacies of the e-commerce industry. By definition e-commerce stands for electronic commerce. Dealing in goods and services through the electronic media and internet is called as E-commerce. E-Commerce or E-business involves carrying on a business with the help of the internet and by using information technology like Electronic Data Interchange (EDI). E-commerce players have progressed significantly in their decision making over time. Earlier simple analysis was used to make recommendation but now customer specific predictive algorithms and models being prepared using data and data analytics.

According to an IBM Marketing Cloud study, 90% of the data in the internet has been created after 2016. People, businesses, and electronic devices have all become data factories that are thrusting out incredible amounts of data to the web every day. The customer examination pays an extraordinary route for the advertisers to distinguish the most deals item and furthermore serves to their image as for the contenders and it gives incredible help to customize the consumer experiences based on their preferences to increase retention rate. Product specific analysis is the most critical to understand the customer satisfaction and to forecast the sales of a product, basically used for prediction of customer preference and demand.

Literature Review

This paper aims to identify definitional perspectives of big data analytics, characteristics of big data in e-commerce, business value of big data in e-commerce, tackling the challenges of big data in e-commerce. This paper is based on literature review and secondary data collected from different sources of literature i.e., articles, journals etc. (BDA) progressively provides value to e-commerce organizations by using the changing aspects of people, processes, and technologies to transform data into valuable insights for healthy decision making and solutions to various business problems. This is a universal process which deals with data, sources, skills, and systems in order to create a sustainable and competitive advantage. Leading e-commerce firms such as Google, ASOS, Amazon, e Bay, Netflix and Facebook have already encompassed Big Data Analytics and acquired huge market experienced enormous growth. The study provides an approach for summarizing all the best practices that build and shape Big Data Analytics capabilities. **(**Akter & Wamba, 2016),

E-Commerce companies by using big data analytics experienced 73% increase in sales by implementing Predictive Analytics compares to other companies not using big data ,45% of online shoppers prefer to online shop in websites which are offering personalized recommendation based on previous purchase history or online buying behavior. This paper focus on identifying the application areas of big data analytics for E-Commerce functions, to identify various different sources of big data or data for usage in E-Commerce, to interpret business value of big data for E- Commerce companies and also to analyze future challenges of big data application. The important application of big data analytics as follows personalization, dynamic pricing, customer service, predicting customer behavior, supply chain visibility, and managing fraud. Big data is generated by machine, people and organization. Machine generated data are likely through real time sensors ex. Page visited, product viewed, website or page traffic etc. People or human generated data such as tweets, likes, comments, reviews, photos, status updates etc. Organization data such as business transaction made, client, financial information etc (Avinashi & Akarsha 2017)

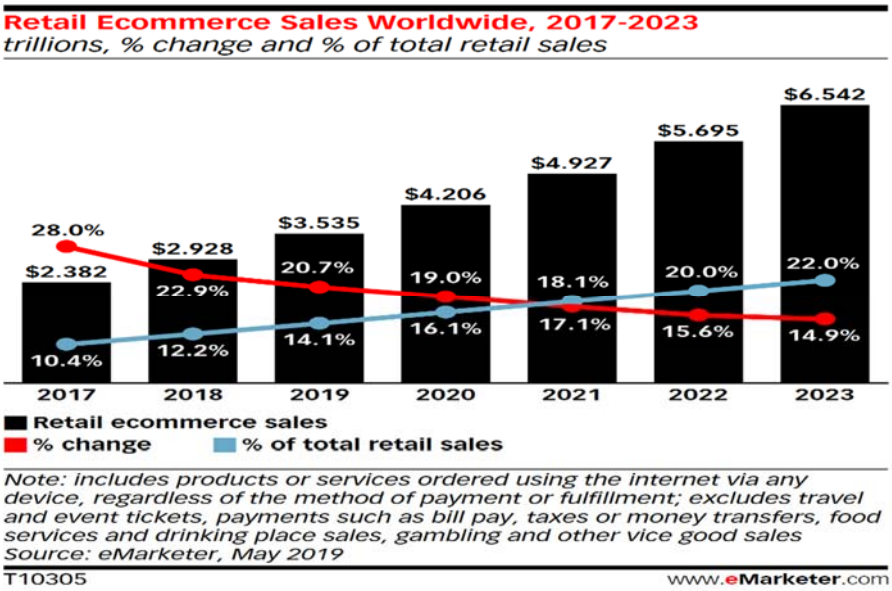
developed a four-level architecture model for business intelligence for e-commerce that helps in transforming raw data collected from the e-commerce customers into meaningful knowledge or insights. It helps to identify customers purchase pattern to improve the relationship between management and customers also for providing better service to customers. Cloud computing is useful to reduce the infrastructure cost and time for storing and analyzing the data. Machine learning plays significant role that machine automatically constructs algorithms for predictions which leads to fast and defined analysis (Pedrosa and Bernardino, 2017).

According to author Business Intelligence helps in predicting market trends and analyzes competitors and also helpful in implementing customer targeted market. A new three tier architecture is proposed, Hadoop Distributed File System (HDFS) used instead of data warehouse for storing data (Moorthi, 2017).

The customer and organization are the information generators for Architecture.

- Tier 1 – Data: Customer information, purchase history, reviews and social media data collected from the e-commerce portal. The noise is removed from the data for further process.
- Tier 2 – Hadoop: The gathered data are next transferred into HDFS then map reduce is applied, the data get input from HDFS. Map phase maps certain set of data to another set of data and gives the final desired data result further for analyzing the data.
- Tier 3- Data Analytics and BI: After applying map reduces the data is arranged for analyzing. In this phase R Statistical tool is used for analyzing the data and get the results. R supports different statistical complex computation and generates wide variety of graphical outputs.

E Commerce in Global and Indian Market



Source: eMarketer, May 2019

Global ecommerce sells goods or services across geographic boundaries from the country of origin of a business usually identified as its establishment or incorporation venue the goods or services are marketed through internet marketing and outreach to non-native markets.

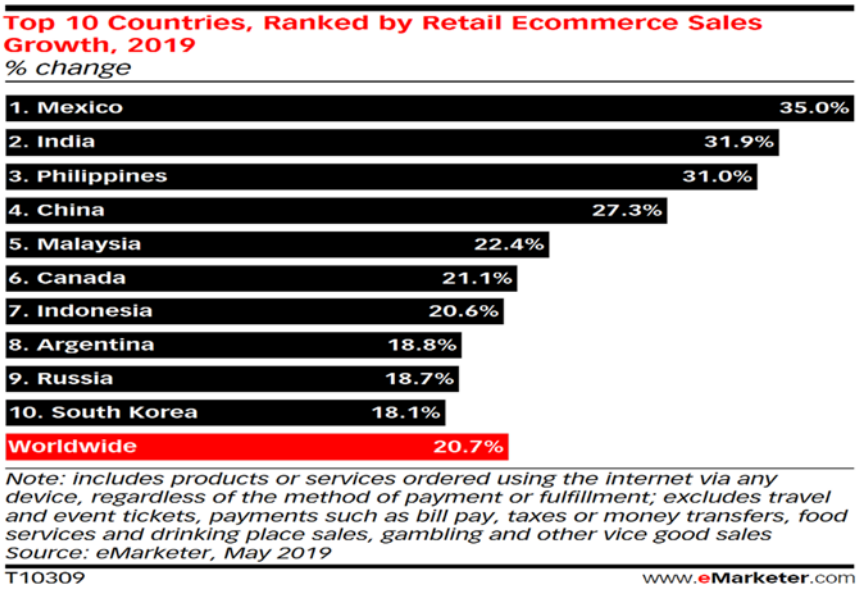
According to the above chart In Global market, Retail Ecommerce in 2019 total sales made is 3.53 trillion US Dollars with the rise of 20.7% to the previous year. Also anticipated 4.206 Trillion US Dollars of sales in 2020 however due to the COVID 19 Pandemic situation we see there is huge demand for groceries, vegetables, daily usage products, health care, personal care and essential products. A Business Standard report citing an official of the online business industry said that 20 to 30 percent of the turnover of e commerce companies like Flipkart, Amazon, Big Basket and Grofers most recent 15 days after the corona virus cases increased in India.

To begin with, going global doesn't necessarily demand a global presence. Shoppers online are constantly searching for transactions beyond their country's borders. In fact, overseas purchases were in the majority on all but one continent during one six-month assessment: North America. None of this requires multiple storefronts for each location, or the establishment of international warehousing and performance. Prioritizing internet ads or social media overseas is one of the easiest ways to get off exploring foreign markets. That needs an international approach by global targeting of Google Advertising, Product Listing Advertising, Twitter, and Instagram by using Big data analytics technology.

Aside from the worldwide figures, the regional ecommerce markets rate as follows:

Region	Total sales
Asia	$831.7 billion
North America	$552.6 billion
Europe	$346.5 billion
Australia	$18.6 billion
Africa and the Middle East	$18.6 billion
South America	$17.7 billion

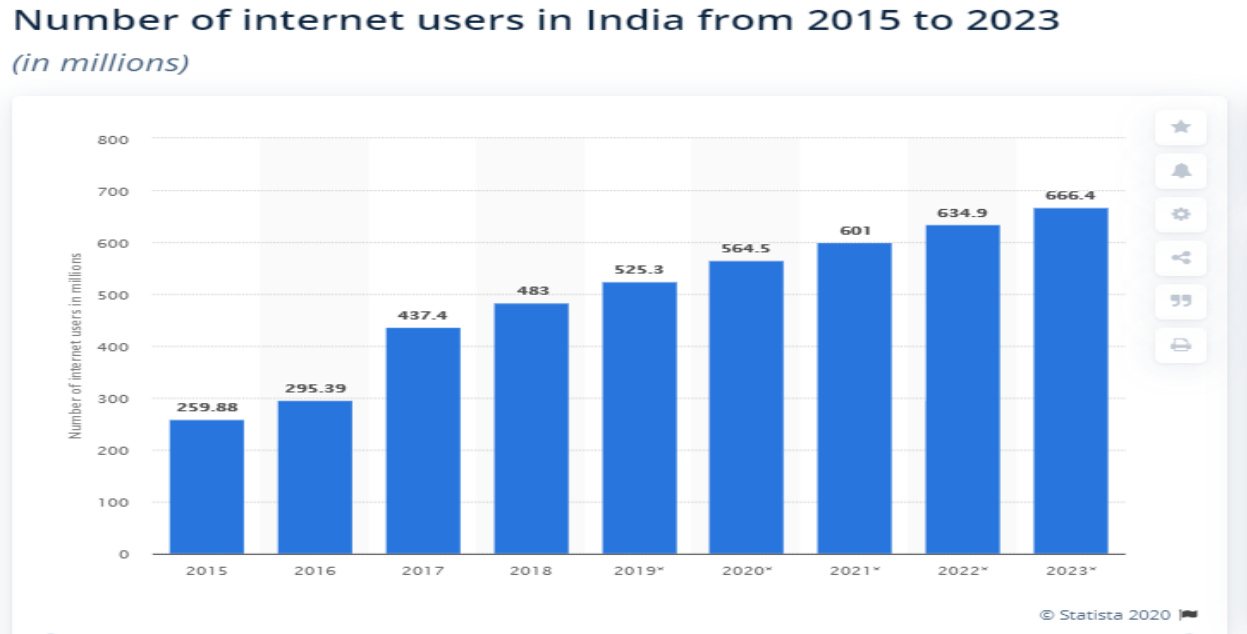
Asia region has highest sales of 831.7 Billion US Dollars followed by North America, Europe, Australia, Africa and Middle East and South America with total sales of 552.6 Billion USD, 346.5 Billion USD, 18.6 Billion USD, 18.6 Billion USD and 17.7 Billion USD respectively.



Source: eMarketer, May 2019

Even though there is rise in growth in retail Ecommerce, Mexico with 35% growth in sales ranked first in 2019, followed by India with 31.9% growth in sales. Worldwide however there is 20.7% growth in sales.

Internet Usage and Data Literacy



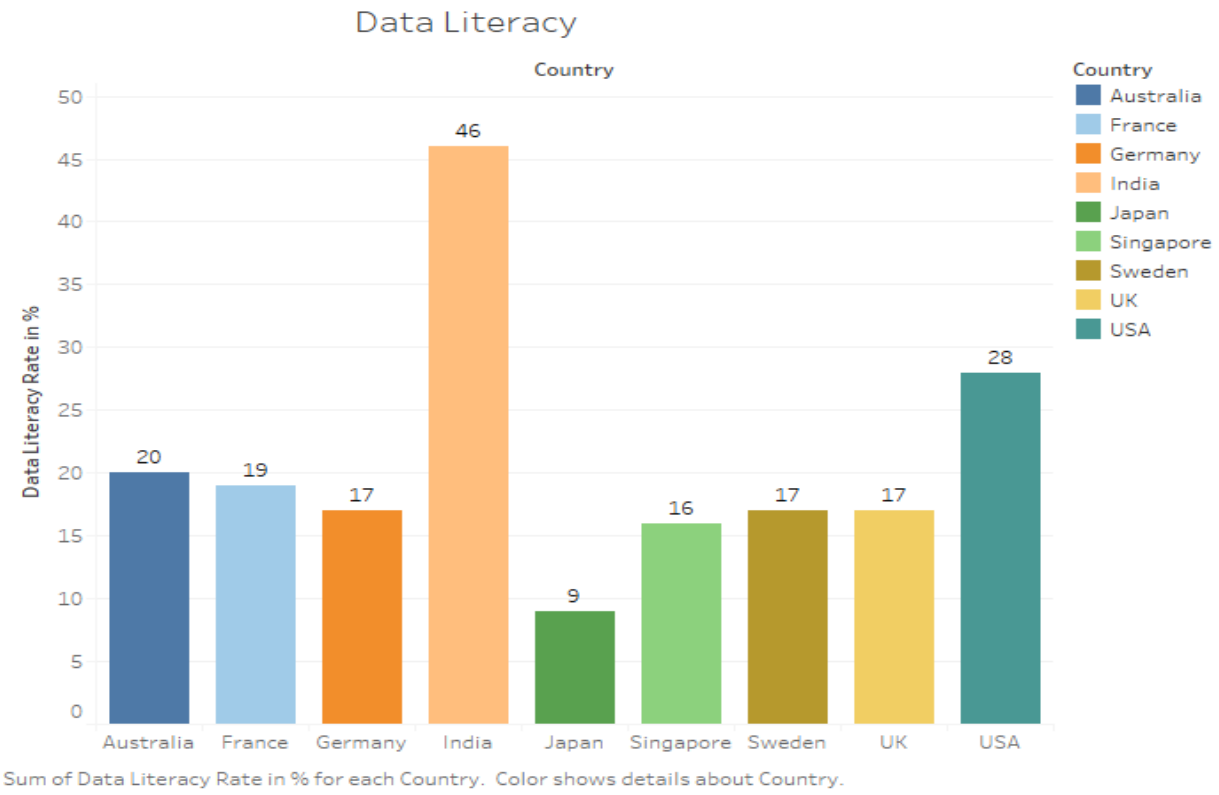
Source: Statista 2020 (https://www.statista.com/statistics/255146/number-of-internet-users-in-india/)

The internet usage or network and device are one of the important factors that influence the E commerce sector as the transaction takes place through internet. The numbers of internet users in India are increasing consistently every. According to a report by Internet and Mobile Association of India (IAMAI). According to the study, of the 451 million monthly active users, 385 million are over 12 years of age and 66 million are between 5 and 11 years of age, who have access to the Internet via the smartphones of family members. In India, almost 1/3 of users accessed the Internet for 1more than an hour in urban India, where a comparable proportion of users accessed the Internet for 15-30 minutes, as in rural India. The study claimed that improved accessibility, quality of service and availability of mobile Internet could lead to an increase in rural consumers spending more time on the Internet in the future. The study reported that 2/3 of Internet users in India are between 12 and 29 years of age. Interestingly, a higher proportion of this age group is present in rural India. This indicates that there is a scope for growth in this market. The biggest challenge apart from data security is transportation and infrastructure availability to provide services for the rural area. It is anticipated that the number of internet users to reach 564.9 million by the end of year 2020.

The ability to learn, interact with, interpret and connect with the data is data literacy. It is an ability that empowers all levels of staff to ask the right questions about data and computers, create awareness, make choices, and communicate meaning to others. The complexity of data analysis, especially in the context of big data, means that some knowledge of mathematics and statistics is needed in data literacy. Many companies employ experts, called data scientists, who have specialized analytical expertise to deal with the difficulty. Some organizations have also added a C-level employee, the Chief Data Officer (CDO), to make sure the company understands the full value of its data. Nevertheless, given that data is so essential to the success of a company, an increasing number of companies require some degree of data literacy from all employees.

Data literacy skills include the following competencies:

- Understanding what data will be used for a particular purpose;
- Interpreting visualizations of data, such as graphics and charts.
- Thinking logically about the information generated by the data analysis.
- Understanding the tools and methods of data analysis and where and when to use them.
- To understand when data is misrepresented or misused.
- Transmitting data information to people without data literacy, an ability often referred to as data storytelling.



Source: Indian express February 2020

India is having highest data literacy rate in workplace compared to any other countries in the world with 46% of literacy rate. Above graph describes the data literacy rate for the top 10 data literate countries. Employees in workplace are keen to learn data related skills. Among them India, USA and Australia given more importance and have highest data literates in workplace.

Application of Big Data in E Commerce:

E-commerce organizations like Amazon and Flipkart using data analytics to identify their loyal and new customers by tracking browsing habits and purchase patterns i.e. machine generated data. Customer targeted market implemented by making use of extracted customer related data to form market segmentation. Data analytics, complex algorithms, machine learning and business analytics is used to boost the sales. Highly personalized advertisements based on the purchase pattern and browsing history is the result of Big data analytics again. E commerce companies keep the track of each click made by the visitor/ customer for predictive analysis to anticipate the next product likely to be purchased by customer. The target customer groups are identified and the online advertisements are displayed in the blogs or the sites they visit through machine learning. According to the customer data the products are recommended like if the customer searches for mobile phone by using data analytics mobile cases are also displayed in recommended products advertisement. As per Flipkart each of its customers has one profile every time customer browses the site the unique session is. All the sessions are then combined to make stronger user profile and journey in the Flipkart.

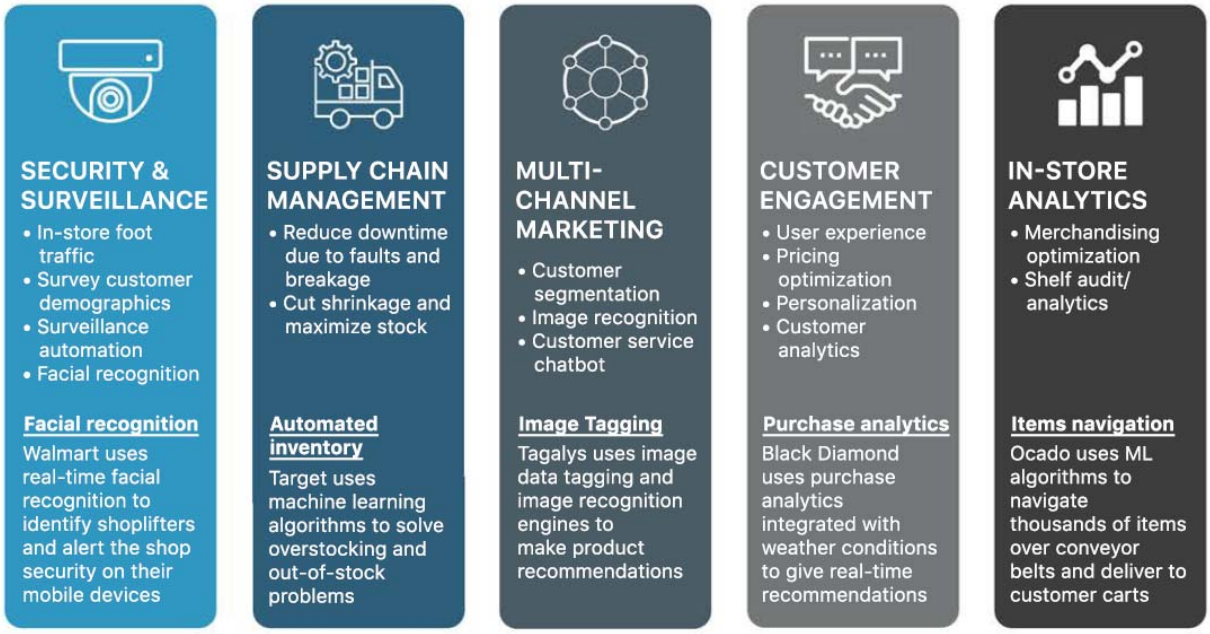
Big Data Analytics uncovers hidden patterns, unknown correlations, market trends, customer preferences and other useful business information by analyzing large set of structured, semi structured and unstructured data.

The analytical findings help in effective marketing strategies, new revenue opportunities, better customer service, personalized services, improved operational efficiency and competitive advantages in market. It also provides solutions for issues and better application related to business intelligence, global supply chain management, global process visibility, CRM, recommendations, personalized targeting, customer targeted market, pricing strategies, discounting etc., there are still gaps that data analytics can fill and enhance performance of e commerce players. According to the Macro level surveys conducted by Accenture, GE, and IBM, makes strong conclusions on data analytics. Companies that are using data analytics, 92% of executives are satisfied with the results provided by data

analytics and it has been found that 89% of executives agree Big Data as “very” or “extremely” important factor for their company.

The retail sector is moving from 'on-shelf supply' to a model of 'on-demand delivery' to offer goods dependent on place and time of the consumers. Customers deserve the lowest expense of high-quality goods and excellent service. Al technologies help retailers attract new consumers, meet buyers across improved networks, perform effective low-cost operations and improve workforce motivation.

One of the biggest challenges retailers face is user experience. Retailers are now using Al-based chatbots as a user-friendly way of making digital shopping experience more engaging. Target launched an in-store bot that can track up and down the aisles. Image recognition technologies are used to search image based products. Cortexica has partnered with John Lewis on an image-recognition platform that enables the visual search functionality allowing individuals to explore fashion products with similar colors, shapes and patterns. Al also offers plenty of opportunities for optimization in warehousing and in-store operations. Online marketplace in the UK, Ocado has embedded Al in the heart of its activities. Al and Big Data Analytics algorithms navigate thousands of items over conveyor belts and transport them to fill their shopping carts with customers. Machine Learning and Computer Vision will help retailers foresee customer demands in the future while automating supplier agreements.



Retail Value Chain

Source: NASSCOM Talent Demand and Supply AI and Big data analytics report, Banking and Industry Report 2018

While big data linked to new technology somehow squeezes the industry, it also offers significant new value-creating opportunities. Business retailers and their rivals are constantly in the process of finding and applying certain big data mechanisms that can give them a market advantage. The volume of data is rapidly growing as retailers not only monitor any consumer purchase and service, but also keep track of new data sources such as RFID chips that control goods, and online customer activity and feeling.

U.S. retail has effectively been leveraging information technology for decades. The transactional point-of - sale data, obtained primarily through the use of bar codes, first appeared in the 1970s. Several leading retailers have been using shop-level and supply chain data since the 1990s to automate delivery and storage, sharpen the preparation and procurement of inventory, and improve store operations. In previous McKinsey Global Institute research on productivity acceleration in general merchandise retail in the 1990s, we found that Wal-Mart caused the bulk of productivity acceleration directly and indirectly through ongoing managerial innovation (e.g., large-box formats, daily low price) that increased competitive intensity and led to the spread of managerial and technological best practices.

To link their supply chain electronically, Wal-Mart pioneered the development of an electronic data exchange network. Wal-Mart has also developed "Retail Link," a tool that provides its suppliers with a view of demand in their stores so that they know when stores should be restored instead of waiting for a Wal-Mart order. This "vendor-managed inventory," when introduced in the late 1980s, was a revolutionary concept. Both of these initiatives enhanced the capital and labor productivity and cost position of the retailers. As other stores tried to mimic what Wal-Mart had pioneered and stay profitable in the 1990's, profitability improved across the sector. Today, leading players are collecting consumer data to notify decisions they make about merchandising and price optimization of their supply chain. Detailed and cost-effective customer tracking by Wal-Mart gives retailers the ability to mine petabytes of customer preferences and buying behavior data and thus gain significant price and distribution concessions from the consumer product company. Branch-wide retailers are becoming more sophisticated in slicing and dicing big data from multiple sales channels, catalogues, stores and online interactions. The widespread use of increasingly granular customer data may allow retailers to improve their marketing and merchandising efficiency. Big technology mechanisms applied to logistics and supply chains will continue to reduce costs and eventually generate new strategic advantages and tactics to raise sales for retailers.

Mckinsey Global Institute Has Identified 16 Big Data Levers in Retail

We defined 16 big data market triggers that can be used by retailers throughout the value chain. These levers fall into five main categories: marketing, merchandising, operations, supply chain, and new business models

Big data retail levers can be grouped by function

Function	Big data lever
Marketing	▪ Cross-selling ▪ Location based marketing ▪ In-store behavior analysis ▪ Customer micro-segmentation ▪ Sentiment analysis ▪ Enhancing the multichannel consumer experience
Merchandising	▪ Assortment optimization ▪ Pricing optimization ▪ Placement and design optimization
Operations	▪ Performance transparency ▪ Labor inputs optimization
Supply chain	▪ Inventory management ▪ Distribution and logistics optimization ▪ Informing supplier negotiations
New business models	▪ Price comparison services ▪ Web-based markets

SOURCE: McKinsey Global Institute analysis

Marketing

1. Cross Selling: State-of-the-art cross-selling uses all the details that can be learned about a buyer, including profiles of the consumer, buying experience, interests, positions in real time, and other information to maximize the overall size of the order. For example, Amazon.com uses collaborative filtering to generate prompts for each product purchased or visited, "you might also want". At one point, Amazon announced that the recommendation engine accounted for 30 per cent of revenue. Another example of this tool is the use of big data processing to automate in-store sales relating complementary goods and packaged products.

2. Location-based marketing: Location-based marketing is focused on the through proliferation of smartphones and other mobile apps allowed for personal location data. It addresses customers who are or are still in store close by. For example, as a customer enters a clothing shop, the shop will deliver a promotional deal to the client's smartphone on a sweater. The startup PlaceCast claims that as a result of such location-based ads, more than 50 per cent of its users made a purchase. Near to 50 per cent of handset owners are using or intending to use their mobile shopping phones.

3. In-store behavior analysis: Analysis of in-store behavior data can help improve the layout of stores, product mix and positioning of shelves. Recent innovations have facilitated retailers to track the buying habits of shoppers (e.g., footpath and time spent in various areas of a store), derive real-time location data from smartphone apps (e.g., shopkick), buying cart transponders, or actively control cell phone positions within a retail setting. For monitoring in-store traffic trends and customer behavior, some retailers use advanced image recognition tools linked to their video surveillance systems.

4. Customer micro-segmentation: The next major data tool contributing to marketing is micro-segmentation of consumers. Big data has allowed enormous creativity in recent years, but this is a common concept in retail. The volume of data available for segmentation has exploded, and the growing complexity of analytical software has allowed the division into ever more granular micro-segments — to the point that certain retailers can claim to be engaging in personalization rather than pure segmentation. In addition to using traditional market research data and historical purchasing data, retailers can now track and leverage data on individual customer behavior — including clickstream web-based data. Retailers can now update this increasingly granular data to suit customer changes in near real time. Neiman Marcus, a high-end retailer, has established both customer segmentation and a multi-tier loyalty incentive scheme, resulting in considerably more sales of higher-margin goods from the richest, higher-margin customers.
5. Sentiment analysis: Sentiment analysis leverages the voluminous streams of consumer-generated data in different forms of social media to help inform a variety of business decisions. For example, retailers can use sentiment analysis to guess the response to marketing campaigns in real time and adjust course accordingly. The emerging area of analyzing social media data plays a key role, as consumers continually rely on peer opinions and suggestions to make buying decisions. Various applications have arisen to track and respond to Web-based user behavior and preferences in real-time.

6. Enhancing the multichannel consumer experience: Improving the multichannel experience can be a strong engine of revenue, customer service and engagement for customers. Retailers may use big data to easily combine shoppers' sales and pricing, whether such customers are online, in-store, or perusing a catalogue. For example, Williams-Sonoma has incorporated consumer databases with information about some 60 million families, monitoring such items as their wages, housing prices and children's numbers. Targeted emails based on this information generates ten to 18 times the response rate of non-target emails, and the organization is able to develop various variations of its catalogs that are customized to the actions and desires of specific client segments.

Merchandising

1. Assortment optimization: Deciding which items to sell in which stores will dramatically improve revenue based on market preferences, customer experience, and other big data so-called variety optimization. For example, one leading drug manufacturer used customer studies, demand and competitive analysis, and comprehensive economic modeling to determine the root causes of its flat and decreasing category-level performance. It that its cumulative stock-keeping unit (SKU) count by 17%, increased private label labels from 10% of the commodity mix to 14%, and reported a 3% increase in earnings as well as a 2% rise in sales.

2. Price optimization: Today, retailers can take advantage of the increasing granularity of pricing and sales data, and use higher analytical horsepower levels to bring price optimization to a new level. Various sources of data can be used to evaluate and inform pricing decisions in near real time. Specific demand-elasticity models analyze past market data to obtain insights into SKU-level dynamics, including markdown prices and pacing. Retailers can use the resulting data to analyze promotional events and evaluate sales lift sources and any underlying costs that could result from them. One food retailer looks at pricing elasticity among its customers for various categories. For example, rural food consumers see butter and rice as a higher purchasing priority and so these products may be less price elastic than they would be for urban shoppers. Meanwhile, urban consumers tend to rank cereals and sweets higher among their priorities.

3. Placement and design optimization: Brick-and - mortar stores may also make significant improvements by improving the positioning of products and graphic structures (e.g., end caps, shelves) by extracting inventory data at the SKU level — in effect, a more local variant of the kinds of optimization that take advantage of foot traffic data. Online retailers are able to change Web site placements based on user activity data such as loading, tapping and mouse over. For example, eBay performed thousands of trials with different facets of its website to determine optimal layout and other features from navigation to images scale.

Operations

1. Performance transparency: Retailers will also execute daily performance analyzes that can be aggregated by retail revenue, SKU revenue, and per employee revenues and recorded by them. Today these systems are increasingly moving closer to real time. Retailers should look to cashiers for consistency and daily sales and customer service efficiency depending on the number of consumer issues handled in a single call, customer feedback, and satisfaction surveys. Although the industry still makes extensive use of performance monitoring at a simple level, the move towards much higher frequency, immediacy, and granular monitoring enables managers to make substantive changes in their activities even more rapid, i.e. there is still headroom to add benefit from this lever.

2. Labor inputs optimization: Another organizational tool that can generate efficiency by reducing costs while retaining service rates is around labor supply management, automatic time and attendance monitoring, and better job scheduling. This lever will make forecasts of staffing needs more accurately, particularly during peak times, so that overcapacity can be avoided. Since store labor accounts for around 30 per cent of the fixed costs of the typical retailer, it is well worth using this lever.

Supply chain

1. Inventory management: Big data will help to enhance inventory control for retailers with the enhanced insight provided by deep analytics mining various databases, Best-in-class inventory management offers full clarity at SKU level, while bar code solutions linked to automated replenishment processes minimize instances of out of stock. Leading retailers are enhancing stock forecasting by combining multiple datasets such as sales history, weather forecasts and seasonal sales cycles. Together, improved inventory management allows retailers to hold a lower stock level because supplies are coupled with demand signals much more closely, while reducing the number of sales lost due to merchandise stock-outs.

2. Distribution and logistics optimization: Leading retailers now automate transportation through the use of GPS-enabled Big Data Telematics (i.e. remote location monitoring, etc.) and route planning to enhance their vehicle management and delivery. By improving fuel efficiency, proactive maintenance, driver behavior and vehicle routing, transport analytics can increase profitability.

3. Informing supplier negotiations: In a world of big data, leading retailers can analyze customer preferences and buy behavior to inform their dealings with suppliers. For example, they will use the data on pricing and sales to target negotiating negotiations on key goods. Use big data in this area is a huge advantage, because the inventory of the products sold by a store represents the greatest portion of the business. Nonetheless, we should remember that manufacturers still realize the value of recognizing customer desires and are constantly gaining access to and reviewing consumer experience data to discover trends that strengthen their hand in retailer negotiations.

New business models

The avalanche of data in the retail industry, along with other market developments, is encouraging new business models to emerge. Such models are the most exciting and creative to mainstream retailers-but perhaps the most potentially dangerous. Two new business models with the most traction today is Web-based and price comparison services.

1. Price comparison services: It is popular for third parties today to provide real-time or near-real-time pricing and associated market control over several retailers on goods. At different retail outlets customers will immediately compare the price of a single commodity. Where those comparisons can be made, prices tend to be lower. Studies show that when consumers can shop using these services, they save an average of 10 per cent. Retailers ought to think carefully about how to respond to these price matching services. Many that are willing to compete on demand would want to make sure they are the most noticeable on these facilities. Retailers who are unable to compete on demand

would need to decide how to separate themselves from rivals in a market-transparent environment, whether it be in customer experience content, differentiated goods, or other value-added service delivery.

2. Web-based markets: Web-based marketplaces like those offered by Amazon and eBay offer searchable lists of items from a wide range of sellers. We provide access to a large number of niche retailers in addition to pricing stability, because we do not have the distribution or selling resources to attract customers. We do have a vast amount of valuable product knowledge, including consumer-generated feedback, and offer customers more access.

Case Study on Amazon

Amazon uses the big data tools to deliver superior customer service. Right from the time it started to emerge as a dominant provider of Internet services in the early 2000s, Amazon had focused primarily on big data to improve its performance. Among many other major Internet companies, it discovered the importance of big data in the early 2000s, and since then has focused on the proper use of the vast databases of people shopping on its e-commerce portals.

Amazon has leveraged its broad data repositories to provide its consumers with strong product reviews and thereby strengthen their interaction with them. It has used its big data tools to meticulously update its well-known consumer recommendations framework. Data on previous transactions made by consumers are used to offer highly personalized product advice. Analysis of previous customer details has allowed Amazon to give feedback to potential consumers who were first ordering from its platform. Big data helped Amazon create 360-degree user accounts and produce hyper-personalized product marketing communications focused on the needs and desires of particular consumers. Amazon has also used its big data capabilities on the customer side to increase the performance of the customer service. Simple access to client profiles and their previous buying / browsing patterns made it possible for the customer service executives of the business to have swift responses to consumer problems. The acquisition of Zappos by Amazon in 2009 further encouraged the use of big data to enhance the level of customer support. Few innovative uses, such as checking fraud at the corporate level, were also put to big data services.

Over the years, Amazon has grown from being a single e-commerce player to a giant Internet service company delivering a wide variety of offerings to consumers and businesses. It began to concentrate extensively on big data and embarked on its transition from a pure online retailer to a giant big data organization. Amazon, along with other big Internet companies such as Yahoo! Inc. (Yahoo) and Twitter, Inc. (Twitter), discovered in the early 2000s that they had large volumes of customer data that they could use in a meaningful manner. While the other businesses did not dwell on the value of big data, Amazon was swift to cash in on the invaluable inventory of customers who shopped on its e-commerce sites around the world. At Amazon, the product recommendations team dreamed of imaginative ways in which it could exploit the company's gathered data. The result was a big data boom that changed the way Amazon did business. Amazon 's reputation as an e-commerce leader has always focused on making the best goods available to its consumers. In reality, making the best products available was focused on knowing the specific products that consumers needed. Understanding consumers' interests and preferences included conducting thorough market analysis and assessing their own customer base. Amazon was known for its product recommendation system that, based on their previous purchase actions, supplied consumers with product reviews. The main driving force behind Amazon's recommendation system was data obtained from its consumers. As the largest e-commerce player, Amazon had a huge set of data on its consumer base 's preferences and previous purchase actions.

In order to improve customer satisfaction and build company loyalty, the organization uses predictive analytics for targeted marketing.

Personalized Recommendation System

In using a robust, collaborative filtering engine (CFE), Amazon is a pioneer. It analyses what goods you have recently bought, what is in your online shopping cart or on your wish list, what things you have tested and assessed, and what items are most searched for. This data was used to recommend additional products purchased by other consumers after ordering the same items.

At the moment, Amazon.com is the biggest online shopping store in the country. Amazon originated as an online book store and quickly diversified into, among other things, the manufacture and selling of luxury products, appliances, and household objects. Recommenders for Amazon items are used to tailor the online shop experience of each user. Things are tailored (on a real-time basis) to suit each client's taste. As Amazon gathers millions of client data, this is a big data challenge.

To aggregate customers based on interests, Amazon makes use of clustering algorithms and collaborative filtering. Customers with their product recommendation scheme organize into clusters / groups focused on:

- Similar search
- Item to Item collaborative filtering.

Search based Product Recommender at Amazon.com: Search-based (also referred to as content-based search) uses a customer order history and a rated item to produce a search query that identifies all items similar to customer preferences (such as authors or related genres). For instance, if a consumer purchases a "God Father" DVD. Movies from similar writers, similar formats, and similar distributors would be recommended by the product recommender (Linden et al., 2003).

Item to Item collaborative filtering at Amazon.com: Not only does Amazon.com use its product reviews in its e-commerce website, but it is often used in the field of email campaigns as a marketing tool. Customers can sort the product by product suggestion based on products currently in their shopping carts by clicking on the link " Your Recommendations". Collaboration with items at Amazon.com operates by comparing one of the items bought or valued by users to an object that is close. The algorithm creates a table of related items by identifying items that consumers choose to buy. By iterating across item pairs, the algorithm also generates a product-to-product matrix and computes a similarity metric for each pair (Linden et al., 2003).

One-Click Ordering

Since big data suggests that while the items are shipped fast, you buy elsewhere, Amazon built One-Click shopping. When you put your first order and enter a shipping address and payment form, One-Click is a proprietary function that is instantly activated. You have 30 minutes in which you can change your mind about the transaction when you want One-Click shopping. After that, the product is automatically paid and delivered to your address through your payment system.

Anticipatory Shipping Model

Big data is also used through the proprietary anticipatory shipment platform of Amazon to forecast the goods you are likely to purchase, where you will purchase them, and when you may like the goods. The goods are shipped to a nearby delivery centre or distributor, so if you order them, they will be ready for shipment. To raise its retail revenue and profit margins, Amazon uses predictive analytics while reducing its distribution time and total expenses.

Supply Chain Optimization

Since Amazon needs to easily satisfy the orders, the company communicates with suppliers and records their product. To cut delivery costs by 10 to 40 percent, Amazon's big data solutions pick the warehouse nearest to the retailer and/or you, the customer. In addition, graph theory helps determine the optimal distribution plan, route, and product groupings in order to further minimize shipping costs.

Price Optimization

Big data were now used to monitor the costs of Amazon to draw more buyers and raise revenues yearly by an average of 25 percent. Prices are calculated according to your website operation, positioning of rivals, supply of goods, tastes of goods, order history, estimated profit margin, and other considerations. Product rates normally adjust every 10 minutes as it monitors and analyses big data. As a consequence, Amazon usually provides discounts on best-selling merchandise and receives higher earnings on less-popular goods. The cost of a novel on the New York Times Best Selling list, for instance, could be 25% cheaper than the market price, while a novel not on the list costs 10% more than the same book published by a competitor.

Dynamic Pricing

It allows the use of historical details such as: past sales, clickstream, cookies, forecasting mechanisms for company capital to automatically set an item's pricing or provide personalized discounts. This technology personalizes the price / discount to meet a single consumer in real time for a particular item. Therefore, it is possible to buy the same thing from an online retailer at two different rates for two different clients. While this technology has its advantages, clients can get worse or feel the sense of discrimination due to the variation in the price offered for the products.

Amazon Web Services

Businesses can build portable big data systems and protect them without using hardware or managing resources through Amazon Web Services (AWS), Amazon's cloud storage service launched in 2006. Cloud-based computing involves big data technologies including clickstream analytics, fraud detection, event-driven ETL, data warehousing, recommendation engines, and Internet-of-Things (IoT) processing. Through using them to evaluate consumer profiles, buying patterns, and other related details to more efficiently cross-sell company goods in ways comparable to Amazon, businesses will benefit from Amazon Web Services. These stores, in other words, will use Amazon to stalk you as well.

Conclusion

Big Data and Data analytics application in business helps in increasing revenue, enhance sales, optimization of process, reduce cost, understand customer needs and preference in better way to find behavioral pattern. Predictive analysis helps in analyzing product demand, preferred price of products in customer mind, product recommendation and model building. One of the most important players in E commerce i.e., Amazon by effective utilization of big data and data analytics technology made huge profit. Data literacy and Increase in the usage of mobiles and Internet are added advantage of E commerce companies. Big data application helps in personal recommendation, one click ordering, Anticipatory shipping model, Supply chain optimization, Price optimization, Dynamic pricing and predictive analysis. The implementation of advanced technology in business needs investment of money however it is competitive advantage in turn helps in running business in smooth way and reduce cost by making optimal utilization of resources and reduce process life cycle as well.

Reference

1. Agarwal, N. (2018, November 16). *How Amazon, Flipkart use data analytics to predict what you are going to buy | Mint*. Mint. https://www.livemint.com/Companies/RX5eOy12n5JFJu617G5GnM/Amazon-Flipkart-data-analytics-ecommerce.html
2. Aggarwal, Kavita. 2022. “E-COMMERCE INDUSTRY.” (March).
3. Akter, Shahriar, and Samuel Fosso Wamba. 2019. “Big Data and Disaster Management: A Systematic Review and Agenda for Future Research.” *Annals of Operations Research* 283(1–2):939–59. doi: 10.1007/s10479-017-2584-2.
4. *Amazon`s Big Data Strategy*. (2014). Amazon`s Big Data Strategy. https://www.icmrindia.org/casestudies/catalogue/IT%20and%20Systems/Amazon’s%20Big%20Data%20Strategy-Excerpts.htm
5. Avinash BM, and Akarsha BM. 2007. “Big Data Analytics for E-Commerce-Its Impact on Value Creation.” *International Journal of Advanced Research in Computer and Communication Engineering ISO* 6(12):181–88. doi: 10.17148/IJARCCE.2017.61235.
6. Balaraman, Premkumar, and Sabarinathan Chandrasekar. 2016. “E-Commerce Trends and Future Analytics Tools.” *Indian Journal of Science and Technology* 9(32):1–9. doi: 10.17485/ijst/2016/v9i32/98653.
7. Basuroy, T. (2021, July 27). *India: Number of internet users 2040*. Statista. Retrieved September 18, 2022, from https://www.statista.com/statistics/255146/number-of-internet-users-in-india/
8. Bhardwaj, Mamta. 2016. “How Ecommerce Analyses Data.”
9. Flipkart, Walmart-backed. 2018. “How Amazon , Flipkart Use Data Analytics to Predict What You Are Going to Buy.”
10. Follow, Sophia Martin. 2020. “Is Big Data Analytics Boon for Your ECommerce Business ? In the Third Quarter of the Year 2018 PayPal , Which Is One of the Most Trusted Payment Gateways , Calculated Approximately 200 Million Registered Users in Total Worldwide .” 1–9.
11. *Global Ecommerce 2019 - Insider Intelligence Trends, Forecasts & Statistics*. (2019, June 27). Insider Intelligence. https://www.emarketer.com/content/global-ecommerce-2019
12. *Global Ecommerce: Stats and Trends to Watch (2022)*. (2022, February 16). Shopify Plus. https://www.shopify.com/enterprise/global-ecommerce-statistics
13. Guldi, Jo. 2021. “Big Data.” *The Routledge Companion to Historical Theory* (October 2012):430–45. doi: 10.4324/9780367821814-27.
14. *How Amazon Uses Big Data In Their Business [Case Study] | Board Infinity*. (2019, May 18). Blog | Board Infinity. https://blog.boardinfinity.com/role-of-big-data-in-business-with-case-studies/

15. *How ecommerce analyses data | Deccan Herald*. (2016, June 5). Deccan Herald. https://www.deccanherald.com/content/550697/how-ecommerce-analyses-data.html
16. K, Moorthi, Srihari K, and Karthik S. 2017. "A Survey on Impact of Big Data in E-Commerce." *International Journal of Pure and Applied Mathematics* 116(21):183–88.
17. Linden, Greg, Brent Smith, and Jeremy York. 2003. "Amazon.Com Recommendations: Item-to-Item Collaborative Filtering." *IEEE Internet Computing* 7(1):76–80. doi: 10.1109/MIC.2003.1167344.
18. Mandavia, M. (2019, September). *India has second highest number of Internet users after China: Report - The Economic Times*. The Economic Times. https://economictimes.indiatimes.com/tech/internet/india-has-second-highest-number-of-internet-users-after-china-report/articleshow/71311705.cms?from=mdr
19. Manyika, James. (2011, May). *Big data: The next frontier for innovation, competition, and productivity | McKinsey*. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation
20. Martin, S. (2019, November 16). *Is Big Data analytics boon for your eCommerce business? | by Sophia Martin | DataDrivenInvestor*. Medium. https://medium.com/datadriveninvestor/is-big-data-analytics-boon-for-your-ecommerce-business-b97c73f449fd
21. Muhtaroglu, F. Canari Pembe, Seniz Demir, Murat Obali, and Canan Girgin. 2013. "Business Model Canvas Perspective on Big Data Applications." *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013* 32–37. doi: 10.1109/BigData.2013.6691684.
22. Paper, Conference. 2016. "Big Data Analytics and Its Application in E- Commerce Name : Uyoyo Edosio Big Data Analytics and Its Application in E-Commerce Using Case Studies of Adidas , Walmart and Amazon . Com." *ResearchGate* (April 2014):0–12.
23. Schultz, J. (2019, August 6). *How Much Data is Created on the Internet Each Day?* Micro Focus Blog. https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/
24. Shows, Qlik, Data Skills Gap, Costing Organizations Billions, and Lost Productivity. 2020. "New Research from Accenture and Qlik Shows the Data Skills Gap Is Costing Organizations Billions." 43–45.
25. Srivastava, T. (2015, August 7). *What is the role of analytics in E-Commerce industry?* Analytics Vidhya. https://www.analyticsvidhya.com/blog/2015/08/role-analytics-e-commerce-industry/
26. *Talent Demand & Supply Report: AI & Big Data Analytics | NASSCOM*. (2018, December 10). NASSCOM. https://www.nasscom.in/knowledge-center/publications/talent-demand-supply-report-ai-big-data-analytics
27. Vázquez-Herrero, Jorge, and Xosé López-García. 2017. *Recent Advances in Information Systems and Technologies V3*. Vol. 571.
28. Δήμος Τρικκαίων. 2018. "Τρίκαλα Open Data." 44(1):7567.